

# 决策树结合混合像元分解的中国竹林遥感信息提取

崔璐, 杜华强, 周国模, 李雪建, 毛方杰, 徐小军,  
范渭亮, 李阳光, 朱迪恩, 刘腾艳, 邢璐琪

1. 省部共建亚热带森林培育国家重点实验室, 临安 311300;

2. 浙江省森林生态系统碳循环与固碳减排重点实验室, 临安 311300;

3. 浙江农林大学 环境与资源学院, 临安 311300

**摘要:** 竹林是中国亚热带地区特殊而重要的森林资源, 现有方法难以实现全国范围竹林时空分布信息快速准确提取。针对此问题, 本研究利用2003年、2008年、2014年MODIS NDVI、反射率产品数据和省域Landsat分类数据, 提出了基于决策树结合混合像元分解的全国竹林信息提取方法。首先, 通过最大似然法获取中国林地分布信息; 然后, 在林地信息的基础上, 构建决策树模型提取中国竹林分布信息; 最后, 采用线性最小二乘法混合像元分解得到中国竹林丰度图, 并计算竹林面积。研究结果表明: (1)最大似然法提取的3个时期中国林地信息的生产者与用户精度均在90%以上, Kappa系数均值为0.93, 为竹林信息提取奠定了基础。(2)C5.0算法构建的决策树模型能够很好的提取中国竹林时空分布信息, 3个时期竹林分类精度均在80%左右。(3)在混合像元分解的基础上, 统计得到的全国各省竹林估算面积与清查面积具有较高的相关性,  $R^2$ 分别为0.98、0.97和0.95, RMSE范围为3.92万—9.58万ha, 说明估算得到全国竹林面积与实际情况较为吻合。本研究所提出基于MODIS遥感数据运用C5.0算法决策树结合混合像元分解的方法, 实现了全国竹林时空分布信息的准确提取, 为全国竹林资源信息动态监测及管理提供了技术手段和数据支撑。

**关键词:** 中国竹林, MODIS数据, C5.0算法, 决策树分类, 混合像元分解

**引用格式:** 崔璐, 杜华强, 周国模, 李雪建, 毛方杰, 徐小军, 范渭亮, 李阳光, 朱迪恩, 刘腾艳, 邢璐琪. 2019. 决策树结合混合像元分解的中国竹林遥感信息提取. 遥感学报, 23(1): 166–176

Cui L, Du H Q, Zhou G M, Li X J, Mao F J, Xu X J, Fan W L, Li Y G, Zhu D E, Liu T Y and Xing L Q. 2019. Combination of decision tree and mixed pixel decomposition for extracting bamboo forest information in China. Journal of Remote Sensing, 23(1): 166–176 [DOI:10.11834/jrs.20187155]

## 1 引言

竹林是亚热带地区特殊的森林类型, 截至第八次森林资源清查(2009年—2013年), 中国竹林面积已达601万ha, 占世界竹林总面积的20%左右, 是名副其实的“竹子王国”。随着竹林生态功能及经济效益的凸显, 竹林资源及其时空动态监测已成为近年研究的热点(周国模等, 2017)。

遥感是大范围监测森林资源及其时空动态有效的工具(Griffiths等, 2014; Lu等, 2016; Stagakis等, 2016), 基于多种卫星遥感数据的竹林信息提

取方法也取得了较大的进展。在高分遥感方面, 多以IKONOS、SOPT 5等为数据源, 采用影像纹理信息或基于面向对象多尺度分割技术实现了竹林信息提取(颜梅春等, 2004; Han等, 2014, 2015; 孙晓艳等, 2013; 高国龙等, 2016), 因影像分辨率高, 且可以综合利用影像的光谱特征、几何形状和纹理细节等空间特征, 竹林信息提取精度一般能达到90%以上。中等分辨率遥感数据方面, 多以Landsat TM作为数据源进行竹林信息提取, 如杜华强等人(2008)构建了基于光谱特征的竹

收稿日期: 2017-05-19; 预印本: 2017-10-31

基金项目: 国家自然科学基金(编号: 31670644, 31370637, 31500520); 浙江省自然科学基金项目(编号: LR14C160001, LQ15C160003); 浙江省与中国林业科学研究院省院合作林业科技项目(编号: 2017SY04); 浙江省竹资源与高效利用协同创新中心开放项目

第一作者简介: 崔璐, 1990年生, 女, 硕士研究生, 研究方向为森林资源遥感监测与信息技术。E-mail: cui\_lu\_199108@163.com

通信作者简介: 杜华强, 1975年生, 男, 教授, 研究方向为竹林资源遥感监测及其碳循环遥感定量估算。E-mail: dhqrs@126.com

林信息提取方法, 竹林提取精度达到88%以上; 徐小军等人(2011)采用人工神经网络从TM影像提取浙江省安吉县毛竹林信息, 其用户精度达到91.94%, 与实地调查样地验证, 精度达到82.19%; Li等人(2016)采用随机森林方法从TM影像中获取浙江省竹林及时空动态信息, 提取精度在85%—90%。而基于低分辨率遥感数据的竹林分类精度较低, 如李俊祥等人(2005)利用NOAA\_AVHRR数据, 采用ISODATA算法对中国东部地区植被进行分类, 其中竹林分类精度仅为28.7%, 闽北西北武夷山、赣东、赣西等竹林分布区的竹林信息均未提取出来。

综上所述, 对于中高分辨率遥感数据, 不同方法均能得到较好的竹林信息提取结果, 但竹林提取范围多局限于局部区域; 低分辨率遥感数据能进行大尺度植被监测, 但受低分辨率遥感数据混合像元或分类方法所限, 竹林信息提取精度不是十分理想。多源遥感能够实现时间分辨率、空间分辨率和光谱分辨率的互补, 在大面积、全方位森林资源遥感监测方面具有明显优势, 如Shang等人(2013)综合MODIS和TM数据, 建立了一种基于匹配滤波技术竹林信息提取方法, 从MODIS像元中分解处理竹林, 实现了长三角地区竹林信息较高精度的提取, 为大尺度竹林信息提取提供可借鉴的方法。

中国竹林资源十分丰富, 但竹林面积仅占中国森林总面积(2.08亿ha)的3%左右, 且广泛分布于浙江、福建、江西、湖南、安徽、湖北、广西、广东等长江以南15个省(市、区), 如何从中国广袤的森林中快速、准确提取竹林信息, 并监测其时空动态是竹林资源信息化管理和高效利用所面临的重要问题。局部区域竹林资源遥感信息提取难以满足全国竹林同步监测及其时空动态分析的需求, 而大范围同步监测需要解决竹林遥感信息提取精度等问题, 因此, 必须发展适合全国竹林遥感信息精确提取方法。

NOAA\_AVHRR、SPOT 4\_VEGETATION和Terra/MODIS等粗分辨率遥感数据都可以进行大尺度植被动态监测。基于机器学习规则的决策树可以通过数据挖掘算法找出隐含在训练样本中模式类的特征, 并自动建立决策树及决策规则实现精确分类, 在遥感信息提取中得到很好的应用(李明诗等, 2006), 但对于特定森林类型, 粗分辨率遥

感数据需要解决混合像元这一问题。基于此, 本研究以全国MODIS NDVI产品和地表反射率产品为数据源, 并结合局部区域Landsat数据和地面样地调查数据, 提出构建决策树结合混合像元分解的中国竹林提取方法, 期望实现中国竹林时空分布信息准确提取。

## 2 研究方法

### 2.1 多源遥感数据与预处理

遥感数据包括中国MODIS NDVI产品(MOD13A2)、反射率产品(MOD09A1)和基于Landsat的浙江省竹林分布信息。2003年、2008年和2014年3个时期的MODIS数据由NASA提供(<https://ladsweb.modaps.eosdis.nasa.gov/>[2017-05-19])。MODIS数据包括MODIS NDVI数据和地表反射率数据, 其中MODIS NDVI数据(MOD13A2)每16天合成一次, 标准产品一年有23个时相数据, 空间分辨率为1000 m; 而MODIS地表反射率数据(MOD09A1)每8 d合成一次, 标准产品一年有46个时相数据, 空间分辨率为500 m。对MODIS NDVI和反射率数据, 重投影到WGS84坐标系统, 并利用最近邻域法将空间分辨率统一重采样到1000 m; 然后, 通过拼接、裁剪获得中国区的影像数据。MODIS数据由于受云层覆盖、大气气溶胶、雪覆盖、传感器故障等因素的影响而存在不同程度的数据噪声。根据云层覆盖少、数据完整、噪点少等筛选条件, 本研究剔除了每期MODIS NDVI数据中质量最差的3个时相数据, 并将剩余时相利用最大值合成法MVC(Maximum Value Composite)(Yang等, 2015)合成得到每期10组的NDVI月最大值(MODIS-NDVI<sub>max</sub>); 针对MODIS反射率数据则选取了第305 d、113 d、273 d的数据(MODIS<sub>Ref</sub>)分别用于3期的竹林分布信息提取。

### 2.2 竹林训练样本数据

竹林样本数据包括234个浙江省森林资源清查数据、157个30 m×30 m的样地调查数据, 以及136个来自公开发表文献的竹林样点数据(2003年20个、2008年35个、2014年81个), 样本数据分布如图1所示。另外, 以2014年浙江全省基于Landsat系列影像的土地利用分类图(孙少波, 2016)作为局

部区域的辅助数据，参与分类训练样本的选择。本研究采用随机不重复抽样从所有竹林样地中选取

50%作为决策树模型构建的训练样本，剩余样本用作精度检验。

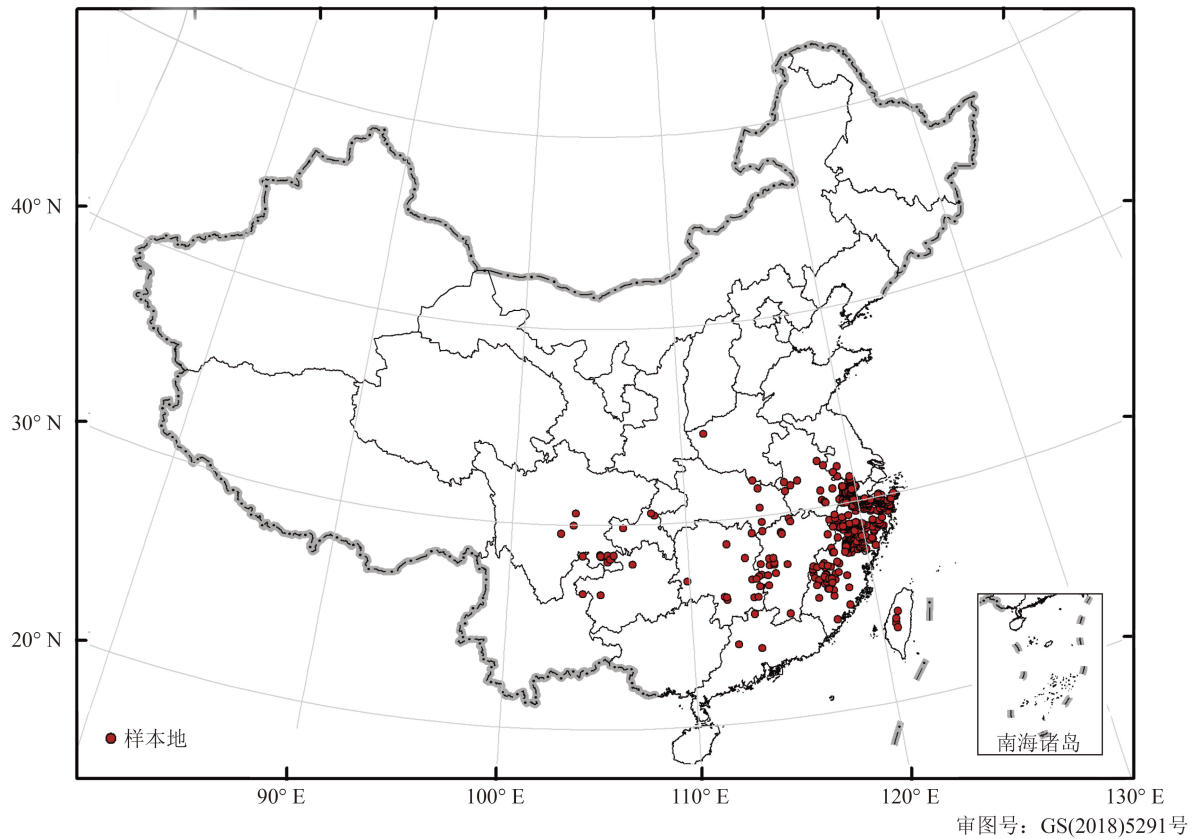


图1 研究区及竹林样本分布

Fig. 1 Study area and the distribution of bamboo forest samples

### 2.3 竹林信息提取方法

本研究利用决策树结合混合像元分解的方法提取中国竹林信息，技术路线如图2所示。即首先通过最大似然法获取中国竹林分布信息；然后，

在林地信息的基础上，构建C5.0决策树模型提取中国竹林分布信息；最后，采用线性最小二乘法混合像元分解得到中国竹林丰度图，并计算中国竹林面积。

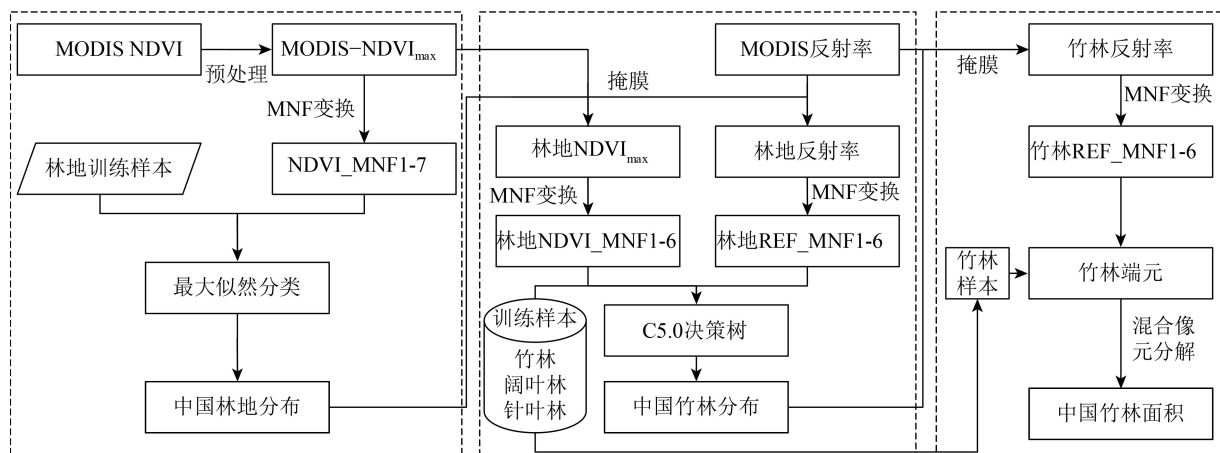


图2 本研究技术路线

Fig. 2 Flow chart of this study

### 2.3.1 林地分布信息提取

从全国尺度看，竹林呈碎片化分布于林地之中，因此本研究首先提取林地信息，然后再从林地中提取竹林信息。为去除噪声、减少数据冗余，采用最小噪声分离MNF (Minimum Noise Fraction)变换对每期10组MODIS-NDVI<sub>max</sub>进行处理，选取累计贡献率达到95%的前7个波段作为全国林地信息提取的输入变量(表1)，具体方法如下：首先，将浙江全省基于30 m土地利用类型分类结果重新划分为林地和非林地两大类；其次，为了与MODIS数据尺度一致，对30 m分类图结果进行尺度上推处理，计算得到1 km×1 km的林地丰度图(Shang等, 2013)；再次，因有林地定义为郁闭度0.2以上的乔木林地、竹林地、疏林地等，因此，选择林地丰度大于20%的像元作为林地信息提取训练样本；最后，采用最大似然法提取全国林地信息。

表1 2003年—2014年MODIS NDVI的MNF变换的波段特征

Table 1 Band eigenvalue of MNF transform of MODIS NDVI during 2003—2014

波段	2003年		2008年		2014年	
	特征值	累积贡献率/%	特征值	累积贡献率/%	特征值	累积贡献率/%
1	155.86	68.10	158.00	69.72	121.62	67.27
2	30.92	81.61	28.24	82.18	24.39	80.75
3	10.49	86.19	10.14	86.66	7.87	85.11
4	7.41	89.43	7.55	89.99	6.06	88.46
5	6.61	92.31	5.70	92.50	5.32	91.40
6	4.45	94.26	4.44	94.46	3.99	93.61
7	4.01	96.01	4.12	96.28	3.79	95.71
8	3.27	97.44	3.13	97.66	3.17	97.46
9	3.00	98.75	2.77	98.88	2.48	98.83
10	2.87	100.00	2.54	100.00	2.11	100.00

### 2.3.2 基于决策树的竹林分布信息提取

决策树是通过特征属性和目标变量构成的训练数据集的循环分析形成的二叉树结构，它由一个根节点、一系列分支及终极节点组成，并通过终极节点实现分类(孙晓艳, 2014; 乔艳雯等, 2013)。C5.0决策树由Quinlan从最初ID3(Duan等, 2016; Ture等, 2009)、C4.5(Quinlan, 1993)算法

中发展起来的最新树算法，它以Boosting方法生成一系列决策树、以信息增益率为标准确定最佳分组变量和最佳分割点，并对决策树各节点进行裁剪或合并，从而得到最佳终极节点。因此相对于其他决策树算法，C5.0能得到更高的分类精度，且适用于大数据集分类(Chao等, 2014; Klein等, 2012)。

本研究采用C5.0算法，以全国林地MODIS-NDVI<sub>max</sub>和MODIS<sub>Ref</sub>作为竹林信息提取的数据。首先，分别对林地的MODIS-NDVI<sub>max</sub>、MODIS<sub>Ref</sub>做MNF变换，并保留累计贡献率大于90%的15个波段作为构建决策树的特征变量。其次，根据尺度转换后的浙江省土地利用分类图，选择阔叶林、针叶林丰度大于60%的像元和全国竹林样点作为训练样本，构建C5.0决策树模型，实现全国竹林分布信息的提取。

### 2.3.3 混合像元分解与竹林面积信息提取

决策树是在像元水平上进行，所提取的竹林信息依然会包含其他森林类型，因此，进一步对决策树初步提取的全国竹林分布信息进行混合像元分解，获取各期竹林丰度，统计竹林面积。端元选择是混合像元分解结果的重要因素(赵英时, 2013)，本研究基于MNF变换(徐小军等, 2011; Shang等, 2013)及图1所示竹林样本获取竹林端元。在此基础上，采用线性最小二乘法模型进行混合像元分解，提取纯净竹林面积信息。线性最小二乘法模型广泛应用于遥感影像混合像元分解(Adams等, 1995; 范渭亮等, 2010; 徐小军等, 2011; 王聪等, 2015)，其一般表达式如下

$$R_i = \sum_{k=1}^n f_k R_{ik} + \varepsilon_i \quad (1)$$

$$\sum_{k=1}^n f_k = 1 \quad (2)$$

式中， $i=1, \dots, m$ ， $m$ 为波段数； $k=1, \dots, n$ ， $n$ 为端元数； $R_i$ 为像元在波段 $i$ 上的光谱反射率； $f_k$ 为像元内端元 $k$ 的丰度值； $R_{ik}$ 为端元 $k$ 在波段 $i$ 上的光谱反射率； $\varepsilon_i$ 为波段 $i$ 上的残余误差。

对竹林面积信息提取精度一方面采用均方根误差RMSE(Root Mean Square Error)对混合像元分解结果进行评价，另一方面，基于各省竹林实际清查面积对结果进行评价。

### 3 结果与分析

#### 3.1 林地提取结果

林地信息提取精度分析如表2所示。可以看

到, 利用最大似然法提取林地信息达到了较高的分类精度, 生产者精度与用户精度均在90%以上, Kappa系数均值为0.93, 为进一步提取竹林信息奠定了重要的数据基础。

表 2 林地提取精度评价

Table 2 Accuracy assessment of forest information extraction

时间/年	检验样本	分类数	正确分类数	生产者精度/%	用户精度/%	Kappa系数
2003	401	382	368	96.34	91.77	0.91
2008	401	388	363	93.56	90.52	0.90
2014	401	397	388	97.73	96.76	0.97

#### 3.2 基于决策树模型的竹林分布信息

在林地信息提取的基础上, 基于MNF变换后的15个MODIS NDVI(NDVI\_MNF1—9)和反射率波段(REF\_MNF1—6), 以及训练样本, 通过C5.0算法得到具有5个终极节点的决策树模型, 如图3所示。由图3可见, 通过B1—B4等4个变量可以将竹林分布信息从林地中提取出来。因MODIS数据获取时间不同及森林信息的动态变化, 使得2003年、2008年和2014年3个时期的B1—B4有所不同, 如表3所示。

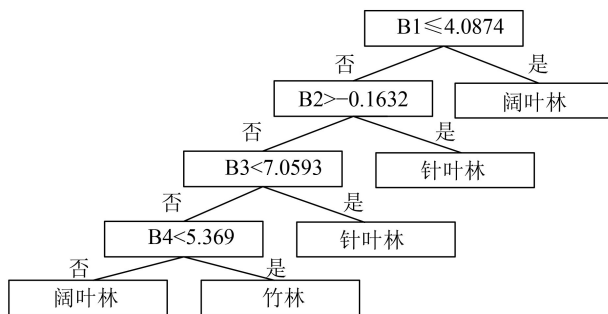


图 3 基于C5.0算法的最优决策树

Fig. 3 The optimal decision tree generated by C5.0 algorithm

表 3 2003年—2014年C5.0决策树特征变量

Table 3 Characteristic variables of decision tree during 2003—2014

时间/年	B1	B2	B3	B4
2003	NDVI_MNF2	REF_MNF3	NDVI_MNF2	REF_MNF4
2008	NDVI_MNF2	NDVI_MNF3	NDVI_MNF1	REF_MNF4
2014	REF_MNF1	NDVI_MNF2	REF_MNF4	REF_MNF1

表4为竹林分布信息提取精度。由表4可以看出, 3个不同时期中国竹林分类的生产者精度分别为79.23%、79.84%和81.16%, 说明在林地信息提

取的基础上, 基于决策树模型提取的竹林遥感信息精度较好, 平均精度为80.07%, 能够反映中国竹林时空分布特征。

表 4 决策树竹林分类精度评价

Table 4 Accuracy assessment of bamboo forest classification using decision tree

时间/年	竹林分类数	竹林	非竹林	生产者精度/%
2003	387	309	78	79.84
2008	414	328	86	79.23
2014	536	435	101	81.16

由决策树模型得到的2003年、2008年和2014年3个时期中国竹林空间分布如图4所示。由图4可见, 所提取的竹林信息主要分布在浙江、福建、江西、湖南、安徽、湖北、广西、广东、四川、重庆、云南、海南、台湾等南方省、市, 与中国竹林的实际分布范围吻合, 而且3个时期竹林的分布在空间具有较好的一致性, 即各个时期竹林在不同省份的空间分布基本相同, 如浙江省竹林主要分布在浙江西北和南部、安徽省竹林主要分布在东南等。当然, 由图4也可以发现, 部分省份的竹林分布在不同时期存在一定的差异。

#### 3.3 基于混合像元分解的全国竹林面积信息

2003年、2008年和2014年3个时期竹林端元选择如图5所示。由图5(a)可见, 2003年较好地提取竹林纯净端元的两个主成分分别为REF\_MNF1和REF\_MNF3, 2008为REF\_MNF1、REF\_MNF2, 2014年REF\_MNF1和REF\_MNF4。由图5(b)可见, 所选取的纯净像元均落在竹林分布区域内, 说明通过MNF变换提取的竹林端元能够代表竹林光谱

特征。3个不同时期纯净竹林像元的平均光谱如图5(c)所示，2008年与2014年纯净像元的光谱特征

较为接近，2003年纯净像元的反射率略大于其他两期。

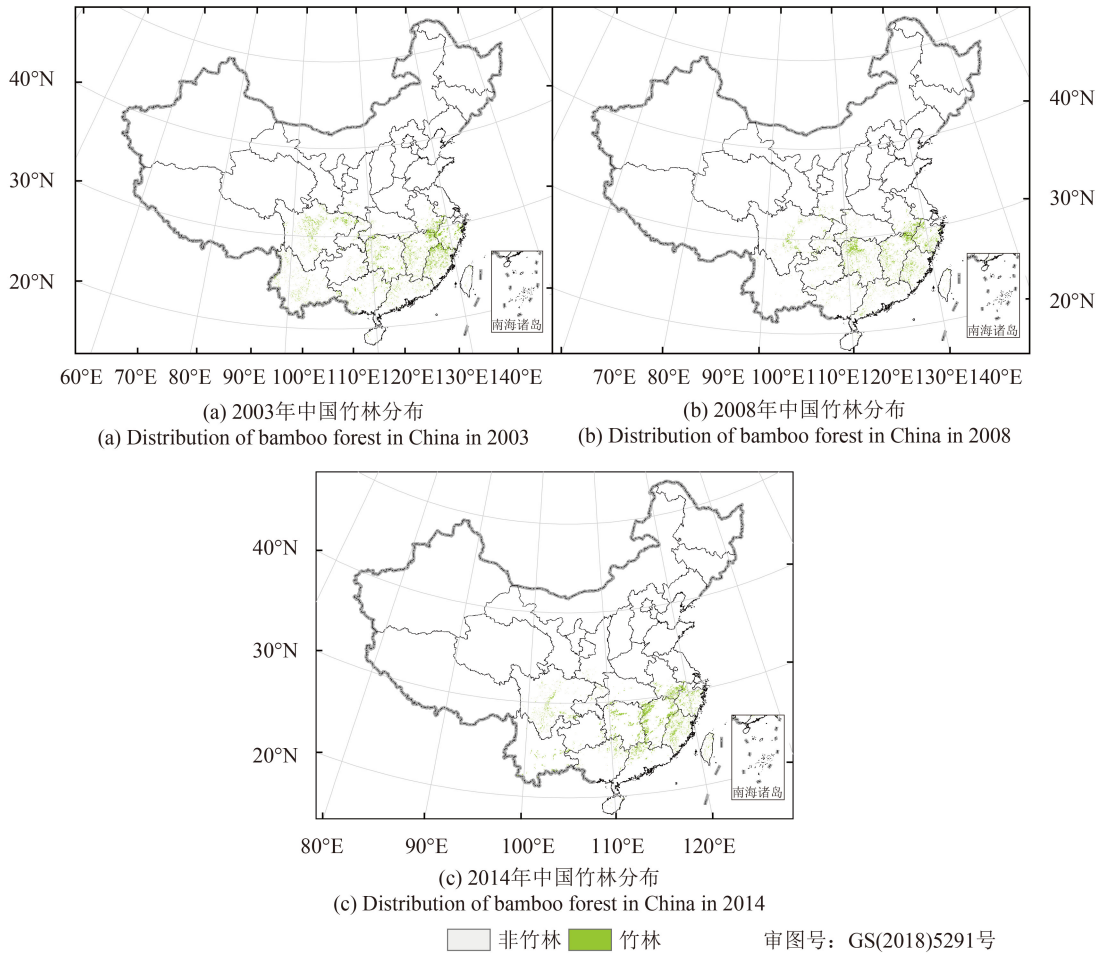


图4 2003年、2008年和2014年中国竹林空间分布

Fig. 4 Distribution of bamboo forest in China in 2003, 2008, 2014

由竹林纯净像元，利用式(1)—(2)对竹林分布(图3)进行混合像元分解，得到全国竹林丰度及RMSE空间分布如图6所示。从图6中可以看到，竹林丰度与分布具有一定的关系，即竹林集中分布的区域其像元丰度普遍较大，反之则丰度也普遍较小。如浙江西北部、江西与湖南交界处、福建西南部等竹林集中分布区丰度相对较大，而其他省份竹林分布非常零散的区域则丰度也相对较小。从各时期的RMSE分布图可以看出，除2014年分布于四川、云南的个别像元大于0.6外，总体上RMSE较低，均在0.3以下。说明通过混合像元分解精度较高，为精确估算竹林面积提供了保障。

基于丰度结果图，统计得到3个时期不同省市竹林面积及其与竹林资源清查面积的对比分别如表5和图7所示。从图7中可以看出，2003年、

2008年和2014年3个不同时期遥感估算面积与森林清查面积之间均具有显著的相关关系， $R^2$ 分别为0.98、0.97和0.95，说明通过决策树模型结合混合像元分解提取的中国竹林面积精度较为理想。

然而，不同时期竹林面积估算还存在一定误差，但RMSE较小，在范围在3.72万—9.35万ha变化。不同时期各省估算竹林面积与调查面积之差(图8)，可以进一步分析竹林面积高估/低估的来源。从图8可以看出，除个别省份外，各时期竹林估算与调查面积之差的波动范围均较小。其中，2003年除安徽、广西、四川、台湾存在一定的高估或低估外，其他省份竹林估算面积与实际面积的差异均在误差变化以内，因此，二者之间的RMSE较小，仅为3.72万ha；2008主要因为江西省竹林面积高估，而到导致RMSE提高到5.73，其他

省市估算面积较好；福建竹林面积低估，江西、广东竹林面积高估，尤其是江西竹林面积高估

32%左右，是造成2014年竹林估算面积RMSE最大的原因。

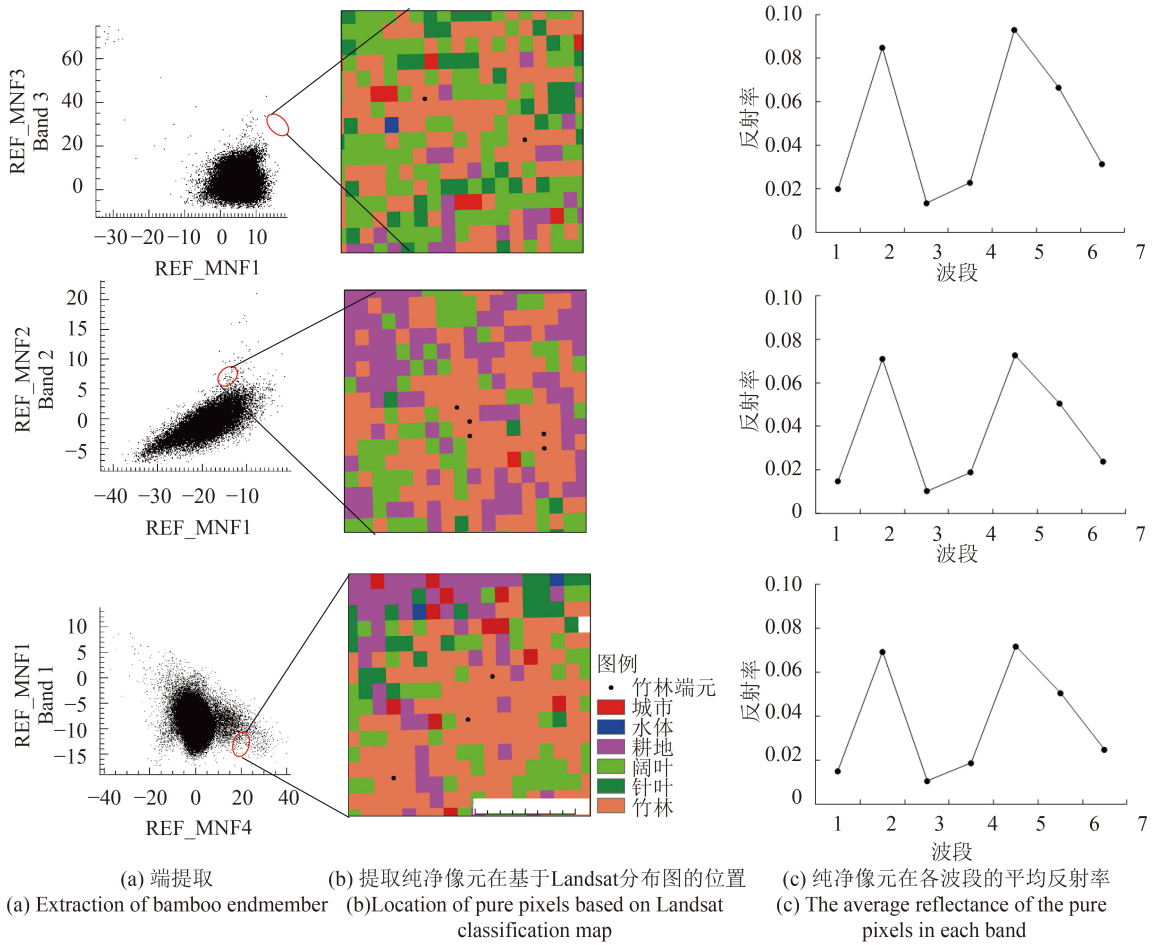
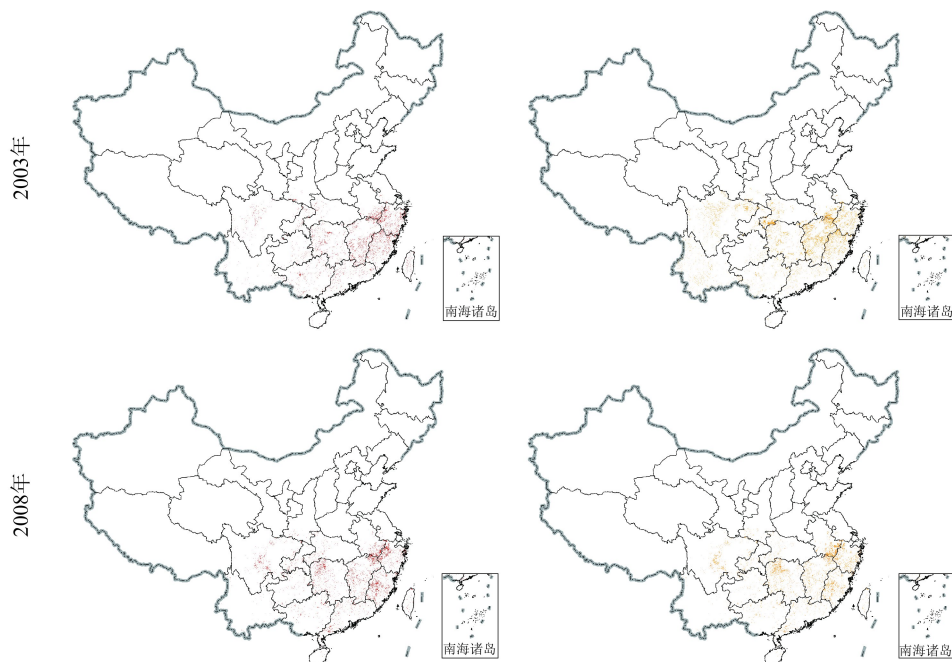


图5 2003年、2008年和2014年竹林端元提取、提取纯净像元在基于Landsat分布图的位置及纯净像元在各波段的平均反射率  
 Fig. 5 Extraction of bamboo endmember, location of pure pixels based on Landsat classification map and the average reflectance of the pure pixels in each band in 2003, 2008 and 2014



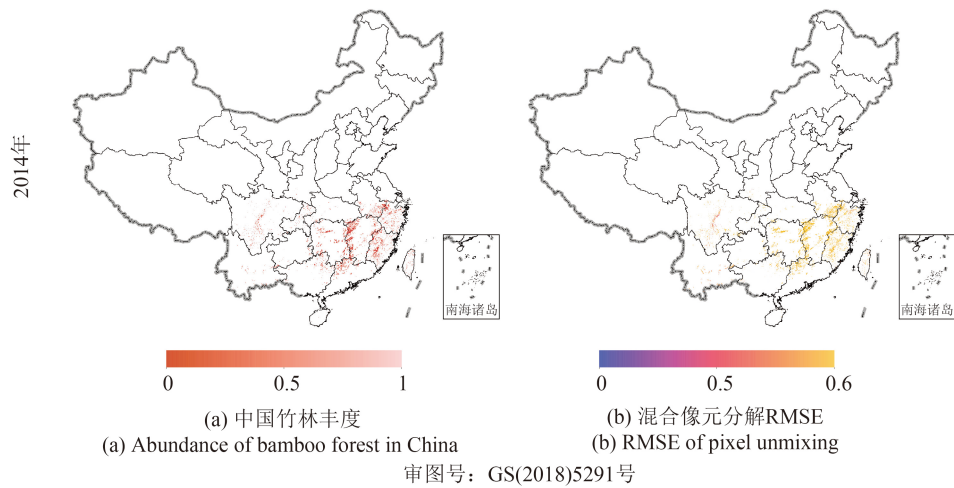


图 6 2003年—2014年中国竹林丰度及RMSE空间分布

Fig. 6 Distribution of abundance and RMSE of bamboo forest in China during 2003—2014

表 5 中国竹林遥感估算面积与调查面积对比

Table 5 Comparison of estimated and observed bamboo forest area in China

地名	2003年		2008年		2014年	
	调查面积	估算面积	调查面积	估算面积	调查面积	估算面积
安徽	26.98	35.04	32.28	35.44	33.72	38.62
重庆	11.71	12.55	12.51	13.64	13.63	14.09
福建	88.52	86.99	99.31	95.61	106.75	95.67
广东	37.42	34.55	40.78	40.25	44.62	56.07
广西	30.74	34.50	29.77	21.95	34.09	35.72
贵州	9.61	11.64	13.14	11.86	15.69	14.11
海南	1.80	1.49	1.56	1.10	1.56	1.02
湖北	13.76	13.58	15.04	14.56	14.40	14.93
湖南	52.20	53.87	62.78	66.08	77.83	80.43
江苏	3.73	4.12	3.61	2.67	3.37	1.87
江西	80.66	78.21	85.16	103.90	99.89	132.61
陕西	4.16	5.58	3.52	3.95	2.56	1.10
四川	37.38	29.48	48.60	43.52	54.90	49.89
台湾	13.26	7.93	13.26	5.66	13.26	10.5
云南	8.64	6.97	9.12	8.58	11.04	10.10
浙江	74.75	70.06	78.29	72.25	83.34	82.50
合计	495.32	486.56	548.73	545.14	610.65	639.22

注：台湾省竹林面积来自台湾第四森林资源调查(2009年—2013年)数据，其他省份2003年、2008年和2014年竹林面积分别来自国家森林资源清查第6次(1999年—2003年)、第7次(2004年—2008年)、第8次(2009年—2013年)清查数据。

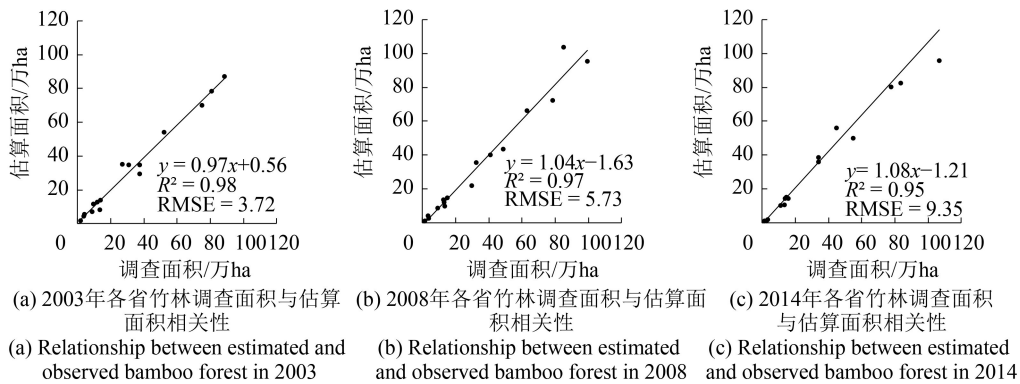


图 7 3期各省竹林调查面积与估算面积相关性

Fig. 7 Relationship between estimated and observed bamboo forest



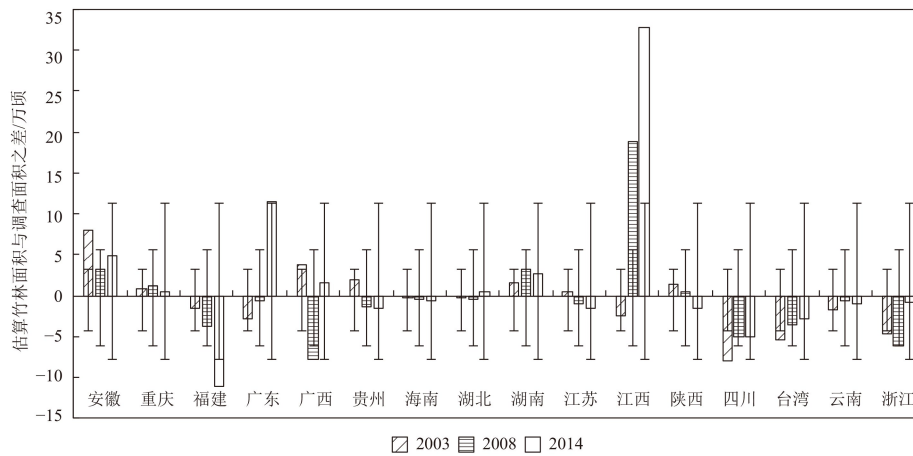


图8 各时期估算竹林面积与实际竹林面积之差及其标准差

Fig. 8 Difference and standard deviation between estimated and observed bamboo forest area

## 4 结 论

本研究基于大尺度范围高精度提取竹林信息的实际需要,采用全国MODIS数据、局部Landsat数据及地面样地数据,创新提出构建决策树结合混合像元分解的中国竹林提取方法,实现了中国竹林空间分布及面积信息准确提取,主要得到以下两方面的结论:

(1) 在林地信息提取基础上,优化后的决策树模型能够准确提取全国竹林时空分布信息。2003年、2008年和2014年竹林生产者精度分别为79.23%、79.84%和81.16%,平均精度为80.07%,在一定程度上反映了中国竹林实际分布情况。

(2) 在决策树分类的基础上,引入线性最小二乘混合像元分解模型,实现了全国竹林面积信息的准确估算,竹林丰度RMSE总体上在0.6以下,估算竹林面积与森林资源清查面积具有较高的相关关系,平均 $R^2$ 为0.97,平均RMSE为6.27万ha。

与传统竹林遥感信息提取方法相比,本研究既考虑竹林光谱特征构建适合竹林信息提取的决策树模型,又兼顾粗分辨率遥感信息提取需要解决混合像元的实际情况,因而实现了大尺度竹林时空分布和面积信息的高精度提取。当然个别省份的估算面积与调查面积仍有一定出入,如高估了江西省2008年和2014年竹林面积,这可能是MODIS影像数据质量、MNF变换及竹林样本分布不均所造成的。对此,我们将继续进行深入分析。尽管在时空分布信息和面积方面还存在一定误差,但本研究所提出的方法在一定程度上实现

了全国竹林时空分布信息的准确提取,为全国竹林资源信息动态监测及管理提供了技术手段和数据支撑。

## 参考文献(References)

- Adams J B, Sabol D E, Kapos V, Almeida Filho R, Roberts D A, Smith M O and Gillespie A R. 1995. Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52(2): 137–154 [DOI: [10.1016/0034-4257\(94\)00098-8](https://doi.org/10.1016/0034-4257(94)00098-8)]
- Chao C M, Yu Y W, Cheng B W and Kuo Y L. 2014. Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree. *Journal of Medical Systems*, 38(10): 106 [DOI: [10.1007/s10916-014-0106-1](https://doi.org/10.1007/s10916-014-0106-1)]
- Du H Q, Zhou G M, Ge H L, Zhao X W and Cui L L. 2008. Bamboo information extraction based on Landsat TM data. *Journal of Northeast Forestry University*, 36(3): 35–38 (杜华强, 周国模, 葛宏立, 赵宪文, 崔林丽. 2008. 基于TM数据提取竹林遥感信息的方法. *东北林业大学学报*, 36(3): 35–38) [DOI: [10.3969/j.issn.1000-5382.2008.03.012](https://doi.org/10.3969/j.issn.1000-5382.2008.03.012)]
- Duan H J, Deng Z D, Deng F F and Wang D Q. 2016. Assessment of groundwater potential based on multicriteria decision making model and decision tree algorithms. *Mathematical Problems in Engineering*, 2016: 2064575 [DOI: [10.1155/2016/2064575](https://doi.org/10.1155/2016/2064575)]
- Fan W L, Du H Q, Zhou G M, Xu X J, Cui R R and Dong D J. 2010. Spectral mixture analysis method based on the simulation of real scenario. *Journal of Remote Sensing*, 14(6): 1250–1265 (范渭亮, 杜华强, 周国模, 徐小军, 崔瑞蕊, 董德进. 2010. 模拟真实场景的混合像元分解. *遥感学报*, 14(6): 1250–1265) [DOI: [10.11834/jrs.20100613](https://doi.org/10.11834/jrs.20100613)]
- Gao G L, Du H Q, Han N, Xu X J, Sun S B and Li X J. 2016. Mapping

- of Moso bamboo forest using object-based approach based on the optimal features. *Scientia Silvae Sinicae*, 52(9): 77–85 (高国龙, 杜华强, 韩凝, 徐小军, 孙少波, 李雪建. 2016. 基于特征优选的面向对象毛竹林分布信息提取. *林业科学*, 52(9): 77–85) [DOI: 10.11707/j.1001-7488.20160909]
- Griffiths P, Kuemmerle T, Baumann M, Radeloff V C, Abrudan I V, Lieskovsky J, Munteanu C, Ostapowicz K and Hostert P. 2014. Forest disturbances, forest recovery, and changes in forest types across the Carpathian ecoregion from 1985 to 2010 based on Landsat image composites. *Remote Sensing of Environment*, 151: 72–88 [DOI: 10.1016/j.rse.2013.04.022]
- Han N, Du H Q, Zhou G M, Sun X Y, Ge H L and Xu X J. 2014. Object-based classification using SPOT-5 imagery for Moso bamboo forest mapping. *International Journal of Remote Sensing*, 35(3): 1126–1142 [DOI: 10.1080/01431161.2013.875634]
- Han N, Du H Q, Zhou G M, Xu X J, Ge H L, Liu L J, Gao G L and Sun S B. 2015. Exploring the synergistic use of multi-scale image object metrics for land-use/land-cover mapping using an object-based approach. *International Journal of Remote Sensing*, 36(13): 3544–3562 [DOI: 10.1080/01431161.2015.1065357]
- Klein I, Gessner U and Kuenzer C. 2012. Regional land cover mapping and change detection in Central Asia using MODIS time-series. *Applied Geography*, 35(1/2): 219–234 [DOI: 10.1016/j.apgeog.2012.06.016]
- Li J X, Da L J, Wang Y J and Song Y C. 2005. Vegetation classification of east China using multi-temporal NOAA-AVHRR data. *Acta Phytocologica Sinica*, 29(3): 436–443 (李俊祥, 达良俊, 王玉洁, 宋永昌. 2005. 基于NOAA-AVHRR数据的中国东部地区植被遥感分类研究. *植物生态学报*, 29(3): 436–443) [DOI: 10.17521/cjpe.2005.0058]
- Li M N, Li C C, Jiang H, Fang C Y, Yang J, Zhu Z L, Shi L, Liu S R and Gong P. 2016. Tracking bamboo dynamics in Zhejiang, China, using time-series of Landsat data from 1990 to 2014. *International Journal of Remote Sensing*, 37(7): 1714–1729 [DOI: 10.1080/01431161.2016.1165885]
- Li M S, Peng S K, Zhou L and Ma Y X. 2006. A study of automated construction and classification of decision tree classifiers based on aster remotely sensed datasets. *Remote Sensing for Land and Resources*, 18(3): 33–36 (李明诗, 彭世揆, 周林, 马以秀. 2006. 基于ASTER数据的决策树自动构建及分类研究. *国土资源遥感*, 18(3): 33–36) [DOI: 10.6046/gtzyyg.2006.03.08]
- Lu M, Pebesma E, Sanchez A and Verbesselt J. 2016. Spatio-temporal change detection from multidimensional arrays: detecting deforestation from MODIS time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117: 227–236 [DOI: 10.1016/j.isprsjprs.2016.03.007]
- Qiao Y W, Zang S Y and Na X D. 2013. The information extraction of freshwater marsh wetland based on the decision tree method: taking Zhalong wetland as an example. *Chinese Agricultural Science Bulletin*, 29(8): 169–174 (乔艳雯, 臧淑英, 那晓东. 2013. 基于决策树方法的淡水沼泽湿地信息提取——以扎龙湿地为例. *中国农学通报*, 29(8): 169–174) [DOI: 10.3969/j.issn.1000-6850.2013.08.032]
- Quinlan J R. 1993. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Shang Z Z, Zhou G M, Du H Q, Xu X J, Shi Y J, Lü Y L, Zhou Y F and Gu C Y. 2013. Moso bamboo forest extraction and above-ground carbon storage estimation based on multi-source remotely sensed images. *International Journal of Remote Sensing*, 34(15): 5351–5368 [DOI: 10.1080/01431161.2013.788260]
- Stagakis S, Vanikiotis T and Sykioti O. 2016. Estimating forest species abundance through linear unmixing of CHRIS/PROBA imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119: 79–89 [DOI: 10.1016/j.isprsjprs.2016.05.013]
- Sun S B. 2016. Retrieval of Canopy Parameters and Estimation of Aboveground Carbon Storage for Typical Types of Subtropical Forest Using Remote Sensing data. Hangzhou: Zhejiang A&F University (孙少波. 2016. 亚热带典型森林冠层参数及碳储量定量估算研究. 杭州: 浙江农林大学)
- Sun X Y. 2014. Object-Based Extraction of Moso Bamboo Forest and Estimation of Moso Bamboo Inventory Factors. Hangzhou: Zhejiang A&F University (孙晓艳. 2014. 面向对象的毛竹林分布遥感信息提取及调查因子估算. 杭州: 浙江农林大学)
- Sun X Y, Du H Q, Han N, Ge H L and Gu C Y. 2013. Multi-scale segmentation, object-based extraction of moso bamboo forest from SPOT5 imagery. *Scientia Silvae Sinicae*, 49(10): 80–87 (孙晓艳, 杜华强, 韩凝, 葛宏立, 谷成燕. 2013. 面向对象多尺度分割的SPOT 5影像毛竹林专题信息提取. *林业科学*, 49(10): 80–87) [DOI: 10.11707/j.1001-7488.20131013]
- Ture M, Tokatli F and Kurt I. 2009. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications*, 36(2): 2017–2026 [DOI: 10.1016/j.eswa.2007.12.002]
- Wang C, Du H Q, Zhou G M, Xu X J, Sun S B and Gao G L. 2015. Retrieval of crown closure of moso bamboo forest using unmanned aerial vehicle (UAV) remotely sensed imagery based on geometric-optical model. *Chinese Journal of Applied Ecology*, 26(5): 1501–1509 (王聪, 杜华强, 周国模, 徐小军, 孙少波, 高国龙. 2015. 基于几何光学模型的毛竹林郁闭度无人机遥感定量反演. *应用生态学报*, 26(5): 1501–1509) [DOI: 10.13287/j.1001-9332.20150302.001]
- Xu X J, Du H Q, Zhou G M, Dong D J, Fan W L and Cui R R. 2011. Spectral mixture analysis based on Erf-BP model and applied in extracting forest information. *Scientia Silvae Sinicae*, 47(2): 30–38 (徐小军, 杜华强, 周国模, 董德进, 范渭亮, 崔瑞蕊. 2011. Erf-BP混合像元分解及在森林遥感信息提取中应用. *林业科学*, 47(2): 30–38) [DOI: 10.11707/j.1001-7488.20110205]

- Yan M C, Zhang Y J and Bao Y S. 2004. Deriving bamboos from IKONOS image by texture information. *Remote Sensing Information*, 15(2): 31–34 (颜梅春, 张友静, 鲍艳松. 2004. 基于灰度共生矩阵法的IKONOS影像中竹林信息提取. *遥感信息*, 15(2): 31–34) [DOI: [10.3969/j.issn.1000-3177.2004.02.009](https://doi.org/10.3969/j.issn.1000-3177.2004.02.009)]
- Yang G, Shen H F, Zhang L P, He Z Y and Li X H. 2015. A moving weighted harmonic analysis method for reconstructing high-quality SPOT vegetation NDVI time-series data. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11): 6008–6021 [DOI: [10.1109/TGRS.2015.2431315](https://doi.org/10.1109/TGRS.2015.2431315)]
- Zhao Y S. 2013. *Analysis Principle and Method of Remote Sensing Applications*. 2nd ed. Beijing: Science Press (赵英时. 2013. *遥感应用分析原理与方法*. 2版. 北京: 科学出版社)
- Zhou G M, Jiang P K, Du H Q and Shi Y J. 2017. *Technology for the Measurement and Enhancement of Carbon Sinks in Bamboo Forest Ecosystems*. Beijing: Science Press (周国模, 姜培坤, 杜华强, 施拥军. 2017. *竹林生态系统碳汇计测与增汇技术*. 北京: 科学出版社)

## Combination of decision tree and mixed pixel decomposition for extracting bamboo forest information in China

CUI Lu, DU Huaqiang, ZHOU Guomo, LI Xuejian, MAO Fangjie, XU Xiaojun, FAN Weiliang, LI Yangguang, ZHU Dien, LIU Tengyan, XING Luqi

1. State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Lin'an 311300, China;
2. Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, Zhejiang A & F University, Lin'an 311300, China;
3. School of Environmental and Resources Science, Zhejiang A & F University, Lin'an 311300, China

**Abstract:** A bamboo forest is a special and important forest resource that is distributed in subtropics of China. However, present methods cannot quickly and accurately extract spatiotemporal distribution information of a bamboo forest at the national scale.

This study presents a method for extracting bamboo forest information in China by combining decision tree and linear spectral unmixing methods using MODIS NDVI, reflectance product, and provincial Landsat classification data in 2003, 2008, and 2014. The steps are as follows: First, forest information was extracted based on the maximum likelihood method; second, a decision tree model was constructed to extract bamboo forest distribution information based on a forest information map; finally, the linear spectral unmixing method was applied to obtain the abundance map of a bamboo forest in China, and the bamboo forest area was calculated.

The main experimental results are as follows: (1) The forest information in China was extracted through the maximum likelihood method. The accuracy of a producer and a user was higher than 90%, and the Kappa coefficient was 0.93, thereby establishing the extraction of bamboo forest information. (2) The decision tree constructed by using a C5.0 algorithm satisfactorily extracted the bamboo forest spatiotemporal distribution information in China, with an average classification accuracy of 80%. (3) The estimated bamboo forest area of each province in China was highly correlated with the observations, with the  $R^2$  values of 0.98, 0.97, and 0.95, correspondingly. The RMSE ranged from  $3.92 \times 10^4$  to  $9.58 \times 10^4$  ha, thus indicating that the estimated bamboo forest area was essentially consistent with the actual situation.

In this study, the C5.0 algorithm decision tree, which is based on the MODIS remote sensing data, combined with the mixed pixel decomposition presents accurate extraction of bamboo forest spatiotemporal distribution information in China. This method provides the technical approach and data support for the dynamic monitoring and management of bamboo forest resources.

**Key words:** Chinese bamboo forest, MODIS data, C5.0 algorithm, decision tree classification, mixed pixel decomposition

**Supported by** National Natural Science Foundation of China (No. 31670644, 31370637, 31500520); National Natural Science Foundation of Zhejiang Province, China (No. LR14C160001, LQ15C160003); Joint Research Fund of Department of Forestry of Zhejiang Province and Chinese Academy of Forestry (No. 2017SY04); Zhejiang Provincial Collaborative Innovation Center for Bamboo Resources and High-efficiency Utilization