

引用格式: Zhao Shuai, Cao Meiqin, Jiang Xiandie, *et al.* Man-made Tree Species Classification in Lixin County, Anhui Province[J]. Remote Sensing Technology and Application, 2022, 37(3): 589-598. [赵帅, 曹美芹, 蒋先蝶, 等. 安徽省利辛县平原区人工林树种分类研究[J]. 遥感技术与应用, 2022, 37(3): 589-598.]
DOI: 10.11873/j.issn.1004-0323.2022.3.0589

安徽省利辛县平原区人工林树种分类研究

赵 帅^{1,2}, 曹美芹^{1,2}, 蒋先蝶^{1,2}, 陈耀亮^{1,2}, 陆灯盛^{1,2}

(1. 福建师范大学 湿润亚热带山地生态国家重点实验室培育基地, 福建 福州 350007;

2. 福建师范大学地理研究所, 福建 福州 350007)

摘要:快速准确地绘制平原区人工林树种分布对研究人工林的生态水文和社会经济效益具有重要的意义。将资源3号(ZY-3)全色波段分别同ZY-3多光谱、哨兵2号多光谱进行融合,在图像分割基础上提取变量,采用分层优化变量组合的随机森林分类方法对安徽省利辛县人工林树种进行分类,并与分类回归树和随机森林相比较。结果表明:利用分层分类方法,平原区的人工林树种分类精度可以达到92%以上;基于哨兵2号和ZY-3融合的光谱特征变量分类精度比ZY-3数据本身的融合提高了2.49%~2.91%;而分别加入纹理变量后,分层分类方法大幅度提高了树种分类精度。因此,基于高分辨率遥感数据的光谱和纹理特征,采用分层分类方法,可以有效提高平原区人工林树种的分类精度。

关键词: 分层分类; 特征优选; 平原区; 资源3号; 哨兵2号

中图分类号: TP79 **文献标志码:** A **文章编号:** 1004-0323(2022)03-0589-10

1 引言

近年来,为推动经济和生态建设协调统一发展,全国各地大力开展人工造林运动,平原区作为我国农业、工业发展的主要集聚地,是人工造林运动的主要发生地之一。平原区人工林建设在防风固沙、水土保持和保障农作物安全等方面发挥着重要作用,而准确、快速地绘制平原区人工林树种的分布对于评价平原区造林成果、监测林木成活率以及为其它相关研究提供基础数据等方面具有重要意义。平原区人工林具有成片/成带的分布规律,又有零星破碎的分布现象。针对平原区树种造林布局分散、种植制度灵活以及林地面积大小不一的空间分布特点,采用传统森林资源调查手段实现其空间分布的准确提取耗时费力^[1-2]。

遥感以成本低、效率高、时效性强的优势成为森林树种分类的主要数据源^[3-4],特别是高分辨率遥

感影像以其空间信息丰富的优势,近年来得到广泛的应用^[5-6]。由于同一森林树种类型的光谱存在高度异质性,传统的基于像元分类方法常导致分类精度低^[7]。面向对象分类方法以影像分割产生的对象作为分类单元,避免了分类结果的“椒盐效应”,已经广泛应用于高空间分辨率的地表覆盖分类研究^[8-9]。王二丽等^[2]利用高分一号全色和4个多光谱波段的融合数据,基于面向对象的多时相光谱和纹理特征对平原造林区进行树种分类,其应用KNN方法获得了最高64%的总体分类精度。李丹等^[1]基于WorldView-2卫星全色和8个多光谱波段的融合数据,基于面向对象的光谱和纹理特征对城市典型乔木树种进行分类,并获得了最高86.5%的总体分类精度。增加遥感数据的光谱波段数和提高空间分辨率在一定程度上改善人工林树种的分类精度。

收稿日期:2021-04-18;修订日期:2022-03-08

基金项目:国家重点研发计划项目“人工林资源监测关键技术研究”(2017YFD0600900)。

作者简介:赵 帅(1996-),男,河南郑州人,硕士研究生,主要从事地表覆盖分类研究。E-mail: ygork123@163.com

通讯作者:陆灯盛(1965-),男,浙江东阳人,教授,主要从事地表覆盖分类、森林生物量估测研究。E-mail: ludengsheng@fjnu.edu.cn

随机森林(Random Forest)作为一种集成学习方法,已广泛地应用于森林树种的分类^[10]。由于随机森林建立的暗箱模型使得每个决策树都是不充分的弱学习器,且只能在分类前进行一次变量的重要性排序,无法确定每个节点的关键变量,难以有效的提高分类精度^[10]。针对这一问题,Chen等^[11]提出一种基于节点变量优化和阈值循环的分类方法,在每一个节点选取最佳变量组合用于树种精细分类,可视化调节关键变量和阈值,在北方林区的人工林树种分类取得了较好的分类精度,但该分类方法在每个节点的多个规则之间应用简单的交运算,对于人为干扰严重、景观破碎程度高的平原区人工林树种,应用简单的交运算不能得到很好的分类结果。因此,本研究针对平原区人工林提出了一种基于随机森林算法的分层分类方法,并以安徽省利辛县为例,利用多源高分辨率遥感数据,在面向对象多尺度分割的基础上,基于分层优化节点变量的随机森林分类算法,对平原区的人工林树种进行分层分类提取,探索该分层分类方法在平原区的应用潜力,以期为平原区人工林树种资源管理与可持续性经营提供科学依据。

2 研究区概况

研究区位于黄淮平原南部、安徽省西北部的利辛县(图1)。该研究区属于暖温带半湿润季风气候区,地势平坦,平均海拔28 m,境内河渠纵横汇入淮河^[12]。自2000年以来,在长江防护林等国家林业项目的带动下,研究区大力发展人工林种植,尤以杨树速生丰产林为主,杨树面积不断扩大,资源总量迅速增加^[13]。但是,由于杨树树种单一、生长速度快导致大量杨絮集中发生。近几年开展杨絮治理工作,开始大规模片状栽种苗木,将成熟杨树逐步更新改造为桃树、梧桐、广玉兰、女贞、石楠、榉树、女贞、朴树等景观树种。

3 研究方法

图2描述了平原区人工林树种分类的技术路线,主要包括4个部分:①数据准备,包括数据收集与预处理;②数据处理,包括全色与多光谱数据融合、图像分割、变量提取和数据组合方案设计;③应用不同的分类方法,包括分类回归树、随机森林、和分层随机分类进行人工林树种分类;④分类产品输出及精度验证分析。

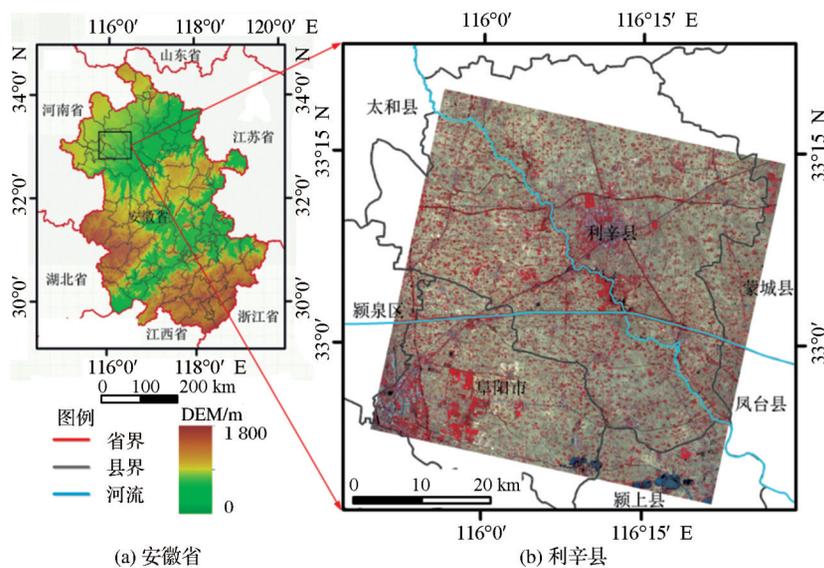


图1 研究区位置(资源三号多光谱标准假彩色影像)

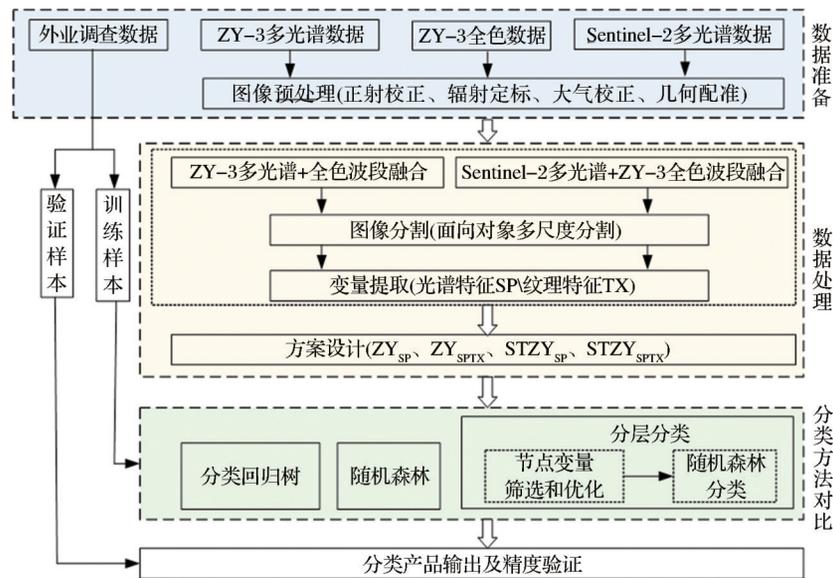
Fig. 1 Location of study area (false color composite of ZY-3 multispectral data)

3.1 数据收集与预处理

3.1.1 外业调查数据的收集和处理

以研究区2019年林地一张图为参考,于2020年7月进行了野外实地地表覆盖类型调查,详细记录了采样点的地理坐标和土地覆盖类型,主要包括杨树、其他阔叶、桃园、苗圃、在耕农田(农作物以玉

米为主)、闲置农田、城镇用地和水体等。结合Google Earth对调查数据进行组织整理,最后保留1 613个地类样本(表1)。按照7:3的比例随机把每种地类样本分为训练样本和验证样本。按照上述8种地类进行初始分类,最后把非人工林地类归并为一类,产生最终的分类系统:杨树、其他阔叶、桃园、



(ZY_{SP}和ZY_{SPTX}分别为ZY-3多光谱和全色波段融合后的光谱变量组合、光谱和纹理特征变量组合;STZY_{SP}和STZY_{SPTX}分别为Sentinel-2A多光谱和ZY-3全色波段融合后的光谱变量组合、光谱和纹理特征变量组合)

图 2 利辛县土地覆盖/人工林树种分类技术路线图

Fig.2 Framework of land cover / tree species classification based on ZY-3 and Sentinel-2 data

苗圃和其他地类。

表 1 不同地表覆盖类型样本数量

Table 1 Number of sample plots for different land cover types

名称	杨树	其他阔叶	桃园	苗圃	在耕农田	闲置农田	城镇用地	水体	共计
样本数量	391	145	150	207	188	168	224	140	1 613

3.1.2 遥感数据的收集和处理

研究选用了2019年6月14日获取的资源3号(ZY-3)影像,包括空间分辨率2.1 m的全色波段和空间分辨率5.8 m的多光谱波段(蓝、绿、红和近红外)。在对影像进行正射校正基础上,利用绝对辐射定标系数将多光谱图像的数字量化值(DN)转换为辐射亮度,然后用暗黑像元法进行大气校正^[14],并采用最邻近插值法将全色和多光谱波段分别重采样至2 m和6 m。

哨兵2号数据(Sentinel-2)的获取日期与资源3号相同,其空间分辨率为10~20 m,包括4个10 m的多光谱波段(蓝、绿、红和近红外)和6个20 m的多光谱波段(3个红边波段、一个窄近红外和两个短波红外)。利用Sen2cor软件对其进行辐射定标和大气校正^[15],使用SNAP软件将10个波段统一重采样为10 m;然后以ZY-3全色影像为基准,对哨兵2号多光谱影像进行几何配准。

3.2 在图像融合和分割基础上的变量提取

以往研究发现Gram-Schmidt(GS)融合方法可以较好地改善原始影像的空间细节特征,且能最大限度地保持原始影像的光谱特性^[16]。因此,本研究使用GS融合算法将已配准的ZY-3多光谱(6 m)和全色(2 m)波段融合成2 m的4个波段新影像,标记为ZY;将Sentinel-2多光谱(10 m)和ZY-3全色(2 m)波段融合成2 m的10个波段新影像,标记为STZY。

采用eCognition软件中的多尺度分割算法,根据地物类型在遥感影像上的光谱和纹理等信息,利用多尺度参数评价工具确定最优分割尺度^[17]。其中,ZY和STZY数据方案分别通过参数的不断优化调整,最终选定分割尺度为100,形状和颜色因子分别设置为0.2和0.8,平滑度和紧致度因子均设置为0.5,最后将影像像素自下而上合并成互不相交、高度同质性斑块的对象。

基于遥感数据提取适宜的变量组合是森林树种分类的关键之一。常用的变量包括光谱变量和纹理变量。光谱变量通常利用光谱波段数据和各种植被指数,而纹理变量常用灰度共生矩阵方法^[18]。本研究以分割单元为基础计算光谱变量的平均值和光谱波段相应的纹理变量,所选变量如表2所示。依据使用的变量,设计了4种数据方案:ZY_{SP}和STZY_{SP}只包括光谱变量ZY_{SPTX}和STZY_{SPTX}则包括光谱变量和纹理变量的组合(表3)。

表2 变量汇总表

Table 2 Summary of extracted variables from fused ZY-3 images and fused Sentinel-2 with ZY-3 images

数据	光谱变量		纹理变量
	波段	植被指数	
ZY-3 多光谱和全色波段融合(ZY)	蓝, 绿, 红, 近红外	归一化植被指数(NDVI): (近红外-红波段)/(近红外+红波段)	均值、标准差、
Sentinel-2 多光谱和ZY-3 全色波段融合(STZY)	蓝, 绿, 红, 近红外,	归一化水体指数(NDWI): (绿波段-近红外)/(绿波段+近红外)	同质性、对比度、
	红边 1~3, 窄近红外, 短波红外 1 和 2	可见光波段差异植被指数(VDVI): (2 * 绿波段 - 红波段 - 蓝波段)/(2 * 绿波段 + 红波段 + 蓝波段)	异质性、熵、 角二阶矩、相关性

表3 数据方案设计

Table 3 Data schemes based on variable combinations for classification

数据方案	变量组合		
ZY _{SP}	4个光谱波段	3个植被指数	
STZY _{SP}	10个光谱波段	3个植被指数	
ZY _{SPTX}	4个光谱波段	3个植被指数	32个纹理变量
STZY _{SPTX}	10个光谱波段	3个植被指数	80个纹理变量

3.3 平原区树种分类及精度验证

分层优化变量组合的随机森林分类方法采用“自上而下,先易后难,先总后分”的分层分类策略,在不同分层中灵活引入不同的变量组合实现最佳的分类结果,降低数据冗余及去除不相关的变量,从而提高计算效率^[11]。

与随机森林只在分类前进行一次变量的重要性排序相比,分层分类方法通过对每个节点均进行一次变量重要性排序,得到每个节点的最佳变量组合^[19]。具体流程如下:利用随机森林的平均下降精度算法评估变量的重要性,对所有变量进行相关分析,当两个变量的相关性系数超过0.9时,重要性等级较低的变量将被删除,最终在每个节点获得不同重要性等级的至多5个变量。

根据地类的光谱特征和纹理特征,通过不断试验,构建图3所示的分层分类树。其中节点1包含所有地表覆盖类型;根据植被和非植被的光谱反应差异,可以将节点1分为植被(节点4)和非植被(节点2);节点2包含所有的非植被类型,利用水体与其他地类的光谱反应差异,从而把节点2分为水体(为叶子节点)和其他非植被地物类型(节点3),最后节点3被区分为闲置农田和城镇用地,两者均为叶子节点。另一侧,节点4包含所有的植被类型,根据纹理差异,可以把节点4区分为桃园与在耕农田(节点5),和苗圃与杨树与其他阔叶(节点6);利用

不同的变量组合依次将节点5分为桃园和在耕农田(两者均为叶子节点),将节点6分为苗圃和杨树与其他阔叶;最后节点7被分为杨树和其他阔叶。由于“其他阔叶节点”的子类型之间混淆严重及缺乏有代表性纯林样本,没有继续区分。

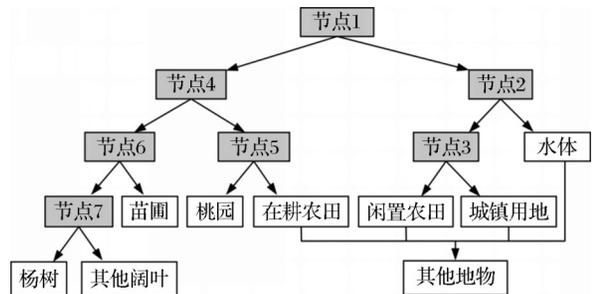


图3 分层分类树设计方案

Fig. 3 Design of hierarchical-based classifier for land cover classification

分层分类针对4种数据方案,确定在分层分类中所用的变量,最后完成所有节点变量优化后对整个研究区进行分类。因为本研究的关注重点为平原人工林,所以将在耕农田、闲置农田、城镇用地、水体归并为其他地类,最后的分类结果包括杨树、其它阔叶、苗圃、桃园和其他地物。

作为比较,选择分类回归树算法和随机森林算法对相同的数据方案进行分类。分类回归树是二叉树形式的决策树结构,每个节点表示区分不同类别的属性规则^[20-21]。该方法易于理解、运行速度快,但分类精度高度依赖训练样本质量,并且类别多时分类错误会快速向下逐渐累积,而当变量较少时易出现过拟合现象。随机森林是由多个决策树构成的集成学习方法,通过定义生长树的数目(ntree)和分裂每个节点上的变量数目(mtree)两个参数,利用自举重采样算法从原始数据集中有放回的抽样以构造子数据集,利用子数据集构建子决策树,其中每棵树进行独立分类运算得到各自的分类结果,再

根据每棵树分类结果的投票众数决定最终结果^[22]。该方法具有较强的容忍度和较低的泛化误差,但在选择最优变量集合时易受到其他关联特征的影响,使得进一步提高分类精度存在限制^[23]。

对分类结果,应用各个地类的验证样本,构建混淆矩阵,计算总体精度, Kappa系数,生产者精度和用户精度,其中总体精度和 Kappa系数用于评价整体分类精度,生产者精度和用户精度用于评价特定地类的分类精度。

4 结果与分析

4.1 节点变量优化结果分析

通过对4种数据方案各节点最终选择的变量对比分析(表4)可以看出:①方案 STZY_{SP}相比方案 ZY_{SP}增加了 B5~B10共6个光谱波段变量,对比两

者选择的变量集可以得出,方案 STZY_{SP}中增加的光谱波段变量在多个节点被选中,这表明方案 STZY_{SP}被选中的新增加光谱波段变量对分类结果是有效的;②方案 ZY_{SPTX}和方案 STZY_{SPTX}分别在 ZY_{SP}和方案 STZY_{SP}的基础上计算了相应的纹理变量,在前3个节点中,排名较为靠前的均为光谱和植被指数变量,表明仅利用光谱特征就可以成功实现植被与非植被,水体与非水体,闲置农田以及建筑用地的区分。在节点4到节点7中,重要性排名靠前的变量为 HomB4、HomB3、HomB8、AngB1,说明纹理变量在这些节点的表现优于光谱变量。这是因为在耕农田、桃园和苗圃通常有一定的空间分布规律。节点5在区分桃园和在耕农田时,方案 ZY_{SPTX}被选中的变量除 B1之外,其余4个变量均为纹理变量,表明桃园和在耕农田的纹理特征具有显著差异。

表4 不同方案节点变量选择

Table 4 The selected variables corresponding to each node for different data scenarios

序号	方案①:ZY _{SP}	方案②:STZY _{SP}	方案③:ZY _{SPTX}	方案④:STZY _{SPTX}
节点1	NDWI、B3、B4、VDVI	NDVI、VDVI、B1、B8、B10	B3、B4、VDVI、NDWI、CorB4	B1、B10、NDVI、VDVI、B8
节点2	B3、B4、VDVI、NDWI	B7、B10、VDVI、B1、B6	B1、NDWI、VDVI、B4、ConB22	B1、NDWI、HomB8、B8、CorB9
节点3	VDVI、NDWI、B1、B4	VDVI、B1、B9、B5、B10	NDWI、VDVI、B1、AngB3、EntB3	B1、B10、NDVI、VDVI、EntB10
节点4	NDVI、B1、B4、VDVI	B2、VDVI、B4、B7、B9	HomB4、MeanB4、NDWI、B1、CorB3	B4、DisB4、B1、DisB3、HomB4
节点5	B1、B4、VDVI、NDWI	B4、B7、VDVI、B6、B8	HomB3、B1、CorB3、AngB4、AngB3	HomB8、B1、B5、AngB4、StdB1
节点6	B1、VDVI、B4、NDWI	B1、B9、B6、VDVI、B8	AngB1、B2、NDVI、VDVI、B4	HomB8、B8、B1、VDVI、NDVI
节点7	B4、B3、VDVI、NDVI	B2、B5、B6、B9、VDVI	AngB1、NDVI、VDVI、B3、StdB3	HomB8、HomB6、HomB5、B5、B2

注:ZY(ZY-3融合)方案B1~B4分别为蓝、绿、红和近红外波段,STZY(Sentinel-2与ZY-3全色波段融合)方案B1~B10分别为蓝、绿、红、3个红边、近红外、窄近红和2个短波红外波段;纹理特征包括均值 Mean、标准差 Std、同质性 Hom、对比度 Con、异质性 Dis、熵 Ent、角二阶矩 Ang和相关性 Cor

4.2 不同分类方法的精度对比分析

不同数据方案和分类方法的精度评价结果(表5)说明利用4种数据方案和3种分类方法对该平原区人工林树种分类都达到了85%的精度;ZY_{SPTX}数据方案利用分层优化变量组合的随机森林分类算法取得了97.30%的最高总体精度。在仅使用光谱变量时,同一数据方案(ZY_{SP}或STZY_{SP})随机森林和分层分类的总体精度没有明显差别;但是,STZY_{SP}(10个光谱波段和3个植被指数)比ZY_{SP}(4个光谱波段和3个植被指数)数据方案分类总体精度提高了2.49%~2.91%,这表明了Sentinel-2A多光谱中的红边、窄近红外和短波红外波段能有效提高分类精度。在纹理特征参与分类后,不同数据方案中分层分类的总体精度最高、分类回归树的总体精度最低,并且不同分类器均是SPTX(光谱+纹理变量组合)提供了最佳分类结果。在分层分类

中,ZY_{SPTX}比ZY_{SP}数据方案分类总体精度提高了10.19%,STZY_{SPTX}比STZY_{SP}数据方案分类总体精度提高了4.99%。

4.3 不同树种的精度对比分析

在不同数据方案中,分层分类通过分层优化节点选取最佳变量组合,可视化调节光谱和纹理变量组合为单树种取得了最佳树种分类精度(表6)。在仅使用光谱波段时,相比ZY_{SP}方案,光谱波段的增加使得STZY_{SP}将杨树、桃园和苗圃的生产者精度分别提高了4.27%、11.36%、和11.48%。在纹理特征参与分类后,ZY_{SPTX}和STZY_{SPTX}的数据方案均比ZY_{SP}和STZY_{SP}提高了树种分类精度。如相较于ZY_{SP},ZY_{SPTX}将其他阔叶、桃园、苗圃的生产者精度分别提高了26.08%、34.09%和21.31%。从树种分类精度角度分析,基于ZY_{SPTX}的分层分类方法为杨树、其他阔叶、苗圃和桃园提供了最佳树种分类精度,

表5 不同分类方法的总体精度对比

Table 5 Comparison of accuracy assessment results among different data schemes

分类器 数据方案	分类回归树		随机森林		分层分类	
	总体精度/%	Kappa	总体精度/%	Kappa	总体精度/%	Kappa
ZY _{SP}	85.65	0.798	87.53	0.824	87.11	0.817
STZY _{SP}	89.40	0.851	90.02	0.859	90.02	0.860
ZY _{SPTX}	90.02	0.861	94.80	0.927	97.30	0.962
STZY _{SPTX}	91.89	0.886	94.59	0.924	95.01	0.930

分别为97.44%、93.37%、92.14%和96.72%，比分类回归树提高了2.35%、11.81%、14.49%和9.98%。

在ZY_{SPTX}数据方案下使用具有分层筛选关键变量组合的分层分类要优于随机森林和分类回归树。

表6 不同树种的分类精度对比(单位:%)

Table 6 Comparison of classification accuracies of tree species among different data schemes (Unit:%)

分类器 数据方案		分类回归树				随机森林				分层分类			
		杨树	其他阔叶	桃园	苗圃	杨树	其他阔叶	桃园	苗圃	杨树	其他阔叶	桃园	苗圃
ZY _{SP}	PA	86.32	63.04	59.09	73.77	89.74	67.39	61.36	77.05	90.60	65.22	59.09	75.41
	UA	88.60	69.05	55.32	69.23	90.52	65.96	61.36	85.45	89.08	73.17	60.47	77.97
STZY _{SP}	PA	88.03	69.57	70.45	85.25	91.45	69.57	68.18	85.48	94.87	58.70	70.45	86.89
	UA	91.15	74.42	70.45	77.61	89.17	76.19	83.33	81.54	88.10	77.14	77.50	84.13
ZY _{SPTX}	PA	85.47	80.43	90.91	81.97	94.87	91.30	88.64	85.25	97.44	91.30	93.18	96.72
	UA	96.15	74.00	66.67	83.33	96.52	80.77	82.98	96.30	97.44	95.45	91.11	96.72
STZY _{SPTX}	PA	88.89	73.91	88.64	86.89	94.87	89.13	86.36	86.89	94.02	89.13	88.64	90.16
	UA	92.04	82.93	72.22	88.33	91.74	91.11	88.37	92.98	94.02	82.00	97.50	90.16

注:PA和UA为生产者精度和用户精度

基于分层分类结果的混淆矩阵(表7)表明,仅利用光谱数据进行的树种分类精度较低,如基于ZY_{SP}数据方案,桃园的精度和生产者精度都只有60%左右,与杨树,其他阔叶林,甚至苗圃都严重混淆。由于不同树种的林分结构有差异,引入空间信息可以显著提高分类精度,如基于ZY_{SPTX},桃园的精度和生产者精度提高到91.1%和93.2%。

图4是基于ZY_{SPTX}数据方案利用分层分类算法产生的人工林树种空间分布图。从图中可以看出,杨树主要沿河或沿路呈条带式分布,苗圃和桃园分

布相对集中,主要分布在研究区中上部和西南角,整体表现为斑块状,面积较大,而且靠近水源。对比分层分类、随机森林和分类回归树(图4(c)~4(e))来看,总体空间分布极其相似,但局部细节上,由于分层分类可针对主要人工林类型杨树、桃园和苗圃进行分类优化而表现更加准确。

5 讨论

5.1 不同变量在人工林树种分类中的作用

光谱变量在平原区人工林树种分类中发挥重

表7 分类混淆矩阵(分层分类方法)

Table 7 Confusion matrix of classification result from hierarchical-based classifier

	ZY _{SP} 数据方案						ZY _{SPTX} 数据方案					
	杨树	其他阔叶	桃园	苗圃	其他地物	用户精度	杨树	其他阔叶	桃园	苗圃	其他地物	用户精度/%
杨树	106	2	5	6	0	89.08	114	1	2	0	0	97.44
其他阔叶	1	30	6	2	2	73.17	0	42	0	1	1	95.45
桃园	7	8	26	2	0	60.47	2	2	41	0	0	91.11
苗圃	3	3	7	46	0	77.97	1	1	0	59	0	96.72
其他地物	0	3	0	5	211	96.35	0	0	1	1	212	99.07
生产者精度	90.60	65.22	59.09	75.41	99.06		97.44	91.30	93.18	96.72	99.53	
总体精度=87.11%, Kappa系数=0.817						总体精度=97.30%, Kappa系数=0.962						

注:ZY为ZY-3多光谱和全色波段融合,ZY_{SP}为仅有光谱变量组合,ZY_{SPTX}光谱和纹理特征变量的组合

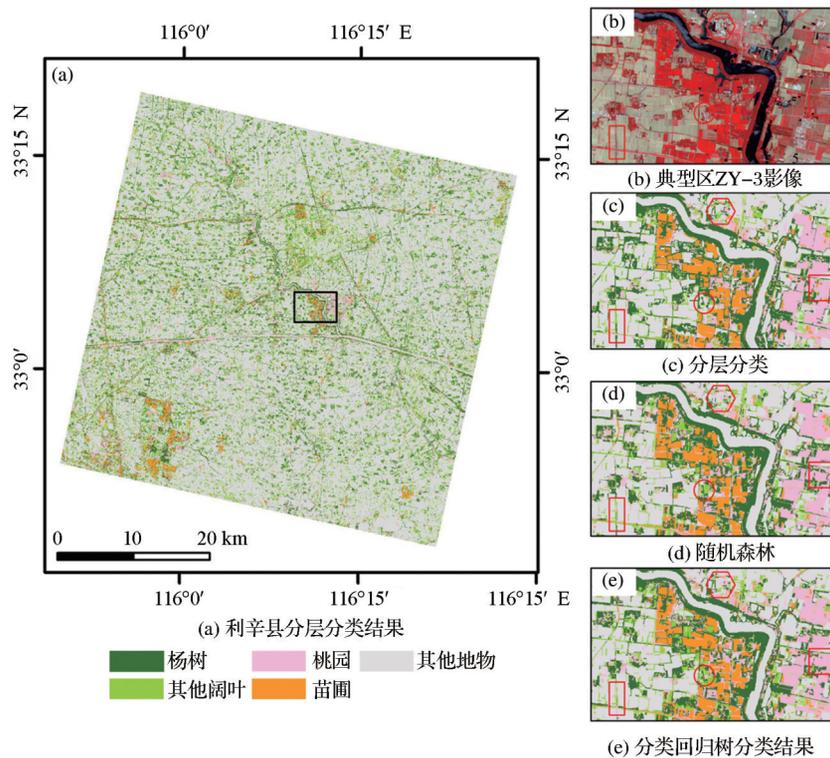


图4 利辛县人工林树种空间分布图

Fig. 4 Spatial distribution of tree species in Linxi County, Anhui Province

要作用,但对不同树种类型或者分层分类树的不同节点所起的作用不同。从表3和表5不同数据方案分层分类树节点变量选择及分类结果可以发现,光谱特征可以较好的区分植被和非植被类型;而在人工林树种分类中 Sentinel-2 数据的红边、窄近红外和短波红外波段对提高杨树、其他阔叶、桃园和苗圃的分类精度发挥了重要的作用,这与以往的研究结论相似^[11]。从图5可以得出,树种间的光谱差异在新增波段上表现明显,尤其是红边和短波红外波段,其中B9波段对苗圃和其他阔叶的区分能力明显优于蓝,绿,红,近红外波段,红边波段对杨树和苗圃的区分能力和B3红波段相近,但B3红波段杨树有过多的异常值,因此红边波段对杨树和苗圃的区分能力略优于红波段。这与本研究的实验结果一致,在节点6和节点7区分苗圃,杨树和其他阔叶时,红边和短波红外均被同时选中。然而,依赖有限光谱信息的分类精度难以保证,协同使用纹理和光谱变量可以显著提高平原树种分类精度。研究区的在耕农田、桃园和苗圃是按一定的空间有规律的分布,其中在耕农田(以玉米为主)种植密度高间距小,而桃园冠幅大且种植间距大,两者具有明显不同的纹理特征。这与表3的变量选择结果一致,节点5在区分桃园与在耕农田时,最终选择的变量

以纹理变量为主。

5.2 分层分类方法在平原区人工林树种分类的优势

相比随机森林和分类回归树,本研究采用的分层优化变量组合的分类模型从分类特征和分类策略上均保证了分类结果的可靠性和稳定性。在分类特征方面,从高维特征空间中有针对性地选择分类树不同层级各类别的最优特征集合参与分类,简单高效地避免了同一分类特征集对分类结果的影响,其中非植被地物主要使用了光谱特征,而纹理特征参与分类后能有效提高树种分类精度。在分类策略上方面,通过分层分类的形式,按照从易到难的顺序层层分解,增强了信息提取能力、分类精度和计算效率。其中,联合纹理信息的资源三号多光谱和全色波段融合的数据方案为单一树种特别是杨树提供了最佳树种分类精度,与李丹等^[1]和王二丽等^[2]杨树提取结果相比,分类精度的稳定性和准确度都有提升,但这可能是本研究区优势树种杨树大多成片种植且较为单一、冠幅大且枝叶茂密,与复杂的城市树种中提取杨树相比分类难度较小。

分类回归树和随机森林方法分类过程中,用户无法了解区分每一个树种的关键变量,而分层分类方法可以反馈分层分类树所用的变量,方便用户进一步的研究。另外,随机森林在训练过程中,精度

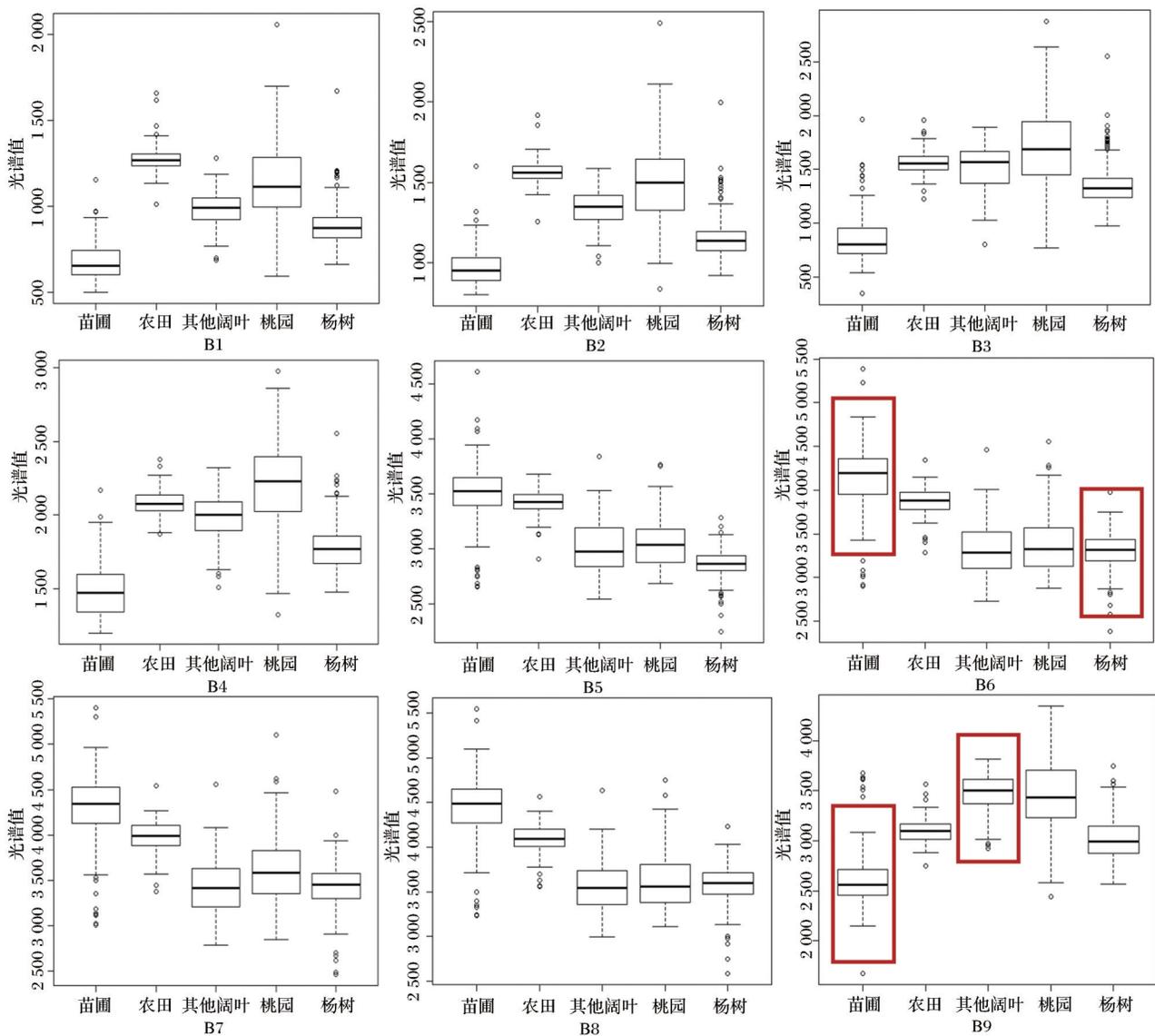


图 5 研究区主要植被类型在 Sentinel-2 和 ZY-3 融合后各波段的光谱反应

Fig.5 Spectral responses of major vegetation types based on fused images of Sentinel-2 and ZY-3 panchromatic data

评价指标为总体精度,其算法内部所使用的变量往往偏向大样本植被类型,忽略小样本树种^[12],而准确绘制小样本树种的分布往往同样重要。分层分类方法在分层分类树的每一层只区分两个类别,可以很好地兼顾小样本树种类型,从而提高树种分类精度。应用分层分类方法的一个关键步骤是设计适合的分层分类树结构,不合理的树结构可能会影响植被分类的准确性^[19]。本研究基于专家知识建立分层分类树结构,该树结构可能不是最优的,未来的研究应该聚焦在如何应用训练样本和相关统计学知识自动设计最适合的分层分类树结构以获得更可靠的分类产品。

地物的光谱反应特征是遥感影像土地覆盖分类的基础,而“同物异谱”和“异物同谱”的存在给图

像的分类带来很大困难,造成错分和误分。比如,同是农田,由于作物的种类不同、生长季不同,其光谱反应曲线也不同;再比如,城市建筑物的屋顶材料不同,颜色不同,在光谱曲线上表现也不同。如果选取这些类别的训练样本时作为一类,由于光谱反应差异大,样本集内部方差过大,分类时往往和其他类别混淆,造成分类精度低。因此在进行地表覆盖分类时,应尽可能的设计较细的分类系统,充分考虑到“同物异谱”现象,分别选取训练样本,产生较为详细的分类结果,最后根据研究目的进行同类归并,达到较好的分类效果。本研究主要关注人工林树种,而农田和其他非植被地类(水体、城镇用地等)不是本研究的目的所在。但是非人工林地类包含的地表覆盖类型多样,既包括了有植被覆盖的

在耕农田,又包含了非植被的休耕农田、水体和城镇用地。如果直接把所有非人工林作为一类,由于样本集内光谱的差异太大,势必增大分类的误差。同为耕地,在耕农田与休耕地的光谱曲线大不相同,在分层分类时,首先把它们分在植被和非植被层中。因此必须对在耕农田和休耕农田分开选取训练样本,分为不同的类别,最后把在耕农田与其他非植被类型合并为其他地类,以区别于人工林树种,从而提高总体的分类精度。

6 结 论

本研究基于面向对象高分辨率遥感数据的光谱和纹理特征,采用分层优化变量组合的随机森林分类算法,探索了不同数据方案对平原区的人工林树种分类精度的影响。结果表明:分层分类方法为平原区的人工林树种提供了大于92%的分类精度,与分类回归树、随机森林相比提高了2.35%~14.49%。利用光谱和纹理信息能有效提高平原区人工林树种分类精度,其中ZYSPTX数据方案通过分层分类方法取得了97.30%的总体精度,相比于其他数据方案的总体精度提高了7.28%~11.65%。

致谢:感谢滁州学院李龙伟和李楠老师在外业样地数据收集过程中的帮助!

参考文献(References):

- [1] Li Dan, Ke Yinghai, Gong Huili, *et al.* Urban tree species classification with machine learning classifier using WorldView-2 imagery[J]. *Geography and Geo-Information Science*, 2015, 32(1):84-89.[李丹,柯樱海,宫辉力,等.基于高分辨率遥感影像的城市典型乔木树种分类研究[J]. *地理与地理信息科学*, 2015, 32(1):84-89.]
- [2] Wang Erli, Li Cunjun, Zhou Jingping, *et al.* Classification of Beijing afforestation species based on multi-temporal images [J]. *Journal of Beijing University of Technology*, 2017, 43(5):710-718.[王二丽,李存军,周静平,等.基于多时相遥感影像的北京平原人工林树种分类[J]. *北京工业大学学报*, 2017, 43(5):710-718.]
- [3] Immitzer M, Atzberger C, Koukal T. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data[J]. *Remote Sensing*, 2012, 4(9):2661-2693. DOI:10.3390/rs4092661.
- [4] Lu D, Li G, Moran E, *et al.* A comparative analysis of approaches for successional vegetation classification in the Brazilian Amazon [J]. *GIScience and Remote Sensing*, 2014, 51(6):695-709. DOI: 10.1080/15481603.2014.983338.
- [5] Kempeneers P, Sedano F, Seebach L, *et al.* Data fusion of different spatial resolution remote sensing images applied to forest-type mapping[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2011, 49(12):4977-4986. DOI: 10.1109/TGRS.2011.2158548.
- [6] Wu Q, Zhong R, Zhao W, *et al.* Land-cover classification using GF-2 images and Airborne Lidar data based on random forest[J]. *International Journal of Remote Sensing*, 2018, (6): 1-17. DOI: 10.1080/01431161.2018.1483090.
- [7] Lu D, Hetrick S, Moran E. Land cover classification in a complex urban-rural landscape with QuickBird imagery[J]. *Photogramm Eng Remote Sensing*, 2010, 76.10: 1159-1168. DOI: 10.14358/PERS.76.10.1159.
- [8] Xie Z, Chen Y, Lu D, *et al.* Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data [J]. *Remote Sensing*, 2019, 11(2): 164-191. DOI:10.3390/rs11020164.
- [9] Yu X, D. Lu, Li G, *et al.* Examining the roles of spectral, spatial, and topographic features in improving land-cover and forest classifications in a subtropical region[J]. *Remote Sensing*, 2020, 12: 2907. DOI:10.3390/rs12182907.
- [10] Belgiu M, Dragut L. Random forest in remote sensing: A review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31. DOI:10.1016/j.isprsjprs.2016.01.011.
- [11] Chen Y, Zhao S, Xie Z, *et al.* Mapping multiple tree species classes using a hierarchical procedure with optimized node variables and thresholds based on high spatial resolution satellite data [J]. *GIScience & Remote Sensing*, 2020, 57(6): 526-542. DOI:10.1080/15481603.2020.1742459.
- [12] Zhou Yang, Li Xunhuan. Geographical pattern and mechanism of poverty differentiation in plain areas: A case study of Lixin county Anhui Province[J]. *Scientia Geographica Sinica*, 2019, 39(10):1592-1601.[周扬,李寻欢.平原农区贫困地理格局及其分异机制——以安徽省利辛县为例[J]. *地理科学*, 2019, 39(10):1592-1601.]
- [13] Li Mingshi. A study on information extraction of poplar resources in Jianhu county using ASTER remotely sensed datasets classification, modeling & retrieval and mapping[D]. Nanjing: Nanjing Forestry University, 2005.[李明诗.基于ASTER遥感数据的建湖县杨树信息提取的研究——分类、建模与反演制图[D].南京:南京林业大学,2005.]
- [14] Chavez P. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data[J]. *Remote Sensing of Environment*, 1988, 24(3):459-479. DOI: 10.1016/0034-4257(88)90019-3.
- [15] Louis J, Debaecker V, Pflug B, *et al.* Sentinel-2 SEN2COR: L2A processor for users [C]// In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 2016
- [16] Karathanassi V, Kolokousis P, Ioannidou S. A comparison study on fusion methods using evaluation indicators[J]. *International Journal of Remote Sensing*, 2007, 28(10): 2309-

2341. DOI:10.1080/01431160600606890.
- [17] Liu Jinli, Chen Zhao, Gao Jinping, *et al.* Research on the method of determining the optimal segmentation scale for tree species classification of high-resolution image[J]. *Scientia Silvae Sinicae*, 2019, 55 (11): 95-104.[刘金丽, 陈钊, 高金萍, 等. 高分影像树种分类的最优分割尺度确定方法[J]. *林业科学*, 2019, 55 (11): 95-104.]
- [18] Lu D, Li G, Moran E, *et al.* The roles of textural images in improving land-cover classification in the Brazilian Amazon [J]. *International Journal of Remote Sensing*, 2014, 35, 8188-8207. DOI:10.1080/01431161.2014.980920
- [19] Zhao S, Jiang X, Li G, *et al.* Integration of ZiYuan-3 multi-spectral and stereo imagery for mapping urban vegetation using the hierarchy-based classifier[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 105: 102594. DOI:10.1016/j.jag.2021.102594.
- [20] Gao Guolong, Du Huaqiang Han Ning, *et al.* Mapping of mo-
so bamboo forest using object-based approach based on the optimal features[J], *Scientia Silvae Sinicae*, 2016, 52(9):77-85. [高国龙, 杜华强, 韩凝, 等. 基于特征优选的面向对象毛竹林分布信息提取[J]. *林业科学*, 2016, 52(9):77-85.]
- [21] Hao Huang, Chen Yongfu, Liu Hua, *et al.* Object-oriented forest classification of Linzhi county based on CART decision tree with texture information[J]. *Remote Sensing Technology and Application*, 2017, 32(2):386-394.[郝洸, 陈永富, 刘华, 等. 基于纹理信息CART决策树的林芝县森林植被面向对象分类[J]. *遥感技术与应用*, 2017, 32(2):386-394.]
- [22] Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45 (1): 5-32.
- [23] Zhou Tianing, Ming Dongping, Zhao Rui. Land cover classification based on algorithm of parameter optimization random forests[J]. *Science of Surveying and Mapping*, 2017, 42(2): 88-94.[周天宁, 明冬萍, 赵睿. 参数优化随机森林算法的土地覆盖分类[J]. *测绘科学*, 2017, 42(2):88-94.]

Man-made Tree Species Classification in Lixin County, Anhui Province

Zhao Shuai^{1,2}, Cao Meiqin^{1,2}, Jiang Xiandie^{1,2}, Chen Yaoliang^{1,2},
Lu Dengsheng^{1,2}

(1.State Key Laboratory for Subtropical Mountain Ecology of the Ministry of Science and Technology and Fujian Province, Fujian Normal University, Fuzhou, China;

2.School of Geographical Sciences, Fujian Normal University, Fuzhou 350007, China)

Abstract: Forest plantations in plain area of China are highly fragmented and dispersed due to flexible planting system. Accurate and timely mapping plantation distribution in plain area plays an important role in plantation management and hydro-ecological functions. In this study, a hierarchical-based classification method, which optimizes variable combinations at each node, was proposed and applied in Linxi County, Anhui Province for mapping tree species in plain area. The object - based classification with combinations of various types of spectral and texture features derived from the fused image of ZY-3 multispectral and panchromatic band (ZY), and the fused image of Sentinel-2 multispectral bands and ZY-3 panchromatic band (STZY) were conducted. The results showed that the proposed method provides higher classification accuracy than regression tree and random forest approaches for both datasets, especially for dominant poplar tree species. When only spectral features were used, STZY offered better results than ZY, overall accuracies increased by 2.49%~2.91%, indicating the importance of spectral information in classification. When textures were integrated with spectral features into classification procedure, overall accuracies increased by 10.19% and 4.99% for STZY and ZY respectively using the hierarchical-based classifier, implying that texture features are essential for classification. Thus, the proposed hierarchical-based classifier with optimized variable combinations is an effective method in tree species mapping in plain area using spectral and texture features derived from high spatial-resolution data.

Key words: Hierarchical-based classifier; Feature optimization; Plain area; ZY-3; Sentinel-2