

罗凤宇, 高艺非, 谢 勇, 等. 基于随机森林算法与多时相 Sentinel-2 影像数据的茶树种植区信息提取[J]. 江苏农业学报, 2024, 40(9): 1671-1680.

doi: 10.3969/j.issn.1000-4440.2024.09.011

基于随机森林算法与多时相 Sentinel-2 影像数据的 茶树种植区信息提取

罗凤宇¹, 高艺非², 谢 勇¹, 邹旭辉¹, 邵 雯¹, 张世雨¹

(1. 南京信息工程大学地理科学学院, 江苏 南京 210044; 2. 江苏师范大学地理测绘与城乡规划学院, 江苏 徐州 221116)

摘要: 茶树是中国重要的木本经济作物, 及时准确地获取茶树种植区面积及空间分布对区域农业经济发展具有重要意义。本研究以安徽省郎溪县为研究区, 首先分析茶树、小麦和红叶石楠时序光谱特征, 其次基于 3 个时相的 Sentinel-2 影像数据提取光谱特征、水体指数及植被指数特征、红边指数特征、纹理特征组成多时相特征变量数据集, 并设置 6 种特征变量组合方案, 利用随机森林算法进行茶树种植区信息提取精度的比较, 筛选得到适宜的特征变量组合方案, 最后基于适宜的特征变量组合方案进行郎溪茶树种植区信息的提取。结果表明, 在光谱特征变量的基础上, 分别融合水体指数及植被指数特征、红边指数特征和纹理特征变量均能有效提高茶树种植区信息的提取精度, 其中, 红边指数特征对茶树种植区信息提取精度的提高效果最好, 其次是水体指数及植被指数特征。基于随机森林-平均精确度下降算法(RF-MDA)优选后的特征变量组合的分类效果最佳, 总体分类精度达 94.95%, Kappa 系数为 0.934 8, 说明特征变量优选能有效地保留重要的地物识别特征变量, 避免冗余信息对分类结果的影响。综上, 基于随机森林算法和茶树多时相 Sentinel-2 影像数据能实现郎溪县茶树种植区信息的高精度提取。

关键词: 茶树种植区; 随机森林; 多时相特征; 面积监测

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-4440(2024)09-1671-10

Extraction of tea plantation area information based on random forest algorithm and multi-temporal Sentinel-2 image data

LUO Fengyu¹, GAO Yifei², XIE Yong¹, ZOU Xuhui¹, SHAO Wen¹, ZHANG Shiyu¹

(1. School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: Tea plant is an important woody cash crop in China, and timely and accurate acquisition of the planting area and spatial distribution of tea plants is of great significance to regional agricultural economic development. In this study, Langxi County in Anhui province was taken as the study area. Firstly, the temporal spectral characteristics of tea trees, wheat and *Photinia fraseri* were analyzed. Secondly, based on the Sentinel-2 image data of three phases, spectral features, water index and vegetation index features, red edge index features and texture features were extracted to form a multi-temporal feature variable data set, and six feature variable combination schemes were set up. The random forest algorithm

was used to compare the extraction accuracy of tea tree planting areas, and the suitable combination scheme of characteristic variables was selected. Finally, the tea tree planting area in Langxi County was extracted based on the suitable combination scheme of characteristic variables. The results showed that on the basis of spectral feature variables, the fusion of water index and vegetation index fea-

收稿日期: 2023-11-10

基金项目: 国家重点研发计划项目(2023YFB3905801)

作者简介: 罗凤宇(1998-), 男, 贵州黔南人, 硕士研究生, 主要研究方向为茶树遥感识别。(E-mail) 20211210039@nuist.edu.cn

通讯作者: 谢 勇, (E-mail) xieyong@nuist.edu.cn

tures, red edge index features and texture feature variables could effectively improve the extraction accuracy of tea planting area information. Among them, red edge index features had the best effect on improving the extraction accuracy of tea planting areas, followed by water index and vegetation index features. The classification effect of the feature variable combination based on the random forest-average accuracy reduction algorithm (RF-MDA) was the best, the overall classification accuracy was 94.95%, and the Kappa coefficient was 0.934 8, indicating that the feature variable optimization could effectively retain the important feature recognition feature variables and avoid the influence of redundant information on the classification results. In summary, based on the random forest algorithm and the multi-temporal Sentinel-2 image data of tea trees, the high-precision extraction of tea planting area information in Langxi County can be realized.

Key words: tea plantation area; random forest; multi-temporal features; area monitoring

茶树(*Camellia sinensis*)发源于中国,山茶科山茶属,是中国重要的木本经济作物,广泛分布于中国的 18 个省市^[1]。安徽省郎溪县是中国著名的产茶区,1996 年被原农业部授予“中国绿茶之乡”称号。及时准确地获取茶树种植面积及其空间分布对于茶园规划、茶产业的可持续健康发展及乡村振兴和区域农业经济发展均具有重要意义。

传统的依靠基层人员对作物种植面积进行实地调查的方法不仅难以获取准确的数据及空间分布,而且人工成本高、时效性差。20 世纪 80 年代以来,卫星遥感技术凭借覆盖范围广、时效性高、获取信息丰富等特点,在农作物种植面积和空间分布的监测方面取得了许多成果^[2-3]。近年来,随着遥感影像资源的日益丰富及人工智能算法的应用,农业遥感监测的范围逐步由大宗作物向小宗作物、种植面积和分布状况向长势、营养状况、病虫害等方向深入^[4-6]。基于多时相遥感数据的作物监测能充分利用地物不同物候期信息,有效减少光谱混淆导致的分类误差。张猛等^[7]基于时间序列 MODIS *EVI*(增强型植被指数)数据提取洞庭湖流域湿地信息,能有效解决“异物同谱”、“同物异谱”的现象。柳文杰等^[8]融合多时相环境卫星(HJ-1A/1B)数据和 MODIS13Q1 数据,获取时间序列 *NDVI*(归一化植被指数)数据,并结合水稻关键物候特征参数,对洞庭湖流域水稻种植面积进行提取,取得了较好的效果,总体分类精度达 91.71%,Kappa 系数达 0.902 4。

由于茶树常与其他灌木、林木、农作物混种,基于遥感影像的茶树种植区信息提取具有一定的难度。但茶树与其他农作物、常绿天然林和灌木林在管理方面存在一定的差异,如定期的修剪以保持茶树活力^[9-11],这为精细化的茶树种植区信息提取提供了可能。王斌等^[12]基于春季、秋季和冬季 Landsat-8 OLI 影像数据,构建单时相和多时相特征变量

数据集,并利用随机森林算法对安吉县茶树种植区信息进行了提取,取得了较好的结果。Xu 等^[9]基于多时相 Landsat ETM 和 OLI 遥感影像数据构建植被指数分析茶树的物候期,并提出一种基于像元和物候的茶树种植区信息的提取方法。

红边指数等光谱衍生因子的应用有利于提高地物分类精度^[13]。Sentinel-2 系列卫星拥有 3 个红边波段,能够对植被生长信息及其健康状况进行监测^[14],此外,双星运行模式使得重访周期缩短至 5 d,极大增强了对地面物体的观测能力^[15]。然而,利用 Sentinel-2 影像对茶树种植区信息的精细化提取方面还未有报道。因此,本研究利用实地采集的典型植被(茶树、红叶石楠、小麦)时序反射率数据,分析其同时期的光谱差异,并以无人机获取郎溪县不同植被类型的样本,结合 3 个时相的 Sentinel-2 影像数据,构建不同的特征变量数据集,利用随机森林算法进行地表植被类型的分类,以期为皖南地区茶树种植面积动态监测提供方法和依据,促进当地茶产业可持续健康发展。

1 材料与方法

1.1 研究区概况

郎溪县(30°48′~31°18′N,118°58′~119°22′E)隶属于安徽省宣城市,地势东高西低、南高中低,低山和丘陵岗地约占 57.2%,土壤类型以红壤、黄壤等酸性土壤为主,气候温和、日照充足,降水丰富,具有发展茶树种植的优良条件(图 1)。郎溪茶树主要种植在具有一定坡度的山坡上,株高低于 1.5 m,以灌木型中小叶绿茶^[16-18]为主,且不同品种的茶树物候期相近。当地还种植以红叶石楠为主的灌木,耕地则以小麦、水稻轮作为主。郎溪是皖南地区重要的茶树种植区,具有较好的代表性,因此本研究以郎溪的茶树为研究对象,进行茶树种植区面积和分布的提取。

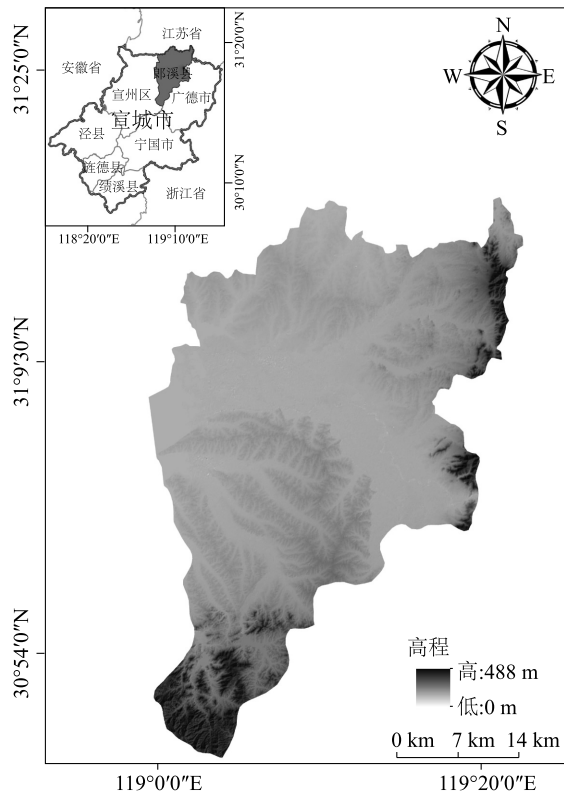


图1 郎溪县地理位置
Fig.1 Location of Langxi County

1.2 数据获取及处理

1.2.1 遥感数据获取与处理 本研究所使用遥感数据均为欧洲航天局(ESA, European Space Agency)提供的 Sentinel-2A 的 L2C 级数据。Sentinel-2 卫星搭载多光谱影像仪,具有 13 个波段,涵盖可见光到近红外,具体参数如表 1 所示。本研究使用的遥感数据为 2023 年 2 月、5 月、6 月共 3 个时相的 Sentinel-2A 影像 B2、B3、B4、B5、B6、B7、B8、B8A、B11、B12 共 10 个波段数据。所有影像数据均基于 ESA 官方提供的软件 SNAP 进行大气校正和重采样至 10 m 空间分辨率等预处理。

1.2.2 典型植被高光谱数据采集与处理 为更准确进行茶树的遥识别,本研究中针对易与茶树光谱产生混淆的红叶石楠和小麦进行高光谱监测,以明确易混淆植被的光谱差异。分别选取茶树、红叶石楠和小麦 60 m×60 m 样地,利用 RS-8800 高性能便携式地物光谱仪(美国 Spectral Evolution 公司产品)进行样方的光谱监测,具体采集时间如表 2 所示。RS-8800 地物光谱仪的光谱范围为 350~2 500 nm,采样间隔 1 nm,视场角为 25°。数据采集时,先

选取植被覆盖度均匀的样方 6 个,样方大小为 1 m×1 m,样方之间距离大于 10 m,并使采样样方均匀分布于样地。选择晴朗、无云、风小的天气条件,于 10:00~14:00 进行数据采集,测量时将探头置于样方垂直上方 0.3~0.5 m 处。为减少试验误差,每测完 1 个样方均重新利用白板进行校正,每个样方采集 5~8 条反射率数据,最后对所有数据进行质量筛选,计算平均值得到不同时相的茶树、红叶石楠和小麦冠层反射率数据。

表 1 Sentinel-2A 影像波段信息统计

Table 1 Statistics of Sentinel-2A image band information		
波段	中心波长 (nm)	分辨率 (m)
B1(海岸波段)	443	60
B2(蓝波段)	490	10
B3(绿波段)	560	10
B4(红波段)	665	10
B5(植被红边波段 1)	705	20
B6(植被红边波段 2)	740	20
B7(植被红边波段 3)	783	20
B8(宽带近红外波段)	842	10
B8A(窄带近红外波段)	867	20
B9(水汽波段)	945	60
B10(卷云波段)	1 375	60
B11(短波红外 1)	1 610	20
B12(短波红外 2)	2 190	20

表 2 样区信息及数据采集日期

Table 2 Sample area information and data collection date		
样区	样区大小	数据采集日期(年-月-日)
小麦	60 m×60 m	2023-01-10、2023-02-27、2023-03-28、2023-05-01
红叶石楠	60 m×60 m	2023-01-09、2023-02-27、2023-03-28、2023-05-01、2023-06-09
茶树	60 m×60 m	2023-01-09、2023-02-27、2023-03-28、2023-05-01、2023-06-09

1.2.3 不同植被类型样本数据的采集 在参考《土地利用现状分类》^[19]以及文献^[20]、^[21]的基础上,结合研究区实际植被情况,将郎溪县土地覆盖类型划分为耕地、林地(灌木、果园、森林等)、茶树种植区、水域(坑塘、河流等)、人造地表、裸土 6 类。于 2023 年 2 月至 6 月在郎溪县域内利用无人机采集样本数据,每

个样本大小控制在50 m×50 m 以上,然后依据实地考察及 Google Earth 高分辨率影像的目视解译,共获取耕地样本 65 个、林地样本 72 个、茶树种植区样本 80 个、水域样本 77 个、人造地表样本 62 个、裸土样本 67 个。最后基于分层抽样法将 70%的样本用于模型训练,30%的样本用于结果验证。

1.3 研究方法

首先对茶树、小麦和红叶石楠的实测光谱特征进行分析,基于光谱差异选择茶树种植区识别的 Sentinel-2 影像构建特征变量数据集,并设置 6 种特征变量组合方案,利用随机森林算法筛选适宜茶树种植区信息提取的特征变量组合方案,最后基于适宜的特征变量组合方案进行郎溪茶树种植区信息的提取。

1.3.1 特征变量提取 首先基于 Sentinel-2 影像的光谱特征,计算水体及植被指数特征、红边特征、纹

理特征,构建特征变量数据集,各类特征的变量名称及算法等描述见表 3。

鉴于光谱的一阶导数能部分消除大气效应以及植被光谱中土壤成分的影响^[22],研究中对红边指数特征中引入红边波段 1 一阶微分反射比($B5_FDR$)、红边波段 2 一阶微分反射比($B6_FDR$)及归一化差异红边一阶微分反射比($NDred_FDR$)3 个变量。 $B5_FDR$ 和 $B6_FDR$ 利用 ENVI 软件的 Image Derivative 插件进行提取。 $NDred_FDR$ 由 $B5_FDR$ 和 $B6_FDR$ 计算得到(表 3)。

纹理特征具有鲁棒性强、计算简单等特点,可基于影像的灰度共生矩阵^[23]提取。研究中对每个时相的 Sentinel-2A 影像 $B2$ 、 $B3$ 、 $B4$ 、 $B5$ 、 $B6$ 、 $B7$ 、 $B8$ 、 $B8A$ 、 $B11$ 和 $B12$ 共 10 个波段数据进行主成分分析后,选择方差贡献率最大的第一主成分,并取 4 个方向的平均值来表示灰度共生矩阵,进而得到纹理特征参数。

表 3 Sentinel-2 特征变量数据集描述

Table 3 Description of the Sentinel-2 feature set

特征类型	特征变量名称	来源或方法	参考文献
光谱特征	$B2$ 、 $B3$ 、 $B4$ 、 $B5$ 、 $B6$ 、 $B7$ 、 $B8$ 、 $B8A$ 、 $B11$ 、 $B12$	Sentinel-2A 影像的 10 个波段	-
水体指数及植被指数特征	地表水分指数($LSWI$)	$(B8-B11)/(B8+B11)$	Xiao 等 ^[24]
	归一化水体指数($NDWI$)	$(B3-B8)/(B3+B8)$	Gao ^[25]
	水分胁迫指数(MSI)	$B11/B8$	Erjr 等 ^[26]
	土壤调节植被指数的优化($OSAVI$)	$1.16 * (B8-B4)/(B8+B4+0.16)$	Rondeaux 等 ^[27]
	比值植被指数(RVI)	$B4/B8$	Birth 等 ^[28]
	归一化差异植被指数($NDVI$)	$(B8-B4)/(B8+B4)$	Tucker ^[29]
	增强型植被指数(EVI)	$2.5 * (B8-B4)/(B8+6 * B4-7.5 * B2+1)$	Miura 等 ^[30]
	植物衰老反射指数($PSRI$)	$(B4-B2)/(B6)$	Gamon 等 ^[31]
	叶绿素植被指数($GCVI$)	$B8/B3-1$	Gitelson 等 ^[32]
	归一化差异耕地指数($NDTI$)	$(B11-B12)/(B11+B12)$	Deventer 等 ^[33]
红边指数特征	归一化植被指数红边 1($NDVIre1$)	$(B8-B5)/(B8+B5)$	Gitelson 等 ^[34]
	归一化植被指数红边 2($NDVIre2$)	$(B8-B6)/(B8+B6)$	张磊等 ^[35]
	归一化植被指数红边 3($NDVIre3$)	$(B8-B7)/(B8+B7)$	张磊等 ^[35]
	归一化差异红边 1($NDre1$)	$(B6-B5)/(B6+B5)$	Gitelson 等 ^[36]
	归一化差异红边 2($NDre2$)	$(B7-B5)/(B7+B5)$	Merzlyak 等 ^[37]
	新型倒红边叶绿素指数($IRECI$)	$(B7-B4) * B6/B5$	Frampton 等 ^[38]
	改良红边土壤调整植被指数($MERre$)	$(B8-B5-1)/[\sqrt{(B8-B5)}+1]$	Wu 等 ^[39]
	红边波段 1 一阶微分反射比($B5_FDR$)	由 ENVI 软件的 Image Derivative 插件提取	-
	红边波段 2 一阶微分反射比($B6_FDR$)	由 ENVI 软件的 Image Derivative 插件提取	-
	归一化差异红边一阶微分反射比($NDred_FDR$)	$(B6_FDR-B5_FDR)/(B6_FDR+B5_FDR)$	-
纹理特征	均值、方差、协同性、对比度、非相似性、熵、角二阶矩、相关性	8 个灰度共生矩阵(GLCM)特征	-

1.3.2 分类及特征优选算法 随机森林(Random Forest, RF)分类算法^[40]能够高效处理高维特征数据,具有较高的泛化能力,在遥感研究中得到了广泛的应用。随机森林-平均精确度减少算法(Random Forest -Mean Decreased Accuracy, RF-MDA)^[41]通过打乱每个特征值的顺序,然后度量这种顺序变化对模型精确度的影响,以此来计算特征值的重要性得分。本研究利用 RF-MDA 算法进行特征选择,并结合随机森林分类算法进行地物分类。

1.3.3 特征值选取方案设计 本研究设计 6 组特征值组合进行茶树种植区信息提取结果的比较,其中方案 1~方案 5 是直接利用表 3 中的特征值进行组合,方案 6 是在方案 5 的基础上,采用 RF-MDA 算法对 3 个时相 114 个特征变量进行重要性计算,选择重要性排序前 20%^[12]的特征构成(表 4)。

表 4 特征值选取方案试验设计

Table 4 Eigenvalue selection test scheme design

分类方案	特征组合	特征变量 (个)
1	多时相光谱特征	30
2	多时相光谱特征+水体指数及植被指数特征	60
3	多时相光谱特征+红边指数特征	60
4	多时相光谱特征+纹理特征	54
5	多时相光谱特征+水体指数及植被指数特征+红边指数特征+纹理特征	114
6	多时相综合特征集前 20%的特征	23

1.3.4 分类结果精度评价方法 本研究基于混淆矩阵^[42]进行分类精度评价,评价指标包括生产者精度(PA)、用户精度(UA)、总体精度(OA)和 Kappa 系数。各指标计算公式如下:

$$PA = \frac{x_{ii}}{x_{i+}} \times 100\% \quad (3)$$

$$UA = \frac{x_{ii}}{x_{i+}} \times 100\% \quad (4)$$

$$OA = \frac{\sum_{i=1}^N x_{ii}}{N} \times 100\% \quad (5)$$

$$Kappa = \frac{N \sum_{i=1}^k x_{ii} - \sum_{i=1}^k (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^k (x_{i+} \times x_{+i})} \quad (6)$$

式中: N 为样本数, x_{i+} 为分类类别为 i 的样本总数, x_{+i} 为验证类别为 i 的样本总数, k 为分类结果类别总数, x_{ii} 为混淆矩阵中第 i 行第 i 列的值,即分类类别和实际类别一致的第 i 类样本数。

2 结果与分析

2.1 不同植被的光谱特征

2023 年 1~6 月小麦、茶树、红叶石楠的反射率特征如图 2 所示。从图中可以看出,1 月份茶树 350~2 500 nm 波段光谱反射率与 2 月份差异不大,而 1~2 月小麦和红叶石楠短波近红外波段(780~1 100 nm)的光谱反射率明显高于茶树,说明此阶段茶树生长发育较为缓慢。3 月,茶树反射率变化仍不明显,红叶石楠 780~1 100 nm 波段的光谱反射率明显低于小麦,其他波段两者基本一致,这可能是红叶石楠生长过程中受到了外界胁迫。5 月,茶树进入新梢生长期,长出新叶,此时茶树近红外波段的光谱反射率高于红叶石楠和小麦。6 月,由于研究区小麦已经收割,且田中蓄水,没有采集到反射率数据;茶树的 800~1 300 nm 波段的光谱反射率较 5 月份出现较大幅度的下降,且低于红叶石楠,可能是茶叶采摘后,茶树修剪导致的。

2023 年 5~6 月茶树修剪前后反射率的变化主要表现在 670~1 300 nm 波段,修剪后的茶树反射率明显低于未修剪茶树;2 000~2 400 nm 波段,修剪后的茶树反射率明显高于未修剪茶树(图 2a)。

从 3 种植被 1~6 月反射率变化来看,茶树由于经历人工修剪,修剪前后反射率发生显著变化,与小麦和红叶石楠存在一定的可分辨性。因此,为更好地提取茶树种植区信息,以 5 月新梢期影像为基础,在 5 月前后各选 1 景的影像,构成多时相数据。实际应用中,应选取邻近实际修剪时间的影像进行茶树种植面积与分布的提取。本研究中,选取的 3 景 Sentinel-2A 影像的成像时间分别为 2023 年 2 月 27 日、2023 年 5 月 1 日、2023 年 6 月 10 日。

2.2 特征选择

基于 RF-MDA 算法,方案 6 筛选出的 23 个特征变量的重要性分布情况如图 3 所示。从图中可以看出,水分胁迫指数_6 月的重要性得分最高,达 4.37,这可能由于茶树修剪导致 6 月份茶树冠层水分变化剧烈,而水分胁迫指数对于植被冠层水分异常敏感。归一化差异红边一阶微分反射比_6 月的特征重要

性得分排在第 6 位,此外红边波段 1 一阶微分反射比_{6 月}、红边波段 2 一阶微分反射比_{6 月}特征变量亦被选出,说明本研究构建的红边波段衍生因子有利于正确分类。优选出的特征变量大部分与 6 月份影像数据相关,因此,6 月份是比较适宜提取茶树种植区信息的主要时相,一方面 6 月份研究区的冬小

麦已经收割,秋收作物如玉米等刚栽种,植被覆盖度较小,另一方面茶树经过人工修剪虽然冠层覆盖度降低,但仍明显高于新播种的玉米等秋收作物,而林地的植被覆盖度在 5-6 月份则不会存在人工修剪导致的冠层覆盖度降低。

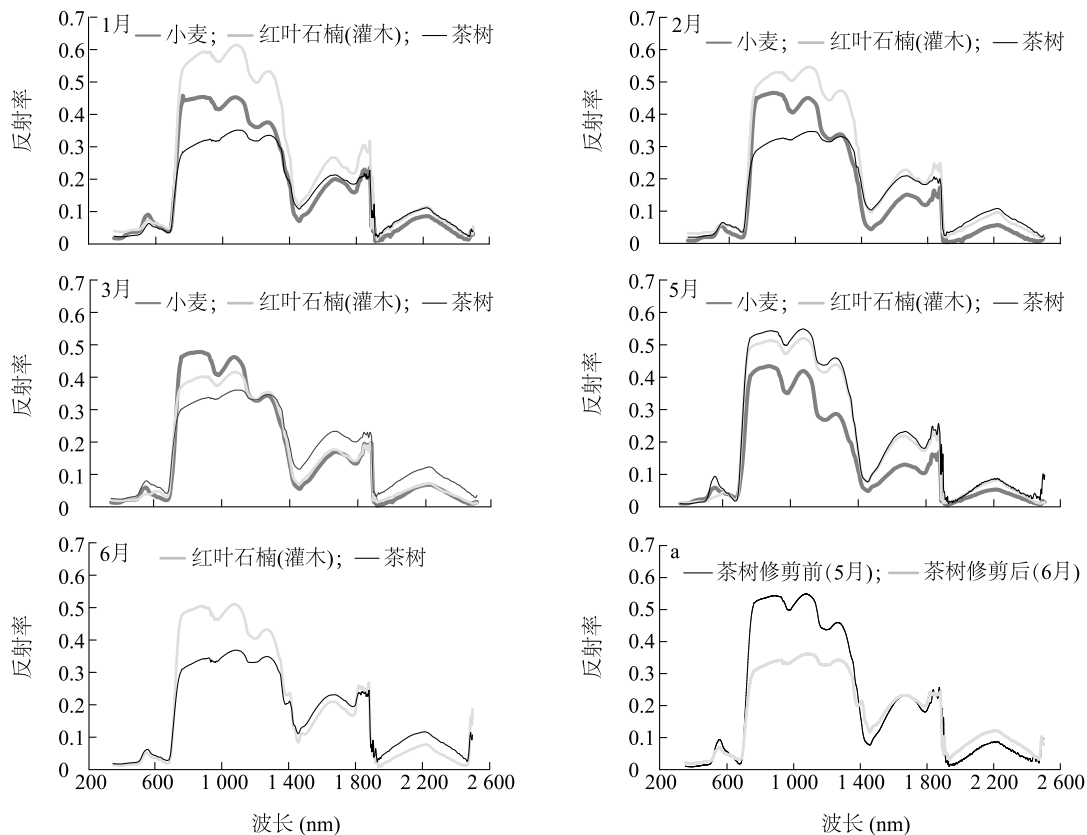


图 2 实测光谱反射率曲线

Fig.2 Measured spectral reflectance curves

根据优选的 23 个特征变量的重要性得分,统计得到各个分段不同类型特征变量的累计重要性得分比重如图 4a 所示。前 6 个特征变量中,光谱特征、水体指数及植被指数特征、红边指数特征各占 2 个,累计重要性得分所占比重差距不大。前 12 个特征变量中,水体指数及植被指数特征累计重要性得分占比明显高于光谱特征、红边指数特征。前 18 个特征中,水体指数及植被指数特征与红边指数特征的累计重要性得分占比相当,均高于光谱特征。23 个特征变量中,红边指数特征的变量累计重要性积分最高,水体指数及植被指数

特征的变量次之,光谱特征的变量最低。23 个特征变量有 9 个变量属于红边指数特征,7 个变量为光谱特征,7 个变量为水体指数及植被指数特征(图 4b)。从上述分析可知红边指数特征相关的变量不仅数量多,累计重要性得分亦高,这说明红边指数特征对植被的分类和茶树信息的提取有重要贡献。优选的特征变量中没有与纹理特征相关的变量,这一方面可能是由于 Sentinel-2 卫星 10 m 空间分辨率的影像数据所呈现的纹理特征不明显,另一方面可能是纹理特征对于地表植被的分类贡献较小。

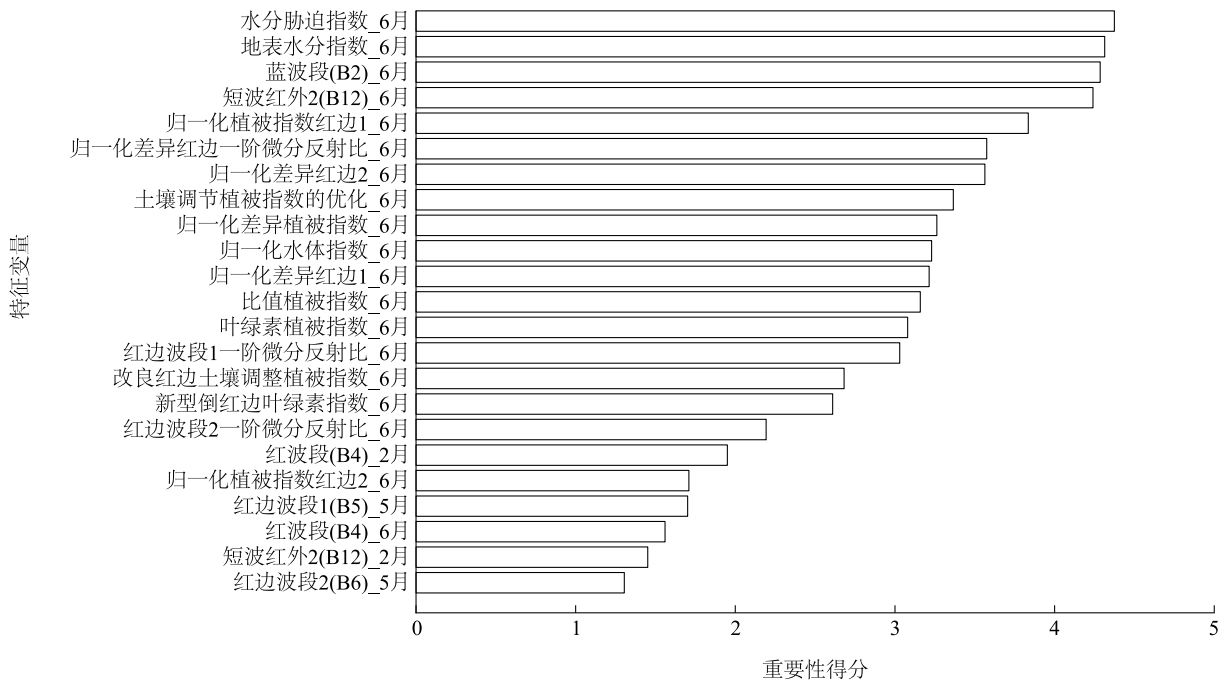
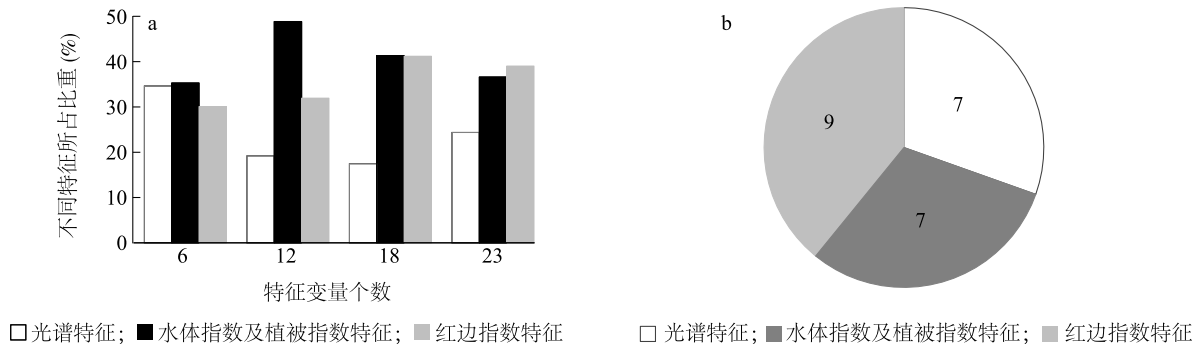


图3 优选特征变量的重要性

Fig.3 Importance of preferred characteristic variables



a: 优选特征集不同特征变量比重统计; b: 优选特征集不同特征数量统计。

图4 特征变量统计图

Fig.4 Statistical charts of characteristic variables

2.3 提取精度对比

基于不同特征变量组合方案、随机森林分类算法及训练集和验证集数据得到各分类方案下的植被分类精度如表5所示。从表5可以看出,由多时相光谱特征构成的分类方案1的总体分类精度(OA)为91.97%,Kappa系数为0.8961,茶树种植区的生产者精度(PA)和用户精度(UA)均超过90%。在光谱特征变量的基础上加入水体指数及植被指数特征变量后,方案2的茶树种植区PA和UA分别比方案1提高1.23个百分点和0.52个百分点。而在光谱特征变量的基础上加入红边指数特征变量后,方案

3的茶树种植区PA和UA分别比方案1提高0.3个百分点和2.46个百分点,提升较明显。在光谱特征变量的基础上加入纹理特征变量后,方案4的茶树种植区PA和UA分别比方案1提高0.17个百分点和0.64个百分点,提升效果相对较差,这也说明10 m分辨率影像的纹理特征对地表植被分类精度的提升作用有限。融合多时相光谱特征、水体指数及植被指数特征、红边指数特征及纹理特征114个变量的方案5茶树种植区PA比方案1增加0.81个百分点,但UA下降0.97个百分点,这可能是由于特征变量冗余导致分类精度的下降,这与张磊等^[35]研

究结果一致。而经 RF-MDA 算法优选后的方案 6 可以获得最佳的分类效果, OA 为 94.95%, $Kappa$ 为 0.934 8, 茶树种植区 PA 和 UA 分别为 93.07%、95.95%。

表 5 研究区地表植被分类精度

Table 5 Accuracy of surface vegetation classification in the study area

方案	OA (%)	$Kappa$ 系数	耕地		林地		茶树种植区		水域		人造地表		裸土	
			PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
1	91.97	0.896 1	95.84	85.42	95.57	87.39	92.39	92.35	93.86	99.66	85.50	96.15	87.75	91.16
2	92.46	0.902 3	96.70	89.78	96.48	85.58	93.62	92.87	96.18	99.72	85.48	96.57	90.69	97.85
3	91.84	0.894 6	96.35	87.96	95.15	86.37	92.69	94.81	95.44	99.36	83.49	97.49	91.17	97.78
4	92.49	0.902 8	95.18	88.57	95.23	87.09	92.56	92.99	97.22	99.33	85.24	96.62	90.85	92.62
5	92.80	0.906 7	96.06	92.03	97.01	85.99	93.20	91.38	96.35	99.42	84.21	97.77	92.44	97.81
6	94.95	0.934 8	93.68	94.95	93.75	97.76	93.07	95.95	97.20	99.51	96.65	93.72	88.46	98.41

方案 1~方案 6 见表 4。 OA 为总体分类精度; PA 为生产者精度; UA 为用户精度。

2.4 茶树种植区空间分布

由方案 6 提取的郎溪县茶树种植区分布如图 5 所示。郎溪县茶树种植区主要分布在海拔 200 m 以下具有一定地势的中部、南部和东北部地区。茶树根系不耐水, 茶树种植对地形具有一定的要求, 而坡度太大不便于茶树种植和人工养护, 因此茶树往往种植在不易蓄水且坡度较小的山坡上。郎溪县茶树种植区空间分布按乡镇可划分为 3 个等级: 十字镇、飞鲤镇、凌笪镇茶树种植较多, 累计茶树种植面积达 1 929.39 hm^2 ; 毕桥镇、姚村镇、涛城镇次之, 累计茶树种植面积为 811.21 hm^2 ; 新发镇、梅渚镇、建平镇和郎溪经济开发区茶树种植面积相对较少。

3 讨论

本研究通过实地采集 1~6 月茶树、小麦及红叶石楠高光谱数据, 明确茶树与其他植被不同的光谱特征, 为茶树信息的提取提供依据。但本研究仍存在一些局限, 如 10 m 分辨率 Sentinel-2 影像难以准确反映和有效利用茶树独特的纹理特征。因此, 后续的研究中应充分利用茶树为行植作物这一特性, 应用更高空间分辨率的遥感数据(如高分系列遥感影像等)提取纹理特征, 以进一步提高茶树种植区信息提取的精度; 另一方面由于 Sentinel-2 单星的重访周期约 10 d, 未来研究中可考虑将 Sentinel-2 与 MODIS 数据进行时空融合得到时间分辨率更高的时序数据, 以实现更为准确的茶树种植面积和分布等信息提取。

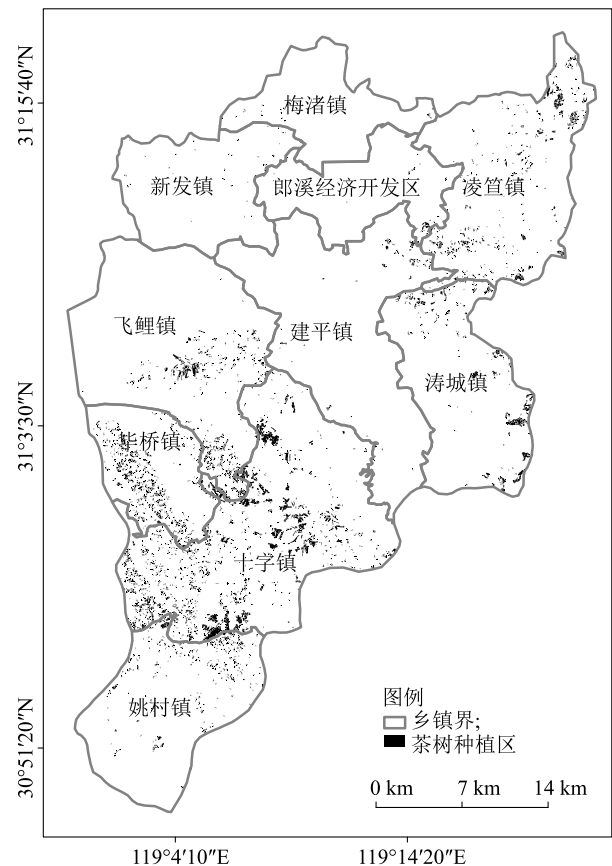


图 5 2023 年郎溪县茶树种植区空间分布

Fig.5 Spatial distribution of tea plantation areas in Langxi County in 2023

4 结论

本研究基于 2023 年 1~6 月多时相 Sentinel-2 影

像提取光谱特征、水体指数及植被指数特征、红边指数特征、纹理特征,设计6组不同的特征变量组合方案并使用随机森林算法对郎溪县地表植被进行分类,主要结论如下:(1)不同类型特征变量对于地物分类精度的贡献具有明显差异,相比较而言,红边指数特征对提升茶树种植区信息提取精度贡献最大,其次是水体指数及植被指数特征,纹理特征贡献最小,表明本研究构建的红边指数特征对茶树修剪前后的冠层变化更为敏感。(2)基于随机森林-平均精确度减少算法(RF-MDA)特征优选后的特征变量进行分类能得到最高的分类精度,总体精度达94.95%,Kappa系数为0.9348,表明RF-MDA能有效地保留识别地物重要的特征变量,避免冗余信息对分类结果的影响。

致谢:感谢国家民用空间基础设施项目对本研究数据采集与分析等方面的资助和支持!

参考文献:

- [1] 肖智,黄贤金,孟浩,等. 2009-2014年中国茶叶生产空间演变格局及变化特征[J]. 地理研究,2017,36(1):109-120.
- [2] 梁继,郑镇伟,夏诗婷,等. 高分六号红边特征的农作物识别与评估[J]. 遥感学报,2020,24(10):1168-1179.
- [3] 巫明焱,董光,税丽,等. 基于Landsat 8影像的济宁市春季主要作物种植面积变化监测[J]. 江苏农业学报,2018,34(3):559-569.
- [4] KHATAMI R, MOUNTRAKIS G, STEHMAN S V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research[J]. Remote Sensing of Environment, 2016, 177:89-100.
- [5] 徐乔,张霄,余绍淮,等. 综合多特征的极化SAR图像随机森林分类算法[J]. 遥感学报,2019,23(4):685-694.
- [6] 王娜,李强子,杜鑫,等. 单变量特征选择的苏北地区主要农作物遥感识别[J]. 遥感学报,2017,21(4):519-530.
- [7] 张猛,曾永年,朱永森. 面向对象方法的时间序列MODIS数据湿地信息提取——以洞庭湖流域为例[J]. 遥感学报,2017,21(3):479-492.
- [8] 柳文杰,曾永年,张猛. 融合时间序列环境卫星数据与物候特征的水稻种植区提取[J]. 遥感学报,2018,22(3):381-391.
- [9] XU W, HUANG S, WU C, et al. The pruning phenological phase-based method for extracting tea plantations by field hyperspectral data and Landsat time series imagery[J]. Geocarto International, 2020,37(7):2116-2136.
- [10] DAVRANCHE A, POULIN B, LEFEBVRE G. Mapping flooding regimes in Camargue wetlands using seasonal multispectral data[J]. Remote Sensing of Environment, 2013,138:165-171.
- [11] 平跃鹏,臧淑英. 基于MODIS时间序列及物候特征的农作物分类[J]. 自然资源学报,2016,31(3):503-513.
- [12] 王斌,何丙辉,林娜,等. 基于随机森林特征选择的茶园遥感提取[J]. 吉林大学学报(工学版),2022,52(7):1719-1722.
- [13] SHOKO C, MUTANGA O. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C2 and C4 grass species[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017,129:22-40.
- [14] KORHONEN L, HADI, PACKALEN P, et al. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index[J]. Remote Sensing of Environment, 2017, 195:259-274.
- [15] 郑阳,吴炳方,张森. Sentinel-2数据的冬小麦地上干生物量估算及评价[J]. 遥感学报,2017,21(2):218-228.
- [16] 陈炳环. 茶树品种分类的探讨[J]. 中国茶叶,1988(2):16-18.
- [17] 金志凤,黄敬峰,李波,等. 基于GIS及气候-土壤-地形因子的浙江省茶树栽培适宜性评价[J]. 农业工程学报,2017,27(3):231-236.
- [18] 李时睿,王治海,杨再强,等. 江南茶区茶叶生产现状和气候资源特征分析[J]. 干旱气象,2014,32(6):1007-1014.
- [19] 国家质量监督检验检疫总局,国家标准化管理委员会. 土地利用现状分类:GB/T 21010-2017[S]. 北京:中国标准出版社,2017.
- [20] 刘纪远. 中国资源环境遥感宏观调查与动态[M]. 北京:中国科学技术出版社,1996.
- [21] 张景华,封志明,姜鲁光. 土地利用/土地覆被分类系统研究进展[J]. 资源科学,2011,33(6):1195-1203.
- [22] 童庆禧,张兵,郑兰芬. 高光谱遥感-原理、技术与应用[M]. 北京:高等教育出版社,2006.
- [23] HARALICK R M, SHANMUGAM K. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973,3(6):610-621.
- [24] XIAO X M, BOLES S, LIU J Y, et al. Mapping paddy rice agriculture in Southern China using multi-temporal MODIS images[J]. Remote Sensing of Environment, 2005,95(4):480-492.
- [25] GAO B C. NDWI: a normalized difference water index for remote sensing of vegetation liquid water from space[J]. Remote Sensing of Environment, 1996,58(2):257-266.
- [26] ERJR H, ROCK B N. Detection of changes in leaf water content using near-and middle-infrared reflectances[J]. Remote Sensing of Environ, 1989,20(1):42-54.
- [27] RONDEAUX G, STEVEN M, BARET F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996,55(2):95-107.
- [28] BIRTH C S, MCVEY G. Measuring the color of growing turf with a reflectance spectrophotometer[J]. Agronomy Journal, 1968, 60: 640-643.
- [29] TUCKER C J. Red and photographic infrared linear combinations for monitoring vegetation[J]. Remote Sensing of Environment,

- 1979, 8(2):127-150.
- [30] MIURA T, HUETE A R, YOSHIOKA H. Evaluation of sensor calibration uncertainties on vegetation indices for MODIS[J]. IEEE Transactions on Geoscience & Remote Sensing, 2000, 38(3): 1399-1409.
- [31] GAMON J A, PEÑUELAS J, FIELD C B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency[J]. Remote Sensing of Environment, 1992, 41(1):25-44.
- [32] GITELSON A A, VIÑA A, ARKEBAUER T J, et al. Remote estimation of leaf area index and green leaf biomass in maize canopies[J]. Geophysical Research Letters, 2003, 30(5):1248.
- [33] DEVENTER V, WARD A D, GOWDA P H, et al. Using thematic mapper data to identify contrasting soil plains and tillage practices[J]. Photogrammetric Engineering & Remote Sensing, 1997, 62(1):87-92.
- [34] GITELSON A A, MERZLYAK M N. Remote estimation of chlorophyll content in higher plant leaves[J]. International Journal of Remote Sensing, 1997, 18(12):2691-2697.
- [35] 张磊, 宫兆宁, 王启为, 等. Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 22(2):212-226.
- [36] GITELSON A A, GRITZ Y, MERZLYAK M N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves[J]. Journal of Plant Physiology, 2002, 160(2):271-282.
- [37] MERZLYAK M N, GITELSON A A, CHIVKUNOVA O B, et al. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening[J]. Physiologia Plantarum, 1999, 106(1):125-141.
- [38] FRAMPTON W J, DASH J, WATMOUGH G, et al. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 82:82-92.
- [39] WU C Y, NIU Z, TANG Q, et al. Estimating chlorophyll content from hyperspectral vegetation indices: modeling and validation[J]. Agricultural and Forest Meteorology, 2008, 148(8/9):1220-1241.
- [40] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1):5-22.
- [41] GENUER R, POGGI J M, TULEAU-MALOT C. Variable selection using random forests[J]. Pattern Recognition Letters, 2010, 31(14):2225-2236.
- [42] CONGALTON R G. A review of assessing the accuracy of classifications of remotely sensed data[J]. Remote Sensing of Environment, 1991, 37(1):35-46.

(责任编辑:石春林)