



文献 CSTR:
32001.14.11-6035.csd.2023.0115.zh



文献 DOI:
10.11922/11-6035.csd.2023.0115.zh
数据 DOI:
10.57760/sciencedb.j00001.00825

文献分类: 地球科学

收稿日期: 2023-06-08
开放同评: 2023-07-13
录用日期: 2024-03-14
发表日期: 2024-06-27

王祎帆¹, 周小成^{1*}, 熊皓丽¹, 吴善群²,

谭芳林³, 郝优壮¹, 田国帅¹

1. 福州大学, 卫星空间信息技术综合应用国家地方联合工程研究中心, 空间数据挖掘与信息共享教育部重点实验室, 福州 350108
2. 福建省大田县林业局, 福建三明 366199
3. 福建省林业科学研究院, 福州 350012

摘要: 福建省作为中国的产茶大省, 快速准确地获取茶园的空间分布对于福建省的农业经济发展以及生态环境建设具有重大的决策意义。本研究在 GEE 云平台调用与处理 Sentinel-1 (S1) 雷达数据和 Sentinel-2 (S2) 多光谱数据, 结合地形数据从中提取光谱特征、纹理特征、地形特征等 98 个特征, 利用递归消除支持向量机算法 (SVM_RFE) 对特征变量进行筛选, 共设计 4 种特征组合方案, 通过支持向量机分类器 (SVM) 进行茶园提取, 并分别对 4 种分类方案进行精度评价, 获得了福建省 2020 年 10 m 分辨率茶园空间分布数据。在此基础上, 利用 GEE 云平台获取福建省 2000–2020 年植被干扰信息, 以 2020 年茶园提取结果掩膜剔除 2000–2015 年影像中非茶园区域, 得到 2000–2020 年每隔 5 年的福建省 10 m 分辨率茶园空间分布数据集。本数据集利用样本点对重点产茶县市进行人工验证, 结果表明: 2020 年茶园提取精度在 92% 以上, 利用干扰数据剔除法获得的 2000 年、2005 年、2010 年、2015 年茶园提取精度均在 80% 以上。提取茶园精度较高, 可为有关部门进行茶园管理提供支持。

关键词: Google Earth Engine; 茶园; 支持向量机; 植被干扰; 福建省

数据库 (集) 基本信息简介

数据库 (集) 名称	福建省 2000–2020 年 10 m 分辨率茶园空间分布数据集
数据作者	王祎帆, 周小成, 熊皓丽, 吴善群, 谭芳林, 郝优壮, 田国帅
数据通信作者	周小成 (zhouxc@fzu.edu.cn)
数据时间范围	2000–2020 年
地理区域	福建省, 地理范围为 23°33′–28°20′N、115°50′–120°40′E
空间分辨率	10 m
数据量	2.25 GB
数据格式	*.shp
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00825
基金项目	福建省科技厅高校产学研合作项目 (2022N5008)
数据库 (集) 组成	本数据集以五年为一个时间节点, 包括 2000 年、2005 年、2010 年、2015 年、2020 年共 5 个年份的茶园空间分布数据, 数据保存为 *.shp 格式。

* 论文通信作者
周小成: zhouxc@fzu.edu.cn

引言

福建省作为重点产茶大省，全省茶园面积约为 341 万亩，茶园面积总量以及茶叶产量居于全国第五^[1]。精准地确定茶园所处地理位置，较为准确地获取茶园的种植面积，因地制宜发展茶园产业，对于相关部门进行茶园总体规划、茶园的估产以及生态评价工作具有至关重要的作用^[2]。

传统的茶园监测手段主要为人工野外实践考察，这种方法不仅成本高、效率低，而且很难获得持续性观测数据^[3]，不能保证茶园监测的时效性。利用遥感技术可进行大面积观测，且基于其时效性的特征，可以更加准确和高效地提取茶园数据。茶园作为多年生的常绿植物，仅依靠光谱特征进行选择提取，易与幼林地、梯田等植被类型相混淆，不能够保证茶园提取的精确度。因此本研究综合考虑提取多源特征，采用适当的特征选择算法筛选出对提取茶园有益的特征，减少因特征冗余而导致的茶园提取精度低、分类效率差的问题。Sentinel-2 光学影像对于植被分类具有重要的作用，但受云层或天气问题影响严重^[4-5]，Sentinel-1 雷达影像则可以穿透云层获得较为精确的地表纹理信息，弥补了光学影像的不足^[6-7]。Google Earth Engine (GEE) 云平台则为用户结合自身的应用需求，进行海量数据的挖掘与在线计算处理提供了可能^[8]。

本数据基于 GEE 平台，综合 Sentinel-1、Sentinel-2 影像以及地形数据，综合构建多源特征，利用递归消除支持向量机算法 (SVM_RFE) 提取最优特征组合，采用支持向量机分类器进行茶园分类提取，得到福建省 2020 年 10 m 分辨率茶园种植区的空间分布图。利用 GEE 云平台获取福建省 2000–2020 年植被干扰信息，在 2020 年茶园数据基础上，掩膜剔除 2000–2015 年影像中非茶园区域，最终得到 2000–2020 年福建省 10 m 分辨率茶园空间分布数据集。

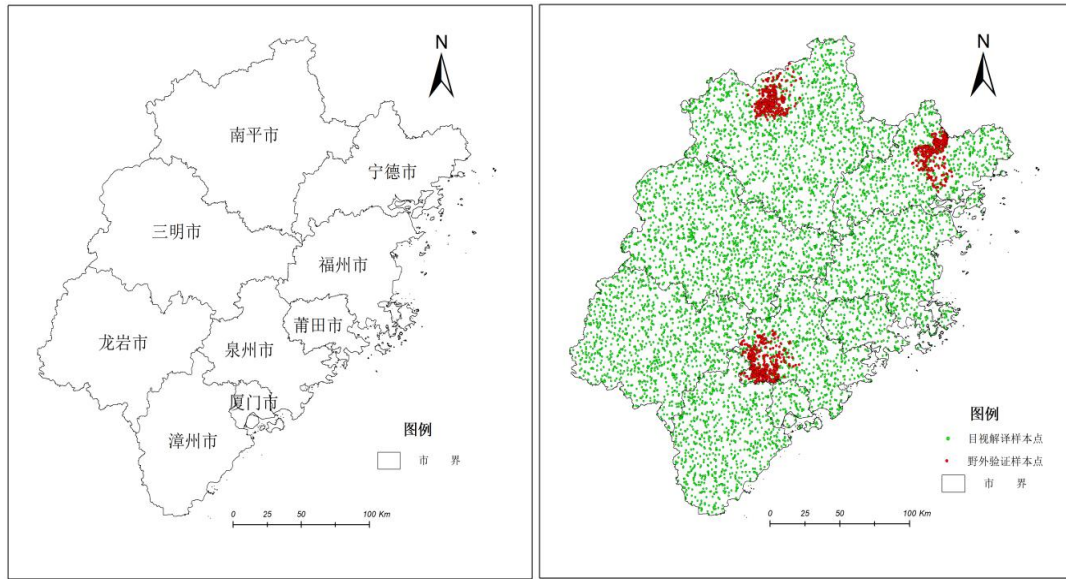
1 数据采集和处理方法

1.1 数据源

在 GEE 云平台上通过在线编写代码实现调用和处理：Sentinel-1 (S1) 雷达数据和 Sentinel-2 (S2) 多光谱数据（为了保证能够有效地观测影像，设置影像的获取时间为 2019 年 9 月 10 日到 2019 年 11 月 10 日，共获得 S1 影像 64 景、S2 影像 112 景）；Landsat 系列影像（筛选出 2000–2020 年以来覆盖福建省的 Landsat 系列影像，包括 Landsat-5、Landsat-7 以及 Landsat-8）。使用 SRTMGL1_003 数字产品作为地形数据，空间分辨率为 30 m。样本数据包括：基于福建省矢量边界随机生成样本点数据 7046 个，通过野外 GPS 实地采集得到武夷山市、福安市和安溪县的样本点数据 1062 个，共计有 8108 个样本点数据，其中人工地表 1120 个（包含 134 个实地样点）、耕地 1030 个（包含 157 个实地样点）、林地 3509 个（包含 315 个实地样点）、茶园 1953 个（包含 406 个实地样点）、水体 496 个（包含 47 个实地样点），将其中的 30% 作为验证样本点，样本点详细分布见图 1。

1.2 数据处理

本数据集处理过程主要包括数据预处理、特征提取、特征组合、茶园分类、干扰数据剔除法获取时序数据集 5 部分，详细的技术流程见图 2。



(a) 空白辅助图

(b) 样本点空间分布

图 1 野外验证样本点和目视解译样本点空间分布

Figure 1 Spatial distribution of field validation sample points and visual interpretation sample points

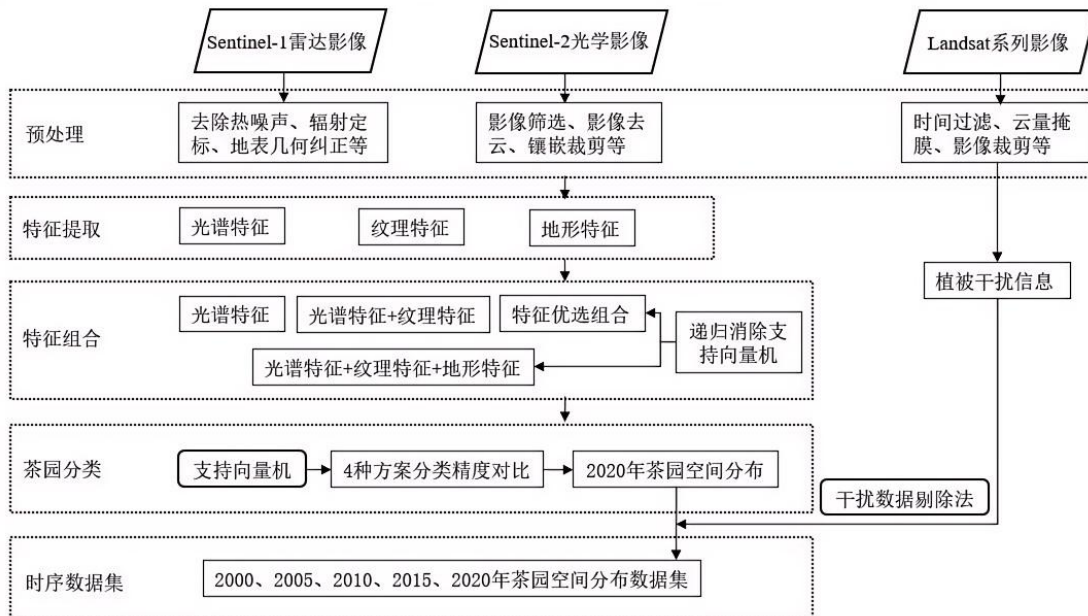


图 2 技术路线图

Figure 2 Technical roadmap

1.2.1 数据预处理

S1 数据使用 Sentinel-1 工具箱完成预处理，即实现标定轨道参数、去除边界噪声和热噪声、辐射定标等^[9]，之后将影像进行合成。S2 数据主要进行去云处理，最终得到 112 景无云的 S2 影像并将其合成。在 GEE 中使用 `resample (mode)` 和 `reproject (crs, crsTransform, scale)` 函数将 30 m 空间分辨率的地形数据转换为 10 m 分辨率，之后按照福建省行政边界进行裁剪处理。在 GEE 云平台上筛选 2000–2020 年 7 至 10 月的 Landsat 长时序影像，并利用 QA 波段进行云掩膜，用邻近月份替换云层覆盖数据。

1.2.2 特征提取

本研究通过分析不同的数据源，共构建了 3 种不同的特征变量，其中包括光谱特征 26 个，纹理特征 68 个，地形特征 4 个。以上在 GEE 平台中计算的特征共有 98 个（表 1）。其中 $B1$ 、 $B2$ 、 $B3$ 、 $B4$ 、 $B5$ 、 $B6$ 、 $B7$ 、 $B8$ 、 $B8A$ 、 $B9$ 、 $B10$ 、 $B11$ 和 $B12$ 分别为超蓝（沿海和气溶胶）波段、蓝色波段、绿色波段、红色波段、植被红边波段 1、植被红边波段 2、植被红边波段 3、宽带近红外波段、窄带近红外波段、短波红外波段 1、短波红外波段 2、短波红外波段 3 和短波红外波段 4 的反射率。 $NDVI$ 、 $NDWI$ 、 $LSWI$ 、 $NDTI$ 、 $MNDVI$ 分别表示归一化植被指数、归一化水体指数、地表水分指数、归一化差异耕作指数、修正型归一化植被指数。 $IRECI$ 、 $MTCI$ 、 $NDVI_{re1}$ 、 $NDVI_{re2}$ 、 $NDVI_{re3}$ 、 ND_{re1} 、 ND_{re2} 、 CI_{re} 分别表示新型倒红边叶绿素指数、地面叶绿素指数、归一化植被指数红边 1、归一化植被指数红边 2、归一化植被指数红边 3、归一化差异红边 1、归一化差异红边 2、红边叶绿素指数。 $Elevation$ 、 $Slope$ 、 $Aspect$ 、 $Hillshade$ 分别表示海拔、坡度、坡向、山体阴影。

表 1 本研究涉及的特征信息汇总

Table 1 Summary of feature information used in this research

数据源	特征名称	特征说明或介绍	特征数目
S2	原始光谱特征	$B1$, $B2$, $B3$, $B4$, $B5$, $B6$, $B7$, $B8$, $B8A$, $B9$, $B10$, $B11$, $B12$	13
	植被指数特征	$NDVI$, $NDWI$, $LSWI$, $NDTI$, $MNDVI$	5
	红边指数特征	$IRECI$, $MTCI$, $NDVI_{re1}$, $NDVI_{re2}$, $NDVI_{re3}$, ND_{re1} , ND_{re2} , CI_{re}	8
	$B8$ 波段纹理特征	NIR 近红外波段的 17 个纹理特征	17
	$NDTI$ 指数纹理特征	$NDTI$ 指数的 17 个纹理特征	17
S1	雷达纹理特征	VV 和 VH 的 17 个纹理特征	34
SRTM	地形特征	$Elevation$, $Slope$, $Aspect$, $Hillshade$	4
合计			98

1.2.3 特征组合

为了能够获得更加精确的茶园提取结果，本研究共设计了 4 种实验方案（表 2），其中方案 4 为特征优选组合，是利用 SVM_RFE 特征选择算法选择出对茶园提取最为重要的特征变量，避免了特征冗余造成的提取精度和效率低的问题^[10]。SVM_RFE 为递归消除支持向量机算法，该特征选择算法的大体流程为：首先将所有特征作为初始特征，利用 SVM 计算特征权重，将特征根据权重值大小排序，每次去除一个权重最小的特征，同时使用 SVM 进行分类，使用总体精度和 Kappa 系数作为精度评价标准，按此步骤进行迭代直到特征集为空即止，最后输出精度评价最高的特征组合^[11]。经过 SVM_RFE 特征筛选之后共得到 18 个特征变量作为方案 4 的特征优选组合，其中包括 13 个光谱特征、4 个纹理特征、1 个地形特征。

表 2 实验方案信息表

Table 2 Details of experimental programs

实验方案	特征组合
1	光谱特征
2	光谱特征+纹理特征

实验方案	特征组合
3	光谱特征+纹理特征+地形特征
4	SVM_RFE 特征优选

1.2.4 茶园分类

利用支持向量机分类器进行分类，提取茶园数据，之后利用混淆矩阵分别对 4 种分类方案进行精度评价（表 3）。主要参考数值为生产者精度、用户精度、总体精度和 Kappa 系数。由精度评价表 3 可知，方案 1 单独利用光谱特征提取茶园信息，可以得到较高的分类精度，但是耕地和茶园的 PA 和 UA 较低，说明仅利用光谱特征不能有效区分茶园和耕地。方案 2 加入纹理特征，使得茶园与耕地的 PA 与 UA 有所提高，说明利用纹理特征可以有效提高茶园与耕地的分类精度。方案 3 在方案 2 基础上又加入了地形特征，使得耕地的 PA 和 UA 进一步提升，进一步减少了茶园错分现象。经过特征优选后的方案 4，在大幅度降低数据维度的同时，又能保证较高的精度，其提取精度最高。最终获得福建省 2020 年 10 m 分辨率的茶园专题空间分布图。

表 3 2020 年不同实验方案的分类精度统计

Table 3 Classification accuracy statistics of various experimental schemes in 2020

实验方案	OA/%	Kappa 系数	人工地表		耕地		林地		茶园		水体	
			PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%
1	89.42	0.85	93.70	96.07	83.28	80.38	92.28	91.74	83.45	84.31	95.33	96.62
2	93.04	0.90	94.79	96.11	87.54	90.20	95.11	94.01	90.27	90.12	96.67	96.03
3	93.87	0.92	94.79	96.11	91.15	91.45	96.09	94.70	90.27	91.52	96.00	96.64
4	94.65	0.93	95.62	96.14	93.11	92.51	96.19	95.44	91.64	92.91	96.67	96.67

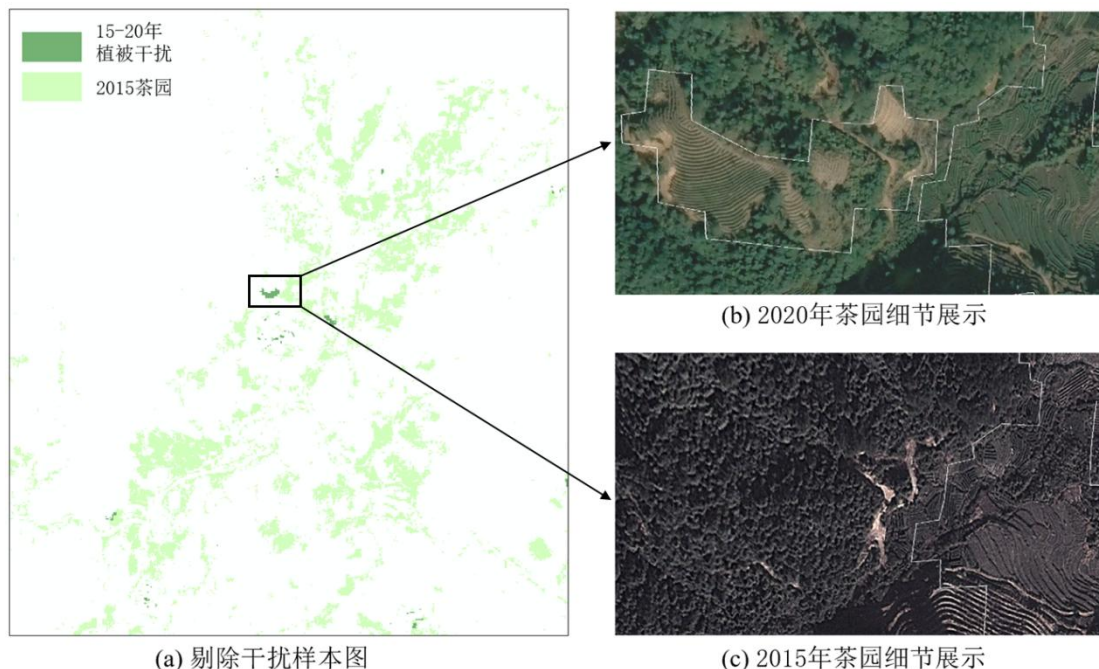
注：表中 OA、PA 和 UA 分别代表总体精度、生产者精度和用户精度。

1.2.5 干扰数据剔除法获取时序数据集

通过实地考察发现，福建省茶园面积在近 20 年里一直处于不断增加的趋势，茶园分布不断扩张。因此本研究采用干扰数据剔除法，利用获取的植被干扰信息对 2020 年以及更早期的茶园分布结果进行掩膜，依次获得 2000 年、2005 年、2010 年、2015 年的茶园空间分布。如下图 3（a）是截取了部分利用干扰数据剔除法对 2020 年茶园提取结果进行掩膜的样本图，其中图 3（b）是利用支持向量机法提取得到的 2020 年茶园结果，图 3（c）是由干扰数据剔除法得到的 2015 年茶园结果。

干扰数据剔除法实现的具体步骤为：基于长时间序列 Landsat 系列卫星数据，在 GEE 云平台中采用 LandTrendr 算法检测 Landsat 时序影像的变化^[12]，得到植被扰动与恢复的时间节点，将福建省植被划分为干扰区和非干扰区，由此获得的数据包括 2000–2020 年每一年的植被干扰信息。每 5 年设置一个干扰节点进行植被干扰信息的合并，分别得到 2000–2004 年、2005–2009 年、2010–2014 年、2015–2019 年植被干扰信息。以获取 2015 年茶园空间分布数据的步骤为例：根据茶园扩张的先验知识，利用 2015–2019 年植被干扰信息与 2020 年茶园专题进行叠加分析，剔除茶园专题范围内与干扰信息重叠的图斑，即剔除掉非茶园范围，进而得到 2015 年茶园空间分布数据。按照上述方式依次对更早期茶园数据进行叠加分析和剔除操作后，得到 2000 年、2005 年、2010 年茶园空间分布时序数据集。

经过实地考察验证发现，茶园虽呈现逐步扩张的趋势，但是每年的变化量很少，仅个别县市茶园变化量相对较多。且本研究将 2020 年 10 m 分辨率的茶园数据作为掩膜对象，掩膜后导入 Google earth 进行验证发现，2020 年以前的茶园数据基本接近 10 m 分辨率。因此，可认为经过干扰数据剔除法得到的 2000–2015 年茶园数据分辨率为 10 m。



注：（a）是截取了部分利用植被干扰信息对 2020 年茶园提取结果进行掩膜的样本图，（b）是利用支持向量机法提取得到的 2020 年茶园结果，（c）是由于干扰数据剔除法得到的 2015 年茶园结果。

图 3 剔除干扰样本图

Figure 3 Sample graph after removing interference

2 数据样本描述

基于上述数据处理方法和流程，最终得到福建省 2000–2020 年 10 m 分辨率的茶园空间分布数据集，其空间分辨率为 10 m，投影坐标系为 WGS_1984_UTM_Zone_50N。将全部结果存放于一个文件夹中，命名为“FJ_tea_10 m”。其中的文件格式是 shp 的面状数据，文件命名方式为“tea_日期.shp”，如 2020 年茶园空间分布数据，命名为“tea_2020.shp”，数据示例如图 4 所示。

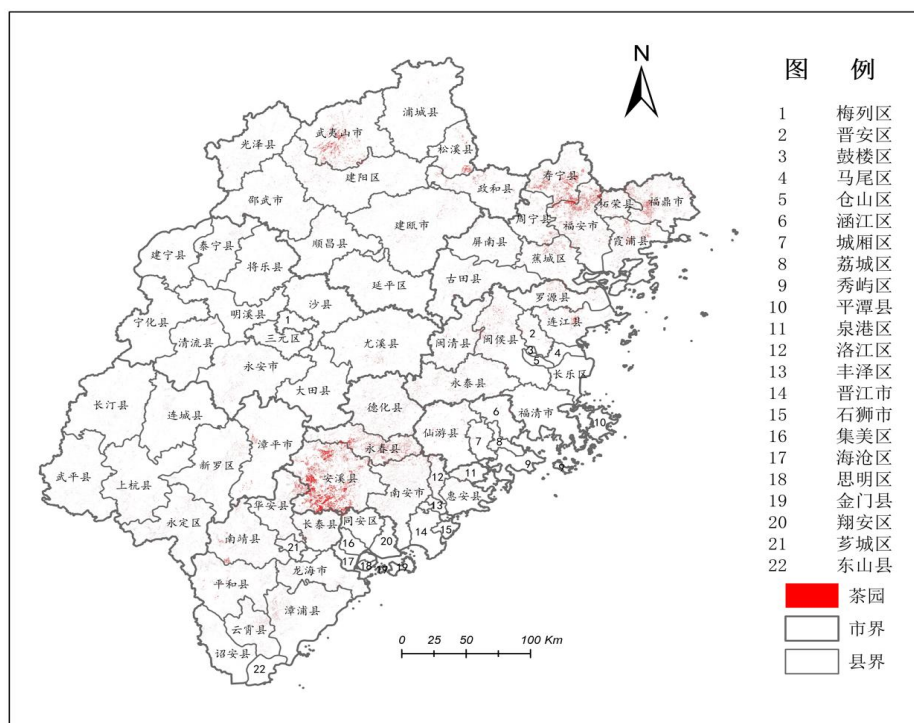
3 数据质量控制和评估

鉴于 2020 年茶园数据与其余年份的茶园时序数据的数据源以及数据处理和获取方法有所差异，在此采用不同的方式进行茶园提取结果的精度验证。

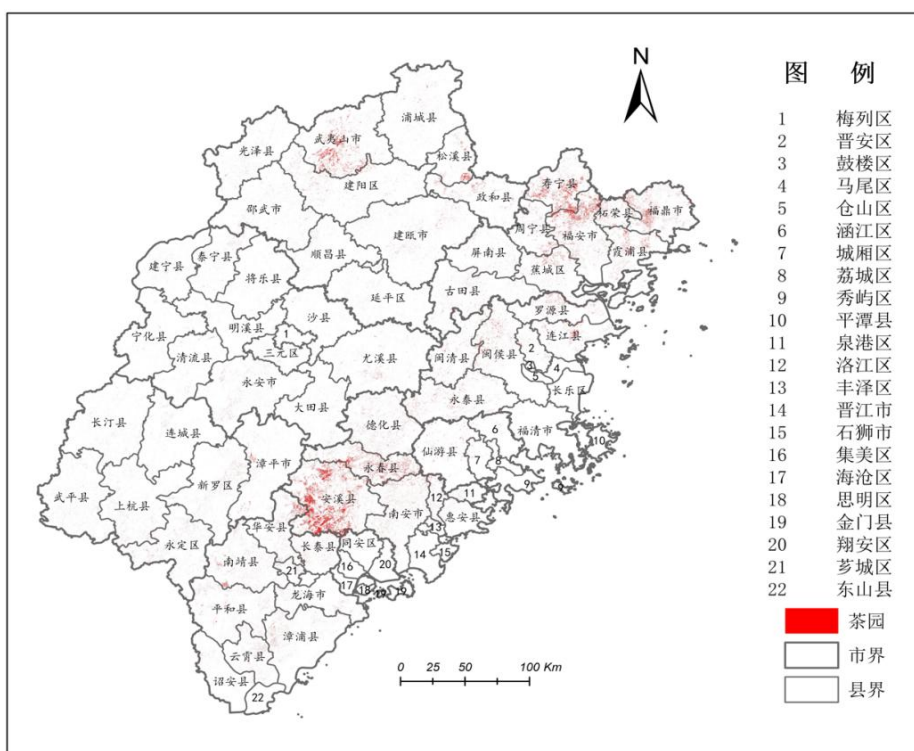
3.1 2020 年茶园空间分布数据精度验证

对利用支持向量机法求得的 2020 年茶园数据进行精度评估，除了在上文中提到的使用混淆矩阵的方法外，本研究还采用人工验证的方法进行更为细致的精度评价。鉴于在提取结果验证的过程中所选择的验证点在各县市分布均匀且数目较多，所验证精度均在 92% 以上，因此不再选择其余验证

点进行验证，而是提取茶园矢量，直接导入至 Google earth 中对相关地区进行人工验证，看是否存在漏分和错分的现象，这样进行的精度验证也能很容易地分析产生提取误差的原因。



(a) 2010 年茶园



(b) 2020 年茶园

图 4 福建省部分年份茶园空间分布 (审图号: GS 京 (2024) 0933 号)

Figure 4 Spatial distribution of tea plantations in 2010 and 2020 in Fujian Province

综合考虑茶园分布面积、地理位置等相关因素，选取了具有代表性的 6 个县市进行进一步的人工验证，6 个县市分别是安溪县、福安市、福鼎市、武夷山市、大田县、建阳区：

(1) 安溪县（茶园提取面积为 68.4525 万亩）：总体验证精度高，存在少数错分的现象。安溪县分布茶园面积相对较大，其中大部分为连片的、整块分布的、面积较大的茶园，对于这些茶园，提取精度很高；对于分布比较破碎且靠近居民地的小块茶园，容易与耕地以及幼林地混淆，发生错分的情况。

(2) 福安市（茶园提取面积为 21.1680 万亩）：总体验证精度高，存在部分错分的现象，也有少数漏分的情况。福安市北部茶园分布密集，对于大块茶园提取的精度很高；而福安市南部茶园面积小，且分布很细碎，则出现较多的错分现象，将采伐迹地以及幼林地错分为茶园，也存在一些漏分的现象，比如小块茶园边界框不全；域内沿河地区会出现错分现象，将幼林地错分为茶园。

(3) 福鼎市（茶园提取面积为 19.9755 万亩）：总体验证精度高，存在少数的错分与漏分现象。福鼎市茶园分布密集，对于大块茶园提取的精度很高；福鼎市南部茶园面积小，分布较为细碎，存在一些漏分的现象，提取茶园不完整；福鼎市北部茶园面积小，分布很细碎，会出现较多的错分现象，易与采伐迹地以及幼林地发生混淆，部分靠近居民地的区域，会把耕地错分为茶园。

(4) 武夷山市（茶园提取面积为 17.9301 万亩）：总体验证精度高，存在少数的错分与漏分现象。武夷山市茶园主要分布在中部和南部，对于面积较大的、连片分布的茶园，提取精度很高，对于面积较小、分布较为细碎的茶园则会出现部分漏分现象；武夷山市茶园主要分布在靠近耕地的平地区域，会出现部分把耕地错分为茶园的现象。

(5) 大田县（茶园提取面积为 6.4005 万亩）：总体验证精度高，存在少数的漏分现象。因在全域范围内茶园分布均非常细碎，几乎全部为小块茶园，所以提取困难较大，会出现部分漏分的现象。但经过验证后发现实际漏分的茶园并不多，数据精度较高。

(6) 建阳区（茶园提取面积为 0.3528 万亩）：总体验证精度较高，存在部分漏分的现象。建阳区西部以及西北部存在较多的漏分现象，这些茶园分布于平原地区，且靠近耕地，因此较多茶园被错分为耕地；中部以及东部茶园分布非常细碎，一些采伐迹地被错分为茶园，还有溪流两岸的一些幼林地也被错分为茶园。

中等分辨率的 Sentinel-2 影像能较为精准识别 15 亩以上的茶园，现以大田县和安溪县为例进行验证：提取面积大于等于 15 亩（1 公顷）的茶园，将其导入至 Google earth 上进行验证。验证结果表明可以较为精确地提取 15 亩左右的茶园。由于大田县是非重点的产茶县市，茶园分布十分破碎，面积大于 15 亩的茶园所占比例很小，因此大部分小面积的破碎茶园（<15 亩）提取精度较差，会出现漏分与错分现象。而安溪县作为重点产茶县市，茶园分布十分广泛，大部分茶园分布面积大于 15 亩，因此经验证发现提取结果较为精确。

3.2 2000–2015 年茶园空间分布数据精度验证

综合考虑茶园分布面积、地理位置等相关因素，选择安溪县、福安市、武夷山市 3 个重点产茶县市进行验证。利用干扰数据剔除法获得了历年的茶园空间分布数据，提取出茶园矢量，并在此基础上利用 ArcGIS 软件随机生成点，将点数据导入至谷歌地球中逐点进行验证并标记，将最后结果导出得到各县市的验证精度（表 4）。

表 4 2000–2015 年茶园验证精度统计

Table 4 Statistics of verification accuracy of tea plantations from 2000 to 2015

县市	茶园验证点/个				验证点总数/个				验证精度/%			
	2015	2010	2005	2000	2015	2010	2005	2000	2015	2010	2005	2000
安溪县	325	311	104	99	359	357	125	118	90.53	87.11	83.20	83.89
福安市	316	311	109	105	346	346	128	124	91.33	87.36	85.15	84.68
武夷山市	350	265	136	131	387	307	168	161	90.44	86.32	80.95	81.37

4 数据价值

本数据是基于 GEE 云平台首次实现对福建省 10 m 分辨率茶园信息的提取。在数据源方面，使用多源数据，结合利用 S1 雷达影像和 S2 光学影像，减少了云雨天气对数据处理的影响；在数据处理方面，利用 GEE 云平台进行数据预处理和 Landsat 时序数据的处理，大大减轻了数据处理的工作量，提高了数据计算的效率；在数据质量控制方面，经过精细的人工验证，使得数据集具有较高的精度保障。本数据集不仅能够为福建省的茶园产业发展和生态环境的保护提供较为及时的预测以及重要的决策支持，而且能够为大范围茶园提取的后续研究提供基础的数据支撑。

5 数据使用方法和建议

本数据集保存为 SHP 格式，能够在 ENVI、ArcGIS 和 QGIS 等 GIS 与遥感主流软件中进行相关读取和处理操作。本研究的提取结果可为福建省茶叶产业发展和预测，为生态环境的评价提供重要的决策支持。也可以结合本研究中采取的数据处理的过程与方法，考虑利用 Landsat 数据对茶园进行长时间序列研究。

数据作者分工职责

王祎帆（1999—），女，河北省唐山市人，在读硕士，研究方向为环境与资源遥感。主要承担工作：数据处理、论文撰写。

周小成（1977—），男，陕西省渭南市人，博士，副研究员，研究方向为环境与资源遥感。主要承担工作：实验思路与方案设计、论文修改。

熊皓丽（1996—），女，安徽省池州市人，硕士生，研究方向为遥感信息处理与应用。主要承担工作：数据收集、数据处理。

吴善群（1973—），男，福建省三明市人，学士，营林高级工程师，研究方向为森林资源保护及培育。主要承担工作：协助完成数据采集工作。

谭芳林（1970—），男，甘肃省定西市人，博士，教授级高级工程师，研究方向为森林生态、湿地生态。主要承担工作：协助完成数据采集工作。

郝优壮（1995—），男，河南省洛阳市人，硕士生，研究方向为环境与资源遥感。主要承担工作：数据收集与整理。

田国帅（1998—），男，河南省三门峡市人，在读硕士，研究方向为环境与资源遥感。主要承担工

作：数据收集与整理。

参考文献

- [1] 国家统计局. 分省年度数据, 2008–2017[EB/OL] (2017-09-04) [2023-06-02]. <https://data.stats.gov.cn/easyquery.htm?cn=C01>. [State Statistics Bureau. Provincial annual data, 2008–2017[EB/OL].(2017-09-04) [2023-06-02] <http://data.stats.gov.cn/easyquery.htm?cn=C01>.]
- [2] SINGH A, DUTTA R, STEIN A, et al. A wavelet-based approach for monitoring plantation crops tea: *Camellia sinensis* in North East India[J]. International Journal of Remote Sensing, 2012, 33(16): 4982–5008. DOI: 10.1080/01431161.2012.657364.
- [3] QIN Y W, XIAO X M, DONG J W, et al. Mapping paddy rice planting area in cold temperate climate region through analysis of time series Landsat 8 (OLI), Landsat 7 (ETM+) and MODIS imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 105: 220–233. DOI: 10.1016/j.isprsjprs.2015.04.008
- [4] XU F, LI Z F, ZHANG S Y, et al. Mapping winter wheat with combinations of temporally aggregated sentinel-2 and landsat-8 data in Shandong Province, China[J]. Remote Sensing, 2020, 12(12): 2065. DOI: 10.3390/rs12122065.
- [5] 郭交, 朱琳, 靳标. 基于 Sentinel-1 和 Sentinel-2 数据融合的农作物分类[J]. 农业机械学报, 2018, 49(4): 192–198. DOI: 10.6041/j.issn.1000-1298.2018.04.022. [GUO J, ZHU L, JIN B. Crop classification based on data fusion of sentinel-1 and sentinel-2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4): 192–198. DOI: 10.6041/j.issn.1000-1298.2018.04.022.]
- [6] 谭深, 吴炳方, 张鑫. 基于 Google Earth Engine 与多源遥感数据的海南水稻分类研究[J]. 地球信息科学学报, 2019, 21(6): 937–947. DOI: 10.11975/j.issn.1002-6819.2018.01.010. [TAN S, WU B F, ZHANG X. Mapping paddy rice in the Hainan Province using both google earth engine and remote sensing images[J]. Journal of Geo-Information Science, 2019, 21(6): 937–947. DOI: 10.11975/j.issn.1002-6819.2018.01.010.]
- [7] 常文涛, 王浩, 宁晓刚, 等. 融合 Sentinel-2 红边波段和 Sentinel-1 雷达波段影像的扎龙湿地信息提取[J]. 湿地科学, 2020, 18(1): 10–19. DOI: 10.13248/j.cnki.wetlandsci.2020.01.002. [CHANG W T, WANG H, NING X G, et al. Extraction of Zhalong wetlands information based on images of sentinel-2 red-edge bands and sentinel-1 radar bands[J]. Wetland Science, 2020, 18(1): 10–19. DOI: 10.13248/j.cnki.wetlandsci.2020.01.002.]
- [8] 徐晗泽宇, 刘冲, 王军邦, 等. Google Earth Engine 平台支持下的赣南柑橘果园遥感提取研究[J]. 地球信息科学学报, 2018, 20(3): 396–404. [XU H Z Y, LIU C, WANG J B, et al. Study on extraction of *Citrus orchard* in Gannan region based on google earth engine platform[J]. Journal of Geo-Information Science, 2018, 20(3): 396–404.]
- [9] GULÁCSI A, KOVÁCS F. Sentinel-1-imagery-based high-resolution water cover detection on wetlands, aided by google earth engine[J]. Remote Sensing, 2020, 12(10): 1614. DOI: 10.3390/rs12101614.
- [10] Wu Y S, Zhang X L. Object-based tree species classification using airborne hyperspectral images and lidar data [J]. Forests, 2019, 11(1): 32.

- [11] Cheng K, Wang J L. Forest type classification based on integrated spectral- spatial- temporal features and random forest algorithm: A case study in the Qinling mountains [J]. *Forests*, 2019,10(7):559
- [12] KENNEDY R E, YANG Z Q, COHEN W B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. Land Trendr—Temporal segmentation algorithms[J]. *Remote Sensing of Environment*, 2010, 114(12): 2897–2910. DOI: 10.1016/j.rse.2010.07.008.

论文引用格式

王祎帆, 周小成, 熊皓丽, 等. 福建省2000–2020年10 m分辨率茶园空间分布数据集[J/OL]. 中国科学数据, 2024, 9(2). (2024-06-27). DOI: 10.11922/11-6035.csd.2023.0115.zh

数据引用格式

王祎帆, 周小成, 熊皓丽, 等. 福建省2000–2020年10 m分辨率茶园空间分布数据集[DS/OL]. V1. Science Data Bank, 2023. (2023-06-13). DOI: 10.57760/sciencedb.j00001.00825.

A dataset of spatial distribution of tea plantations at 10 m resolution in Fujian Province from 2000 to 2020

WANG Yifan¹, ZHOU Xiaocheng^{1*}, XIONG Haoli¹, WU Shanqun²,
TAN Fanglin³, HAO Youzhuang¹, TIAN Guoshuai¹

1. National Engineering Research Center of Satellite Spatial Information Technology and Applications, Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou University, Fuzhou 350108, P. R. China

2. Datian Forestry Bureau of Fujian Province, Sanming 366199, P. R. China

3. Fujian Academy of Forestry, Fuzhou 350012, P. R. China

*Email: zhouxc@fzu.edu.cn

Abstract: Being a major tea producing province in China, Fujian Province is badly in need of quick and accurate spatial distribution of tea plantation for decision-making in both the agricultural economic development and the ecological environment construction for the province. This study retrieved and processed Sentinel-1 (S1) radar data and Sentinel-2 (S2) multispectral data on the GEE cloud platform; and it extracted 98 features, such as spectral features, texture features, and terrain features from the terrain data. The recursive elimination support vector machine algorithm (SVM RFE) is used to screen the characteristic variables, resulting in the creation of 4 feature combination schemes. Using a support vector machine (SVM) classifier to extract the distribution data of tea plantations and assessing the accuracy of the 4 feature combination schemes, we obtained the spatial distribution data of tea plantations at 10 m resolution in Fujian Province in 2020. On this basis, we used the GEE cloud platform to access the vegetation disturbance information in Fujian Province from 2000 to 2020. We finally obtained a dataset of spatial distribution of tea plantations at 10 m resolution in Fujian Province from 2000 to 2020 by excluding the

non-tea plantation areas from the images between 2000 and 2015 with the mask generated from the 2020 tea garden extraction results. This dataset has been manually validated using sample points from key tea-producing counties and townships. The results indicate an extraction accuracy of over 92% for tea plantations in 2020. The extraction accuracy of tea plantations obtained using interference data removal method in 2000, 2005, 2010, and 2015 was all above 80%. The dataset with a high accuracy in tea plantation extraction can provide support for relevant departments in tea plantation management.

Keywords: Google Earth Engine; tea plantation; support vector machines; vegetation disturbance; Fujian Province

Dataset Profile

Title	A dataset of spatial distribution of tea plantations at 10 m resolution in Fujian Province from 2000 to 2020
Data authors	WANG Yifan, ZHOU Xiaocheng, XIONG Haoli, WU Shanqun, TAN Fanglin, HAO Youzhuang, TIAN Guoshuai
Data corresponding author	ZHOU Xiaocheng (zhouxc@fzu.edu.cn)
Time range	2000–2020
Geographical scope	Fujian Province (23°33'–28°20' N, 115°50'–120°40' E)
Spatial resolution	10 m
Data volume	2.24 GB
Data format	*.shp
Data service system	< https://doi.org/10.57760/sciencedb.j00001.00825 >
Source of funding	Fujian Provincial Department of Science and Technology University Industry Student Cooperation Project (2022N5008)
Dataset composition	This dataset is organized with five years as a time node, including the spatial distribution data of tea gardens in 2000, 2005, 2010, 2015 and 2020, stored in *.shp format.