



文献 DOI:

10.11922/csdata.2018.0091.zh

数据 DOI:

10.11922/sciencedb.711

文献分类: 地球科学

收稿日期: 2018-12-14

开放同评: 2019-01-16

录用日期: 2019-04-15

发表日期: 2019-05-09

基于多源遥感数据的海南岛天然林分类数据集

史健康¹, 宫晨^{2*}, 李新武², 万祥星², 孙中昶²

1. 海南省环境科学研究院, 海南海口 571126

2. 中国科学院遥感与数字地球研究所, 北京 100094

摘要: 热带森林是全球物种最为丰富的森林生态系统, 主要分布在赤道附近的热带气候区域, 在全球气候变化中发挥着重要作用。因此, 对热带森林进行资源调查及监测具有十分重要的意义, 其中天然林的分类研究不仅能够为热带森林的调查提供数据支持, 也能够为森林的物种多样性研究提供依据。本数据集利用 Sentinel-1A 卫星提供的双极化 SAR 数据和 Landsat-8 卫星提供的光学遥感数据对海南岛热带天然林进行了分类研究。首先通过分析光学遥感数据的单波段特征、多波段特征、归一化植被指数 (NDVI) 以及 SAR 数据的单时相、多时相后向散射特征, 选取了适合天然林分类的光学特征和后向散射特征。随后利用选取的光学特征和后向散射特征, 采用支持向量机 (Support Vector Machine, SVM) 分类算法对海南岛的天然林范围进行提取, 在此基础上对其内部林型进行分类, 将其分为典型热带雨林、热带季雨林、常绿针叶林、落叶阔叶混交林以及海岸林 5 种森林类型。此外, 结合野外实地采集的验证数据以及海南林业调查资料对分类结果进行精度验证和评价, 其中利用支持向量机分类精度达到了 80% 以上。本数据集为海南岛热带天然林分类提供了可靠的遥感分类方法, 同时对于其他地区的热带天然林分类研究具有一定的参考价值。

关键词: 多源遥感数据; 热带森林; 遥感分类; 支持向量机

数据库 (集) 基本信息简介

数据库 (集) 名称	基于多源遥感数据的海南岛天然林分类数据集
数据作者	史健康、宫晨、李新武、万祥星、孙中昶
数据通信作者	宫晨 (gongchen@radi.ac.cn)
数据时间范围	2015年
地理区域	北纬18°10'–20°10', 东经108°37'–111°03'
空间分辨率	30 m
数据量	1.20 GB
数据格式	*.bat, *.tif, *.shp, *.xml
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/711
基金项目	海南省重大科技计划项目 (ZDKJ2016021); 海南省自然科学基金面上基金项目 (20164177)。

* 论文通信作者

宫晨: gongchen@radi.ac.cn

数据库（集）组成	数据集由2个压缩包组成，文件以产品名命名： （1）Classification_of_Hainan_Island_Land_cover.zip是海南岛土地初级利用分类产品数据，数据量约1.20 GB； （2）Classification_of_Hainan_Island_Natural_Forest.zip是海南岛天然林分类产品数据，数据量约15.5 MB。
----------	--

引言

森林是全球生态系统的重要组成部分，也是国家可持续发展的重要物质基础和人类社会赖以生存的宝贵资源，随着人类活动的破坏及全球气候的变化，使得森林资源遭到了不同程度的破坏，并且面临着越来越多的威胁^[1-2]。目前，传统的森林资源调查手段在森林资源的调查和监测中依然占据着重要地位。但是随着遥感技术的进步，利用其对森林资源进行调查正逐渐成为林业调查的一种重要技术手段。相比于传统林业调查，遥感技术可以快速、宏观、准确地完成对森林地区信息的收集，通过遥感图像的处理、解译和分析等手段可以对目标区域的森林资源进行定性或者定量的调查。遥感图像分类是遥感技术在林业领域应用的一个重要内容，针对不同来源、不同形式的遥感图像进行分类不仅为森林资源调查和监测提供了有力支持，而且在森林制图、林业规划、森林火灾及病虫害的监测等方面都发挥着重要作用^[3-4]。当前，针对海南岛人工林的遥感监测研究已经取得了很大进展^[5-6]。从分类精度上看，中尺度的橡胶林遥感识别精度较高，而大尺度的橡胶林分类精度较低，小尺度的油棕榈、柚木林、桉树林等遥感分类精度则达 90%左右。对于海南岛热带天然林分类研究而言，当前采用的主要还是野外调查、实地勘测等方法，采用遥感手段对其进行分类和类型识别的研究和探讨较少。

对于热带地区的气候来说，多云多雨是影响光学遥感数据的最主要因素，海南地区亦是如此。而微波遥感由于波长较长，穿透能力强，能够有效穿透云雾，减少云雾对观测效果的影响，因此微波遥感对于热带地区进行遥感森林监测来说是一个有力的补充^[7]。光学遥感中，主要利用数据中的红外、近红外波段对植被疏密程度的敏感性，以及影像中的色彩、纹理等特征来进行森林资源的调查和监测。而微波遥感中，利用短波段（C 波段和 X 波段）对于森林顶部冠层中不同叶片形状的后向散射特性的不同来进行识别。利用光学遥感数据和微波遥感数据结合，利用两者的优势开展热带森林资源的调查和监测正逐渐发展成为一种趋势。在热带森林多云多雨的气候条件下，选择合适的遥感数据源，并利用野外实地采集数据以及其他林业调查资料，研究适用于我国热带地区森林分类的方法，使其更好地服务于当地林业部门森林资源调查、林业规划以及合理开发，是当前的研究热点之一。

海南岛热带森林主要分布于中部的山区，主要有尖峰岭、霸王岭、黎母山、五指山和吊罗山 5 大热带森林分布区。其中尖峰岭是海南岛面积最大的热带森林分布区，由于地形和气候影响，尖峰岭热带森林类型丰富，植被种类繁多，具有垂直分布结构，植被类型的优势种并不明显。尖峰岭地区主要森林类型有典型热带雨林、热带季雨林、常绿苔藓林以及常绿落叶阔叶混交林。尖峰岭地区的植被种类多达 2800 多种，主要植被科目有芸香科、蝶形花科、菊科、乔本科、桑科、樟科、龙脑香科、棕榈科、桃金娘科、杜英科以及灰木科等。其他 4 个主要的森林分布区由于地势、气候以及土壤类型等因素的影响，造成了森林类型分布的一些差异，但总体上的森林类型以及植被种类分布情况基本相似。

因此,本数据集以海南岛为研究区,利用光学遥感数据和多时相 SAR 数据提取海南岛天然林主要林型的空间分布。本文研究可以为其他地区的热带天然林分类研究提供一定的参考,同时也为海南岛天然林的保护和规划提供决策支持和依据。

1 数据采集和处理方法

1.1 数据源

1.1.1 遥感数据

考虑到数据分辨率的可分性与大区域应用的经济适用性之间的协同问题,本研究选取中高空间分辨率的遥感数据作为数据源,包括 Landsat-8 卫星的 OLI 数据和 Sentinel-1A 卫星提供的干涉宽幅地距多视产品(IW-GRD),既保证了能有效提取热带森林中的天然林类型,也节约了数据收集的成本。采用的投影均为基于 WGS84 的 UTM 投影。具体情况如下:2015 年 11 月至 2016 年 2 月期间的 Landsat-8 卫星 OLI 数据共 4 景(云量较少,获取时间相近),2015 年 6 月至 2016 年 4 月期间的 Sentinel-1A 卫星 SAR 数据 10 期共计 20 景数据(成像模式为干涉宽幅地距多视产品,GRD)。数据详细信息见下表 1。

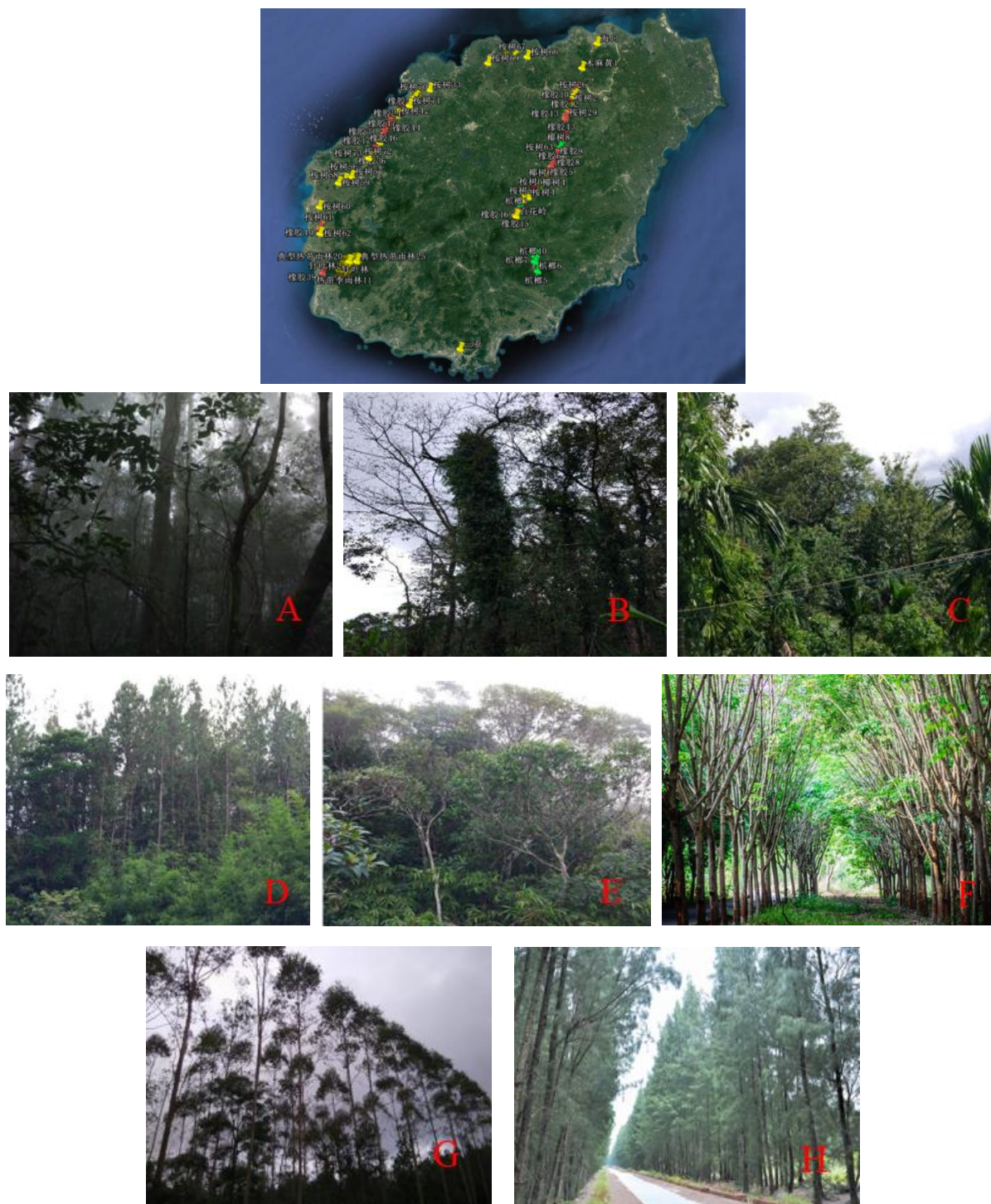
表 1 光学遥感与 SAR 数据参数

参数	SAR 数据	光学数据
卫星	Sentinel-1A	Landsat-8
获取时间	2015/6-2016/4	2015/11-2016/2
极化方式	VV/VH	
幅宽/km	250	185
空间分辨率/m	5×20	30×30
重访周期/d	12	16

1.1.2 森林调查数据

除遥感数据外,为了保证研究工作的准确性和可靠性,还在海南岛进行了林业调查,得到了天然林类型采样数据集。在海南岛热带森林野外数据采集中,针对海南岛热带森林典型热带雨林、热带季雨林、常绿苔藓林、常绿落叶阔叶混交林等天然林类型和橡胶林、桉树林、槟榔等人工林类型进行实地数据采集。数据采集实验主要测量参数有:森林类型、地理坐标和森林类型特征(森林冠层特征、代表性树种)。

野外数据采集在考虑到样本典型性以及采样地点可达性行原则的基础上,选取了儋州、屯昌、乐东附近的人工林分布密集区,以及尖峰岭、百花岭、东寨港等天然林分布较为典型的区域。其中,人工林种植区多分布于人类活动易干扰区,为便于样本快速采集,人工林样本点基本沿国道或乡村道路采集;天然林多分布于山区,车辆不易通行,需要徒步进入采集,因此采样区域有限,此次天然林样本点采集主要集中在尖峰岭、百花岭等地区。本次数据采集样点分布及样本信息如图 1 所示。



天然林: A 典型热带雨林 B 常绿落叶阔叶混交林 C 热带季雨林 D 常绿针叶林 E 常绿苔藓林

人工林: F 橡胶林 G 桉树林 H 木麻黄

图 1 野外数据采集样本实例

野外实地数据共包含了 300 个样本点, 数据采集后根据数据记录对采集的森林样本进行总结, 海南岛热带森林类型具体特征如下。

(1) 人工林: 主要分布在海拔较低的平原和丘陵地区, 基本为大面积连续的单一树种, 种植分布具有明显的规律, 绝大多数为经济林种植区。

(2) 典型热带雨林: 植被茂盛, 树种类型丰富, 基本没有受到人为影响, 空间结构分层比较明显, 一般分层达到 5–7 层, 主要为草本层、灌木层、幼小乔木层、一般乔木层以及高大乔木层。

(3) 常绿落叶阔叶混交林: 遭受到的人为影响很大, 喜阳植物较多、分层不明显, 一般只有 1–2 层, 主要为灌木层和乔木层。其中落叶树种主要为楹树、枫香、海南菜豆树等。

(4) 热带季雨林: 受到一定的人为影响, 分层一般有 3–4 层, 有一定的季相变化, 其中变化树种主要为蒲桃、榕树等。

(5) 常绿针叶林: 叶片形状为针形, 树种主要为加勒比松、南亚松等松树, 分布单块面积较小且零星分散, 基本没有大面积连续的针叶林树种。

(6) 常绿苔藓林: 主要分布海拔 1200 m 以上的山顶, 树木矮小, 整个冠层高度较矮, 且分布面积很小。

1.2 数据处理方法

由于光学遥感和微波遥感成像机理不同, 影响影像数据质量的因素也不同。所以本研究将数据预处理分成 2 个部分: SAR 数据预处理和光学遥感数据预处理。SAR 数据预处理包括雷达辐射定标、图像滤波和地形辐射校正; Landsat-8 卫星的 OLI 数据预处理包括辐射定标和大气校正。在完成数据预处理后, 考虑到不同空间分辨率数据的图像配准问题, 将 SAR 数据重采样到 Landsat-8 多光谱数据一样的 30 m 空间分辨率。在此基础上再进行分类特征提取, 并利用支持向量机 (Support Vector Machine, SVM) 分类算法对天然林进行提取和分类, 并结合野外实地采集数据以及林业调查资料, 对天然林分类的精度进行分析, 得到海南岛天然林分类图。具体流程如图 2 所示。

1.2.1 分类特征提取

在进行分类特征选取时, 除了要考虑分类地物的典型光谱特征外, 还要考虑波段提供的信息量以及不同波段在分类中作用的重复程度^[8-9]。本研究分别对 OLI 数据的单波段光谱特征、多波段光谱特征以及 SAR 数据的 VV、VH 极化后向散射系数进行统计分析, 提取适合分类特征。

(1) 光谱特征选取

经过对 Landsat-8 卫星数据的单波段统计特征和多波段统计特征的分析得知:

1) 中近红外波段 (Band5) 的标准差最大, 2 个短波红外波段 (Band6 和 Band7) 的标准差同样较大, 蓝光波段 (Band4) 标准差最小 (表 2)。说明蓝光波段能够反映的地表遥感信息较少, 近红外或短波红外包含的信息量更多;

表 2 Landsat-8 卫星 OLI 数据各波段统计特征

统计特征	Band1	Band2	Band3	Band4	Band5	Band6	Band7
最小值	-3099	0	-914	-545	0	-42	-26
最大值	5 676	5 220	5 349	5 404	6 829	8 704	12 862
均值	353.54	309.56	404.78	313.52	1949.68	1155.86	616.20
标准差	468.39	323.82	415.78	394.33	1847.69	1124.38	664.83
亮度差	8 775	5 220	6 263	5 949	6 829	8 746	12 888

2) 近红外波段和前 4 个波段之间的协方差最大 (表 3), 表明深蓝波段、蓝光波段、短波红外 2

波段与其他波段之间的同向性比较弱;

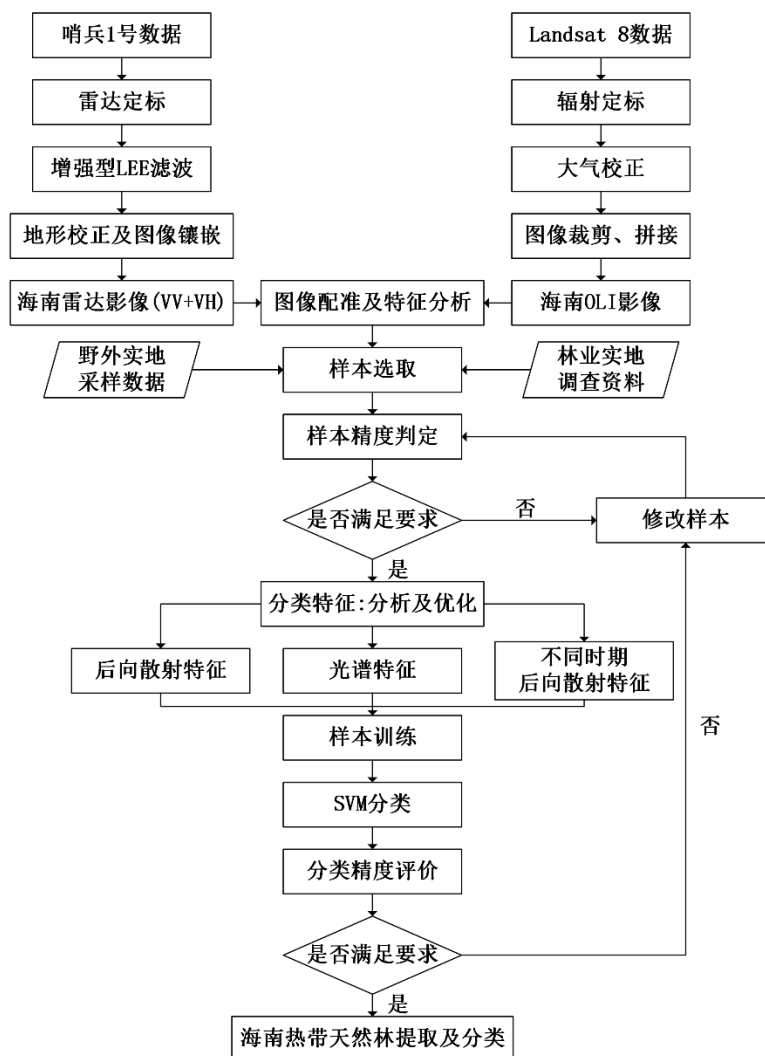


图2 海南天然林分类流程图

表3 Landsat-8 卫星 OLI 数据各波段间的协方差

波段	Band1	Band2	Band3	Band4	Band5	Band6	Band7
Band1	219 394.5						
Band2	121 423.7	104 865.7					
Band3	155 728.8	133 642.0	172 876.5				
Band4	138 853.3	121 917.0	156 226.3	155 503.6			
Band5	602 995.6	500 286.5	663 677.9	499 214.8	3 413 971.7		
Band6	410 048.2	338 228.0	442 356.0	371 290.5	1 969 113.5	1 264 235.0	
Band7	247 885.7	206 489.8	267 330.9	244 302.7	1 034 735.6	719 112.9	442 006.2

3)通过表4各个波段的相关系数,可以看出蓝光波段、短波红外2波段与其他波段相关性更强,其独立性相对较弱。

表 4 Landsat-8 卫星 OLI 数据各波段间的相关系数

波段	Band1	Band2	Band3	Band4	Band5	Band6	Band7
Band1	1.000						
Band2	0.801	1.000					
Band3	0.799	0.992	1.000				
Band4	0.751	0.954	0.952	1.000			
Band5	0.696	0.836	0.863	0.685	1.000		
Band6	0.778	0.928	0.946	0.837	0.947	1.000	
Band7	0.796	0.959	0.967	0.931	0.842	0.961	1.000

由于深蓝波段比较适合海岸带观测，本研究又要对海岸林进行观测，所以深蓝波段也是分类需要考虑的波段。最终将 Landsat-8 卫星数据中除蓝光波段和短波红外 2 波段的其他波段作为优选波段，进行分类。

(2) SAR 后向散射强度特征选取

通过对多时期双极化方式的 SAR 数据进行统计特征的分析（表 5 和表 6），可以知道：

1) 从 6 月份到 10 月份的雨季到 11 月至来年 2 月份的旱季标准差整体趋势是增加的，从侧面说明在旱季对地表进行的探测，往往比雨季得到的地表信息更丰富；

2) VH 极化 SAR 数据的标准差普遍高于 VV 极化的，说明交叉极化模式的 SAR 数据在进行观测时，比同极化 SAR 数据得到的信息更多；

3) 2 种极化 SAR 数据都是在 2 月份达到标准差最大值，说明如果考虑分类数据量的大小，只选择一期 SAR 数据进行分类时优先考虑，即在旱季成像的 SAR 数据。

经过光学数据和 SAR 数据的统计特征分析后，综合考虑数据量大小，波段的有效性和季相变化等因素，最终发现 OLI 数据的 1、3、4、5、6 波段，归一化植被指数以及 2015 年 8 月 12 日和 2016 年 2 月 8 日的双极化 SAR 数据是下一步分类研究需要优先考虑的分类数据组合。

表 5 多时期 VH 极化后向散射系数统计特征

统计特征	最小值	最大值	均值	标准差	亮度差
2015 年 6 月 25 日	-80.20	24.03	-8.24	7.27	104.23
2015 年 7 月 7 日	-93.24	17.29	-8.42	7.38	110.54
2015 年 8 月 12 日	-79.13	22.99	-8.25	7.33	102.12
2015 年 9 月 29 日	-83.78	23.06	-8.45	7.48	106.84
2015 年 10 月 23 日	-76.19	19.06	-8.42	7.40	95.25
2015 年 11 月 4 日	-80.10	21.22	-8.59	7.68	101.32
2016 年 1 月 3 日	-68.75	17.82	-8.71	7.64	86.57
2016 年 2 月 8 日	-83.18	19.34	-8.83	7.74	102.52
2016 年 4 月 20 日	-92.15	21.11	-8.81	7.72	113.26
2016 年 5 月 14 日	-111.51	21.85	-8.56	7.53	133.36

表 6 多时期 VV 极化后向散射系数统计特征

统计特征	最小值	最大值	均值	标准差	亮度差
2015 年 6 月 25 日	-74.45	36.62	-4.49	4.46	111.07
2015 年 7 月 7 日	-71.87	28.79	-4.73	4.58	100.67
2015 年 8 月 12 日	-71.60	37.12	-4.39	4.46	108.72
2015 年 9 月 29 日	-75.76	37.26	-4.52	4.55	113.02
2015 年 10 月 23 日	-68.69	31.48	-4.56	4.53	100.17
2015 年 11 月 4 日	-73.81	36.35	-4.57	4.68	110.17
2016 年 1 月 3 日	-64.99	27.25	-5.27	5.05	92.24
2016 年 2 月 8 日	-75.50	32.40	-5.20	4.92	107.90
2016 年 4 月 20 日	-85.88	33.87	-5.31	5.00	119.75
2016 年 5 月 14 日	-76.37	31.26	-5.03	4.80	107.623

1.2.2 分类算法选择

本研究基于 SVM 算法进行天然林范围提取和天然林分类。SVM 的基础是统计学习理论的 VC 维理论和结构风险最小化原理，主要是利用有限的样本所提供的信息对模型的学习能力和复杂性进行最优的选择，以便得到最好的泛化能力^[10]。SVM 的基本原理是将训练样本数据通过一定的变换，映射到高维特征空间，并在高维特征空间内寻找平面即最优超平面，然后利用该平面使得不同类别的数据区分开来^[11]。SVM 模型遥感领域中得到了较为广泛的应用。

本研究中的 SVM 分类主要是通过 ENVI 软件中的支持向量机模块进行的，其中对于 SVM 分类建模中主要涉及的参数有核函数选择和最优参数选择。在 ENVI 支持向量机分类模块中包含的主要核函数有线性核函数（Linear）、多项式核函数（Polynomial）、Sigmoid 核函数和高斯径向基核函数（Radial Basic Function, RBF）。

以上几种核函数中，使用最为普遍的是 RBF 核函数。该核函数具有较好学习能力，是一种采用较多，且分类结果较为理想的分类核函数^[12]。因此本研究分别选择 RBF 核函数进行建模，在 ENVI 软件中的 SVM 模块中需要选择分类器参数，经过多次试验最终确定 SVM 分类参数。提取天然林范围时，SVM 的参数设置是 RBF 核函数的 γ 值为 0.083，惩罚系数为 100；但是在进行天然林分类时，SVM 的参数设置为 RBF 核函数的 γ 值为 0.077，惩罚系数为 100。

1.2.3 海南热带天然林分类体系

由于海南岛热带天然林分布十分复杂，为避免其他地物类型对天然林分类的影响，本研究的分类方法采用分层分类的策略，首先对天然林范围进行提取，然后在天然林范围提取的基础上对其内部林型进行分类。

（1）海南岛土地初级分类体系

对于海南岛土地进行初级分类主要目的是对天然林的范围进行提取，因此只进行基本地物分类。参照遥感数据情况、海南岛的实际情况以及初级分类的用途，制定了海南岛土地初级分类体系。同时，对于初级分类中的地物类型，通过对 Landsat-8 卫星遥感影像进行解译，同时参考了野外实地采样数据，建立了影像解译标志。表 7 为海南岛土地初级分类体系及解译标志。图 3 为主要地物类型

的解译标志示意图。

本研究利用这些遥感影像解译标志，进行了样本的选取和训练，为接下来利用 SVM 分类模型进行海南岛初级分类奠定了基础。

表 7 海南岛土地初级分类体系及解译标志

一级地类	二级地类	解译标志
森林	海岸林	暗绿色，位置在海边，中间有暗纹
	人工林	深绿色，大多数有规则的纹理特征
	天然林	深绿色偏亮，无规则纹理特征
非森林	人工地表	亮灰色，有些为蓝色
	水体	深青色，有的水体会偏绿色
	裸地	棕黄色，有些具有比较规则的纹理
	其他植被	浅绿色或者亮绿色，形状比较规则

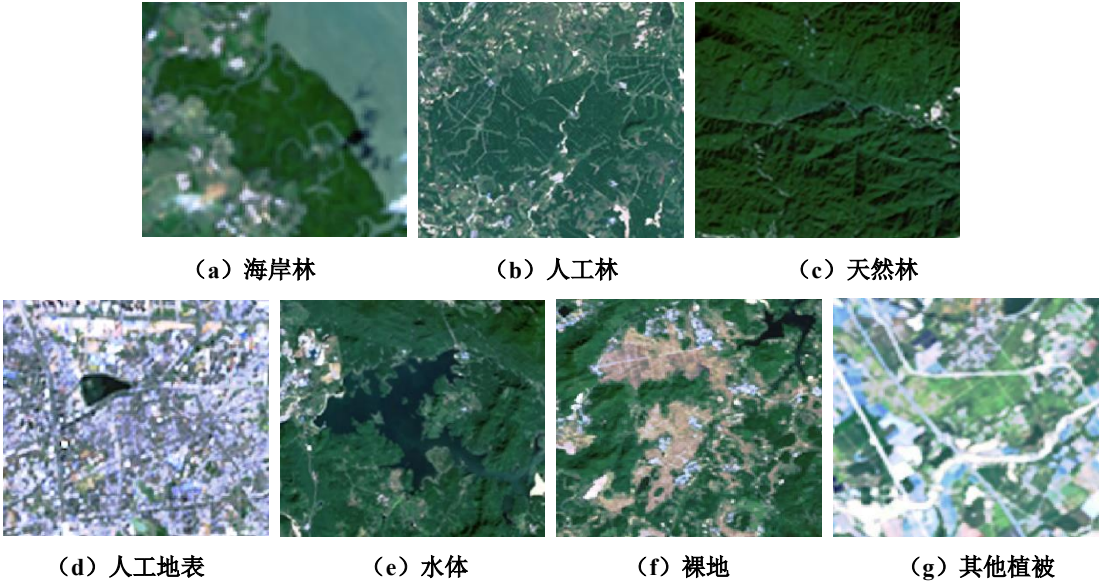


图 3 海南岛土地初级分类解译标志示意图

(2) 海南热带天然林分类体系

海南岛地处热带，是我国热带森林植被的重要分布区，特别是海南岛中南部地区被普遍认为是具有国际意义的包括植被多样性在内的生物多样性最丰富的区域之一。对于海南岛热带森林的分类体系还存在一定的分歧。在国家 2010 年发布的森林资源规划设计调查技术规程^[13]中，将森林类别按照用途或树种进行划分，树种类别较多，遥感解译尚存在较大困难。宋永昌^[14]在 2011 年提出的植被分类系统中，森林被划分为一个独立的植被型纲，这样的划分更有利于森林系统性的识别。因此，在宋永昌分类系统的基础上，利用所选取的多源遥感数据，选择海南岛常见林型进行天然林内部类型的分类研究。

本研究在宋永昌 2011 年提出的植被分类系统（表 8）的基础上^[14-15]，结合海南岛野外调查的实际经验和遥感机理的分析，将海南岛热带天然林分为典型热带雨林、热带季雨林、针叶林、海岸林以及常绿阔叶林。由于海南地区的常绿落叶林只是零散的分布于常绿阔叶林之中，两者之间并没有

明显的区别，所以在这里将常绿阔叶林和常绿落叶阔叶混交林分成一类，统一归类为常绿阔叶林。

表 8 宋永昌提出的森林分类系统

植被型纲	植被型亚纲	植被型组	植被型
森林	I. 针叶林	1. 常绿针叶林	1) 热性常绿针叶林
	II. 阔叶林	2. 常绿落叶阔叶混交林	2) 次生常绿落叶阔叶混交林
		3. 常绿苔藓林	3) 山地常绿苔藓林
		4. 常绿阔叶林	4) 典型常绿阔叶林
			5) 季节（季风）常绿阔叶林
		5. 热带季雨林	6) 热带落叶季雨林
			7) 热带半落叶季雨林
		6. 热带雨林	8) 热带（典型）雨林
			9) 热带季节性雨林
		7. 海岸林	10) 红树林
			11) 热带珊瑚礁海岸林
	III. 竹林与竹丛	8. 竹林	12) 丛生竹林
			13) 混生竹林

2 数据样本描述

2.1 海南岛土地初级分类

首先利用 SVM 分类算法进行海南岛的土地初级分类，参照制定的土地初级分类体系，将海南岛分为水体、其他植被、人工林、天然林、人工地表、裸地以及海岸林 7 种地类。海南岛土地初级分类结果显示如图 4 所示。

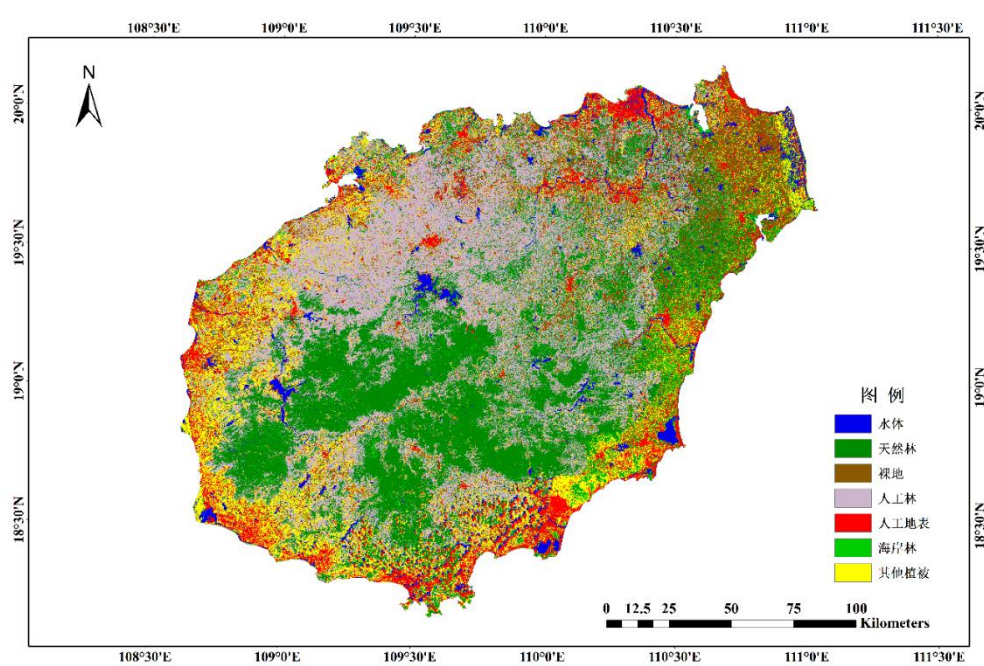


图 4 基于 SVM 的海南岛土地初级分类结果

2.2 海南岛天然林范围

在海南岛土地初级分类的基础上，基于 SVM 方法对海南岛土地初级分类中天然林的范围进行提取。并以此为基础，依据野外实地调查数据、海南林业部门提供的森林抽样统计调查数据以及对高空间分辨率影像的目视解译，对 SVM 分类结果中的天然林零碎图斑进行识别，剔除非天然林斑块，最终得到天然林的总体范围（图 5）。

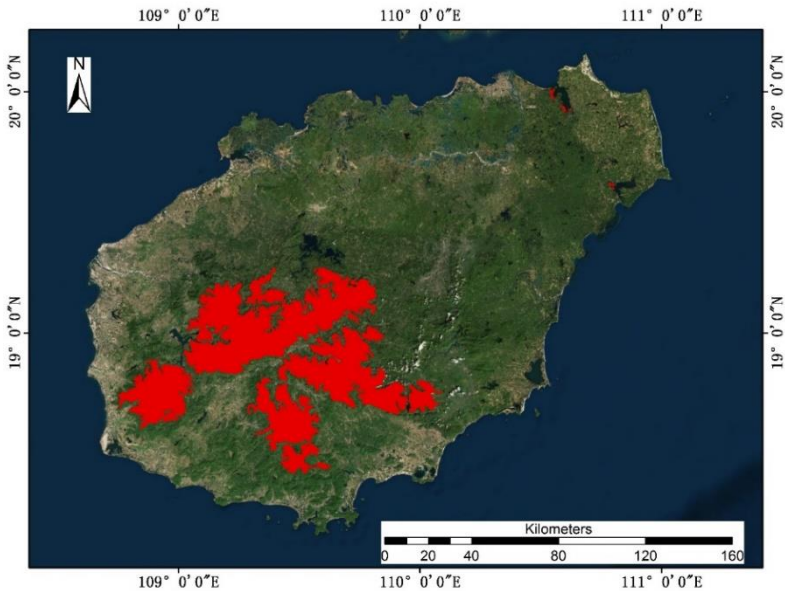


图 5 海南岛天然林范围（红色区域）

2.3 基于 SVM 的天然林分类结果

为了避免其他地类信息干扰天然林分类，本研究在进行天然林范围提取后，对分类所用的遥感数据进行了裁剪，使得在天然林范围内只包含天然林林型，没有其他地类。由于各种天然林林型在遥感影像上的区别较小，仅凭肉眼很难识别。因此，对于天然林分类的样本选取主要依靠野外实地采样点和海南林业部门提供的森林抽样调查资料进行选择。结合 2 种资料最终选取了各类天然林样本 280 块，随机抽取 70%进行样本的训练和分类，其余 30%用于精度验证。

同时，本研究利用 ENVI 软件中提供的计算 ROI 可分离性（Separability）工具来确定 2 个类别间的差异程度。类别间的差异主要通过计算 Jeffries-Matusita 距离和转换分离度（Transformed Divergence）来确定任意类别间的统计距离，从而衡量训练样本（ROI）的可分离性。表 9 为海南岛天然林类型样本可分离性统计。

表 9 海南岛天然林类型样本可分离性统计

可分离性	典型热带雨林	热带季雨林	常绿针叶林	常绿阔叶林	海岸林
典型热带雨林					
热带季雨林	1.936 5				
常绿针叶林	1.922 0	1.915 7			
常绿阔叶林	1.923 6	1.929 1	1.904 8		
海岸林	1.984 6	1.986 9	1.934 5	1.978 0	

训练样本的可分离性值在 0–2.0 之间，大于 1.9 说明样本之间可分离性好，属于合格样本；小于 1.8，需要编辑样本或者重新选择样本；小于 1，考虑将 2 类样本合成一类样本。从表 9 中可以看出任意天然林类型之间的可分离性均大于 1.9，证明了 5 种天然林类型在此类遥感特征中的可分离性较好。

基于 SVM 模型的分类结果如图 6 和图 7 所示。其中图 7 中有一条黄色线状条带，这是因为原始 Sentinel-1A 极化 SAR 数据缺失导致，在拼接过程中线状条带为数据缺失部分。

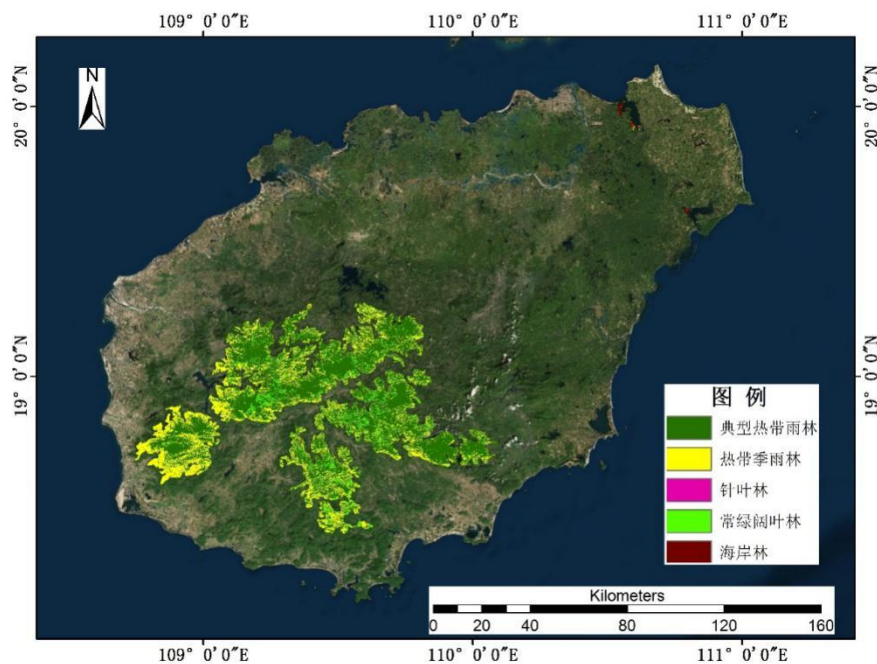


图 6 基于 SVM 的天然林分类结果

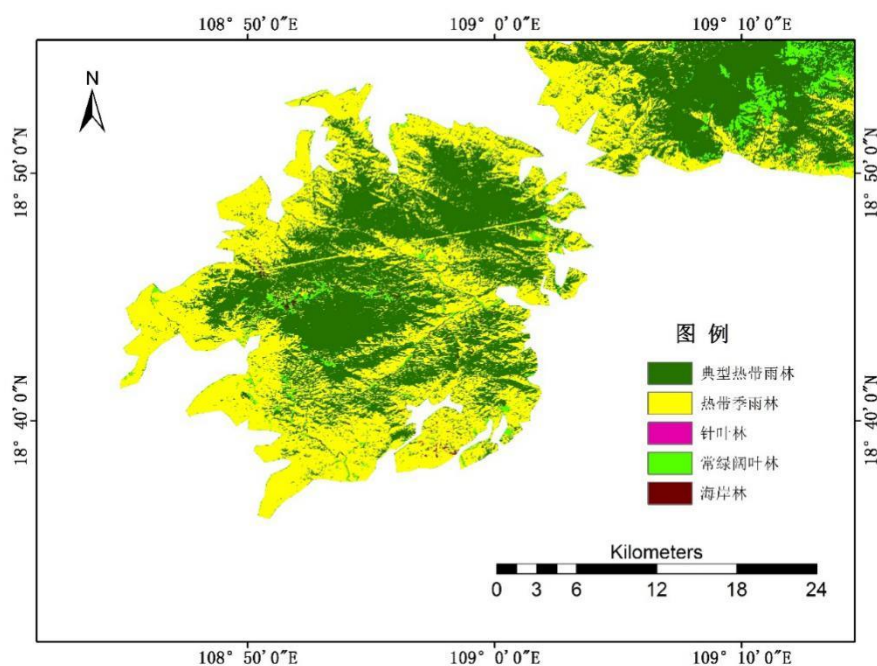


图 7 基于 SVM 的天然林分类结果细部图（尖峰岭）

3 数据质量控制和评估

3.1 海南岛土地初级分类精度分析

本研究分类结果评价均采用混淆矩阵作为分类精度的评价方法。对于精度评价的指标主要利用了生产者精度、用户精度、总体精度以及 Kappa 系数 4 个精度评价指标^[16]。表 10 为海南岛土地初级分类精度评价。

表 10 基于 SVM 的海南岛土地初级分类精度评价

真实数据										
分类数据	基本地类	水体	天然林	裸地	人工林	人工地表	海岸林	其他植被	总计	用户精度
	水体	2 498	17	0	0	2	2	0	2 519	99.17%
	天然林	1	4 525	1	393	0	85	0	5 005	90.41%
	裸地	1	0	2 487	0	6	0	0	2 494	99.72%
	人工林	0	279	0	3 611	0	17	3	3 910	92.35%
	人工地表	0	0	10	0	2 494	0	3	2 507	99.48%
	海岸林	0	206	0	3	0	2 396	0	2 605	91.98%
	其他植被	0	8	2	2	3	0	2 494	2 609	99.40%
	总计	2 500	5 035	2 500	4 009	2 505	2 500	2 500	21 549	
	生产者精度	99.92%	89.87%	99.48%	90.07%	99.56%	95.84%	99.76%		
		总体精度：95.15%							Kappa 系数：0.943	

通过利用混淆矩阵对于 SVM 模型分类结果的精度评价可以看出，基于 SVM 模型的分类方法在海南土地初级分类中得到了很好的效果，总体精度达到了 95.15%，Kappa 系数为 0.943。

3.2 海南岛天然林分类精度分析

同样采用混淆矩阵作为天然林分类精度的评价方法，从生产者精度、用户精度、总体精度以及 Kappa 系数 4 个方面进行评价。表 11 为天然林分类精度评价。

表 11 基于 SVM 的天然林分类精度评价

真实数据								
	森林地类	典型热带雨林	热带季雨林	常绿针叶林	常绿阔叶林	海岸林	总计	用户精度
分类数据	典型热带雨林	1389	37	76	207	0	1709	81.28%
	热带季雨林	108	1354	0	66	0	1528	88.61%
	常绿针叶林	0	0	90	0	0	90	100%
	常绿阔叶林	0	4	0	1352	0	1356	99.71%

	海岸林	3	0	2	0	0	1 080	99.54%
	总计	1500	1400	168	1625	1075	5768	
	生产者精度	92.60%	96.71%	53.37%	83.20%	100%		
		总体精度: 91.19%			Kappa 系数: 0.883 7			

通过表 11 可以看出 SVM 分类模型在天然林分类中取得了较好的精度。对于天然林分类而言, 分类结果的总体精度虽然很高, 但是主要由于海岸林等较为容易区分的林型拉升了总体精度, 将海岸林等林型排除后, 天然林分类中有些林型的分类精度还是比较低, 例如常绿针叶林。

对于常绿针叶林而言, 基于 SVM 算法的分类, 生产者精度只有 53.57%。经过分析, 常绿针叶林分类精度不高的原因有 2 点: (1) 在野外实地调查中, 可以发现常绿针叶林的分布极为零散, 有的甚至以棵的形式存在, 即使有成片的常绿针叶林, 其面积也很小, 根本不足一个像元的面积, 这样就造成了混合像元, 在进行样本的训练和分类时会造成误差, 从而影响分类精度。(2) 本文研究用的是 C 波段雷达数据, C 波段雷达虽然无法穿透茂盛的树木冠层, 但是对于稀疏分布的常绿针叶林来说, 则就有一定的穿透性。并且在热带天然林中常绿针叶林冠层以下均有阔叶类型的灌木丛分布, 这也从一定程度上增加了常绿针叶林和常绿阔叶林的区分难度。因此, 利用 C 波段雷达对于常绿针叶林的叶片形状并不敏感, 这也降低了分类精度。

4 数据价值

本研究以海南岛为研究区域, 利用多源遥感数据, 进行了多源遥感的数据预处理并对数据特征进行了分析, 在此基础上对海南岛的热带天然林进行了提取, 并对天然林进行分类研究。利用信息量较大的光学数据波段、旱季雨季的 SAR 数据和归一化植被指数组成的数据组合对海南岛进行土地初级分类和天然林分类, SVM 的总体分类精度分别达到了 95%和 91.2%, 结果表明本文的分类方法在天然林分类中具有很好的效果。天然林的分类结果对进行森林调查和监测具有重要的意义, 在生态层面和经济层面也都有重要的价值。

致 谢

感谢海南省环境科学研究院协助海南热带森林主要参数(森林类型、地理坐标和森林类型特征)的采集工作, 感谢海南林业部门对海南热带森林相关辅助数据(森林资源抽样统计实地调查资料和海南森林植被分布资料)的采集工作。

数据作者分工职责

史建康(1980—), 男, 安徽省亳州市人, 硕士, 教授级高级工程师, 研究方向为生态环境遥感。主要承担工作: 技术指导和野外数据采集。

宫晨(1993—), 男, 甘肃省兰州市人, 硕士研究生, 研究方向为生态环境遥感。主要承担工作: 算法研究和数据处理。

李新武(1973—), 男, 四川省内江市人, 博士, 研究员, 研究方向为生态环境遥感。主要承担工作: 技术指导和算法研究。

万祥星（1991—），男，山东省聊城市人，博士研究生，研究方向为生态环境遥感。主要承担工作：算法研究和数据处理。

孙中昶（1983—），男，山东省济宁市人，博士，副研究员，研究方向为生态环境遥感。主要承担工作：算法研究和数据处理。

参考文献

- [1] THAPAR B, ITOH T, SHIMADA M, et al. Evaluation of ALOS PALSAR sensitivity for characterizing natural forest cover in wider tropical areas[J]. Remote Sensing of Environment, 2014, 155: 32-41.
- [2] 吴楠, 李增元, 廖声熙, 等. 国内外林业遥感应用研究概况与展望[J]. 世界林业研究, 2017, 30(6): 34-40.
- [3] 陈顺. Landsat-7ETM+和 PALSAR 数据联合反演森林生物量方法研究[D]. 徐州: 中国矿业大学, 2018.
- [4] 廖凯涛, 齐述华, 王成, 等. 结合 GLAS 和 TM 卫星数据的江西省森林高度和生物量制图[J]. 遥感技术与应用, 2018, 33(04): 713-720.
- [5] 张京红, 张明洁, 刘少军, 等. 风云三号气象卫星在海南橡胶林遥感监测中的应用[J]. 热带作物学报, 2014, 35(10): 2059-2065.
- [6] 张京红, 陶忠良, 刘少军, 等. 基于 TM 影像的海南岛橡胶种植面积信息提取[J]. 热带作物学报, 2010, 31(4): 661-665.
- [7] 周建民, 李震, 李新武. 基于 ALOS/PALSAR 雷达干涉数据的中国西部山谷冰川冰流运动规律研究[J]. 测绘学报, 2009, 38(4): 341-347.
- [8] 贾坤, 李强子. 农作物遥感分类特征变量选择研究现状与展望[J]. 资源科学, 2013, 35(12): 2507-2516.
- [9] 柏延臣, 王劲峰. 基于特征统计可分性的遥感数据专题分类尺度效应分析[J]. 遥感技术与应用, 2004, 19(6): 443-449.
- [10] 郑恩辉. 基于支持向量机的代价敏感数据挖掘研究与应用[D]. 杭州: 浙江大学, 2006.
- [11] 王丽艳. 基于视觉词包算法的 SAR 图像分类方法及应用研究[D]. 北京: 中国科学院大学, 2015.
- [12] 崔东文, 郭荣. 基于几种参数优化的支持向量机在径流预报中的比较分析[J]. 水资源研究, 2013, 34(2): 34-38.
- [13] GB/T 26424—2010, 森林资源规划设计调查技术规程[S]. 北京: 国家林业局调查规划设计院, 2011.
- [14] 宋永昌. 对中国植被分类系统的认知和建议[J]. 植物生态学报, 2011, 35 (08): 882-892.
- [15] 宋永昌, 阎恩荣, 宋坤. 中国常绿阔叶林 8 大动态监测样地植被的综合比较[J]. 生物多样性, 2015, 23 (02): 139-148.
- [16] 杨冉冉. 基于 ALOS 数据的遥感森林分类研究[D]. 北京: 首都师范大学, 2013.

论文引用格式

史建康, 宫晨, 李新武, 等. 基于多源遥感数据的海南岛天然林分类数据集[J/OL]. 中国科学数据,

2019, 4(2). (2019-03-29). DOI: 10.11922/csdata.2018.0091.zh.

数据引用格式

史建康, 宫晨, 李新武, 等. 基于多源遥感数据的海南岛天然林分类数据集[DB/OL]. Science Data Bank, 2018. (2018-12-14). DOI: 10.11922/sciencedb.711.

Classification of Hainan island natural forests based on multi-source remote sensing data

Shi Jiankang¹, Gong Chen^{2*}, Li Xinwu², Wan Xiangxing², Sun Zhongchang²

1. Hainan Academy of Environmental Science, Haikou 571126, P. R. China;

2. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, P. R. China

*Email: gongchen@radi.ac.cn

Abstract: Widely distributed in the vicinity of the equator, tropical forest is one type of forest with the most abundant species worldwide which has a profound effect on global climate change. Therefore, it is of great significance for a country to develop the forest resources inventory and perform dynamic monitoring. Research on the classification of natural forests not only supports the investigation of tropical forests, but also provides the basis for the study of forest species diversity. The dual-polarized SAR data from Sentinel-1A sensor and the optical remote sensing data from Landsat-8 sensor were used for classification of Hainan island tropical natural forest. First, we analyzed the single-band, multi-band, normalized difference vegetation index (NDVI) characteristics of optical data, and the single-phase, multi-temporal backscattering characteristics of SAR data. Then, optical and backscattering characteristics were selected for natural forest classification whereby the natural forest range of Hainan Island was extracted by using support vector machine (SVM). The natural tropical forest was classified into five types: tropical rain forest, tropical monsoon forest, evergreen coniferous forest, deciduous broad-leaved mixed forest and coastal forest. The accuracy of classification results was verified and evaluated based on a combination of field survey data and Hainan forestry survey data. The overall accuracy of the classification exceeded 80%. The results provide a reliable remote sensing classification method for Hainan island tropical forest classification. This dataset also has some reference value for the study of tropical natural forest classification in other areas.

Keywords: multi-source remote sensing data; tropical forest; remote sensing classification; support vector machine (SVM)

Dataset Profile

Title	Classification of Hainan island natural forests based on multi-source remote sensing data
Data corresponding author	Gong Chen (gongchen@radi.ac.cn)
Data authors	Shi Jiankang, Gong Chen, Li Xinwu, Wan Xiangxing, Sun Zhongchang
Time range	2015

Geographical scope	18°10'N – 20°10'N, 108°37'E – 111°03'E
Spatial resolution	30 m
Data volume	1.20 GB
Data format	*.bat, *.tif, *.shp, *.xml
Data service system	< http://www.sciencedb.cn/dataSet/handle/711 >
Sources of funding	Major Science and Technology Program of Hainan Province (ZDKJ2016021); Natural Science Foundation of Hainan Province (20164177).
Dataset composition	The dataset consists of 2 compressed packages, each of which is named after respective product name: (1) Classification_of_Hainan_Island_Land_cover.zip stores data on the classification of Hainan land cover, with a data volume of about 1.20 GB; (2) Classification_of_Hainan_Island_Natural_Forest.zip stores data on the classification of Hainan natural forest, with a data volume of about 15.5 MB.