

# 基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集 (2020)

张磊<sup>1</sup>, 周广胜<sup>2\*</sup>, 任鸿瑞<sup>1</sup>, 吕晓敏<sup>2</sup>, 汲玉河<sup>2</sup>

1. 太原理工大学测绘科学与技术系, 太原 030024;

2. 中国气象科学研究院, 北京 100081

**摘要:** 西藏自治区的植被类型复杂多样, 其植被分布对于理解和监测生态系统的健康状况具有重要意义。作者基于 GEE (Google Earth Engine) 平台, 采用地形、气候和 FY-3D (风云三号 D 星) NDVI 数据, 通过随机森林分类算法研发了西藏自治区植被分类 250 m 栅格数据集 (2020), 总体精度为 81.5%, Kappa 系数为 0.79。结果表明, 风云遥感数据可以满足高精度区域植被制图的需求。该数据集为理解西藏自治区的植被特征、生态变化及其对气候的响应提供了宝贵的数据资源。数据集内容包括: (1) 植被分类体系表; (2) 西藏自治区 250 m 分辨率植被分布数据 (2020)。数据集存储为.xlsx 和.tif 格式, 由 2 个数据文件组成, 数据量为 3.16 MB (压缩为 1 个文件, 2.73 MB)。

**关键词:** 西藏自治区; GEE; 风云卫星; 植被图; 随机森林

**DOI:** <https://doi.org/10.3974/geodp.2024.01.05>

**CSTR:** <https://cstr.escience.org.cn/CSTR:20146.14.2024.01.05>

**数据可用性声明:**

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2024.04.07.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2024.04.07.V1>。

## 1 前言

植被是覆盖地球表面所有植物群落的总称, 在维护生物多样性、保持土壤稳定性和调节水循环等方面发挥着关键作用, 被称为地球环境健康的基石。植被分布数据作为地表植被类型的直观展现, 不仅是揭示地表植被分布规律及其驱动因素的基础数据, 同时也为土地管理、环境保护、自然灾害预防等提供了必要信息。

近年来, 遥感数据在土地覆盖/土地利用信息获取中的应用日益广泛, 特别是 Sentinel、Landsat 和 MODIS 系列的卫星因其可以提供高质量的时空分辨率数据而备受青睐。其中,

收稿日期: 2023-11-01; 修订日期: 2024-02-26; 出版日期: 2024-03-25

基金项目: 中华人民共和国科学技术部 (2019QZKK0106); 风云卫星先导专项 (FY-APP-2022.0309)

\*通讯作者: 周广胜, 中国气象科学研究院, zhoughs@cma.gov.cn

数据引用方式: [1] 张磊, 周广胜, 任鸿瑞等. 基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集 (2020) [J]. 全球变化数据学报, 2024, 8(1): 42–50. <https://doi.org/10.3974/geodp.2024.01.05>. <https://cstr.escience.org.cn/CSTR:20146.14.2024.01.05>.  
[2] 张磊, 周广胜, 任鸿瑞等. 基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集 (2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.04.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.04.07.V1>.

基于 Sentinel 系列卫星的 ESA WorldCover 10m<sup>[1]</sup>、Sentinel-2 10m Land Use/Land Cover<sup>[2]</sup>等产品展现了精细的土地覆盖细节,而依托于 Landsat 系列卫星的 GlobeLand30<sup>[3]</sup>、CLCD<sup>[4]</sup>、GLC\_FCS30<sup>[5]</sup>等产品也被用于全球或区域尺度的土地覆盖研究。此外,基于 MODIS 的 MCD12Q1 产品<sup>[6]</sup>也在长期全球变化监测方面发挥了关键作用。风云三号作为我国的第二代极轨气候卫星<sup>[7]</sup>,提供的 FY-3D 遥感数据在空间分辨率、光谱分辨率以及时间范围上相较于 Sentinel、Landsat 和 MODIS 的遥感数据没有明显优势,同时又缺乏 GEE (Google Earth Engine)、PIE-Engine (Pixel Information Expert Engine) 等云平台的支持,因此该数据在区域植被制图领域应用较少,其有效性及潜在应用价值尚待进一步验证与探索。

西藏自治区位于青藏高原腹地,以其复杂多样的地形、寒冷干燥的气候和独特的自然环境而闻名。研究其植被分布不仅有利于当地的可持续发展,而且对青藏高原乃至全国生态系统的保护具有重大意义。本研究旨在通过风云遥感数据刻画西藏自治区的精细植被分布,以此探究该区域的植被特征及其分布模式,并验证风云卫星在区域植被制图领域的潜力。

## 2 数据集元数据简介

《基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集 (2020)》<sup>[8]</sup>的名称、作者、地理区域、数据年代、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

## 3 数据研发方法

### 3.1 研究区介绍

西藏自治区地处 26°50'N 至 36°53'N、78°25'E 至 99°06'E 之间,平均海拔在 4000 m 以上,总面积约 120.28 万 km<sup>2</sup>。作为世界上平均海拔最高的地区之一,其山脉纵横、高原广阔,整体地势自西北向东南的倾斜。受南部喜马拉雅山脉的影响,该区域南北降水量差异极大,造成了东南温暖湿润,而西北严寒干燥的气候特点。西藏自治区植被类型复杂多样,包括阔叶林、落叶林、灌丛、草甸、草原、荒漠等。其中,草甸和草原面积广阔,主要分布在中部高原。受益于温暖湿润的气候条件,林地成为东南部的的主要植被类型,而灌丛则多见于中部拉萨河谷及其周边地区。此外,西藏自治区的西北区域分布着广袤的荒漠植被,形成了独特的荒漠生态环境<sup>[10, 11]</sup>。

### 3.2 数据来源

本研究结合《中华人民共和国植被图 (1:100000000)》<sup>[12]</sup>,将西藏自治区的植被划分为阔叶林、针叶林、针阔混交林、灌丛、高寒草甸、高寒草原、高山植被、高寒荒漠、栽培植被 9 种类型,此外也划分出湿地、水体和其他 3 种类型。样本数据来源于 2020 年高分辨率的谷歌在线影像,在确保每个样本周围 250 m 范围内植被类型基本一致的前提下,通过目视解译选取 2,774 条样本数据 (图 1)。这些样本尽可能分布于研究区域且包含不同植被类型。样本数量如表 2 所示,其中 90%用于训练,余下 10%用于验证。

表 1 《基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集（2020）》元数据简表

条 目	描 述
数据集名称	基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集（2020）
数据集短名	VegetationXizang2020
作者信息	张磊，太原理工大学，zhanglei1136@link.tyut.edu.cn 周广胜，中国气象科学研究院，zhougs@cma.gov.cn 任鸿瑞，太原理工大学，renhongrui@tyut.edu.cn 吕晓敏，中国气象科学研究院，lvxm@cma.gov.cn
地理区域	西藏自治区
数据年代	2020
空间分辨率	250 m
数据格式	.xlsx、.tif
数据量	3.16 MB (压缩后 2.73 MB)
数据集组成	由 2 个数据文件组成，包括（1）植被分类体系表；（2）西藏自治区 250 m 分辨率植被分布数据（2020）
基金项目	中华人民共和国科学技术部（2019QZKK0106）
出版与共享服务平台	全球变化科学研究数据出版系统 <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 <sup>[9]</sup>
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

表 2 西藏自治区 250 m 分辨率植被分布数据（2020）的训练样本与验证样本数量统计表

植被覆盖类型	训练样本	验证样本	总计
阔叶林	239	27	266
针叶林	120	12	132
针阔混交林	79	5	84
灌丛	111	9	120
高寒草甸	409	44	453
高寒草原	490	49	539
高山植被	166	22	188
高寒荒漠	193	26	219
栽培植被	102	10	112
湿地	82	8	90
水体	252	35	287
其他	256	28	284
总计	2,499	275	2,774

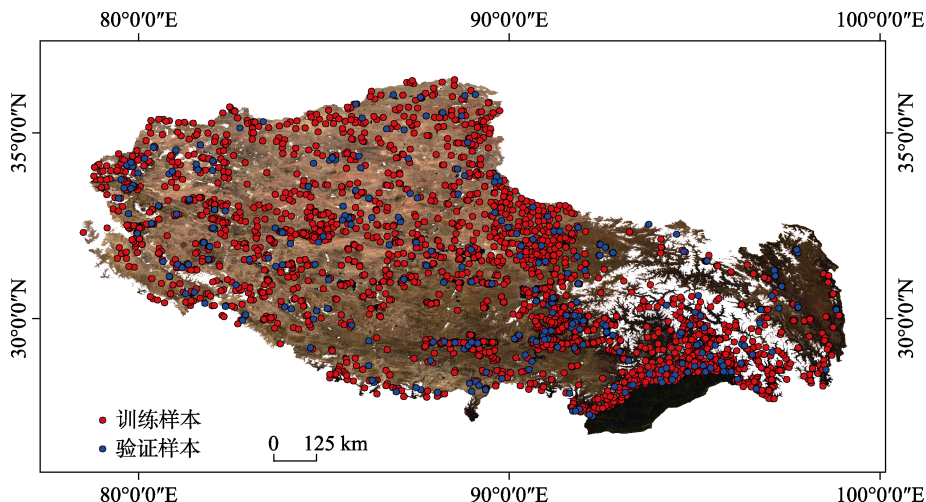


图 1 西藏自治区 250 m 分辨率植被分布数据（2020）的训练样本与验证样本空间分布图  
（依据审图号 GS(2019)3333 号标准地图制作）

遥感数据采用风云三号的 FY-3D MERSI-II 植被指数旬产品数据集,该数据集来自风云卫星遥感数据服务网<sup>1</sup>,时间分辨率为 10 天,空间分辨率为 250 m,包括蓝、绿、红、近红、NDVI、EVI 共 6 个波段。本研究选取了 2020 年的 36 张 NDVI 影像作为植被制图特征。该数据在水体、冰川积雪等区域出现空值,因此通过红、近红波段计算了对应区域的 NDVI 值进行填补（式 1）。

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

式中, NIR 为 FY-3D 遥感数据的近红外波段, Red 为红波段。

地形数据采用美国地质调查局（USGS）提供的 SRTM（Shuttle Radar Topography Mission）数据<sup>[13]</sup>,空间分辨率为 30 m。基于该数据,本研究计算了西藏自治区的高程、坡度和坡向作为植被制图特征。气候数据则来源于国家青藏高原科学数据中心<sup>2</sup>,空间分辨率为 1,000 m,包含 2020 年的年均温<sup>[14]</sup>、年降水<sup>[15]</sup>。为保证数据分辨率一致,本研究在 GEE 平台中将高程、坡度、坡向、年均温、年降水重采样至 250 m。

### 3.3 研究方法

#### 3.3.1 云平台与随机森林算法

GEE 云平台集成了丰富的卫星影像,拥有大规模地理空间数据处理能力,在气候变化、环境监测等领域得到广泛应用<sup>[16]</sup>。该平台支持多种机器学习模型,包括支持向量机、随机森林、梯度提升树等。本研究采用 GEE 云平台中的随机森林分类模型<sup>[17]</sup>进行植被制图。该模型通过随机选择样本和特征来训练各个决策树,并结合多个决策树的预测结果来提升分类的准确性。这种方法可以有效减少过拟合风险,并增强模型对新数据的泛化能力,因此被广泛应用于各种复杂的分析任务中。

<sup>1</sup> 风云卫星遥感数据服务网. <https://satellite.nsmc.org.cn>.

<sup>2</sup> 国家青藏高原科学数据中心. <https://data.tpdc.ac.cn>.

### 3.3.2 特征重要性评估与优选

本研究基于地形、气候和遥感数据构建了 41 个用于植被制图的关键特征，包括高程、坡度、坡向、年均温、年降水，以及 36 个来自 FY-3D MERSI-II 植被指数旬产品数据集的 NDVI 特征。过多的特征可能导致模型过拟合，因此采用递增特征筛选和袋外误差最小化原则筛选最优制图特征。首先，计算这 41 个植被制图特征在随机森林模型中重要性，并由高到低进行排列。随后，从最重要的特征开始，逐渐加入次要特征，分别构建包含首位、前两位、前三位等的随机森林分类模型，并计算每个模型的袋外误差 (out-of-bag error, OOB error)，即随机森林分类模型的每棵决策树在训练过程中未被抽样的样本误分类率。最终，基于袋外误差最小化原则得到最优制图特征，并将其应用于植被制图。

### 3.3.3 精度评价

本研究基于 GEE 平台，利用最优特征组合构建随机森林分类模型，同时结合地形、气候和 FY-3D NDVI 数据，研发了西藏自治区植被分类 250 m 栅格数据集 (2020)。为评价植被分布数据的准确性，本研究采用混淆矩阵来计算总体精度 (式 2)、Kappa 系数 (式 3)、制图精度 (式 4) 和用户精度 (式 5)。

$$OA = \frac{\sum_{i=1}^n m_i}{N} \quad (2)$$

$$Kappa = \frac{N \times \sum_{i=1}^n m_i - \sum_{i=1}^n (G_i \times C_i)}{N^2 - \sum_{i=1}^n (G_i \times C_i)} \quad (3)$$

$$PA = \frac{m_i}{G_i} \quad (4)$$

$$UA = \frac{m_i}{C_i} \quad (5)$$

式中， $m_i$  为第  $i$  类分类正确的样本数； $n$  为分类数； $N$  为分类像元总数； $C_i$  为第  $i$  类的分类像元总数； $G_i$  为第  $i$  类的真实像元总数； $OA$  为总体精度； $Kappa$  为 Kappa 系数； $PA$  为制图精度； $UA$  为用户精度。

## 3.4 技术路线

本数据集的主要研发过程如图 2 所示。首先，收集 2020 年地形、气候和 FY-3D NDVI 数据，通过重采样将其统一至 250 m 分辨率并得到 41 个植被制图特征，包括高程、坡度、坡向、年均温、年降水以及 36 个来自 FY-3D MERSI-II 植被指数旬产品数据集的 NDVI 特征。然后，通过目视解译，在高分辨率谷歌地图影像上选取 2,774 条有代表性的植被类型样本，将 90% 作为训练样本，剩余 10% 作为验证样本。基于训练样本，利用递增特征筛选和袋外误差最小化原则得到最优特征集合并构建随机森林分类模型。该模型的性能通过验证样本上的总体精度、用户精度、制图精度和 Kappa 系数来评估。最终，基于构建的随机森林分类模型，结合地形、气候和 FY-3D NDVI 遥感数据，制作了西藏自治区 250 m 分辨率植被分布数据 (2020)。

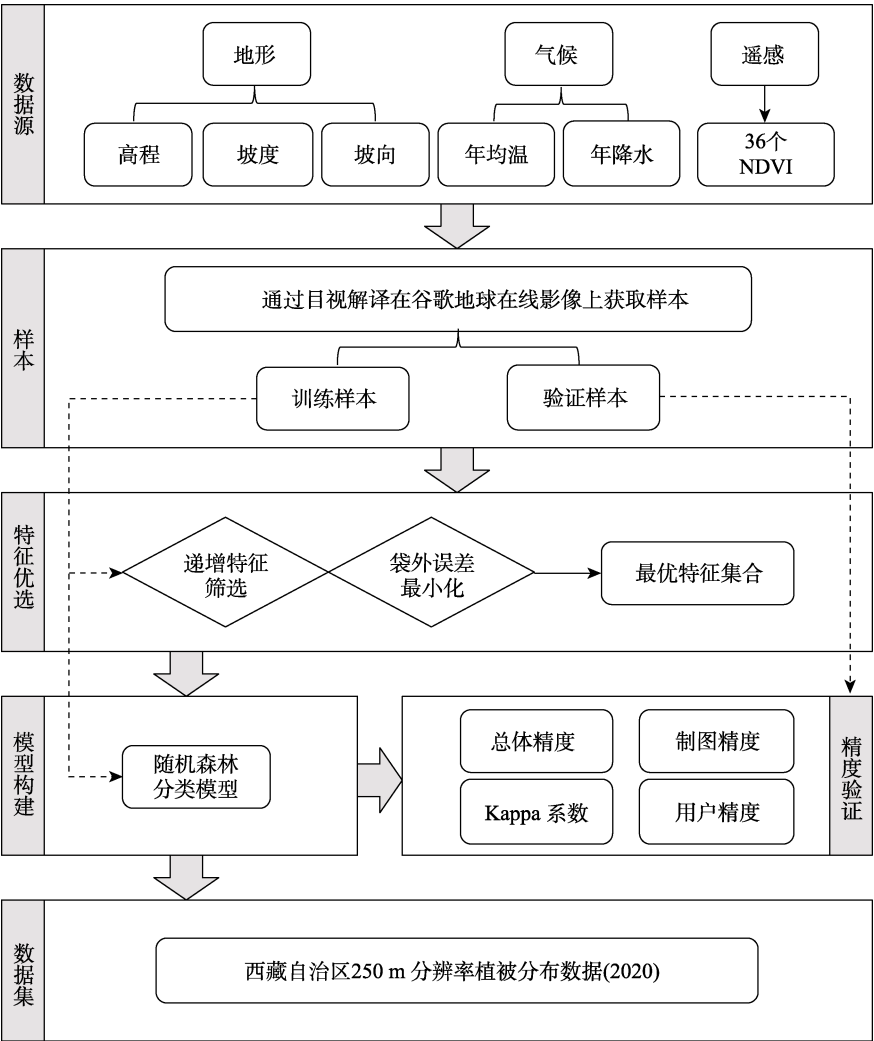


图 2 数据集研发技术路线图

4 数据结果与验证

4.1 数据结果

本研究采用递增特征筛选和袋外误差最小化原则获取最优制图特征。结果显示，随着特征数量的增加，袋外误差逐渐降低，当数量为 14 时，误差降至最低 0.16。但随后，特征数量增至 41 时，误差反而提升至 0.17，表明特征数量超过 14 个后，模型性能未见提高。据此，本研究确定了包含 14 个特征的最优特征组合，即高程、坡度、年均温、年降水、NDVI\_0720、NDVI\_1020、NDVI\_0110、NDVI\_0220、NDVI\_0310、NDVI\_0229、坡向、NDVI\_0930、NDVI\_1031、NDVI\_0120。最后，基于 GEE 中的随机森林分类模型和最优制图特征，制作了西藏自治区 250 m 分辨率植被分布数据（2020）（图 3）并统计了西藏自治区 2020 年各类型植被分布面积（表 3）。

表 3 西藏自治区 2020 年各类型植被分布面积统计表

植被覆盖类型	面积 (km <sup>2</sup> )	植被覆盖类型	面积 (km <sup>2</sup> )
阔叶林	49,039.6	高山植被	136,594.9
针叶林	49,870.5	高寒荒漠	154,924.0
针阔混交林	7,163.1	栽培植被	3,834.1
灌丛	10,386.3	湿地	4,259.3
高寒草甸	292,323.9	水体	32,169.5
高寒草原	404,775.0	其他	63,814.6

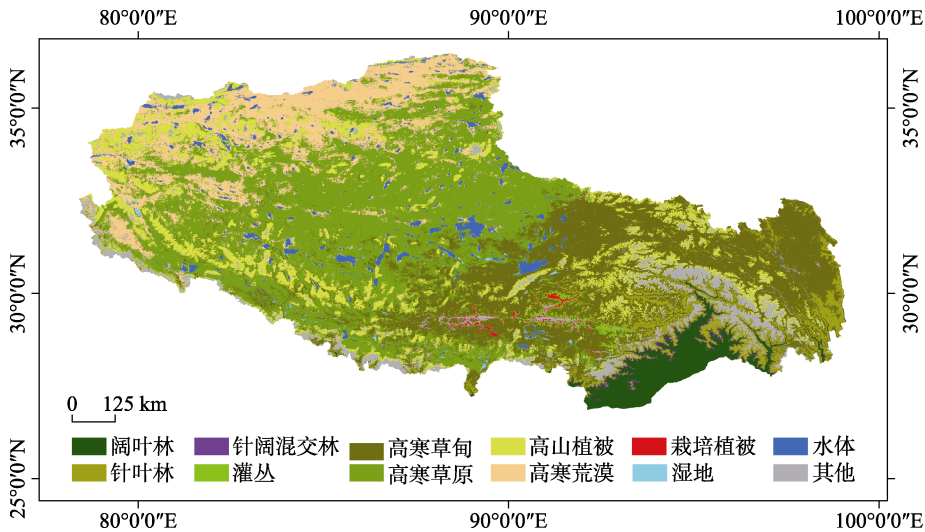


图 3 西藏自治区 250 m 分辨率植被数据可视化图（2020）

4.2 数据验证

西藏自治区 250 m 分辨率植被分布数据（2020）的总体精度为 0.81，Kappa 系数为 0.79，其混淆矩阵、制图精度和用户精度如表 4 所示。针阔混交林、灌丛、高寒草甸、高山植被、湿地的制图精度均低于 80%，其中针阔混交林和高山植被的制图精度分别为 40.0%和 54.5%。针阔混交林、高寒草原、高寒荒漠、栽培植被的用户精度低于 80%，其中针阔混交林的用户精度为 50.0%。

基于混淆矩阵可知，阔叶林与针阔混交林存在混淆，且高寒草原与高寒草甸存在混淆。首先，针阔混交林由阔叶树种和针叶树种共同构成，其光谱特征介于阔叶林和针叶林之间。特别是在春季和秋季，阔叶树的叶片生长和脱落使得针阔混交林的 NDVI 值与阔叶林类似，导致分类模型容易将阔叶林误判为针阔混交林。其次，高寒草甸和高寒草原均由草本植物组成，这两种类型在其过渡区域的空间分布上相互交错而形成生态过渡带。这种特性使得 NDVI 在该区域表现出显著的重叠，进而影响了植被类型的准确分类。最后，高山植被作为分布于林线或灌丛带以上至常年积雪带下限之间的植被类型，通常呈现出低覆盖且稀疏的分布特征，准确提取高山植被具有挑战。

表 4 西藏自治区 250 m 分辨率植被分布数据（2020）的混淆矩阵、制图精度与用户精度统计表

	阔叶林	针叶林	针阔混交林	灌丛	高寒草甸	高寒草原	高山植被	高寒荒漠	栽培植被	湿地	水体	其他	样本总和	制图精度
阔叶林	24	1	2	0	0	0	0	0	0	0	0	0	27	88.9%
针叶林	0	12	0	0	0	0	0	0	0	0	0	0	12	100.0%
针阔混交林	3	0	2	0	0	0	0	0	0	0	0	0	5	40.0%
灌丛	0	0	0	6	3	0	0	0	0	0	0	0	9	66.7%
高寒草甸	0	0	0	1	35	7	0	0	0	1	0	0	44	79.5%
高寒草原	0	0	0	0	0	43	0	5	0	0	1	0	49	87.8%
高山植被	0	0	0	0	2	3	12	0	0	0	0	5	22	54.5%
高寒荒漠	0	0	0	0	0	2	1	22	0	0	1	0	26	84.6%
栽培植被	0	0	0	0	0	0	0	0	9	0	0	1	10	90.0%
湿地	0	0	0	0	0	0	0	0	0	6	2	0	8	75.0%
水体	0	0	0	0	1	3	0	2	0	0	29	0	35	82.9%
其他	0	0	0	0	0	0	1	0	3	0	0	24	28	85.7%
样本总和	27	13	4	7	41	58	14	29	12	7	33	30	275	
用户精度	88.9%	92.3%	50.0%	85.7%	85.4%	74.1%	85.7%	75.9%	75.0%	85.7%	87.9%	80.0%		

5 讨论和总结

本文基于 GEE 云平台和随机森林算法，采用地形、气候和风云遥感数据制作了西藏自治区 250 m 分辨率植被分布数据（2020），总体精度为 81.5%，Kappa 系数为 0.79。在构建随机森林分类模型的 14 个最优制图特征中，高程、坡度、年均温、年降水依次位于重要性的前 4 位，表明西藏自治区的植被分布受地形与气候的影响显著。由东南到西北，依次分布着森林、灌丛、高寒草甸、高寒草原和高寒荒漠；随着海拔的升高，植被从高寒草原、灌丛逐渐过渡至高山植被。此外，与依赖于 Sentinel、Landsat、MODIS 等卫星数据制作的植被分布数据相比，本研究的遥感数据仅来自 FY-3D MERSI-II NDVI 数据，并达到了较高的精度，充分证明了国产风云卫星在植被制图领域的潜力。

综上所述，本数据集不仅揭示了西藏自治区更为精细的植被分布规律，还验证了国产风云卫星在区域植被制图领域的潜力。需要指出的是，在植被过渡区、植被分布离散和破碎化严重的地区，250 m 空间分辨率可能导致光谱信息重叠和混合像元现象。

作者分工：周广胜和任鸿瑞对数据集的开发做了总体设计；吕晓敏和汲玉河做了数据验证；张磊采集和处理了数据并撰写了数据论文等。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

[1] Venter, Z. S., Barton, D. N., Chakraborty, T., *et al.* Global 10 m land use land cover datasets: a comparison of dynamic world, world cover and esri land cover [J]. *Remote Sensing*, 2022, 14(16): 4101.

[2] Karra, K., Kontgis, C., Statman-Weil, Z., *et al.* Global land use/land cover with Sentinel 2 and deep learning



- [C]. 2021 IEEE international geoscience and remote sensing symposium IGARSS. Institute of Electrical and Electronics Engineers, 2021: 4704–4707.
- [3] Chen, J., Ban, Y., Li, S. Open access to Earth land-cover map [J]. *Nature*, 2014, 514(7523): 434–434.
- [4] Yang, J., Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019 [J]. *Earth System Science Data*, 2021, 13(8): 3907–3925.
- [5] Zhang, X., Liu, L., Chen, X., *et al.* GLC\_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery [J]. *Earth System Science Data*, 2021, 13(6): 2753–2776.
- [6] Friedl, M. A., McIver, D. K., Hodges, J. C., *et al.* Global land cover mapping from MODIS: algorithms and early results [J]. *Remote sensing of Environment*, 2002, 83(1/2), 287–302.
- [7] 高浩, 唐世浩, 韩秀珍. 风云气象卫星发展及其应用[J]. 科技导报, 2021, 39(15): 9–22.
- [8] 张磊, 周广胜, 任鸿瑞, 吕晓敏. 基于 FY-3D NDVI 的西藏自治区植被分类 250 m 栅格数据集 (2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.04.07.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.04.07.V1>.
- [9] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策 [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [10] 周广胜, 任鸿瑞, 刘通等. 一种基于地形-气候-遥感信息的区域植被制图方法及其在青藏高原的应用 [J]. 中国科学: 地球科学, 2023, 53(2): 227–235.
- [11] 周广胜, 任鸿瑞, 刘通等. 青藏高原 10 米分辨率植被图(2020)[DB/OL]. 国家青藏高原数据中心, 2022.
- [12] 中国科学院中国植被图编辑委员会. 中国植被及其地理格局——中华人民共和国植被图(1 : 1000000) 说明书[M]. 北京: 地质出版社, 2007.
- [13] Rabus, B., Eineder, M., Roth, A., *et al.* The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2003, 57(4): 241–262.
- [14] 彭守璋. 中国 1km 分辨率逐月平均气温数据集(1901–2022)[DB/OL]. 国家青藏高原数据中心, 2019.
- [15] 彭守璋. 中国 1km 分辨率逐月降水量数据集(1901–2022)[DB/OL]. 国家青藏高原数据中心, 2020.
- [16] Johansen, K., Phinn, S., Taylor, M. Mapping woody vegetation clearing in Queensland, Australia from Landsat imagery using the Google Earth Engine [J]. *Remote Sensing Applications: Society and Environment*, 2015, 1: 36–49.
- [17] Biau, G. Analysis of a random forests model [J]. *The Journal of Machine Learning Research*, 2012, 13: 1063–1095.