



基于机器学习和经验知识的青藏高原多时段植被制图

周继华^{1,2†}, 来利明^{1,2†}, 陈巧娥^{1,3}, 宋长青^{4*}, 高培超⁴, 叶思菁⁴, 沈石⁴, 杨刚刚^{1,3}, 郝海霞^{1,3}, 王贵豪^{1,3}, 熊喆⁵, 郑元润^{1,2*}

1. 中国科学院植物研究所, 植物多样性与特色经济作物全国重点实验室, 北京 100093

2. 国家植物园, 北京 100093

3. 中国科学院大学, 北京 100049

4. 北京师范大学地理学部, 北京 100875

5. 中国科学院大气物理研究所, 北京 100029

† 同等贡献

* 联系人, E-mail: songcq@bnu.edu.cn; zhengyr@ibcas.ac.cn

2024-05-08 收稿, 2024-08-19 修回, 2024-09-12 接受, 2024-09-13 网络版发表

第二次青藏高原综合科学考察研究(2019QZKK0608)资助

摘要 理解植被长期演变趋势及其对气候变化与人类活动的响应需要多时段植被分布资料. 目前青藏高原植被图主要为2000、2020年前后两个时段, 以及1990~2020年时段; 主要是基于地面调查数据的人工制图以及基于地面调查数据、遥感数据、地形与气候资料, 采用机器学习模型分类的自动制图, 存在部分植被类型混淆、制图精度有待提高等问题, 同时, 缺乏更早期时段的植被分布资料. 本文发展了整合地面调查数据、遥感数据、地形数据、气候数据等, 基于随机森林机器学习模型的人工制图技术, 获得了1980s、2020年青藏高原植被分布图; 基于植被地带性分析和1980s植被图, 获得青藏高原近似复原植被图; 以2020年青藏高原植被图为基础, 采用随机森林模型模拟了2035、2050年青藏高原植被分布图. 近似复原植被图、1980s和2020年青藏高原植被图的空间分辨率可达到100 m, 与公认较为可靠的《中华人民共和国植被图(1:100 万)》相比, 本研究获得的1980s和2020年植被图对植被类型边界的刻画更为准确, 植被斑块数目增加较多, 有利于更加准确理解和模拟青藏高原植被对气候变化的响应.

关键词 青藏高原, 植被制图, 机器学习模型, 多源数据, 植被分类

植被是—定地区中植物群落的总体. 由于植被在固定太阳能, 为其他生物提供第一性生产, 以及为人类社会提供最基本的供给、支撑、调节和文化服务功能等方面所发挥着巨大作用, 因而, 研究植被时空分异规律, 对于更好保护和利用植被资源具有重要意义. 植被本底数据的观察、获取和记录是理解植被基本过程的基础^[1]. 在准确获取植被数据的基础上, 采用合适的植被分类体系和符号, 以图件的形式反映植被空间分布是直观呈现植被分布数据的主要方式. 过去主要通过

人工地面调查的方式获取植被分布数据. 这种方法的优点是数据相对准确; 缺点是需要耗费大量人力、物力, 需时较长, 对于一些人工难以到达的区域数据获取困难. 随着遥感技术的发展, 间接获取植被分布数据的方法得到了越来越多的重视. 著名的遥感数据较多, 如Landsat TM数据、Landsat ETM+、SPOT、MODIS、AVHRR、IKONOS、QuickBird、ASTER、AVIRIS、Hyperion^[2]. 遥感数据的优点是能够快速且在较短时间内对大范围植被进行重复观察, 从而获得不同时间尺

引用格式: 周继华, 来利明, 陈巧娥, 等. 基于机器学习和经验知识的青藏高原多时段植被制图. 科学通报, 2025, 70: 134–144

Zhou J, Lai L, Chen Q, et al. Multi temporal vegetation mapping of the Tibetan Plateau via machine learning model simulation and experiential knowledge (in Chinese). Chin Sci Bull, 2025, 70: 134–144, doi: [10.1360/TB-2024-0496](https://doi.org/10.1360/TB-2024-0496)

度的观察数据。缺点是对一些难以区分的植被类型辨识效果欠佳。因而,如何整合包括野外调查数据和遥感数据在内的多源数据,准确识别多样化的植被成为植被模拟和制图的关键。

早期植被制图主要是以地形图为底图,采用地面调查数据,结合植被分布与气候、地形、地貌关系的分析进行人工制图^[1]。代表性的如1954年出版的1:400万《苏联植被图》、联合国粮农组织和联合国植被制图委员会于20世纪60年代陆续出版的1:500万《世界植被图》^[3]、1980年出版的1:400万《中国植被图》、2007年出版的《中华人民共和国植被图(1:100万)》^[1]。之后伴随遥感数据的日益丰富,发展了基于遥感数据的植被信息提取方法,包括非监督分类、监督分类和目视解译等^[4-9]。上述遥感数据植被信息提取方法有各自的优点,但其结果仍存在一定局限,在区分相当于中国植被分类体系中的植被型组、植被型等高级分类单位时效果较好,但在区分群系等中级单位时精度明显降低^[6]。近年来得到广泛运用的各类机器学习模型在进行植被信息提取和分类时亦存在较大不确定性。Zhou等人^[6]在进行黑河流域上游1:10万植被图制图和模拟研究时发现,随机森林、最大似然法、分类树等主流机器学习模型均无法精确区分群系的分布,对植被型组的模拟精度较高,对植被型的模拟精度次之。3种模型中模拟精度最高的是随机森林,在使用最佳数据组合的情况下,模拟精度为75%^[6]。虽然这一结果是在黑河上游山区地形高度复杂的情况下的植被模拟结果,但也表明仅采用机器学习模型方法进行植被制图存在较大误差。Yi等人^[10]采用随机森林、最大似然法、分类树、支持向量机模型模拟了京津冀地区的植被分布,发现在采用合适的变量数据组合时,随机森林模型的模拟精度为72%。由此可见,采用合适的数据组合,或在地形变化不大时,随机森林模型可以较好地模拟植被分布,但与植被制图的精度要求仍存在较大差距。因而,机器模型模拟与人工制图相结合可能是目前提高植被制图精度的有效途径。

青藏高原为地球第三极,孕育着极为丰富的生物多样性和独特的高寒植被,是中国重要的生态安全屏障。但全球变化和人类活动给青藏高原带来深刻影响,准确把握青藏高原过去不同时间段的植被分布,预测未来植被可能变化,有助于更加有效地应对和减缓全球变化的影响,筑牢青藏高原生态屏障。目前,青藏高原植被分布资料主要有《中华人民共和国植被图(1:100

万)》^[1],主要采用地面调查数据制图,反映20世纪末的植被状况;更新的《中华人民共和国植被图(1:100万)》^[11],采用地面调查和遥感数据制图;每10年1期的1990~2020年青藏高原植被图,使用随机森林模型和遥感数据编制^[12];基于地形-气候-遥感信息的青藏高原植被图^[8],采用地面调查、遥感、地形及气候数据,采用随机森林模型制图,考虑了多源数据及地形影响,结果较为可靠。但由于采用随机森林模型进行植被自动分类,尽管植被制图总体精度为89.5%,但高寒灌丛草甸的制图精度为56.9%,湿地的制图精度为64.6%,且存在湿地错分为高寒草甸和水体,落叶阔叶林与针阔混交林之间的混淆,以及高寒草甸与湿地、高寒灌丛草甸之间的混淆^[8]。其次,上述青藏高原植被调查资料来源为20世纪末和2020年左右,缺乏更早的植被分布资料,制约了对青藏高原植被动态的分析。本文基于地面调查数据、各类遥感数据、地形数据、气候数据、土地利用数据等目前各类可用植被图件资料,采用随机森林模型进行青藏高原植被分布模拟,再利用前述各类植被分布资料进行交叉验证分析,对青藏高原模拟植被图进行人工修正,获得更为可靠的1980s、2020年青藏高原植被分布图;结合植被地带性分析,以1980s植被图为基础,获得青藏高原近似植被图;以本文获得的2020年青藏高原植被图为基础,采用随机森林模型模拟了2035、2050年青藏高原植被分布图,为青藏高原植被研究和利用提供可靠的植被分布序列资料。

1 资料与方法

1.1 研究区域

由于知识积累和技术发展等多方面因素,学者们对青藏高原范围的理解和认知处于不断发展的过程中,其中张懿铨等人^[13]发布的2002版青藏高原范围为学界广泛接受,青藏高原在中国境内部分面积为257.24万km²。张懿铨等人^[14]根据地貌特征和高原山体的完整性等特征,发布了2021版青藏高原范围,行政范围涉及中国、印度、巴基斯坦、塔吉克斯坦、阿富汗、尼泊尔、不丹、缅甸、吉尔吉斯斯坦等9个国家,面积为308.34万km²,较2014版青藏高原面积增加了54.11万km²。本文青藏高原近似复原植被图(主要反映20世纪50年代,人为干扰相对较少时的植被分布状况)、2020年、2035年、2050年植被图采用张懿铨等发布的2021版青藏高原范围中国境内部分^[14],1980s青藏高原

植被图采用张懿铨等发布的2021版青藏高原范围^[14]。

1.2 数据资料

本文使用的数据包括青藏高原地貌、地形、气候数据、土地利用数据、土壤数据、植被类型调查数据、遥感数据等。

1.2.1 地貌、地形与气候数据

ASTER GDEM(30 m分辨率)下载自地理空间数据云(<https://www.gscloud.cn/>), 由ASTER GDEM 得到海拔, 使用ArcGIS10.0 (ESRI 2010)的spatial analysis 工具计算坡向和坡度, 生成地形图和粗略地貌数据。其中, 地形数据用于采用随机森林模型进行植被分布模拟。地貌数据用于辅助推断植被分布。

根据数据可获得性和数据时段尽量接近植被制图年份的原则获取气候数据, 气候数据空间分辨率约5 km, 下载自<http://www.worldclim.org/>, 覆盖1980~1989年、2010~2018年2个时段, 通过计算得到19个气候因子, 包括: 年平均气温、平均气温日较差、等温性、气温季节性变动系数、最热月最高温度、最冷月最低温度、气温年较差、最湿季平均温度、最干季平均温度、最暖季平均温度、最冷季平均温度、年降水量、最湿月降水量、最干月降水量、降水量季节性变化、最干季降水量、最湿季降水量、最暖季降水量、最冷季降水量。这些气候变量对植被分布有重要影响, 并在植被模型中经常被用作生物气候限制因子^[15]。由于上述数据缺乏2016~2020、2031~2035、2046~2050年3个时段的数据, 这3个时段的数据来自CESM数据集^[16], 空间分辨率约4 km。上述数据用于植被分布模拟。

1.2.2 土壤数据

全球土壤数据库(HWSD2.0)(<https://gaez.fao.org/pages/hwsd>)是由联合国粮食和农业组织与合作方共同开发的数据库, 于2023年1月发布。该数据包括7个土壤深度的数据, 包括40多种指标, 如酸碱度、含沙量、含水量、粉沙粒含量、有机碳含量等。数据分辨率约为1 km。本文仅使用其中的土壤类型属性, 用于根据土壤与植被分布的关系, 辅助推断和验证植被分布, 不用于模型模拟。

1.2.3 土地利用数据

1980~2020年栅格土地利用数据来自中国科学院资源环境科学数据平台(<http://www.resdc.cn/>), 分辨率为1 km。主要参考其中二级分类系统中的土地利用类型, 用于辅助推断和验证植被分布, 不用于模型模拟。同时

参考国际上较新的2020年左右10 m分辨率土地覆盖数据, 如Worldcover (<https://esa-worldcover.org/en>)和ESRI Land Cover (<https://livingatlas.arcgis.com/landcover/>)数据, 用于辅助推断和验证植被分布。

1.2.4 植被类型调查和观察数据

2020年植被类型调查数据来自2019~2021年的青藏高原野外考察记录数据564个; 各类文献报道的青藏高原植被分布调查数据, 数据选取的原则是尽可能向2020年集中, 前后时间相差不超过5年, 获得196个数据。1980S和2020年植被类型观察数据来自于遥感数据。1980~1985年Landsat3影像(分辨率90 m), 1986~1987年Landsat 4-5 TM影像(分辨率30 m), 2016~2020年Landsat 8 OLI影像(分辨率30 m), 该数据下载自美国地质勘探局(United States Geological Survey, USGS)网站(<http://www.usgs.gov/>)和地理空间数据云(<http://www.gscloud.cn/>)。通过目视解译, 分别在1980~1987、2016~2020年两个时段各选取2900个数据, 数据选取的原则是尽可能向1980或2020年集中, 充分考虑不同植被类型和在研究区域均匀分布。合计获得1980s植被数据2900个, 2020年植被数据3660个(图1)。上述数据分别用于1980S和2020年植被分布模拟模型的输入数据, 也用于基于2020年植被图生成2035和2050年植被分布随机森林模型的训练和验证。

1.2.5 遥感数据

前述提到的青藏高原1980~1987、2016~2020年Landsat遥感影像经地形图几何校正后镶嵌而成。主要包括波段组合、地形图几何配准(采用多项式几何校正方法), 再经过投影变换、数字镶嵌等处理工作。利用均方根误差法评价单景影像几何配准精度, 采样法评价数字镶嵌影像精度, 完成青藏高原遥感影像数字镶嵌图, 得到植被制图基本数据, 即1980s时段和2020年时段遥感数据。选择遥感数据时, 尽量向1980和2020年靠近。如有云雾影响, 则选择临近日期无云雾影响的数据。预处理工具为ENVI 5.1。上述数据用于植被分布模拟和后续植被图人工修正。在数据可得的情况下, 同时参考Google Earth数据、欧洲空间局Sentinel-2系列卫星遥感影像等高分辨率遥感数据, 这两类数据不用于植被分布模拟, 仅用于植被分布图后续人工修正。

1.3 植被指数与植被分布模型

1.3.1 植被指数

Zhou等人^[6]研究表明, 使用气候、地形变量以及

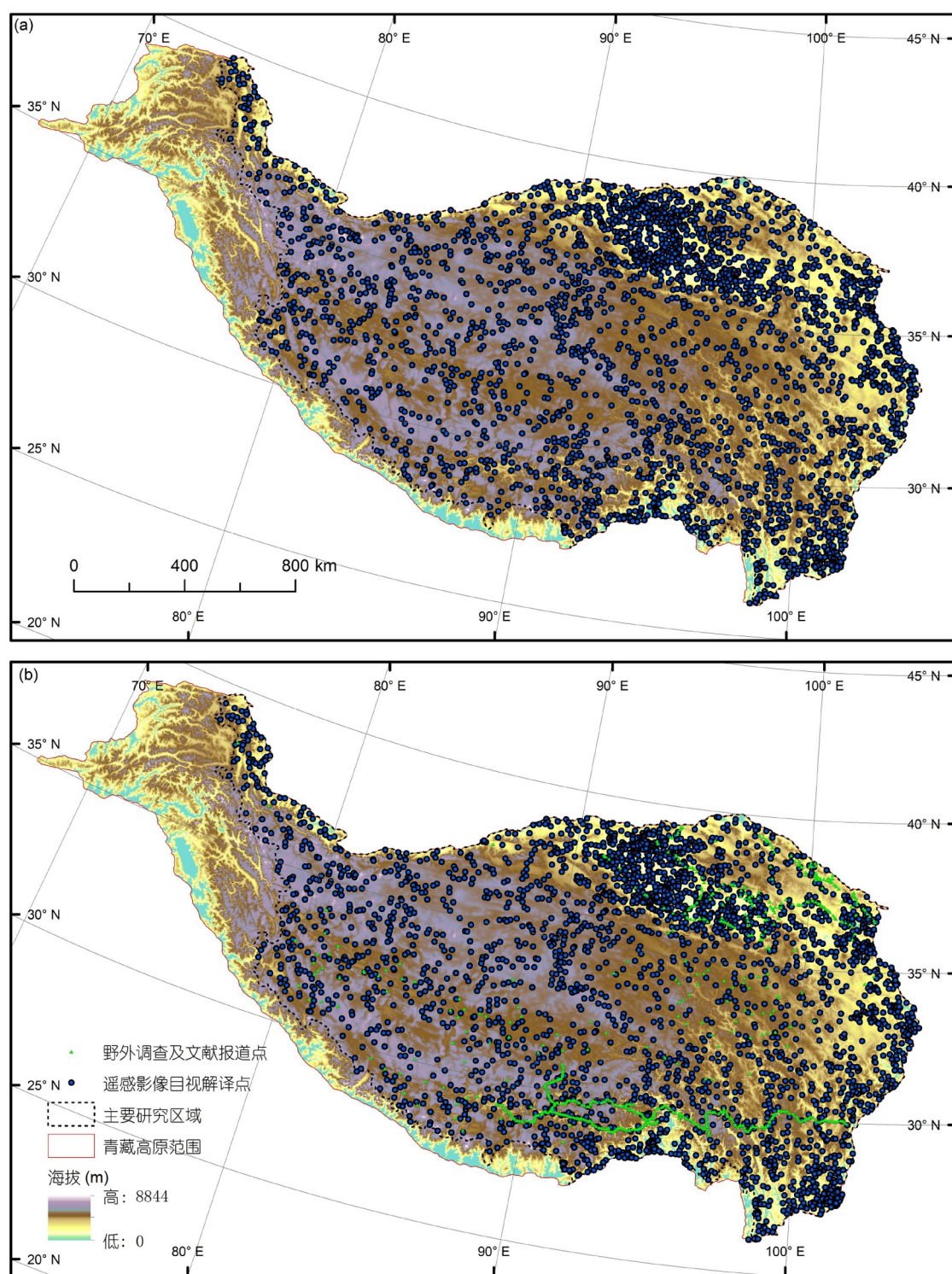


图 1 青藏高原植被类型调查和观察数据空间分布. (a) 1980s; (b) 2020年

Figure 1 Field data and remote sensing observed data in Tibetan Plateau. (a) 1980s; (b) 2020

14个夏季(7月)光谱植被指数, 运用随机森林模型可以较好模拟青藏高原东北部黑河流域上游祁连山区的植

被分布. 因此, 本文使用了14个夏季植被指数参与植被分类. 使用的植被指数包括比值植被指数(RVI)、亮度

指数(BI)、绿度植被指数(GVI)、湿度指数(WI)、差值植被指数(DVI)、绿度比值(GR)、近红外比值(MIR)、土壤校正植被指数(SAVI)、优化土壤校正植被指数(OSAVI)、大气阻抗植被指数(ARVI)、归一化植被指数(NDVI)、增强植被指数(EVI)、归一化耕作指数(NDTI)、归一化衰败植被指数(NDSVI), 其计算方法见Zhou等人^[6].

1.3.2 植被分布模型

本文使用随机森林模型进行青藏高原植被分类^[6]. 随机森林模型使用EnMAP Box^[17]的默认设置生成, 包含100个树, 节点计算使用GINI系数. EnMAP Box 是一款由德国环境制图与分析计划(environmental mapping and analysis program) 项目组基于IDL(interactive data language)开发的处理高光谱遥感数据的工具包. 工具包提供了数据归一化、支持向量机和随机森林分类和回归、滤波等功能.

1.4 植被制图方法

1.4.1 基于机器学习模型的植被模拟制图

采用随机森林模型, 气候数据、地形数据及1980s时段和2020年时段Landsat影像夏季光谱植被指数, 1980s 2900个植被类型观察数据, 2020年3660个植被类型调查和观察数据全部用于模型训练, 模拟生成1980s和2020年植被图, 作为后续人工植被制图的基础.

1.4.2 植被图人工修正

植被图人工修正数据主要是通过地面调查数据、各类发表文献中有植被分布记录的数据、各类遥感数据、1:100万植被图、气候、地形、地貌、土壤数据进行交叉验证, 编制1980s和2020年植被分布图. 以2020年植被图为例, 阐述植被图修正过程, 所使用数据尽量靠近2020年, 除前述指明的时间外, 其他数据与2020年相比, 前后时间相差不超过5年.

(1) 植被分类标准、图例单位和系统. 采用《中华人民共和国植被图(1:100万)》^[1]的分类标准、图例单位和系统, 包括植被型组、植被型两个单位^[1].

主要底图地形图的分辨率为30 m, 主要遥感数据的分辨率为30 m, 编制1980s植被图时参考的少量1986~1987年的遥感数据分辨率为90 m, Google Earth数据分辨率最高为1 m, 最低为30~100 m. 综合考虑, 本图的分辨率可达到100 m. 根据制图精度, 图斑面积 $\geq 1000 \text{ m}^2$, 对于面积 $< 1000 \text{ m}^2$ 的图斑, 使其与相邻图斑融合.

(2) 目视解译标志建立. 依据地面调查数据, 确定植物群落与特殊地物在遥感影像中的具体位置, 判读其在遥感影像中展现的形状、色调及纹理特征, 进而建立目视解译标志.

(3) 特殊地物与植物群落提取和制图. 使用30 m地形图、Landsat遥感影像、Google Earth, 以及前述各类数据, 在随机森林模型植被预分类的基础上, 提取青藏高原易于识别的地物, 主要包括冰川、裸露沙漠、裸露戈壁、裸露盐碱地、水系及水体等地物, 以及栽培植被等.

在随机森林模型植被预分类的基础上, 优先利用地面调查资料修正每个单元的植被类型. 对于缺乏实地数据的区域, 结合《中华人民共和国植被图(1:100万)》^[1]、Landsat遥感数据, Google Earth、气候、土壤类型数据等, 分析每个植物群落分布单元, 修正每一个单元的植被群落类型, 形成初步的青藏高原植被图.

(4) 一致性检查. 针对初步的青藏高原植被图, 结合现有资料, 包括《中华人民共和国植被图(1:100万)》^[1]、土地利用图、土壤、地形图、遥感影像资料进行一致性检查, 确保植被图和气候、土壤、地形、现有相关图件和遥感资料逻辑相符.

(5) 植被图整饰. 与《中华人民共和国植被图(1:100万)》^[1]一致, 运用图斑和数字相结合的方法, 表示不同植被类型和制图单位, 使之满足清晰易读、重点突出的要求.

1.4.3 近似复原植被图编制

复原植被图是指未受人类经济活动破坏以前的原始植被分布状况的植被图. 本文近似复原植被图指20世纪50年代前后受人类干扰较少的植被分布状况的植被图, 近似反映原始植被分布状况, 不对次生植被进行复原. 以1980S植被图为基础, 考虑气候、地形、地貌、土壤特点和周边残存自然植被, 将栽培植被和居民地用邻近区域自然植被替代, 形成青藏高原近似复原植被图.

1.4.4 未来植被图模拟

在完成的2020年植被图中, 忽略面积小于2021版青藏高原^[14]国内部分面积0.02%的植被型(植被亚型), 共6个. 在每个面积大于 5000 km^2 的植被型(植被亚型)和无植被地段中提取100个以上植被样本数据; 在面积小于 5000 km^2 的植被型(植被亚型)和无植被地段中, 视面积大小提取10~50个植被样本数据, 共4690个, 且这些数据尽量均匀分布在青藏高原. 连同2020年地面调

查数据760个,共5450个数据。其中,4360个数据用于随机森林模型训练,1090个数据用于随机森林模型验证。以2016~2020年平均气候数据^[15]及其计算得到的19个气候因子和地形数据训练随机森林模型,模型总体精度为58.5%,Kappa系数为0.55,训练生成的随机森林模型能够较好模拟2020年青藏高原植被类型分布。尽管2020~2050地形变化不大,但在青藏高原地形变化剧烈的区域,地形作为辅助控制因子,可以提高植被模拟精度。运用此模型,采用2031~2035年平均气候数据和2046~2050年平均气候数据及其计算得到的19个气候因子,以及地形数据,分别模拟得到2035年和2050年青藏高原植被图。

2 结果

2.1 青藏高原不同时段植被图及其时间变化

近似复原植被图、1980s、2020年植被图分别包含11个植被型组,43个植被型,2个植被亚型。受大气、气候、地形、土壤等条件影响,5个时段青藏高原由东南向西北呈现森林、灌丛、草甸、草原、荒漠的植被分布格局未发生变化,但各类型植被的面积和分布范围发生变化。根据主要制图数据的空间分辨率,近似复原植被图、1980s和2020年青藏高原植被图的空间分辨率可达到100 m(图2)。

在近似复原植被到2050年青藏高原(2021年张德铨等人^[14]发布实际范围的中国境内部分)植被型(植被亚型)中,面积最大的是高寒禾草、苔草草原,为58.84万 km^2 ,其次是高寒嵩草、杂类草草甸,面积为57.47万 km^2 ,面积超过1万 km^2 的其他植被型还包括:高山稀疏植被,亚高山硬叶常绿阔叶灌丛,亚热带和热带山地针叶林,高寒垫状矮半灌木荒漠,亚高山落叶阔叶灌丛,温带半灌木、矮半灌木荒漠,温带丛生矮禾草、矮半灌木荒漠草原,温带丛生禾草典型草原,高山垫状植被,温带禾草、杂类草盐生草甸,亚高山常绿针叶灌丛,亚热带硬叶常绿阔叶林和矮林,温带灌木荒漠,亚热带、热带常绿阔叶、落叶阔叶灌丛(常含稀树),亚热带季风常绿阔叶林。无植被地段中冰川积雪面积最大,其次是水体,裸露石山、裸露戈壁、风蚀残丘、裸露盐碱地的面积也超过1万 km^2 (表S1)。

高寒禾草、苔草草原,高寒嵩草、杂类草草甸,高山稀疏植被,高寒垫状矮半灌木荒漠,高山垫状植被,寒温带和温带山地针叶林,高寒沼泽,寒温带、温带沼

泽等高寒性质的植被型面积降低,其中,寒温带和温带山地针叶林降低幅度最高,达36.9%,寒温带、温带沼泽降低7.2%(表S1)。

温带禾草、苔草及杂类草沼泽化草甸,温带禾草、杂类草草甸,亚热带、热带草丛,温带落叶阔叶林,亚热带落叶阔叶林,温带针叶林,亚热带硬叶常绿阔叶林和矮林,温带草原化灌木荒漠,温带禾草、杂类草盐生草甸,亚热带常绿、落叶阔叶混交林面积降低,其中,温带禾草、苔草及杂类草沼泽化草甸面积降低97.1%,温带禾草、杂类草草甸,亚热带、热带草丛降低50%以上(表S1)。

亚热带常绿阔叶林,温带矮半乔木荒漠,温带落叶灌丛,亚热带、热带旱生常绿肉质多刺灌丛,亚热带、热带常绿阔叶、落叶阔叶灌丛(常含稀树),亚热带和热带山地针叶林,温带落叶小叶疏林,温带灌木荒漠,亚热带针叶林,亚高山常绿针叶灌丛,亚高山落叶阔叶灌丛,亚热带山地针叶、常绿阔叶、落叶阔叶混交林,温带禾草、杂类草草甸草原面积增加,其中,亚热带常绿阔叶林,温带矮半乔木荒漠,温带落叶灌丛分别增加109.9%、50.5%、29.0%(表S1)。

无植被地段中,冰川积雪面积降低2.8%,水体面积增加34.3%,裸露盐碱地增加9.1%,盐壳增加7.4%(表S1)。

2.2 青藏高原植被制图结果评价

通过使用高分辨率遥感数据,与《中华人民共和国植被图(1:100万)》^[1]相比,本文编制的青藏高原植被图对植物群落边界描绘更为准确,如白郎县、江孜县附近对栽培植被(淡棕色)、长芒草草原(黄色)、小蒿草高寒草甸(淡绿色)边界的描绘,本文的1980S植被图明显要比《中华人民共和国植被图(1:100万)》^[1]更为精细,表明本文的青藏高原植被图精度更高(图3)。

从植被型(植被亚型)和非植被地段斑块数目来看(中国境内部分),《中华人民共和国植被图(1:100万)》^[1]斑块数目为15068个。本文完成的近似复原、1980S、2020年(中国境内部分)植被图斑块数分别为64402、69904、125960个,分别增加3.3、3.6、7.4倍。进一步说明本文所完成的植被图在植被类型斑块划分时更为精细。

3 讨论

随着气候变化的日益明显,分析和预测植被对气

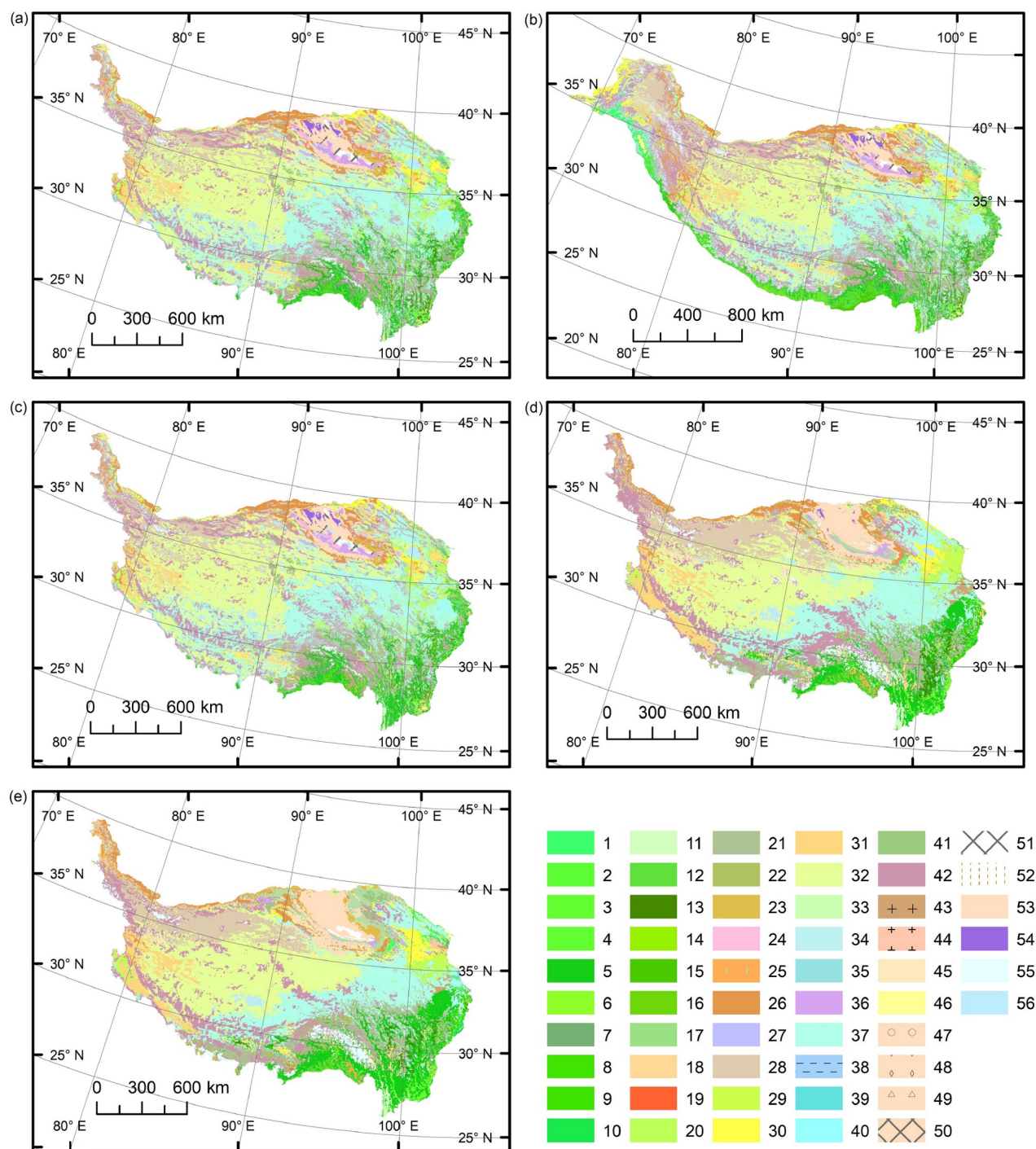


图 2 青藏高原植被图. (a) 近似复原; (b) 1980s; (c) 2020年; (d) 2035年; (e) 2050年. 图例编号见表S1

Figure 2 Vegetation maps of Tibetan Plateau. (a) Approximately natural; (b) 1980s; (c) 2020; (d) 2035; (e) 2050. Codes for legend were shown in Table S1

候变化的影响成为人类可持续发展的重大需求, 而高分辨率、多时段植被图就成为实现这一目标的基础. 研究表明包括高山林线在内的植物群落过渡区是对气

候变化最敏感的区域^[18~20], 也是植被制图的难点区域. 因此, 急需提升植被边界制图的精度. 受限于技术原因和制图区域的可达性, 青藏高原现有植被图对部分植

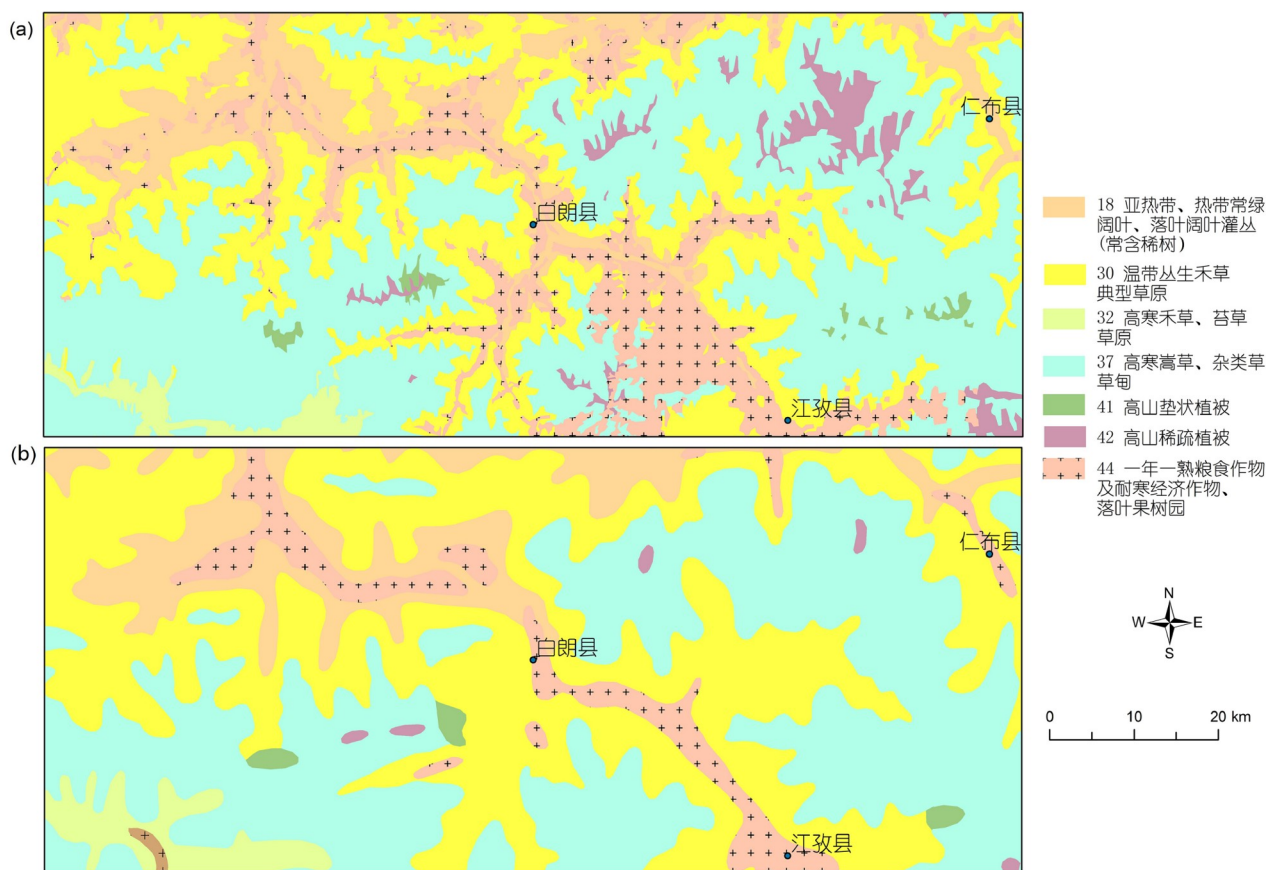


图3 青藏高原植被图局部。(a) 1980s; (b) 《中华人民共和国植被图(1:100 万)》
Figure 3 Vegetation maps of Tibetan Plateau in specific areas. (a) 1980s; (b) *Vegetation Map of China* (1:1000000)

被类型边界的刻画不够准确(图3)。本文基于高分辨率遥感数据和地面数据,对植被类型的边界进行细致分析,植被类型边界的刻画精度提升明显,特别是对于过去难以到达的青藏高原高山植被,以及冰川等非植被地段资料匮乏区的刻画较为准确,有助于更好理解和模拟气候变化对青藏高原植被的影响。但由于植物识别的复杂性,主流遥感数据仍难以准确识别易混淆植物种及其组成的植物群落,针对研究和应用需求,未来应发展激光雷达^[21]、人工智能与无人机结合的高分辨率植被数据获取和分析技术,进一步提高植物群落过渡区的植被制图精度。

植被制图斑块数量可在较大程度上衡量植被制图精度。在分析气候变化和植被关系时,较为详细的植被空间分布信息能够更好反应植被分布特征,有助于准确把握植被与气候的关系,并预测未来植被的可能变化。本文1980s植被图与《中华人民共和国植被图(1:100 万)》^[1]相比,斑块数目提升较多,更加精细地反

应了青藏高原植被分布特征,特别是地形复杂区域的植被特征。随着高性能计算模拟工具的发展,高分辨率的植被分布数据将会更好满足地球表层系统的模拟需求。

多时段植被分布数据是了解植被动态过程的基础。现有青藏高原植被图主要集中在2000年前后和2020年,以及1990~2020年时段^[1,8,11,12],限制了对青藏高原植被动态的准确认识。本文编制的青藏高原多时段植被图理解青藏高原的植被变化提供了科学数据。近似复原植被图到2020年植被图的分析显示,青藏高原植被发生了明显变化,高寒性质的植被类型面积显著减少。这可能是由于气温升高导致高寒植被向其他类型植被转变有关,这种可能变化也可以从青藏高原气候变暖导致植被返青期总体提前,高原树线位置上升,高寒草原植物物种丰富度和多样性下降的相关研究中得到佐证^[22]。其次,由于气温升高,温带矮半乔木荒漠等受干旱气候控制的植被面积增加也较多。由上可见,由于青

藏高原独特的特征, 人类活动相对较弱^[8], 植被变化主要由气候变化引起. 未来应以青藏高原多时段植被图为基础, 结合过去观察气候数据和未来气候变化情景数据, 分析青藏高原过去植被变化规律, 预测未来可能变化, 服务于青藏高原生态屏障建设.

从1980S到2020年, 非植被地段中冰川面积降低, 相关研究也显示气候持续变暖加速了冰川存量的急速减少, 2000~2019年间, 整个亚洲水塔的冰川质量以约 -21.1 Gt/a 的速率减少^[23,24]. 多时段植被图显示从1980S到2020年水体面积增加较大, 达34.3%, 研究表明1995年之后青藏高原湖泊数量和面积都呈现增加趋势, 且与气温上升密切相关^[24-26]. 上述结果表明青藏高原多时段植被图也可为理解冰川、湖泊等非植被单元的动态过程提供基础数据.

本文基于气候和地形数据, 采用随机森林模型模拟了2020年的植被分布, 但模拟精度只有58.5%, 这主要是由于我们使用的气候数据空间分辨率只有约4 km, 由于满足我们未来植被分布模拟时段要求的气候数据空间分辨率只有4 km, 为使二者空间分辨率匹配, 使用

了相同分辨率的数据. 其次, 我们模拟的植被型级别的植被类型和无植被地段数量较多, 达到49个, 因此, 导致模拟精度降低.

4 结论

本文整合植被地面调查数据、遥感数据, 在机器学习模型对植被预分类的基础上, 通过人工修正的方式获得1980S和2020年植被图; 考虑植被地带性规律, 基于1980S植被图, 编制了青藏高原近似复原植被图; 以2020年植被图为基础, 采用随机森林模型模拟了青藏高原2035、2050年植被图. 比较分析显示本文1980s植被图在植被类型边界刻画, 植被型(植被亚型)斑块数目上均明显优于《中华人民共和国植被图(1:100万)》^[1], 显示了本文综合遥感植被制图、机器学习模型分类和传统基于调查数据制图的优越性. 但由于目前大尺度遥感数据仍不足以支持部分植物种类的精确识别, 未来仍需要在人工智能和无人机等新技术的支持下获得精度更高的植被数据. 本文青藏高原多时段植被图可为认识和研究青藏高原植被动态提供科学数据.

数据可用性声明

青藏高原植被图(近似复原, 1980s, 2020)已在国家青藏高原科学数据中心发布, 数据下载链接: <https://data.tpdc.ac.cn/zh-hans/data/0b32907c-9f64-407b-8099-849db1900005>; <https://data.tpdc.ac.cn/zh-hans/data/0eb0a1c1-ccff-41b4-80da-883497dfbb01>; <https://data.tpdc.ac.cn/zh-hans/data/3a49fac7-bf9d-4b69-a855-bd2380ebaa0b>.

参考文献

- 1 Editorial Committee of Vegetation Map of China, the Chinese Academy of Sciences. The Vegetation Map of the People's Republic of China (1:1000000) (in Chinese). Beijing: Geological Publishing House, 2007 [中国科学院中国植被图编辑委员会. 中华人民共和国植被图(1:1000000). 北京: 地质出版社, 2007]
- 2 Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: A review. *J Plant Ecol*, 2008, 1: 9–23
- 3 European Environment Agency. Terrestrial Habitat Mapping in Europe: An Overview. Luxembourg: Publications Office of the EU, 2014
- 4 Colgan M S, Baldeck C A, Féret J B, et al. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. *Remote Sens*, 2012, 4: 3462–3480
- 5 Roy P S, Behera M D, Murthy M S R, et al. New vegetation type map of India prepared using satellite remote sensing: Comparison with global vegetation maps and utilities. *Int J Appl Earth Observ Geoinf*, 2015, 39: 142–159
- 6 Zhou J, Lai L, Guan T, et al. Comparison modeling for alpine vegetation distribution in an arid area. *Environ Monit Assess*, 2016, 188: 408
- 7 Dayaram A, Powrie L, Rebelo T, et al. Vegetation Map of South Africa, Lesotho and Swaziland 2009 and 2012: A description of changes from 2006. *Bothalia*, 2017, 47, <http://doi.org/>
- 8 Zhou G S, Ren H R, Liu T, et al. A new regional vegetation mapping method based on terrain-climate-remote sensing and its application on the Qinghai-Xizang Plateau (in Chinese). *Sci China Earth Sci*, 2023, 66: 237–246 [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. 中国科学: 地球科学, 2023, 53: 227–235], <http://doi.org/>
- 9 Zhang X L, Zhou J H, Cai W T, et al. Vegetation mapping (1:100 000) in Heihe River Basin using 3S technology (in Chinese). *J Northwest Norm Univ (Nat Sci)*, 2018, 54: 95–101 [张晓龙, 周继华, 蔡文涛, 等. 基于3S技术的黑河流域1:100 000植被制图. 西北师范大学学报(自然科学版), 2018, 54: 95–101]

- 10 Yi S, Zhou J, Lai L, et al. Simulating highly disturbed vegetation distribution: the case of China's Jing-Jin-Ji region. *PeerJ*, 2020, 8: e9839
- 11 Su Y, Guo Q, Hu T, et al. An updated Vegetation Map of China (1:1000000). *Sci Bull*, 2020, 65: 1125–1136
- 12 Wu F, Ren H, Zhou G. The 30 m vegetation maps from 1990 to 2020 in the Tibetan Plateau. *Sci Data*, 2024, 11: 804
- 13 Zhang Y L, Li B Y, Zheng D. A discussion on the boundary and area of the Tibetan Plateau in China (in Chinese). *Geogr Res*, 2002, 21: 1–8 [张懿锂, 李炳元, 郑度. 论青藏高原范围与面积. *地理研究*, 2002, 21: 1–8]
- 14 Zhang Y L, Li B Y, Liu X L, et al. Redetermine the region and boundaries of Tibetan Plateau (in Chinese). *Geogr Res*, 2021, 40: 1543–1553 [张懿锂, 李炳元, 刘林山, 等. 再论青藏高原范围. *地理研究*, 2021, 40: 1543–1553]
- 15 Sitch S, Smith B, Prentice I C, et al. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob Change Biol*, 2023, 9: 161–185
- 16 Karger D N, Schmatz D R, Dettling G, et al. High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Sci Data*, 2020, 7: 1
- 17 Van der Linden S, Rabe A, Held M, et al. The EnMAP-Box—A toolbox and application programming interface for EnMAP data processing. *Remote Sens*, 2015, 7: 11249–11266
- 18 Danby R K, Hik D S. Variability, contingency and rapid change in recent subarctic alpine tree line dynamics. *J Ecol*, 2007, 95: 352–363
- 19 Paulsen J, Körner C. A climate-based model to predict potential treeline position around the globe. *Alp Bot*, 2014, 124: 1–12
- 20 Mu H, Han F, Zhang B, et al. Characteristics of timberline and treeline altitudinal distribution in Mt. Namjagbarwa and their geographical interpretation. *J Mt Sci*, 2022, 19: 2846–2860
- 21 Su Y, Guo Q, Fry D L, et al. A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery. *Can J Remote Sens*, 2016, 42: 1–15
- 22 Piao S L, Zhang X Z, Wang T, et al. Responses and feedback of the Tibetan Plateau's alpine ecosystem to climate change (in Chinese). *Chin Sci Bull*, 2019, 64: 2842–2855 [朴世龙, 张宪洲, 汪涛, 等. 青藏高原生态系统对气候变化的响应及其反馈. *科学通报*, 2019, 64: 2842–2855], <http://doi.org/>
- 23 Hugonnet R, McNabb R, Berthier E, et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature*, 2021, 592: 726–731
- 24 Zhang Q, Wang G, Zhao J Q, et al. Water circulation and water resources of Asia's water tower: The past and future (in Chinese). *Chin Sci Bull*, 2023, 68: 4982–4994 [张强, 王港, 赵佳琪, 等. 亚洲水塔水循环和水资源研究进展与展望. *科学通报*, 2023, 68: 4982–4994], <http://doi.org/>
- 25 Xu X D, Dong L L, Zhao Y, et al. Effect of the Asian water tower over the Qinghai-Tibet Plateau and the characteristics of atmospheric water circulation (in Chinese) *Chin Sci Bull*, 2019, 64: 2830–2841 [徐祥德, 董李丽, 赵阳, 等. 青藏高原亚洲水塔效应和大气水分循环特征. *科学通报*, 2019, 64: 2830–2841]
- 26 Zhang G, Luo W, Chen W, et al. A robust but variable lake expansion on the Tibetan Plateau. *Sci Bull*, 2019, 64: 1306–1309

补充材料

表 S1 青藏高原不同时段植被型(植被亚型)面积(万 km²)

本文以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “基于机器学习和经验知识的青藏高原多时段植被制图”

Multi temporal vegetation mapping of the Tibetan Plateau via machine learning model simulation and experiential knowledge

Jihua Zhou^{1,2†}, Liming Lai^{1,2†}, Qiaoe Chen^{1,3}, Changqing Song^{4*}, Peichao Gao⁴, Sijing Ye⁴, Shi Shen⁴, Ganggang Yang^{1,3}, Haixia Hao^{1,3}, Guihao Wang^{1,3}, Zhe Xiong⁵ & Yuanrun Zheng^{1,2*}

¹ State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

² China National Botanical Garden, Beijing 100093, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

⁵ Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

† Equally contributed to this work

* Corresponding authors, E-mail: songcq@bnu.edu.cn; zhengyr@ibcas.ac.cn

Understanding long-term trends in vegetation distribution and its responses to climate change and human activities requires vegetation distribution data from multiple periods. At present, the vegetation maps of the Tibetan Plateau mainly exist for periods of time surrounding 2000 and 2020, during which time they were based on field survey data, and 10-year periods from 1990 to 2020, during which time they were based on random forest classifiers and Landsat imagery. These maps were created through manual mapping based on ground observation data and automatic mapping based on ground observation data, remote sensing data, terrain data and climate data, with machine learning models used for classification. There are problems, such as confusion among some vegetation types, and map accuracy needs to be improved. Moreover, there is a lack of vegetation distribution data from earlier periods. This study developed artificial mapping techniques based on random forest machine learning models, integrating observation data, remote sensing data, terrain data, climate data, etc., and obtained vegetation distribution maps of the Tibetan Plateau in the 1980s and 2020. On the basis of zonal vegetation analysis and the 1980s vegetation map, an approximated original vegetation map of the Tibetan Plateau was obtained. On the basis of the vegetation map of the Tibetan Plateau from 2020, a random forest model was used to simulate the vegetation distribution maps of the Tibetan Plateau in 2035 and 2050. The spatial resolution of the approximated original vegetation map, the 1980s vegetation map and the 2020 vegetation map was approximately 100 m. The approximated original vegetation map, the 1980s vegetation map and the 2020 vegetation map contain 11 vegetation groups, 43 vegetation types, and 2 vegetation subtypes. Owing to the influences of climate, terrain, soil and other conditions, the vegetation distribution patterns of forests, shrubs, grasslands, and deserts on the Tibetan Plateau from southeast to northwest did not change across the five periods, but the areas and distribution ranges of various types of vegetation changed. In the approximated original vegetation map, the vegetation type with the largest area was alpine grass and *Carex* steppe, totaling 588400 km², followed by alpine *Kobresia* spp. and forb meadows, totaling 574700 km². There were more than 15 other vegetation types with areas exceeding 10000 km². In areas without vegetation, the areas covered by glaciers and snow were the largest, followed by water bodies. The areas of bare rocky hills, bare gobi hills, yarden hills, and bare salt licks also exceeded 10000 km². The areas of alpine vegetation types decreased, with the greatest reduction in cold-temperate and temperate mountain needleleaf forests, reaching 36.9%, and a decrease of 7.2% in cold-temperate and temperate swamps. In addition, the areas of temperate grass, *Carex* and forb swamp meadows decreased by 97.1%, whereas those of temperate grass and forb meadows and subtropical and tropical grass forb communities decreased by more than 50%. The areas of subtropical broadleaf evergreen forests, temperate dwarf semiarboreal deserts, and temperate broadleaf deciduous scrubs increased by 109.9%, 50.5%, and 29.0%, respectively. The area of glaciers and snow decreased by 2.8%, the area of water bodies increased by 34.3%, the area of bare salt licks increased by 9.1%, and the area of salt crust increased by 7.4%. Compared with the widely recognized and reliable *Vegetation Map of the People's Republic of China* (1:100000), the vegetation map in the 1980s and 2020 obtained in this study provides a more accurate depiction of vegetation type boundaries, increasing the number of vegetation patches several folds, which is conducive to a more accurate understanding and simulation of the response of vegetation on the Tibetan Plateau to climate change.

Tibetan Plateau, vegetation mapping, machine learning model, multi-source data, vegetation classification

doi: [10.1360/TB-2024-0496](https://doi.org/10.1360/TB-2024-0496)