

引用格式: 卢德彬, 石彰, 索超, 韩畅, 耿敬杰. 基于 GEE 和双层随机森林模型的土地利用分类——以黔中城市群为例 [J]. 山地学报, 2025, 43(5): 749-762.

LU Debin, SHI Zhang, SUO Chao, HAN Chang, GENG Jingjie. Land-use classification based on GEE and a two-layer random-forest model: A case study of the central Guizhou urban agglomeration [J]. Mountain Research, 2025, 43(5): 749-762.

基于 GEE 和双层随机森林模型的土地利用分类 ——以黔中城市群为实证

卢德彬^{1,2}, 石彰¹, 索超², 韩畅¹, 耿敬杰^{1,3}

(1. 铜仁学院 a. 旅游与地理系; b. 贵州省高等学校山地国土空间智能监测与政策仿真工程研究中心, 贵州 铜仁 554300;

2. 贵州大学 公共管理学院, 贵阳 550025; 3. 四川大学 法学院, 成都 610207)

摘要: 在山地多云雾环境下, 地表辐射信号畸变严重, 地物识别困难, 亟需高精度土地利用/覆被变化(LUCC)数据支撑城市群国土空间精细化治理。然而, 现有适宜样本、影像与分类方法所导致的本地化高精度制图仍然不足。本文以黔中城市群为研究区, 基于 Google Earth Engine 平台获取 Sentinel-1/2 影像, 经去云、阴影和雪预处理, 提取光谱、指数与 SAR 波谱三类特征变量; 借鉴 Stacking 思想, 构建双层随机森林框架: 第一层输出地类概率, 第二层以高概率区重采样生成新特征并二次分类。结果表明: (1) 双层模型总体精度达 0.941, Kappa 系数为 0.929, 比单层随机森林分别提升 11.75% 与 14.55%; (2) 耦合样本迁移法快速生产 2018—2024 年度尺度 LUCC 序列数据, 平均总体精度 0.92, Kappa 系数为 0.91, 采用多月份集成方式, 有效利用了稀疏影像, 显著提升年图空间连续性与完整性, 降低噪声影响; (3) 样本点概率纯化策略显著提高样本质量与分类可靠性。本研究为多云雾山地城市群提供了一种高效、稳健的 LUCC 制图新框架, 可为基础数据生产与国土空间治理提供科学参考。

关键词: Sentinel 1/2 影像; 随机森林模型; 土地利用分类; 黔中城市群

中图分类号: F301.2

文献标志码: A

城市群是中国新型城镇化的主体形态和核心经济增长极^[1-2]。城市群土地利用/覆被变化客观记录了人类通过利用土地资源改变地球表面状况的过程, 进而影响区域乃至全球的水循环、环境质量、生物多样性及陆地生态系统生产力与适应能力^[3-5], 在全球环境变化与可持续发展研究中占据重要地位^[6-8]。连续、精细地监测城市群土地利用变化, 准确把握其用地发展趋势, 是实现土地资源精准调控的基础性工作。

遥感云计算平台提供了海量卫星影像与强大算

力, 为城市群土地利用/覆被变化监测提供了高效、可扩展的解决方案^[9-11]。基于 2015 年欧洲航空局 Sentinel-2A 数据源, FROM-GLC^[12]、World Cover 10 m^[13]、ESRI Land Cover^[14] 及 Dynamic World^[15] 等全球尺度产品相继问世。然而, 这些大尺度产品在异质性强的区域泛化能力不足, 尤其在中国中西部高海拔地区精度显著下降, 直接引用该产品数据易引入较大误差^[16-17]。因此, 针对多云、多雾、地形复杂的中国西南山地城市群, 亟需结合适宜样本、影像与分类方法, 开展本地化高精度制图, 以支撑可持续

收稿日期 (Received date): 2025-05-12; 改回日期 (Accepted date): 2025-09-22

基金项目 (Foundation item): 国家自然科学基金 (42361042); 贵州省基础研究 (自然科学) 项目 (黔科合基础-ZK[2023]一般 461); 铜仁学院博士人才启动基金 (trxyDH2218)。[National Natural Science Foundation of China (42361042); Guizhou Provincial Basic Research Program (Natural Science) (ZK(2023)-461); Doctoral Research Start-up Fund of Tongren University (trxyDH2218)]

作者简介 (Biography): 卢德彬 (1987-), 男, 博士, 教授, 主要研究方向: 土地利用变化监测与政策仿真。[LU Debin (1987-), male, Ph. D., professor, research on land use change monitoring and policy simulation] E-mail: sooluo@163.com

管理与社会经济发展。

传统光学遥感在山地城市群受云雾和地形阴影影响,有效影像获取量少,时序插值(如线性插值)虽可补缺,但主要补偿传统分类器缺失值处理能力,对分类精度提升有限^[18-23]。密集影像叠加方法不遵循传统变化检测的周年日期原则,是将特定时段内所有可用影像(含部分云覆盖)叠加,利用其他像元替代云区,生成更完整的多模态数据集。

传统基于支持向量机、随机森林模型、分类回归树和梯度提升树等单一机器学习方法开展土地利用/覆被分类工作^[24-27],受限于单一模型本身的分类能力,难以充分挖掘高维、异质特征,分类精度存在上限。Stacking 集成框架通过双层模型结构,以基础模型概率输出作为元模型输入,可充分融合不同算法优势,降低方差并生成稳定概率图;基础模型与元模型选择灵活,特别适用于高变异性土地覆盖分类^[28-31]。尽管增加模型层级会带来额外计算成本,但可换取显著精度与稳定性提升。

基于此,本文借鉴 Stacking 思想,面向黔中城市群构建双层土地利用分类框架:(1)以随机森林(Random Forest, RF)作为基础模型与元模型,验证 Stacking 对山地城市群土地利用/覆被分类精度的提升效果;(2)对第一层地类概率 >0.5 的稳定区重采样,纯化训练样本;(3)采用密集影像叠加,在月尺度生成无云影像,逐月分类并遴选高精度月份结果,合成年度土地覆盖;(4)结合样本迁移与空间一致性检验,完成 2018—2024 年黔中城市群年度土地利用/覆被分类产品生产。本研究基于 Google Earth Engine 平台,融合 Sentinel-1/2 影像,提出一套适用于多云雨、多雾山地城市群的土地利用/覆被分类高精度监测框架,可以显著提升土地利用数据的获取与分析能力,为山地城市群国土空间治理提供科学工具与基础数据。

1 研究区概况

黔中城市群位于贵州省中部,总面积 $5.38 \times 10^4 \text{ km}^2$,覆盖贵阳市、遵义市、安顺市、毕节市、黔东南州、黔南州 6 个州市及贵安新区的 33 个县(市、区),是贵州省经济最发达的地区(图 1)。黔中城市群作为中国实施新型城镇化的先行示范区、长江经

济带战略重要节点和新一轮西部大开发战略的核心区域,同时也是贵州建设国家生态文明试验区、大数据综合试验区和内陆开放型经济试验区的重要空间场所,对带动贵州经济持续快速增长、促进区域协调发展具有重要意义。截至 2022 年,黔中城市群常住人口达 1885 万人,地区生产总值达到 12 134 亿元,分别占全省总量的 48.5% 和 60.2%。该区域地形地貌以山地为主,山地面积占比高,土地资源紧缺,生态环境承载压力大,为城市群建设带来诸多挑战和限制。

2 数据来源与预处理

2.1 Sentinel-1 SAR 影像

欧洲航天局研制的 Sentinel-1 SAR 卫星,由 A、B 两颗卫星组成,搭载了 C 波段合成孔径雷达(Synthetic Aperture Radar, SAR)仪器,工作频率为 5.405 GHz,单星重访周期为 12 d,双星访周期最短为 6 d,有 4 种工作模式:条带模式(Stripmap Model, SM)、干涉宽幅(Interferometric Wide Swath, IW)模式、超宽幅(Extra-Wide Swath, EW)模式和波模式(Wave Mode, WM)。本文选择数据获取模式为 IW 模式,幅宽为 250 km,空间分辨率为 $5 \text{ m} \times 20 \text{ m}$,具有较高的几何精度。在 2018 年 1 月 1 日—2024 年 12 月 31 日研究期间, Sentinel-1 双极化(Vertically polarized, VV 和 Horizontally polarized, VH)可用像素频率如图 2a 所示,共获取了双极化(VV 和 VH)影像数据 1570 景用于土地利用分类(图 2b)所示。Sentinel-1 是一套合成孔径雷(SAR)系统,可在任何天气条件下昼夜连续获取数据。

2.2 Sentinel-2 MSI 影像

Sentinel-2 系列卫星包含 A、B 2 颗星,分别于 2015 年 6 月和 2017 年 3 月发射升空。卫星采用太阳同步轨道,携带包含 13 个光谱波段的多光谱成像仪覆盖可见光、近红外(NIR)、红边(RE)和短波红外(SWIR)光谱区,幅宽 290 km,单颗卫星重访周期为 10 d,双星组网重访周期为 5 d,专注于地表监测。在本研究中,使用了由欧洲航天局(ESA) Sentinel 科学数据中心提供的 Level 1C 数据。所有 Sentinel-2 图像均经过大气校正,转化为地表反射率数据,并结合质量评估(Quality Assessment Band, QA)波段筛选出高质量观测数据。研究中选取了时间范围为

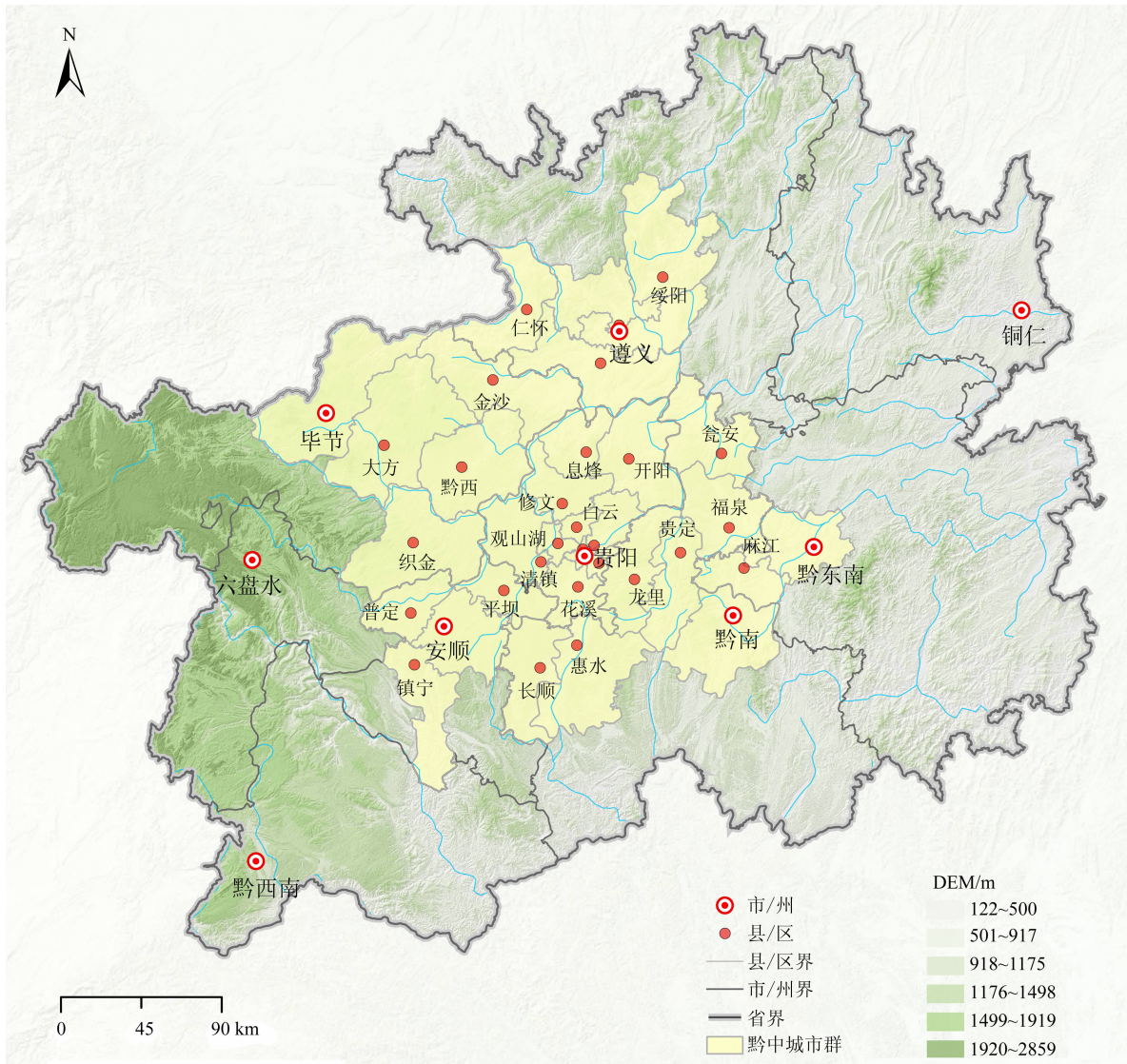


图 1 研究区区位图

Fig. 1 Location of research

(审图号:黔 S(2023)009 号)

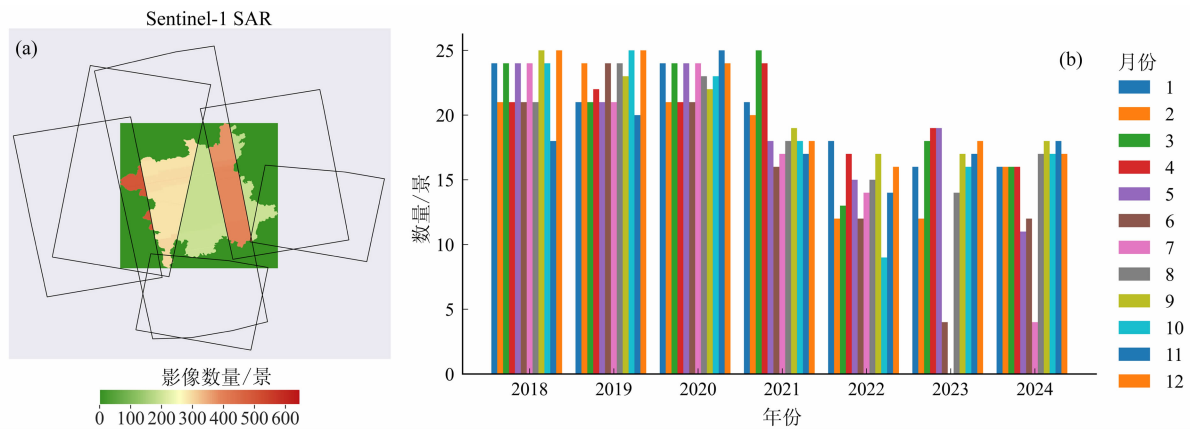


图 2 覆盖研究区 Sentinel-1 可用影像频率与数量统计:(a)可用影像频率;(b)可用影像数量

Fig. 2 Frequency and numbers of Sentinel-1 image scenes covering study area; (a) frequency; (b) numbers of available images

2018年1月1日—2024年12月31日,覆盖黔中城市群区域的 Sentinel-2 可用影像频率(图 3a),共获取 9804 景影像(图 3b),包括蓝、绿、红及近红外波段(空间分辨率 10 m),以及红边和短波红外波段(空间分辨率 20 m)9 个波段用于土地利用分类。以上所有数据均在 Google Earth Engine(GEE)平台上获取并预处理,以确保影像质量适用于进一步的土地利用制图。

本文融合 Sentinel-1 SAR 与 Sentinel-2 MSI 影像 2 种数据,可以充分发挥激光雷达和光学遥感的互补优势,为山地城市群土地利用变化监测提供强大的数据和技术支持。

2.3 土地分类系统与样本

土地利用分类系统是根据人类土地利用行为的目的、方式等不同,将土地利用行为分为若干种类型。本文根据区域的特征和影像数据的可获取性,参照欧洲航天局和联合国粮农组织的土地利用分类体系,将本区域的土地利用类型划分为 8 个类型,分别为林地、灌木地、草地、耕地、建成用地、裸地、水体和湿地。

为了更加客观准确地选择地面样本点,本文以欧洲航天局的 World Cover 10 m 2020 v100 土地利用数据集为辅助参考数据,通过 Google Earth Pro 软件查看 2020 年的研究区域的历史影像。在对每个地类进行样本点选择时,利用 Google earth 历史影像对每个感兴趣区进行目视解译,获取了 8 个地类初始样本,共 3527 个样本点(图 4)。其中,林地为 629 个,灌木林地为 442 个,草地为 453 个,耕地为

424 个,建设用地为 833 个,裸地为 199 个,水体为 485 个,湿地为 62 个。样本点空间分布如图 2 所示。所有满足条件的样本随机分为 70% 和 30%,分别作为土地利用分类模型的训练样本和验证样本数据。

3 研究方法

本文提出了一种双层随机森林土地利用分类模型(图 5),主要包括 4 个步骤。首先,对时序序列 Sentinel-1/2 影像进行预处理,生成样本点和提取光谱特征、指数特征和激光雷达指数特征变量作第一层随机森林土地利用分类模型的变量,并输出地类的概率;其次,重新使用样本点对第一层模型输出的概率大于 0.5 的区域进行采样和纯化样本点,并输入第二层随机森林土地利用分类模型,进行土地利用分类;第三,将输出的 1—12 个月的土地利用分类结果进行叠加分析,提取像元出现次数最多的地类作为该年度的地类;最后,对土地利用分类结果进行精度验证和分类后处理,输出最终的土地利用分类结果图。

3.1 土地利用分类特征变量

构建有效的特征变量数据集对土地利用分类至关重要。为提高不同地物类型之间的可分辨性,本文选取 3 类特征变量:光谱波段、光谱指数与 SAR 波段数据。其中,光谱波段采用 B2、B3、B4、B5、B6、B7、B8、B11、B12 共 9 个波段;光谱指数选取 NDVI、SAVI、MNDWI、NDBI、BSI、CIG 共 6 种指数^[32-34];SAR 数据则选取 HH、HV 以及 HH/HV 3 个指数。

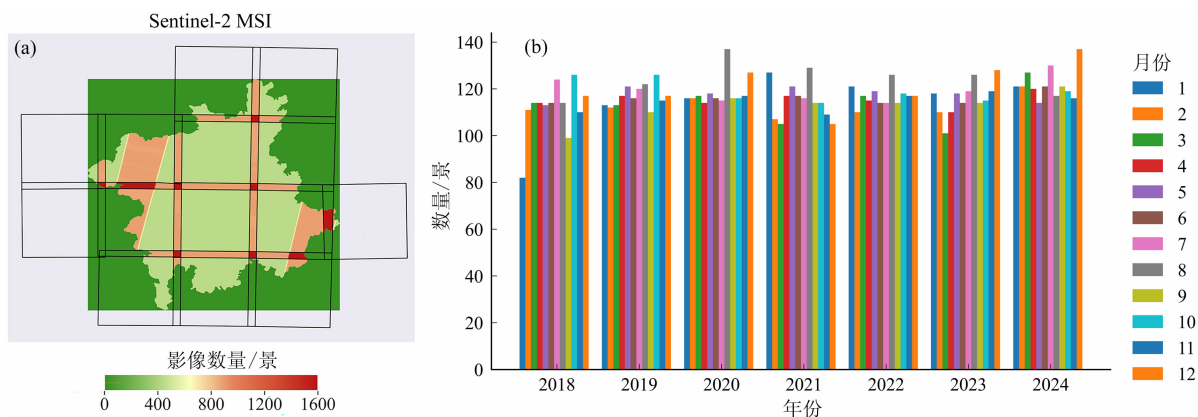


图 3 覆盖研究区 Sentinel-2 可用影像频率与数量统计:(a)可用影像频率;(b)可用影像数量

Fig. 3 Frequency and numbers of Sentinel-2 image scenes covering study area;

(a) frequency; (b) numbers of available images

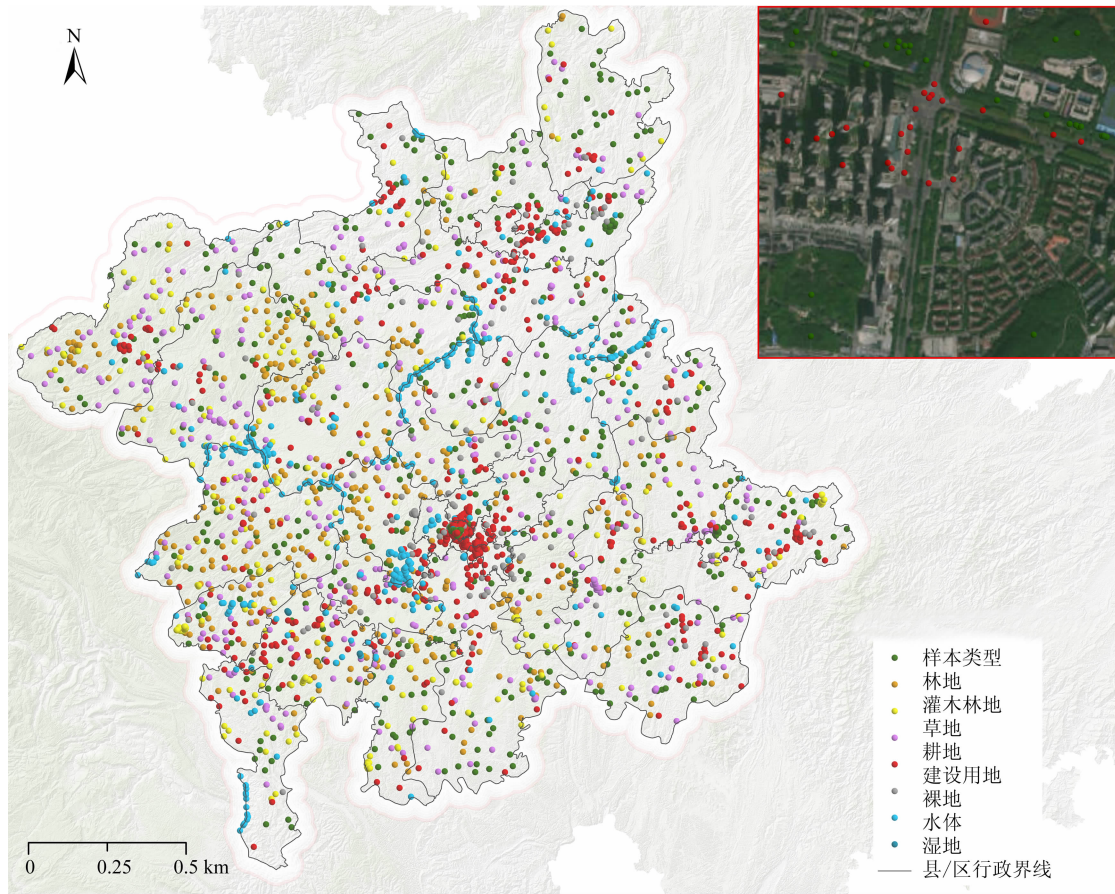


图 4 土地利用类型样本点空间分布

Fig. 4 Spatial distribution of land use type sample points

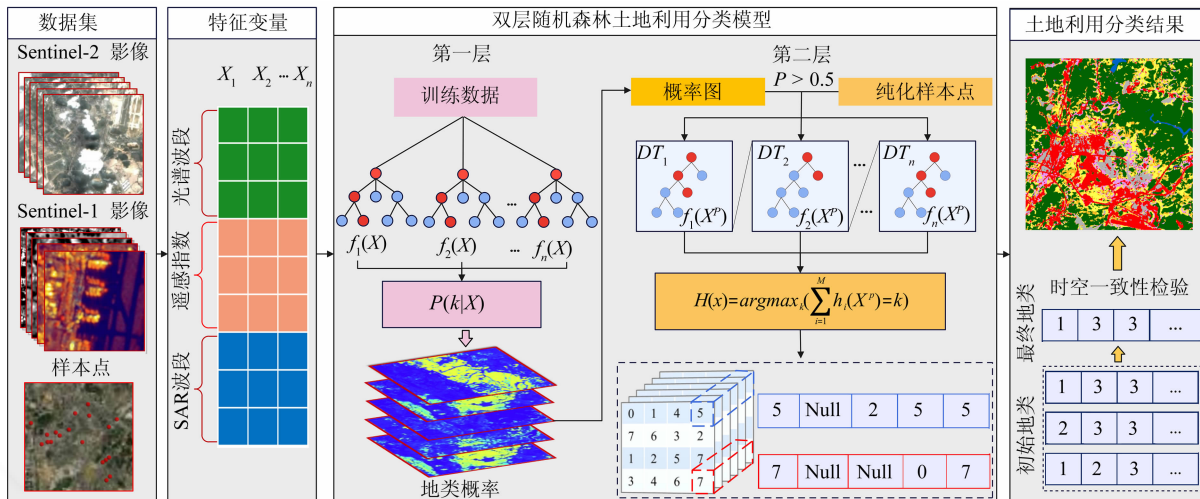


图 5 双层随机森林土地利用分类模型框架

Fig. 5 Framework of double-layer random forest land use classification model

3.2 样本可分离度

Jeffries-Matusita 距离是基于条件概率理论的光谱可分性指标计算某一特征 2 类样本间的距离, 衡量 2 类样本间的可分离度^[35], 其计算公式为:

$$J_{ij} = 2(1 - e^{-B_{ij}}) \quad (1)$$

$$B_{ij} = \frac{1}{8} (m_i - m_j)^2 \frac{2}{\sigma_i^2 + \sigma_j^2} + \frac{1}{2} \ln\left(\frac{\sigma_i^2 + \sigma_j^2}{2\sigma_i\sigma_j}\right) \quad (2)$$

式中, J_{ij} 为地类 i 与地类 j 的可分离性度量; B_{ij} 为地类 i 与地类 j 的 Bhattacharyya 距离; m_i 和 m_j 分别为地类 i 与地类 j 的特征均值; σ_i 和 σ_j 分别为地类 i 与地类 j 的特征标准差。J-M 距离为 0 ~ 2, J-M 距离越大, 可分离性越高。

3.3 样本迁移模型

在不同时间序列内获取训练样本是一个很大的挑战, 利用 2 期多波段影像进行变化检测来确定基准年与目标年之间样本点是否发生变化, 已经被证实是一个行之有效的方法^[36]。其原理是把图像中的每个像元的光谱视为一个高维向量, 通过计算两个光谱角距离 (Spectral angle distance, SAD) 和欧氏距离的平方 (Squared Euclidean distance, SED), 用两个光谱向量距离来度量光谱间的相似性, 夹角或者距离越小, 说明两个光谱越相似, 属于同类地物的可能性越大。反之, 表示两个光谱之间的差异越大, 属于同类地物的可能性越小, SAD 和 SED 分别记为 D_{SAD} 和 D_{SED} , 计算公式分别为:

$$D_{\text{SAD}} = \cos \theta$$

$$\theta = \cos^{-1} \frac{\sum_{i=1}^N X_{i(t_1)} Y_{i(t_2)}}{\sqrt{\sum_{i=1}^N (X_{i(t_1)})^2 \sum_{i=1}^N (Y_{i(t_2)})^2}} \quad (3)$$

$$D_{\text{SED}} = \sqrt{\sum_{i=1}^N (X_{i(t_1)} - Y_{i(t_2)})^2} \quad (4)$$

式中, X 为参考年份多光谱影像 i 在 t 时刻上地物的光谱向量; Y 为目标年份影像 i 在 t 时刻上同一地物的光谱向量。

在采集到研究区域某一年份土地利用分类样本后, 本文利用光谱角距离和光谱距离计算不同年份样本数据的相似性, 通过计算发现当 D_{SED} 值 ≤ 0.2 且 D_{SAD} 值 ≥ 0.95 时, 研究期内黔中城市群基准年与目标年份对应区域的影像未发生变化。将分布在影像未发生变化区域基准年份的样点筛选出来^[31], 形成对应目标年份的样本点, 完成年际样本迁移工作, 大大降低了采集分类样本的时间和成本, 提高了分类工作效率。

3.4 双层随机森林土地利用分类模型

集成堆叠模型是一种机器学习集成策略^[37], 本文借鉴这个思想, 采用随机森林模型构建两层结构的随机森林土地利用分类模型。

首先, 构建输入模型的特征向量 $X = \{x_1, x_2, \dots, x_n\}$, 样本 X 属于地类 k 的概率 P 计算公式为:

$$P(k | X) = \frac{1}{M} \left(\sum_{j=1}^M h_j(X) = k \right) \quad (5)$$

式中, M 为随机森林中的决策树总数; $h_j(X) = k$ 为第 j 棵决策树 T_j 输出结果; k 为地类标签。

其次, 使用样本点对随机森林输出 $P(k | X)$ 值大于 0.5 的区域进行采样, 纯化样本点, 构建新的特征向量 $X^p = \{x_1^p, x_2^p, \dots, x_n^p\}$, 再采用随机森林重新训练模型, 结合所有决策树的分类结果, 通过投票机制确定像元的地类类型, 计算公式为:

$$H(x) = \operatorname{argmax}_k \left(\sum_{j=1}^M h_j(X^p) = k \right) \quad (6)$$

式中, $H(x)$ 为最终输出的地类标签; $\operatorname{argmax}_k(\cdot)$ 为从所有类别中选取获得最多投票数的类别。

最后, 对于像元 l 逐个统计它在 T 个时间点上属于每个地类 k 的次数, 并将出现次数最多的地类作为最终的地类, 计算公式为:

$$F(l) = \operatorname{argmax}_k \left(\sum_{t=1}^T H(x) \right) \quad (7)$$

式中, $F(l)$ 为第 l 个像元的地类最终分类结果。

3.5 精度评估与后分类处理

3.5.1 精度评估

评估土地利用分类图的质量对于验证其是否达到预期目标并了解相关像元的误差至关重要。分类精度的定量分析采用混淆矩阵及其衍生的精度指标, 包括总体精度 (Overall Accuracy, OA)、生产者精度 (Producer's accuracy, PA)、用户精度 (User's accuracy, UA) 和 F1-score 分别记做 P_{OA} 、 P_{PA} 、 P_{UA} 和 F_1 , 所有指标均在 GEE 平台内计算, 计算公式分别为:

$$P_{\text{OA}} = \frac{\sum_i^K n_{ii}}{N} \quad (13)$$

$$P_{\text{PA}} = \frac{n_{ii}}{\sum_i^K n_{ij}} \quad (14)$$

$$P_{\text{UA}} = \frac{n_{ii}}{\sum_i^K n_{ji}} \quad (15)$$

$$F_1 = \frac{2 \times P_{\text{PA}} \times P_{\text{UA}}}{P_{\text{PA}} + P_{\text{UA}}} \quad (16)$$

式中, K 为类别总数; n_{ii} 为第 i 类被正确分类的样本数; N 为所有验证样本总数; $\sum_i^K n_{ii}$ 为第 i 类被正确分类的样本数; $\sum_i^K n_{ij}$ 为 i 类样本的总数; $\sum_i^K n_{ji}$ 为第 i 类的样本总数。

3.5.2 后分类处理

为了减少原始图像中的随机噪声和模型不准确

预测带来的误差,同时减少细碎的栅格像元,本文采用 4×4 的卷积核,通过滑动窗口,将单独的小图斑分配到窗口内中心像素的类别。另外,由于建筑阴影的影响,部分建设用地会被错误分为水体。因此,本文通过暗像元和建筑指数的生成掩膜,对既分布在建筑指数大于 0 的区域也分布在暗像元区域的水体赋值为建设用地。采取分类后处理的方式能够提高时间序列土地利用分类制图的一致性。

4 结果与分析

4.1 样本的可分离性

根据 Jeffries-Matusita 距离的公式,计算所选的样本中不同地物类型的可分离性。如图 6 所示,在特征变量空间上,样本点 8 个地类上可分离度均较高。仅有裸地与建设用地之间、林地与疏林地之间的 Jeffries-Matusita 距离值低于 1.94,分别为 1.83 和 1.87。山区地形起伏大,坡度、坡向和海拔变化显著,植被分布呈现连续渐变特征,导致林地和疏林地边界不明显,存在自然的生态过渡带,而建设用地(如混凝土、沥青路面、未铺装地面)与裸地(如裸露的岩石、沙土、黄土)在遥感影像的光谱特征上非常相似,特别是在可见光和近红外波段,导致分类时容易混淆。虽然,这两对地类之间的 Jeffries-Matusita 距离值较低,但也能满足土地利用模型对样点分离度的要求,说明本文所选择的分类样本点和特征指标能够很好地用于识别土地利用的类型。

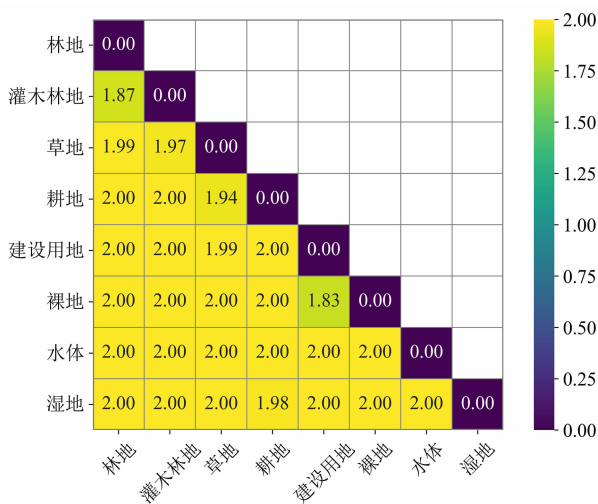


图 6 样本的 Jeffries-Matusita 距离矩阵

Fig. 6 Jeffries-Matusita distance matrix of feature samples

4.2 土地利用分类制图精度评估

为了验证分类精度,采用验证样本和 UA、PA、F1-score、TA 和 Kappa 4 个指标对分类结果进行分类精度评价(表 1)。2020 年黔中城市群的单层随机森林和双层随机森林土地利用分类模型的整体精度分别为 0.842、0.941, Kappa 系数分别为 0.811、0.929, 双层随机森林模型比单层随机森林模型的整体精度和 Kappa 系数分别提高了 11.75% 和 14.55%。整体上,双层比单层的随机森林土地利用分类模型在各地类上的 UA、PA 和 F1-score 上精度均有提升。其中,提升幅度最小的为水体、建设用地和林地,这 3 个地类在 2 个模型上的分类精度均较高,UA 均超过了 0.86, PA 和 F1-score 都超过了 0.911;精度提升幅度最大的为草地、湿地;疏林地精度提升幅度也超过了 16%,说明本文构建的双层随机森林土地利用分类模型显著提高了分类的精度,能够减少光谱特征相似性的地类分类错误。很显然,采用集成堆叠思想构建的双层随机森林土地利用分类模型整体精度明显高于单层随机森林土地利用分类模型,通过概率进行纯化样本,能够提升模型识别光谱特征相似性的地类能力,为土地利用提供了一个新的分类框架。

表 1 土地利用分类模型精度评估

Tab. 1 Accuracy assessment of land use classification models

地类	单层随机森林模型			双层随机森林模型		
	UA	PA	F1-score	UA	PA	F1-score
林地	0.863	0.939	0.899	0.938	0.949	0.943
灌木林地	0.768	0.686	0.725	0.909	0.925	0.917
草地	0.722	0.727	0.724	0.913	0.934	0.924
耕地	0.742	0.827	0.782	0.927	0.905	0.916
建设用地	0.911	0.947	0.92	0.955	0.973	0.964
裸地	0.783	0.527	0.63	0.913	0.763	0.831
水体	0.993	0.955	0.97	0.999	0.986	0.993
湿地	0.666	0.615	0.64	0.833	0.999	0.909

在土地利用制图中,细节的识别效果也是衡量分类结果质量的重要标准之一。良好的土地利用分类工作不仅需要具有较高的模型精度,同时在细节上也需保持较高的辨识能力。为了更直观地评估模型分类的效果,本文选取了 3 个典型区域,并将分类结果与 Google Earth 高分辨率遥感影像进行对比分

析,这一方法能够直观地检验模型对不同地类的分类效果,评估其在复杂地表信息提取方面的能力。2020 年黔中城市群的土地利用分类结果与 Google Earth 的高分辨率遥感影像相比较,双层随机森林土

地利用分类模型不仅在整体上具有较高的分类精度,在边界细节、地类一致性等方面均表现优异,能够较为准确地识别出不同的土地利用类型,并且在细节刻画方面表现良好(图 7)。

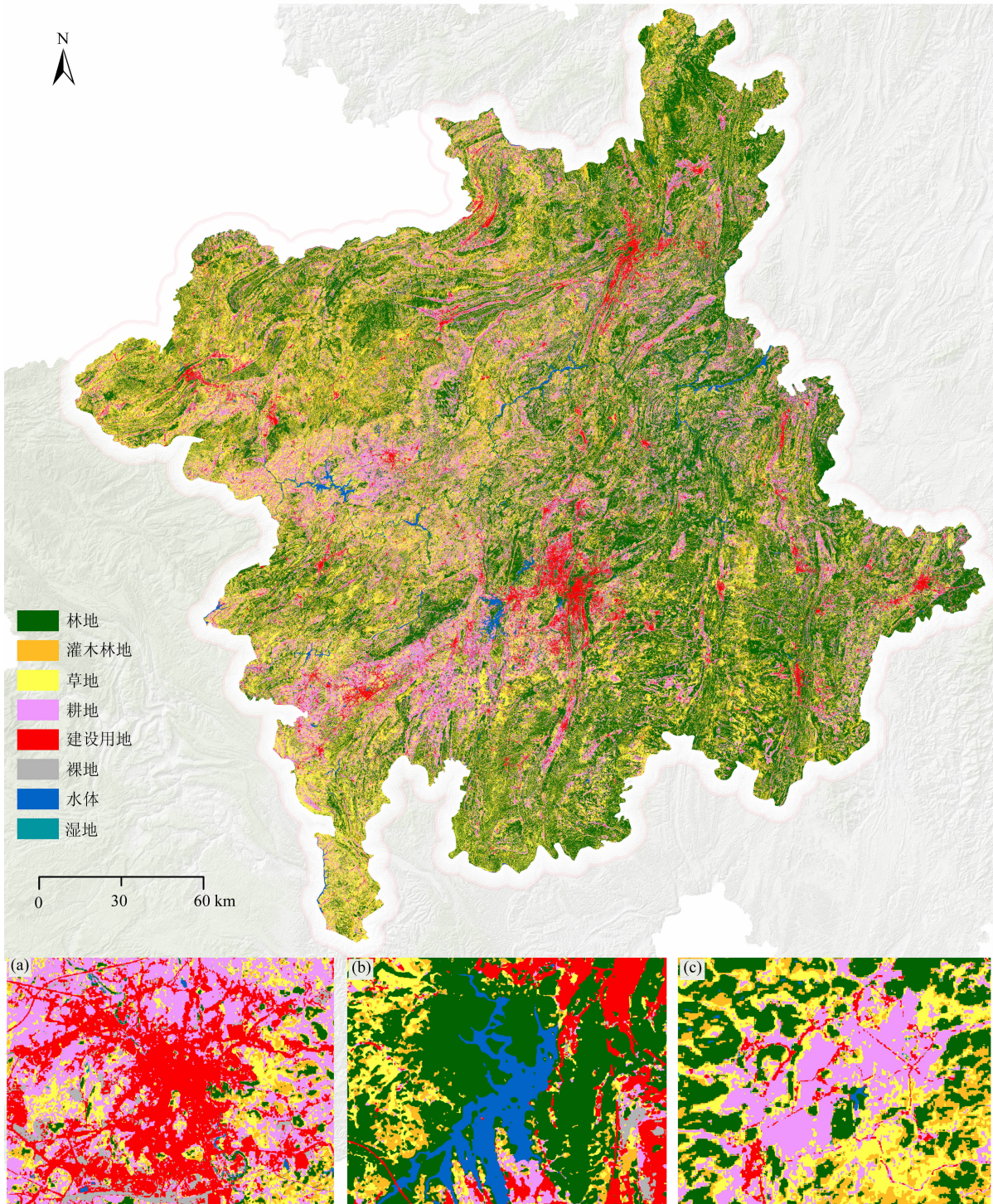


图 7 2020 年黔中城市群土地利用分类:(a) 黔西县城城区;(b) 阿哈湖;(c) 惠水县旧寨村

Fig.7 Land use classification of central Guizhou urban agglomeration in 2020;

(a) urban area of Qianxi County; (b) Aha Lake; (c) Jiuzhai Village, Huishui County

(审图号:黔 S(2023)009 号)

4.3 样本迁移与年际土地利用制图

本文以 2020 年为基准年,以 2018、2019、2021、2022、2023 和 2024 年为目标年份,利用 Google Earth Engine 平台,根据光谱角距离(SAD)与欧氏距离平方(SED)的计算公式,分别计算基准年和目标年的 SAD 和 SED 指数,确定基准年和目标年份之间样本点对应的区域是否发生变化。同时,为了确保迁移样本的准确性,本研究还利用了 Google earth 高分辨率遥感图像进行人工目视交叉评估。如表 2 所示,除了 2023 年样本数量为 3067 个外,其他年份的总样本数量保持在 3300 个左右,各年份总的地类样本数量差距低于 300 个点,地类之间的样本数量除了湿地在 2021 年为 38 个以外,数量波动均很小,说明所采集的基准年份的样本点在时间尺度上具有较好的样本数量稳定性。这也说明利用光谱距离的差异确定基准年和目标年份之间采样点是否发生变化,进而实现样本迁徙^[35],可以极大地降低采集土地利用分类样本点的人力成本和时间,为年际山地城市群土地利用分类制图样本点的制作提供新的途径。

基于采用样本迁移方法生成的新样本点,本文使用构建的双层随机森林土地利用分类模型对不同年份土地利用进行分类。为了保证土地利用分类结果的时空一致性,本文对 2018—2024 年分类结果进行逻辑推理。首先,根据 SAD 和 SED 指数判断,如果研究期内某像元均未发生变化,那么整个研究期间内年度土地利用分类结果保持地类一致;其次,发生变化的区域按照模型分类结果确定,最终实现

2018—2024 年黔中城市群土地利用分类制图(图 8)。经过统计得出基于 GEE 平台的 2018—2024 年黔中城市群逐年土地利用分类各地类用户精度及制图精度均超过 0.93,平均总体精度和 Kappa 系数均为 0.92,不仅分类精度高,且多年土地利用分类结果的总体精度和 Kappa 系数也比较平稳,没有太大的起伏差异,表明本文在选取的样本点数据和特征变量及构建的双层随机森林土地利用分类模型具有较强的可靠性和稳定性,可以为山区城市群土地利用分类制图提供新的方法框架。

5 讨论

本研究基于全年所有可用的 Sentinel-1/2 影像,完成去云、阴影和雪的基础上按照月度进行合成,开展黔中城市群的年度土地利用制图。此外,对时序土地利用分类结果进行了后分类处理(post-classification processing),以提高时序一致性并滤除噪声,从而提高了年度土地利用制图的精度和鲁棒性。将基于双层随机森林土地利用分类研究结果与目前可公开获得的 3 种 10 m 空间分辨率年度土地利用分类产品(欧洲航空局 WorldCover、Esri Land Cover 和 Google Dynamic World)进行对比。土地利用分类结果总体精度达到了 0.929(图 8),优于欧洲航空局 WorldCover 的 0.767、Esri Land Cover 的 0.85 和 Google Dynamic World 的 0.738^[38]。但是需要指出的是,这 3 个土地利用产品报告的精度均为

表 2 基于样本迁移生成的土地覆盖样本数量

Tab. 2 Numbers of land cover samples based on sample transfer

地类	样本数量/个					
	2018 年	2019 年	2021 年	2022 年	2023 年	2024 年
林地	582	599	600	587	546	604
灌木林地	397	411	402	405	330	424
草地	421	414	383	423	388	432
耕地	406	385	362	361	364	407
建设用地	783	792	788	792	778	790
裸地	178	176	174	168	162	168
水体	478	480	463	468	446	474
湿地	62	62	38	50	53	62
合计	3307	3319	3210	3254	3067	3361

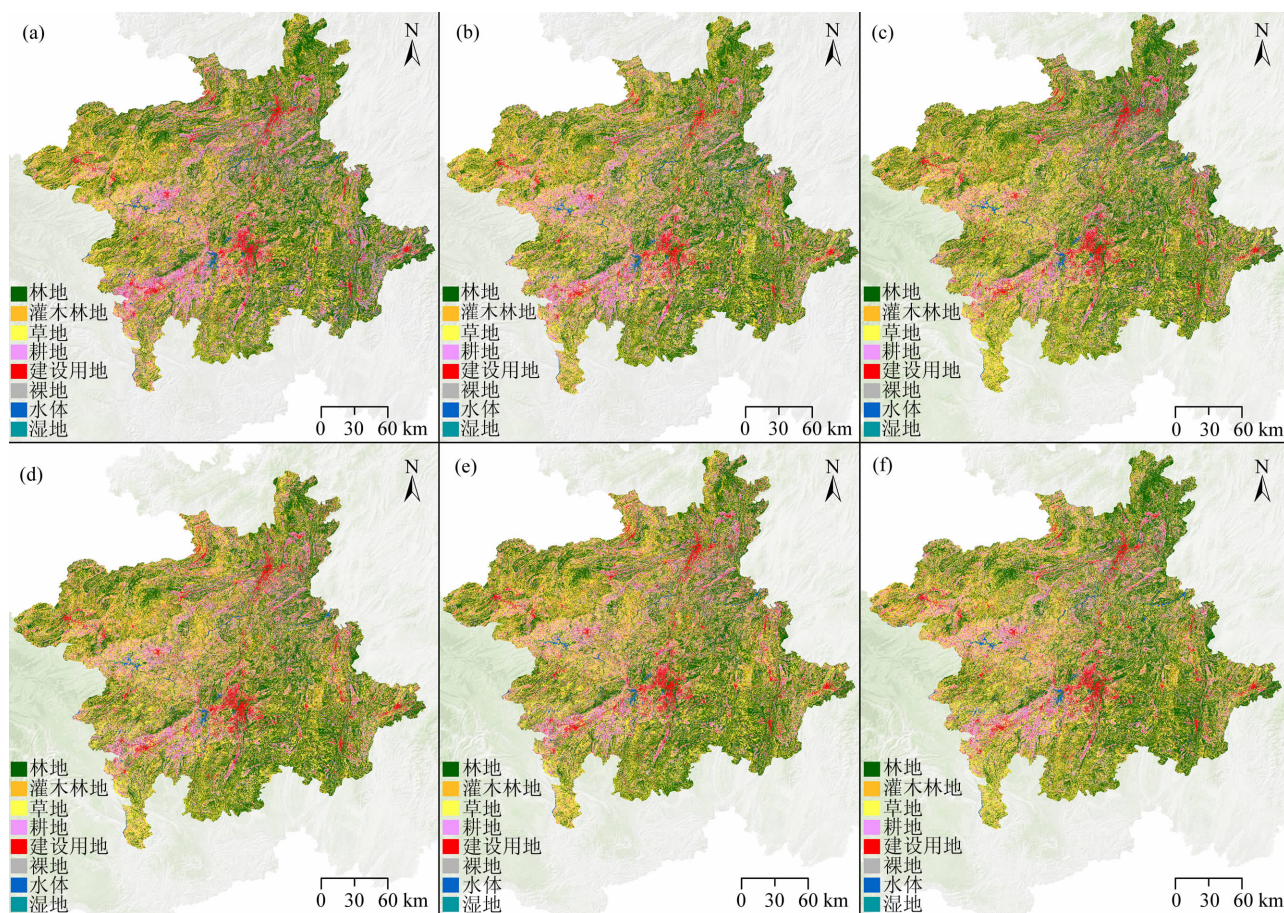


图8 2018—2024年黔中城市群土地利用分类:

(a) 2018年;(b)2019年;(c)2021年;(d)2022年;(e)2023年;(f)2024年

Fig. 8 Land use classification of central Guizhou urban agglomeration from 2018 to 2024:

(a) 2018;(b)2019;(c)2021;(d)2022;(e)2023;(f)2024

全球尺度,可能与本产品不具有直接可比性。因此,进一步在中国西南地区采用目视解译样本集对上述3个产品进行评估发现:Google Dynamic World的OA和Kappa系数分别为0.6242和0.56;欧空局WorldCover的OA和Kappa系数分别为0.645和0.58;Esri Land Cover的OA和Kappa系数分别为0.6194和0.56^[39],明显低于本文的分类结果。此外,不同土地利用产品土地类型的局部对比图(图9)显示土地利用细节刻画优于其他3种土地利用产品,说明异质性是影响全球产品土地分类产品的区域泛化性的重要因素。因此,受大气环境、下垫面构成、土地利用景观的复杂性影响,在进行大范围城市群土地利用分类制图工作时,需要特别考虑区域差异所带来的异质性。在高异质性的区域,需要单独采集合适的样本点和构建符合区域地理环境特征的土地利用分类模型以获得更好的土地利用结果。

通过与不同土地利用产品的精度指标对比,本文所提出的双层随机森林土地利用分类模型在多云雾山地城市群的土地利用制图精度具有显著优势。首先,双层随机森林土地利用分类模型不同于传统的随机森林模型或GEE内置的其他机器学习模型。双层随机森林土地利用分类模型采用了一种并行模式,利用全年月度合成影像进行土地分类,每个月份的单独特征被输入到独立的随机森林模型中,并分别输出多个月份的土地类型概率图,最终的土地利用分类是基于这些土地概率图进行二次分类采样和分类,并且进行集成投票生成的。由于每个月份的土地概率分类图都是独立生成的,云层噪声只影响某一特定月份的分类,而不会阻塞整个信息流,从而可以确保最终的土地分类结果具有较强的空间完整性。这种专门设计的双层随机森林土地利用分类模型能够更有效地利用稀疏的遥感影像数据,在一定

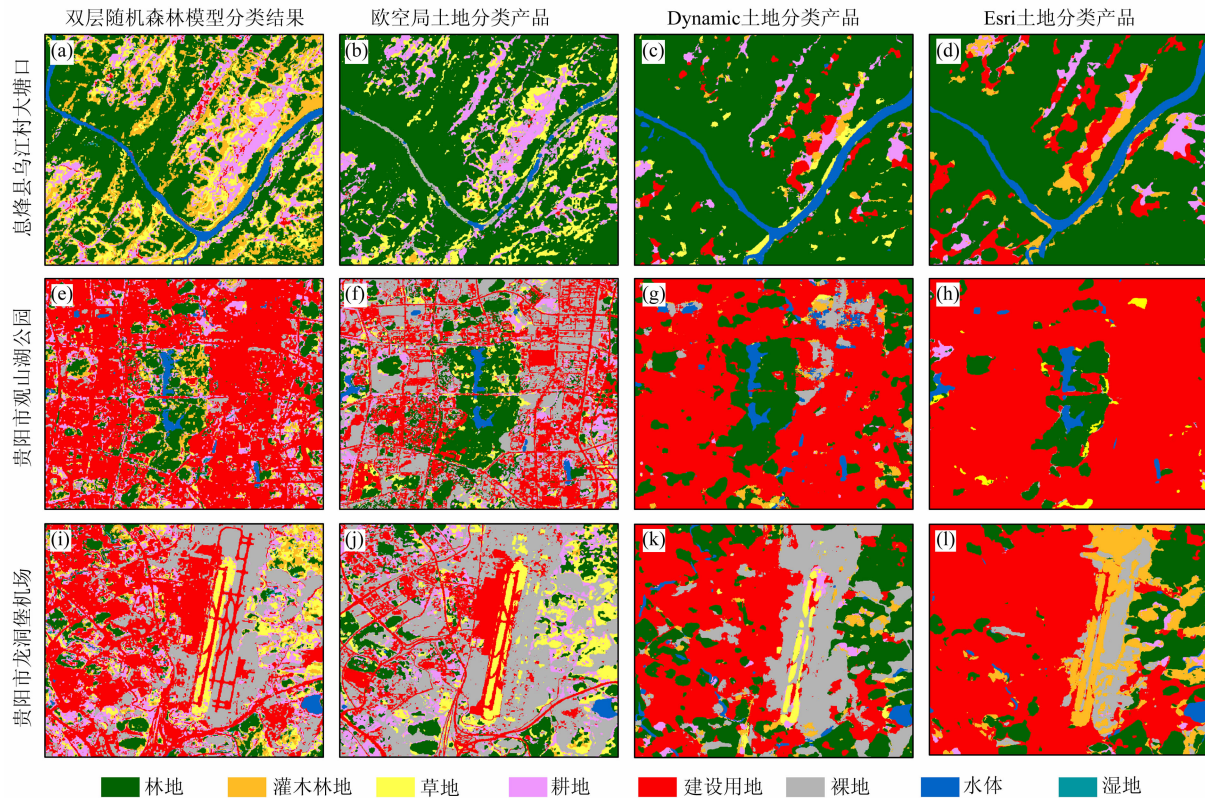


图9 不同土地利用分类产品局部对比

Fig. 9 Localized comparison among different land use classification products

程度上降低了数据噪声,能够克服云层、阴影和雪对山地城市群土地利用分类的影响,从而提升土地利用分类的精度和效率。其次,基于概率值对样本点进行纯化,改变了传统土地利用分类工作依靠线性光谱对混合像元进行解混的方式^[40],更容易获取更纯像元的地物光谱样本点,极大地提高了工作效率。

本文构建的双层随机森林土地分类模型在山地城市群的年度土地利用分类制图具有较高的精度,但是需要根据初始样本点生成不同月份新的样本点和分类影像数据集,相比单层随机森林土地利用分类模型,大大增加了计算工作量。依靠传统的遥感分类工作模式很难实现如此巨大的数据加载和土地分类计算,需要依托类似 Google Earth Engine 遥感云计算平台实现。山地城市群土地利用分类不仅受分类模型的制约,同时也受分类样本和影像质量的影响,如果用于土地分类的遥感影像质量本身很低,样本点光谱不纯,那么通过新模型来提升分类精度则很难实现分类精度提升的目的。因此,要想提高山地城市群土地利用分类的制图精度,需要同时考虑样本点纯度、遥感影像质量和模

型分类能力。

6 结论

本文依托 Google Earth Engine 云计算平台,提出了一种面向山地城市群的双层随机森林土地利用分类模型框架,并融合 Sentinel-1 SAR 和 Sentinel-2 MSI 时间序列遥感影像实现了黔中城市群年度土地利用制图。

(1) 基于双层随机森林土地利用分类模型的黔中城市群土地利用分类总体精度(OA)为 0.941, Kappa 系数分别为 0.941 和 0.929。

(2) 本文构建的双层随机森林模型比单层随机森林模型的总体精度和 Kappa 系数分别提高了 11.75% 和 14.55%。

(3) 采用样本迁移方法,快速完成了黔中城市群 2018—2024 年的土地利用制图,平均总体精度 Kappa 系数均为 0.92,说明本文的所提出的双层土地利用分类框架具有较强的稳定性和可靠性。

(4) 本文双层随机森林土地利用分类模型利用

多个月度合成影像进行分类,采用了一种并行计算的策略,有效地利用稀疏的遥感影像数据,降低了云层、阴影和雪数据噪声对山地城市群土地利用分类制图的影响,从而确保最终的分类结果具有较强的空间完整性。

(5)基于概率值筛选和纯化了分类样本点,显著提高了土地利用制图精度,为山地城市群土地利用制图提供了新的技术框架。本研究可为黔中城市群的国土空间变化动态监测、山地城市群的国土空间治理和土地可持续发展提供理论依据和技术参考。

参考文献 (References)

- [1] 方创琳, 张国友, 薛德升. 中国城市群高质量发展与科技协同创新共同体建设[J]. 地理学报, 2021, **76**(12): 2898–2908. [FANG Chuanglin, ZHANG Guoyou, XUE Desheng. High-quality development of urban agglomerations in China and construction of science and technology collaborative innovation community [J]. Acta Geographica Sinica, 2021, **76**(12): 2898–2908] DOI: 10.11821/dlxb202112003
- [2] CAO Haojie, LI Yu, WENG Min, et al. Deciphering the evolving trajectories of China's megaregions from 1992 to 2020: A novel morphological approach based on global land cover products [J]. Applied Geography, 2024, **164**: 103205. DOI: 10.1016/j.apgeog.2024.103205
- [3] FOLEY J A, DEFRIES R, ASNER G P, et al. Global consequences of land use [J]. Science, 2005, **309**(5734): 570–574. DOI: 10.1126/science.111177
- [4] NEWBOLD T, HUDSON L N, HILL S L L, et al. Global effects of land use on local terrestrial biodiversity [J]. Nature, 2015, **520**(7545): 45–50. DOI: 10.1038/nature14324
- [5] LU Debin, XU Jianhua, YUE Wenze, et al. Response of PM_{2.5} pollution to land use in China [J]. Journal of Cleaner Production, 2020, **244**: 118741. DOI: 10.1016/j.jclepro.2019.118741
- [6] TURNER B L, MEYER W B, SKOLE D L. Global land-use/land-cover change: Towards an integrated study [J]. Ambio, 1994, **23**(1): 91–95.
- [7] WINKLER K, FUCHS R, ROUNSEVELL M, et al. Global land use changes are four times greater than previously estimated [J]. Nature Communications, 2021, **12**(1): 2501. DOI: 10.1038/s41467-021-22702-2
- [8] 杨钟贤, 苏春江. 平原与山区土地利用/覆被变化对比—以双流县和米易县为例[J]. 山地学报, 2009, **27**(5): 585–592. [YANG Zhongxian, SU Chunjiang. A comparative study between plain and mountain on LUCC: Case study of Shuangliu County and Miyi County, Sichuan [J]. Mountain Research, 2009, **27**(5): 585–592]
- [9] 梁顺波, 孙林, 杜永明, 等. 基于 GEE 平台的每年夏冬两期土地覆盖分类方法研究[J]. 北京师范大学学报(自然科学版), 2024, **60**(6): 831–842. [LIANG Shunbo, SUN Lin, DU Yongming, et al. Land cover classification in summer and winter based on the Google Earth Engine platform [J]. Journal of Beijing Normal University (Natural Science), 2024, **60**(6): 831–842] DOI: 10.12202/j.0476-0301.2023250
- [10] 毛丽君, 李海涛, 薛晓明, 等. 基于 GEE 平台的国家公园土地覆盖变化遥感检测方法构建[J]. 南京林业大学学报(自然科学版), 2022, **46**(2): 213–220. [MAO Lijun, LI Haitao, XUE Xiaoming, et al. Developing remote sensing based methods for land cover change detection in national parks from GEE platform: A case study from the Qianjiangyuan National Park pilot area [J]. Journal of Nanjing Forestry University (Natural Sciences Edition), 2022, **46**(2): 213–220] DOI: 10.12302/j.issn.1000-2006.202101013
- [11] 李培林, 刘小平, 黄应淮, 等. 基于 GEE 平台的广州市主城区不透水面时间序列提取[J]. 地球信息科学学报, 2020, **22**(3): 638–648. [LI Peilin, LIU Xiaoping, HUANG Yinghuai, et al. Mapping impervious surface dynamics of Guangzhou Downtown based on Google Earth Engine [J]. Journal of Geoinformation Science, 2020, **22**(3): 638–648] DOI: 10.12082/dqxxkx.2020.190047
- [12] GONG Peng, LIU Han, ZHANG Meinan, et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017 [J]. Science Bulletin, 2019, **64**(6): 370–373. DOI: 10.1016/j.scib.2019.03.002
- [13] ZANAGA D, VAN DE KERCHOVE R, DE KEERSMAECKER W, et al. ESA WorldCover 10 m 2020 v100 [DB]. Zenodo. <https://doi.org/10.5281/zenodo.5571936>
- [14] KARRA K, KONTGIS C, STATMAN-WEIL Z, et al. Global land use/land cover with Sentinel 2 and deep learning [C]// Proceedings of 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels, Belgium: IEEE, 2021: 4704–4707. DOI: 10.1109/IGARSS47720.2021.9553499
- [15] BROWN C F, BRUMBY S P, GUZDER-WILLIAMS B, et al. Dynamic world, near real-time global 10 m land use land cover mapping [J]. Scientific Data, 2022, **9**(1): 251. DOI: 10.1038/s41597-022-01307-4
- [16] WANG Jun, YANG Xiaomei, WANG Zhihua, et al. Consistency analysis and accuracy assessment of three global ten-meter land cover products in rocky desertification region: A case study of southwest China [J]. ISPRS International Journal of Geo-Information, 2022, **11**(3): 202. DOI: 10.3390/ijgi11030202
- [17] 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. DOI: 10.3390/rs14164101

- [18] 李恒凯, 王利娟, 肖松松. 基于多源数据的南方丘陵山地土地利用随机森林分类[J]. 农业工程学报, 2021, **37**(7): 244 – 251. [LI Hengkai, WANG Lijuan, XIAO Songsong. Random forest classification of land use in hilly and mountainous areas of southern China using multi-source remote sensing data [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, **37**(7): 244 – 251] DOI: 10.11975/j.issn.1002-6819.2021.07.030
- [19] INGLADA J, VINCENT A, ARIAS M, et al. Operational high resolution land cover map production at the country scale using satellite image time series [J]. Remote Sensing, 2017, **9**(1): 95. DOI: 10.3390/rs9010095
- [20] GRIFFITHS P, NENDEL C, HOSTERT P. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping [J]. Remote Sensing of Environment, 2019, **220**: 135 – 151. DOI: 10.1016/j.rse.2018.10.031
- [21] GARNOT V S F, LANDRIEU L, CHEHATA N. Multi-modal temporal attention models for crop mapping from satellite time series [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, **187**: 294 – 305. DOI: 10.1016/j.isprsjprs.2022.03.012
- [22] TANG Yunwei, QIU Fang, WANG Bangjin, et al. A deep relearning method based on the recurrent neural network for land cover classification [J]. GIScience & Remote Sensing, 2022, **59**(1): 1344 – 1366. DOI: 10.1080/15481603.2022.2115589
- [23] CHE Xianghong, ZHANG Hankui, LI Zhongbin, et al. Linearly interpolating missing values in time series helps little for land cover classification using recurrent or attention networks [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, **212**: 73 – 95. DOI: 10.1016/j.isprsjprs.2024.04.021
- [24] MAXWELL A E, WARNER T A, FANG Fang. Implementation of machine-learning classification in remote sensing: An applied review [J]. International Journal of Remote Sensing, 2018, **39**(9): 2784 – 2817. DOI: 10.1080/01431161.2018.1433343
- [25] MAO Wanliu, LU Debin, HOU Li, et al. Comparison of machine-learning methods for urban land-use mapping in Hangzhou City, China [J]. Remote Sensing, 2020, **12**(17): 2817. DOI: 10.3390/rs12172817
- [26] HOSSEINI M M, ZOEJ M J V, DEHKORDI A T, et al. Cropping intensity mapping in Sentinel-2 and Landsat-8/9 remote sensing data using temporal transfer of a stacked ensemble machine learning model within Google Earth Engine [J]. Geocarto International, 2024, **39**(1): 2387786. DOI: 10.1080/10106049.2024.2387786
- [27] 潘娇, 李超, 彭文忆, 等. 基于随机森林和支持向量机的云南省土地利用分类[J]. 科学技术与工程, 2024, **24**(17): 7043 – 7051. [PAN Jiao, LI Chao, PENG Wenyi, et al. Land use classification in Yunnan Province based on random forest and support vector machines [J]. Science Technology and Engineering, 2024, **24**(17): 7043 – 7051] DOI: 10.12404/j.issn.1671-1815.2306050
- [28] DAS B, RATHORE P, ROY D, et al. Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies [J]. Catena, 2022, **217**: 106485. DOI: 10.1016/j.catena.2022.106485
- [29] RAHMATI A, ZOEJ M J V, DEHKORDI A T. Early identification of crop types using Sentinel-2 satellite images and an incremental multi-feature ensemble method (Case study: Shahriar, Iran) [J]. Advances in Space Research, 2022, **70**(4): 907 – 922. DOI: 10.1016/j.asr.2022.05.038
- [30] ABDALI E, ZOEJ M J V, DEHKORDI A T, et al. A parallel-cascaded ensemble of machine learning models for crop type classification in Google Earth Engine using multi-temporal sentinel-1/2 and landsat-8/9 remote sensing data [J]. Remote Sensing, 2023, **16**(1): 127. DOI: 10.3390/rs16010127
- [31] ZHANG Chengkang, ZHANG Hongyan, TIAN Sijing. Phenology-assisted supervised paddy rice mapping with the Landsat imagery on Google Earth Engine: Experiments in Heilongjiang Province of China from 1990 to 2020 [J]. Computers and Electronics in Agriculture, 2023, **212**: 108105. DOI: 10.1016/j.compag.2023.108105
- [32] GITELSON A A, KAUFMAN Y J, MERZLYAK M N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS [J]. Remote Sensing of Environment, 1996, **58**(3): 289 – 298.
- [33] MATURANO J G, SANTANA J D M, AGUILAR-GARCIA A L, et al. Remote sensing of illegal dumps through supervised classification of satellite images: Application in Oaxaca, Mexico [J]. Geographical Research Letters, 2024, **50**(2): 157 – 177. DOI: 10.18172/cig.6273
- [34] CHEN Junyi, CHEN Suozhong, YANG Chao, et al. A comparative study of impervious surface extraction using Sentinel-2 imagery [J]. European Journal of Remote Sensing, 2020, **53**(1): 274 – 292.
- [35] WANG Yongji, QI Qingwen, LIU Ying. Unsupervised segmentation evaluation using area-weighted variance and Jeffries-Matusita distance for remote sensing images [J]. Remote Sensing, 2018, **10**(8): 1193. DOI: 10.3390/rs10081193
- [36] HUANG Huabing, WANG Jie, LIU Caixia, et al. The migration of training samples towards dynamic global land cover mapping [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, **161**: 27 – 36. DOI: 10.1016/j.isprsjprs.2020.01.010
- [37] WOLPERT D H. Stacked generalization [J]. Neural Networks, 1992, **5**(2): 241 – 259. DOI: 10.1016/S0893-6080(05)80023-1
- [38] MOHARRAMI M, ATTARCHI S, GLOAGUEN R, et al. Integration of Sentinel-1 and Sentinel-2 data for ground truth

- sample migration for multi-temporal land cover mapping [J]. Remote Sensing, 2024, **16**(9): 1566. DOI: 10.3390/rs16091566
- [39] XU Panpan, TSENDBAZAR N E, HEROLD M, et al. Comparative validation of recent 10 m-resolution global land cover maps [J]. Remote Sensing of Environment, 2024, **311**: 114316. DOI: 10.1016/j.rse.2024.114316
- [40] WANG Jun, YANG Xiaomei, WANG Zhihua, et al. Consistency analysis and accuracy assessment of three global ten-meter land cover products in rocky desertification region: A case study of southwest China [J]. ISPRS International Journal of Geo-Information, 2022, **11**(3): 202. DOI: 10.3390/ijgi11030202

Land-Use Classification Based on GEE and a Two-Layer Random-Forest Model: A Case Study of the Central Guizhou Urban Agglomeration

LU Debin^{1,2}, SHI Zhang¹, SUO Chao², HAN Chang¹, GENG Jingjie^{1,3}

- (1. a. Department of Tourism & Geography; b. Guizhou Provincial Engineering Research Center for Intelligent Monitoring & Policy Simulation of Mountain Territorial Space, Tongren University, Tongren 554300, Guizhou, China;
2. School of Public Administration, Guizhou University, Guiyang 550025, China;
3. School of Law, Sichuan University, Chengdu 610207, China)

Abstract: In cloudy and foggy mountain environments, severe distortion of surface radiation signals makes land-cover identification difficult, and high-accuracy land-use/land-cover change (LUCC) data are urgently needed to support precise territorial-space governance of urban agglomerations. However, appropriate samples, imagery and classification methods for local high-precision mapping were still insufficient.

This study took the urban agglomeration in central Guizhou as the research area. Sentinel-1/2 images were acquired via the Google Earth Engine platform and preprocessed to remove clouds, shadows, and snow. Three types of feature variables—spectral, index, and SAR spectral—were extracted. Drawing on the Stacking concept, a double-layer random forest framework was constructed: the first layer outputs land class probabilities, while the second layer resamples high-probability areas to generate new features and performs secondary classification.

(1) Results show that the two-layer model achieved an overall accuracy of 0.941 and a Kappa coefficient of 0.929, increasing by 11.75% and 14.55% compared with the single-layer random forest.

(2) Coupled with a sample-transfer scheme, monthly-to-annual LUCC sequences from 2018 to 2024 were produced rapidly; the average overall accuracy was 0.92 and the Kappa coefficient was 0.91. Multi-month integration effectively utilized sparse images, significantly improving the spatial continuity and integrity of annual maps and reducing noise impacts.

(3) The above probability purification framework notably improved sample quality and classification reliability.

This study provides an efficient and robust LUCC-mapping framework for cloudy and foggy mountain urban agglomerations, and offers scientific reference for basic-data production and territorial-space governance.

Key words: Sentinel-1/2 imagery; random-forest model; land-use classification; central Guizhou urban agglomeration