

# 综合遥感信息识别棉花空间分布算法阿克苏-阿拉尔区域实验结果数据集（2020）研发

张 萍<sup>1,2</sup>, 范敬龙<sup>1\*</sup>, 丁 刚<sup>3</sup>, 李生字<sup>1</sup>

1. 中国科学院新疆生态与地理研究所国家荒漠-绿洲生态建设工程技术研究中心, 乌鲁木齐 830054;  
2. 中国科学院大学, 北京 100049; 3. 新疆维吾尔自治区应急管理厅风险监测和综合减灾处, 乌鲁木齐 830011

**摘 要:** 阿克苏-阿拉尔区域指阿克苏地区与阿拉尔市, 位于新疆中部。作者基于 Google Earth Engine (GEE) 平台, 利用 2020 年 Landsat8、Sentinel-2 和 MOD13Q1 遥感影像数据, 采用随机森林方法对阿克苏地区与阿拉尔市棉花种植区域进行提取, 并进行分类处理, 得到该区域的棉花分布数据集 (2020)。作者对各县域影像总体分类精度都在 0.9 以上, Kappa 系数都在 0.8 以上。该数据集内容包括: (1) 研究区棉花种植分布, 空间分辨率为 250 m; (2) 样本点数据。该数据集存储为 .tif、.shp 格式, 由 17 个数据文件组成, 数据量为 385 KB (压缩为 1 个文件, 134 KB)。

**关键词:** 阿克苏地区; 阿拉尔市; 棉花; 随机森林

**DOI:** <https://doi.org/10.3974/geodp.2023.04.05>

**CSTR:** <https://cstr.escience.org.cn/CSTR:20146.14.2023.04.05>

## 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.2024.02.10.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2024.02.10.V1>.

## 1 前言

棉花在中国农作物中的地位日益凸显, 成为仅次于粮食的第二大农作物, 中国已形成了长江流域、黄河流域和以新疆为主的西北内陆三大棉区<sup>[1]</sup>。特别是新疆棉花, 作为世界棉花产业链的关键一环, 不仅在国内经济中扮演着重要的支柱产业角色, 同时也在国际市场上具有重要影响力。阿克苏-阿拉尔区域作为新疆最重要的棉花产区之一, 对整个新疆棉花的贡献不可忽视。据 2020 年统计数据, 阿克苏-阿拉尔区域棉花种植面积已达  $664 \times 10^3$   $\text{hm}^2$ , 总共占新疆棉花总种植面积的 26.52%<sup>[2,3]</sup>。因此全面了解阿克苏-阿拉尔区域棉花的空间分布格局对于有效规划该地区棉花的种植空间至关重要。然而, 利用遥感影像提取棉田

收稿日期: 2023-10-08; 修订日期: 2023-12-20; 出版日期: 2023-12-24

基金项目: 中华人民共和国科学技术部 (2021xjkk0305)

\*通讯作者: 范敬龙, 中国科学院新疆生态与地理研究所, [fanjl@ms.xjb.ac.cn](mailto:fanjl@ms.xjb.ac.cn)

数据引用方式: [1] 张萍, 范敬龙, 丁刚等. 综合遥感信息识别棉花空间分布算法阿克苏-阿拉尔区域实验结果数据集 (2020) 研发 [J]. 全球变化数据学报, 2023, 7(4): 375-381.  
<https://doi.org/10.3974/geodp.2023.04.05>. <https://cstr.escience.org.cn/CSTR:20146.14.2023.04.05>.

[2] 张萍, 范敬龙, 李生字. 基于多源遥感影像的阿克苏-阿拉尔区域棉花分布数据集 (2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.02.10.V1>.  
<https://cstr.escience.org.cn/CSTR:20146.11.2024.02.10.V1>.

空间分布格局时,往往需要大量地面调查数据作为训练样本,但野外调查消耗大量人力和物力,在大区域难以保证获得足够的训练样本<sup>[4]</sup>,这成为一个制约因素。在这一背景下,利用 Google Earth Engine (GEE) 成为解决这一问题的有效途径。GEE 是一个基于云的平台,用于行星规模的地理空间分析,它使 Google 的庞大计算能力能够应对各种高影响力的社会问题,包括森林砍伐、干旱、灾难、疾病、粮食安全、水管理、气候监测和环境保护。在遥感图像分类中,许多机器学习算法都能实现,如人工神经网络<sup>[5]</sup>、决策树<sup>[6,7]</sup>和支持向量机<sup>[8]</sup>。而随机森林具有高分类精度、可以处理大量的输入变数、可以平衡误差等优点,已被广泛用于土地覆盖分类<sup>[9,10]</sup>。遥感云计算技术和随机森林方法也广泛应用于新疆棉花的提取。例如,周磊<sup>[11]</sup>基于遥感大数据计算服务 (PIE Engine Studio) 平台和 GEE 平台,将 NDVI 和 EVI 数据作为特征指数,并采用随机森林方法提取了新疆石河子的棉花空间分布。吕邵伦<sup>[12]</sup>基于 PIE 平台,并利用 Sentinel-2 数据提取了 2020 年阿拉尔市棉花空间分布。王汇涵<sup>[13]</sup>基于 GEE 平台 Sentinel-2 数据,采用随机森林、支持向量机和决策树方法对莫索湾垦区进行了棉花提取。然而,阿克苏-阿拉尔区域棉花的空间分布仍有待进一步研究。

本文基于 GEE 平台,利用 Landsat8、Sentinel-2 和 MOD13Q1 等数据,采用随机森林分类方法,构建了 2020 年新疆阿克苏-阿拉尔区域棉花空间分布数据集,以期规划阿克苏-阿拉尔区域棉花种植空间分布格局提供参考。

## 2 数据集元数据简介

《基于多源遥感影像的阿克苏-阿拉尔区域棉花分布数据集 (2020)》<sup>[14]</sup>的名称、作者信息、地理区域、空间分辨率、数据年代、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 数据来源

本研究基于 GEE 平台,选用 2020 年空间分辨率 30 m 的 Landsat8 数据和空间分辨率 10 m 的 Sentinel-2 数据等高分辨率遥感影像作为目视解译基础数据。EVI 数据来源于 GEE 平台 MOD13Q1 数据。遥感数据基本参数见表 2。土地利用数据来源于《中国 30 米年土地覆被数据集 (CLCD)》<sup>[16]</sup>。数字高程模型空间分布数据来源于美国奋进号航天飞机的雷达地形测绘 SRTM (Shuttle Radar Topography Mission, SRTM) 数据。

### 3.2 研究方法

本研究所用的数据基于 GEE 平台。首先,以阿克苏-阿拉尔区域各县域耕地范围的矢量图为边界,地理坐标系为 GCS\_WGS\_1984,利用 2020 年棉花生长期 Sentinel-2 和 Landsat8 高空间分辨率遥感影像,通过目视解译获取各区域棉花和非棉花样本点。将获取的各县域的样本点数据存储为.shp 文件。研究采用随机森林 (RF) 作为分类器,因为随机森林在处理大量训练样本和高维数据时效率较高,对于训练样本的容错能力强<sup>[17]</sup>。RF 模型由多个分类树构成<sup>[18]</sup>。在训练 RF 模型时,使用训练样本总数的 2/3 构建每颗决策树,剩余的训练样本用于验证每颗决策树的分类结果。分类时,随机森林中的每颗决策树获得各自

的分类结果，在通过最大投票法获得 RF 最终的分类结果。其中将 Sentinel2 NDVI 数据和 MOD13Q1 EVI 数据作为随机森林分类的特征值。搭建随机森林分类器，按县域进行分类，得到各县域棉花种植分布。并在 GEE 平台根据 connectedPixelCount 方法去除了小斑块的影响。最终，得到了阿克苏-阿拉尔区域的棉花空间分布。研究流程图如图 1 所示。

表 1 《基于多源遥感影像的阿克苏-阿拉尔区域棉花分布数据集（2020）》元数据简表

条目	描述
数据集名称	基于多源遥感影像的阿克苏-阿拉尔区域棉花分布数据集（2020）
数据集短名	Aksu_Alaer_Cotton_2020
作者信息	张萍，中国科学院新疆生态与地理研究所，z1571824849@163.com 范敬龙，中国科学院新疆生态与地理研究所，fanjl@ms.xjb.ac.cn 李生字，中国科学院新疆生态与地理研究所，oasis@mx.xjb.ac.cn
地理区域	阿克苏-阿拉尔区域
数据年代	2020 年
空间分辨率	250 m
数据格式	.tif、.shp
数据量	385KB（压缩为 1 个文件，134 KB）
数据集组成	（1）棉花种植分布；（2）样本点数据
基金项目	中华人民共和国科学技术部（2021xjkk0305）
数据计算环境	GEE 平台，ArcGIS
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>[15]</sup>
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

表 2 卫星数据源及基本参数表

研究数据	数据时间	GEE 中影像名称	空间分辨率	时间分辨率	轨道号
Landsat8	2020.3.1–2020.10.31	USGS Landsat 8 Collection 2 Tier 1 TOA Reflectance	30 m	16 d	145032 145033
					146031 146032
					146033 147031
					147032
Sentinel-2	2020.3.1–2020.10.31	Sentinel-2 MSI: MultiSpectral Instrument, Level-2A	10 m	5 d	44SNJ 44TLK
					44TLL 44SMJ
					44TNK 44TML
					44TMK 44TNL
MOD13Q1	2020.3.1–2020.10.31	MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m	250 m	16 d	44TMN 44TMM
					h23v04 h24v04
					h23v05 h24v05

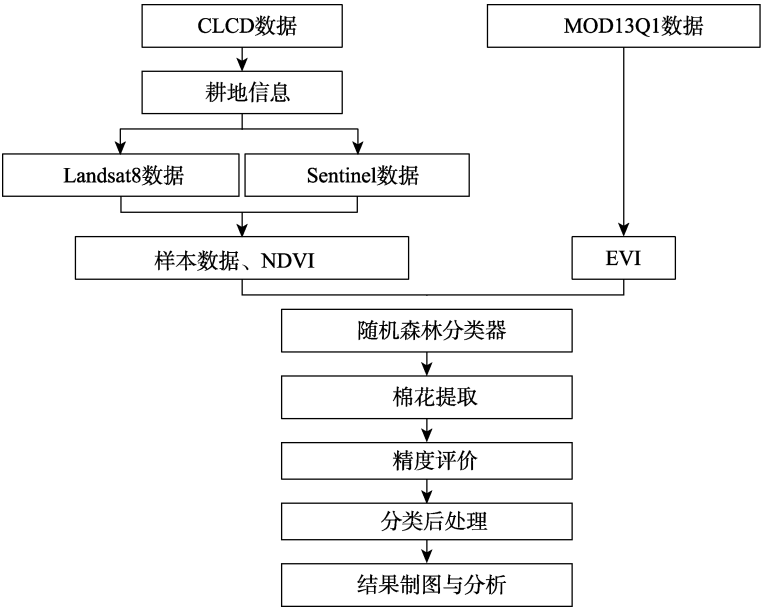


图 1 阿克苏-阿拉尔区域棉花分布数据集技术流程图

4 数据结果与验证

4.1 数据集组成

数据集包含 2 个数据文件：（1）阿克苏-阿拉尔区域棉花空间分布，以.tif 格式储存，空间分辨率为 250 m，时间为 2020 年；（2）阿克苏-阿拉尔区域棉花与非棉花样本点数据，以.shp 格式储存。

4.2 数据结果

阿克苏地区位于新疆维吾尔自治区，下辖两个县级市和七个县，分别是：阿克苏市、库车市、温宿县、沙雅县、新和县、阿瓦提县、柯坪县、拜城县和乌什县。在考虑研究范围时，由于一师阿拉尔建设兵团位于阿克苏地区中部，亦将阿拉尔纳入研究范围，而乌什县和拜城县基本无棉花分布，则不考虑这两个县的棉花提取。研究基于 GEE 平台，利用目视解译选取棉花样本点，如图 2 所示。选取棉花样本点 1,706 个，非棉花样本点 1,277 个，共计选取 2,893 个样本点。将阿克苏-阿拉尔区域按县域进行监督分类，利用随机森林方法得到棉花分布。从阿克苏-阿拉尔区域棉花空间分布格局来看（如图 3 所示），阿克苏-阿拉尔区域棉田集中在中部，主要在阿拉尔市、阿瓦提县的北部、温宿县的南部，沙雅县的北部和库车市的南部。此外，阿克苏-阿拉尔区域各县域棉花面积占耕地的比重差异明显，如表 3 所示。其中，一师阿拉尔比重最大为 67.33%，其次是沙雅县，棉花占耕地比重为 64.63%。库车市和阿瓦提县的比重也超过一半，分别是 58.82%和 52.41%。

4.3 数据结果验证

研究还对各县域提取结果进行了精度检验，采用总体分类精度、kappa 系数、生产者精度、用户精度对数据集进行验证，详细结果如表 4 所示。各县域验证结果表明总体分类

精度都在 0.9 以上，Kappa 系数都在 0.8 以上。其中精度最高的是温宿县，总体分类精度高达 0.99, kappa 精度高达 0.97, 精度最低的是柯坪县, 总体分类精度 0.94, kappa 精度为 0.83。

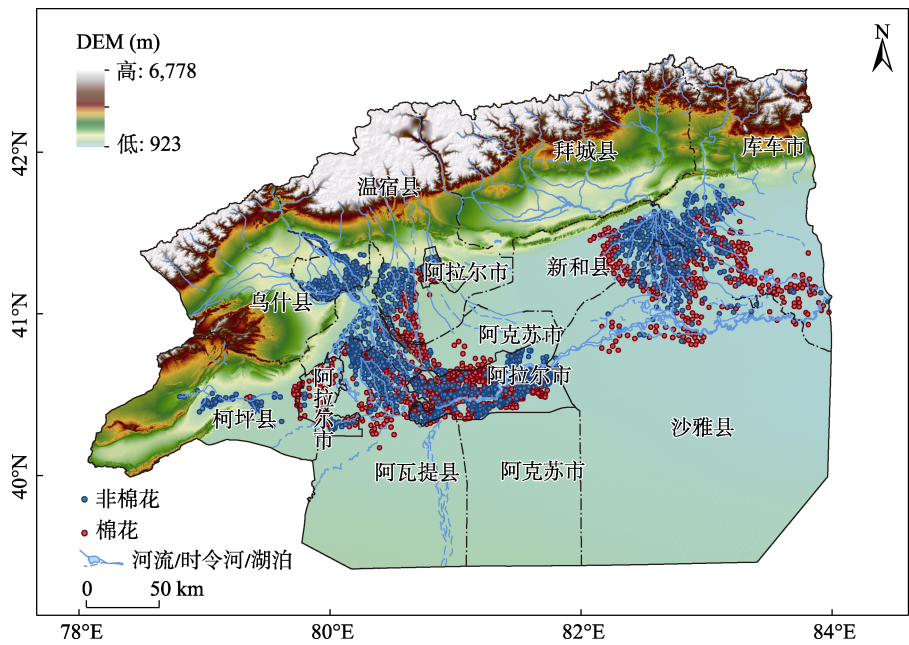


图 2 阿克苏-阿拉尔区域海拔高度与棉花样本点与非棉花样本点分布图

（注：基于新疆维吾尔自治区自然资源厅标准地图服务网站新 S(2023)149 号的标准地图制作，底图边界无修改）

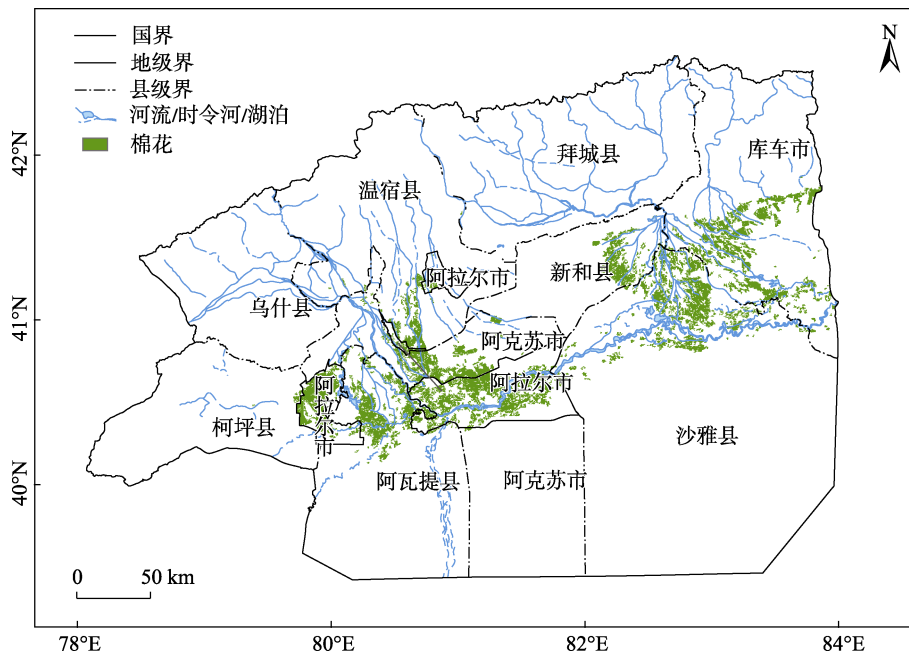


图 3 阿克苏-阿拉尔区域棉花空间分布图

表 3 阿克苏-阿拉尔区域棉花占比

地名（2020 年）	耕地面积（10 <sup>3</sup> hm <sup>2</sup> ）	棉花面积（10 <sup>3</sup> hm <sup>2</sup> ）	占比（%）
阿克苏市	156	74	47.44
库车市	238	140	58.82
阿瓦提县	166	87	52.41
柯坪县	16	2	12.50
沙雅县	229	148	64.63
温宿县	180	37	20.56
新和县	114	47	41.23
阿拉尔市	303	204	67.33

表 4 阿克苏-阿拉尔区域各县域精度

地区（2020 年）	总体分类精度	Kappa 系数	用户精度	生产者精度
阿克苏市	0.97	0.94	[0.94, 1]	[1], [0.94]
库车市	0.98	0.95	[0.98, 0.97]	[0.98], [0.97]
阿瓦提县	0.93	0.86	[0.92, 0.94]	[0.92], [0.94]
柯坪县	0.94	0.83	[0.88, 0.95]	[0.88], [0.95]
沙雅县	0.93	0.85	[0.88, 1]	[1], [0.85]
温宿县	0.99	0.97	[0.95, 1]	[1], [0.99]
新和县	0.97	0.93	[0.92, 1]	[1], [0.95]
一师阿拉尔市	0.95	0.87	[0.95,0.93]	[0.97],[0.88]

5 讨论与总结

本研究基于 GEE 平台，对阿克苏-阿拉尔区域的棉花分布进行了深入分析。利用 Sentinel-2、Landsat8 数据高空间分辨率遥感影像目视解译获取了充分的棉花和非棉花样本点，总计 2,893 个。采用随机森林方法进行监督分类，得到了各县域的棉花分布图。最终构建了阿克苏-阿拉尔区域棉花空间分布数据集，绘制了阿克苏-阿拉尔区域 2020 年棉花的空间分布。研究结果显示阿克苏-阿拉尔区域棉田集中在阿克苏-阿拉尔区域的中部，主要在阿拉尔市、阿瓦提县的北部、温宿县的南部，沙雅县的北部，库车市的南部。近年来，对阿克苏-阿拉尔区域棉花的研究主要将阿拉尔市与阿克苏地区作为一个整体<sup>[19]</sup>研究，或是仅对阿拉尔市<sup>[12]</sup>进行棉花提取，或是一个团<sup>[20]</sup>的精细棉花提取。本研究将阿克苏-阿拉尔区域按县域进行棉花提取，通过验证，各县域的分类准确性表现良好，总体分类精度和 Kappa 系数均在 0.8 以上。研究结果为阿克苏-阿拉尔区域的农业规划和资源管理提供了重要的空间信息支持。未来棉花的提取中可再提高空间分辨率，做到更精、更准。

作者分工：范敬龙、丁刚、李生宇对数据集做了总体设计；张萍采集和处理了数据并撰写了数据论文。

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] 喻树迅, 张雷, 冯文娟. 棉花生产规模化、机械化、信息化、智能化和社会服务化发展战略研究[J]. 中国工程科学 2016, 18: 137–148.
- [2] 中华人民共和国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2021: 4–5.
- [3] 新疆生产建设兵团统计局, 国家统计局兵团调查总队. 新疆生产建设兵团统计年鉴[M]. 北京: 中国统计出版社, 2021: 5.
- [4] Hao, P. Y., Wang, L., Zhan, Y. L., *et al.* Using Moderate-resolution temporal NDVI profiles for high-resolution crop mapping in years of absent ground reference data: a case study of Bole and Manas counties in Xinjiang, China [J]. *ISPRS International Journal of Geo-Information*, 2016, 5: 23. DOI: 10.3390/ijgi5050067.
- [5] Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., *et al.* Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks [J]. *Remote Sensing*, 2015, 7: 2627–2646. DOI: 10.3390/rs70302627.
- [6] Berhane, T. M., Lane, C. R., Wu, Q. S., *et al.* Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory [J]. *Remote Sensing*, 2018, 10: 26. DOI: 10.3390/rs10040580.
- [7] Hubert-Moy, L., Thibault, J., Fabre, E., *et al.* Mapping grassland frequency using decadal MODIS 250 m time-series: towards a national inventory of semi-natural grasslands [J]. *Remote Sensing*, 2019, 11: 21. DOI: 10.3390/rs11243041.
- [8] Xiong, J., Thenkabail, P. S., Tilton, J. C., *et al.* Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine [J]. *Remote Sensing*, 2017, 9: 27. DOI: 10.3390/rs9101065.
- [9] Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., *et al.* An assessment of the effectiveness of a random forest classifier for land-cover classification [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 67: 93–104. DOI: 10.1016/j.isprsjprs.2011.11.002.
- [10] Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., *et al.* Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture [J]. *Remote Sensing of Environment*, 2012, 121: 93–107. DOI: 10.1016/j.rse.2011.12.003.
- [11] 周磊, 林志树, 玉林海等. 基于 PIE 平台的棉花种植面积动态监测研究[J]. 航天返回与遥感 2023, 44: 108–118.
- [12] 吕绍伦, 赵阳, 陈万基等. 基于遥感云计算的阿拉尔市棉花种植面积提取[J]. 棉花科学 2022, 44: 19–25.
- [13] 王汇涵, 张泽, 康孝岩等. 基于 Sentinel-2A 的棉花种植面积提取及产量预测[J]. 农业工程学报 2022, 38: 205–214.
- [14] 张萍, 范敬龙, 李生字. 基于多源遥感影像的阿克苏-阿拉尔区域棉花分布数据集(2020)[J/DB/OL]. 全球变化数据仓储电子杂志, 2024. <https://doi.org/10.3974/geodb.2024.02.10.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2024.02.10.V1>.
- [15] 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).
- [16] 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: 3907–3925. DOI: 10.5194/essd-13-3907-2021.
- [17] Immitzer, M., Vuolo, F., Atzberger, C. First Experience with Sentinel-2 data for crop and tree species classifications in Central Europe [J]. *Remote Sensing*, 2016, 8: 27. DOI: 10.3390/rs8030166.
- [18] Breiman, L. Random forests [J]. *Machine Learning*, 2001, 4: 5–32. DOI: 10.1023/a:1010933404324.
- [19] 刘传迹, 金晓斌, 徐伟义等. 2000–2020 年南疆地区棉花种植空间格局及其变化特征分析[J]. 农业工程学报, 2021, 37: 223–232.
- [20] 张楠楠, 张晓, 白铁成等. 基于无人机可见光影像的新疆棉田田间尺度地物识别[J]. 农业机械学报 2023, 54: 199–205.