



文献 CSTR:

32001.14.11-6035.noda.2022.0020.zh



文献 DOI:

10.11922/11-6035.noda.2022.0020.zh

数据 DOI:

10.57760/sciencedb.j00001.00475

文献分类: 地球科学

收稿日期: 2022-07-28

开放同评: 2022-08-16

录用日期: 2023-02-20

发表日期: 2023-09-20

漓江流域 2016 - 2020 年土地利用类型时空分布数据集

颜业峰^{1,2,3}, 朱岚巍^{2,3,1*}, 吴美青^{1,2,3}, 孙希延¹, 邱玉宝^{2,3,1}

1. 桂林电子科技大学, 信息与通信学院, 桂林 541004

2. 中国科学院空天信息创新研究院, 数字地球重点实验室, 北京 100094

3. 可持续发展大数据国际研究中心, 北京 100094

摘要: 土地利用监测是土地资源管理的重要基础。本数据集基于 Sentinel-2 遥感影像数据, 经过原始数据定标、镶嵌、裁剪等多项预处理手段, 并建立遥感解译标志作为分类训练样本, 采用最大似然法对桂林市漓江流域 2016-2020 年土地利用类型情况进行分类。结合原始随机检验样本, 利用混淆矩阵分析方法对数据集进行精度评估验证。分类结果表明, 漓江流域土地利用类型分类效果较为良好, 总体分类精度均达到 95% 以上, Kappa 系数均在 0.8 以上。总体上, 漓江流域土地利用类型变化较小, 主要的土地利用类型为林地, 占比在 65.2% 左右。空间异质特征表现为建设用地呈现不断扩张的趋势, 2016 - 2020 年建设用地面积占比由 4.17% 扩张到 5.19%。本数据集可以为土地格局时空变化、土地利用转移演变等研究提供数据支持。

关键词: 土地利用; 遥感影像; 遥感解译; 监督分类

数据库(集)基本信息简介

数据库(集)名称	漓江流域2016-2020年土地利用类型时空分布数据集
数据作者	朱岚巍 (zhulw@aircas.ac.cn)
数据通信作者	颜业峰、朱岚巍
数据时间范围	2016-2020年
地理区域	地理范围为24°38'03"-25°54'56"N, 110°02'26"-110°44'38"E, 其中包括漓江流域。地理区域范围涉及的区县有秀峰区、叠彩区、象山区、七星区、雁山区、临桂区、阳朔县、灵川县、兴安县、永福县、资源县、平乐县。
空间分辨率	10 m
数据量	50.3 MB
数据格式	*.tif
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00475
基金项目	广西创新驱动发展专项(桂科AA20302022-3-2); 中国科学院战略性先导科技专项(A类)(XDA19090130)。
数据库(集)组成	本数据集包含1个桂林市漓江流域2016-2020年土地利用类型数据集压缩包文件, 由5个以年份命名的tif格式文件组成。

* 论文通信作者

朱岚巍: zhulw@aircas.ac.cn

引言

土地是人类生产生活的宝贵财富，影响着每一个人的日常生活，同时也是动植物赖以生活的物质基础^[1]。土地利用的监测是土地资源管理、生态环境保护研究以及区域可持续发展的重要研究基础，同时也是人类影响自然，人为活动干扰最直接的表现形式^[2]。

漓江流域属于世界级旅游胜地，具有典型的喀斯特地貌，生态环境较为脆弱。近年来，随着桂林市漓江流域城市化的不断发展，旅游活动的扩张，漓江流域面临着生态景观的破坏，生物多样性的降低，土地退化等诸多环境问题^[3]。目前，关于漓江流域的相关研究仅限于水资源、植被覆盖、生态风险等方面，土地利用类型研究较为欠缺^[4]；并且在较少的漓江流域土地利用分类变化中，使用的数据集分辨率与分类方法不太适应于漓江流域的土地利用分类科学研究^[5]，漓江流域研究所使用的数据集大部分来源于中华人民共和国自然资源部提供的 30 m 全球地表覆盖数据的 globeland 30 数据源^[6]。

本数据集选择桂林市漓江流域作为研究区域，以 2016–2020 年 Sentinel-2 卫星遥感数据为原始数据源，经过一系列的遥感数据预处理后，采用监督分类方法中的最大似然分类法，获取了漓江流域多时相的 10 m 分辨率土地利用分类空间分布数据集。本数据集空间分辨率精度优于其他同类数据集^[6]，且选择的分类方法在漓江流域这类多云多雨地区更为适合，可为漓江流域之后的土地利用评估分析提供更高精度的数据基础。

1 数据采集和处理方法

1.1 数据采集

本数据集以欧洲航天局 (<https://scihub.copernicus.eu>) 提供的 Sentinel-2 遥感影像数据，从 2016 年至 2020 年的 5 期 10 景图像作为数据源。考虑到漓江流域地处中国南方地区，多云多雨季节较多，选择的影像生成时间尽量在冬季，因此影像受到云层和大气的干扰较小，对后续的图像数据处理产生误差的影响也相应减小。表 1 为所用遥感影像数据表。

表 1 遥感影像数据表

Table 1 The table of remote sensing image data

序号	年份	成像日期	影像序列号
1	2016 年	2016-12-09	S2A_MSIL1C_20161209T031122_N0204_R075_T49RDJ_20161209T031514 S2A_MSIL1C_20161209T031122_N0204_R075_T49RDH_20161209T031514
2	2017 年	2017-12-19	S2B_MSIL1C_20171219T031119_N0206_R075_T49RDH_20171219T064338 S2B_MSIL1C_20171219T031119_N0206_R075_T49RDJ_20171219T064338
3	2018 年	2018-11-29	S2A_MSIL1C_20181129T031051_N0207_R075_T49RDJ_20181129T060514 S2A_MSIL1C_20181129T031051_N0207_R075_T49RDH_20181129T060514
4	2019 年	2019-11-04	S2A_MSIL1C_20191104T030911_N0208_R075_T49RDH_20191104T060224 S2A_MSIL1C_20191104T030911_N0208_R075_T49RDJ_20191104T060224
5	2020 年	2020-11-08	S2A_MSIL1C_20201108T030941_N9999_R075_T49RDJ_20201108T171611 S2A_MSIL1C_20201108T030941_N9999_R075_T49RDH_20201108T063040

1.2 数据处理方法

1.2.1 监督分类方法

图像分类方法有很多种，例如支持向量机、最大似然法以及面向对象分类等。支持向量机法与最大似然法能够清晰分辨出块状的图斑信息，但是界限分析并不清楚。面向对象法能够清晰提取出地物类型的边界，但是内部可能会出现明显“飞地”现象^[7]。

根据土地利用分类的需求，以及南方地区云层覆盖下，分类的像元几乎是以块状的形式存在，所以本数据集采用最大似然法。最大似然法又称贝叶斯分类法，是目前监督分类方法中使用比较多的一种，它的分类依据主要是在贝叶斯准则的基础上，对遥感数据进行分类时，将波段合成好的图像数据当作满足多为正态分布的数据，从而进行分类函数的构造^[8]，其原理如下所示：

根据训练区得到各类别的均值、方差等特征参数，假设各类别服从多维正态分布，其条件概率密度 $p(x/w_i)$ 为：

$$p(x/w_i) = (2\pi)^{\frac{-n}{2}} \times \left| \sum_i \right|^{-0.5} \times \exp \left\{ \frac{-(x - \mu_i)'}{2 \sum_i (x - \mu_i)} \right\} \quad (1)$$

式中， x 表示待分类样本， w_i 表示第 i 土地利用分类类别， n 表示波段数， μ_i 表示 w_i 类的均值向量， \sum_i 表示 w_i 类的协方差矩阵。

由贝叶斯函数计算后验概率 $p(w_i/x)$ ：

$$p(w_i/x) = \frac{p(x/w_i) \times p(w_i)}{p(x)} \quad (2)$$

式中， $p(w_i)$ 为 w_i 类的先验概率， $p(x)$ 表示与类别无关的公共项。

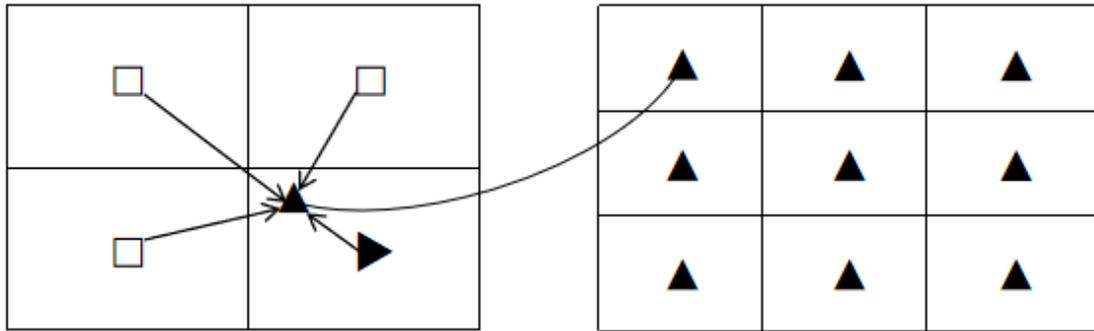
由判别规则确定类别的归属，实际过程中，主要通过改变概率阈值来生成分类影像。若 $p(w_i/x) \times p(w_i) \geq p(w_j/x) \times p(w_j)$ ，则 $x \in w_i$ 。其中， w_i 表示第 i 类土地利用类别， w_j 表示第 j 类土地利用类别。

本数据集依据 2017 年国务院颁布的中华人民共和国《土地利用分类现状国标 GB/T 21010—2017》的土地类型分类标准^[9]，结合桂林市漓江流域当地实际土地动态变化和生态主要影响因素，选取一共 6 类漓江流域较为典型的地物类型特征，本数据集将其分为：裸地、建设用地、耕地、草地、林地、水域。

1.2.2 波段合成方法

彩色合成实际上是在选择一个最佳波段组合基础上，运用三基色原理，进行波段组合的赋色方

案，有利于地物信息的判读。波段合成的方法为最邻近插值法，最邻近插值法是将原始图像数据中最邻近的像素灰度值作为输入值，输出到重采样图像数据的相应像素，其原理方式如图 1 所示^[10]。



□为原始图像像素，▶为需要进行计算合成的像素，▲为重采样后的像素。

图 1 最邻近插值法原理图

Figure 1 Schematic diagram of nearest neighbor interpolation method

利用遥感图像多光谱假彩色合成，可以充分显示自然环境信息。本数据集选取了遥感图像中蓝色波段、绿色波段和近红外波段组合的赋色方案，该方案对于植被、建筑、水体和裸地的区分度较高，适用于漓江流域土地类型的判别。经过波段合成之后，形成漓江流域标准假彩色影像，利用地物在不同波段中像元色差的综合反应，结合原始遥感影像，在标准假彩色影像中，选择出与 6 种土地利用类型一一对应的感兴趣区作为彩色图像中目标地物识别的基本标志，形成遥感解译标志像元集，解译标志像元表如表 2 所示。

表 2 解译标志

Table 2 Interpretation flags

序号	类型	说明	原始图像	标志像元
1	裸地	指土质裸露、植被覆盖几乎没有的土地利用类型，包括岩石、沙地、空闲地等		
2	建设用地	指人为建造的建筑居民点或者道路交通等土地利用类型		
3	耕地	指种植农作物的土地利用类型，其中包含一些零星果树或其他树木等		
4	草地	指生长的主要以草本植物为主的土地利用类型，包含郁闭度低于 0.1 的灌木地		
5	林地	指生长乔木、竹林、灌木等树木郁闭度较高的土地利用类型		
6	水域	指陆地水域、沟渠等水面覆盖的土地利用类型，包括河流水面、湖泊水面、水库水面		

1.2.3 数据处理步骤

Sentinel-2 卫星遥感影像首先也要对原始影像进行辐射定标和大气校正，该过程主要是在插件 Sencor-20.09.00 中完成，完成处理后得到一幅辐射定标和大气校正后的图像^[11]。接着，在欧空局的 SNAP 软件中，对图像进行重采样和格式转化，把图像格式设置为适合 ENVI 打开的图像格式以及把

波段的分辨率调整至 10 m，处理之后得到一幅经过预处理 Sentinel-2 卫星图像^[12]。由于漓江流域占用的影像为两幅，需要把两张影像进行镶嵌，再根据漓江流域的矢量范围对影像进行裁剪，把裁剪好的 Sentinel-2 波段图像集添加至 ENVI 软件中，在 ENVI 的波段合成窗口，选择绿、蓝、近红外三个波段的基本图像，并采用最邻近插值法的波段合成方法，调整三个波段的通道选择，R 通道选择绿色波段，G 通道选择蓝色波段，B 通道选择近红外波段，形成漓江流域的标准假彩色图像。基于该影像解译的标志，选取监督分类所需要的训练样本，选好训练样本后采用最大似然法进行监督分类，得到初步的分类结果图。由于初步的分类结果图存在一定的“椒盐”现象，需要进行后期的优化处理，即对出现偏差较大的像元进行人工修正，采用聚类分析法去除较小范围的误差像元，得到初步的分类结果图。利用混淆矩阵法对分类结果进行精度验证，获取较高精度的分类结果图。图 2 为数据处理技术流程图。

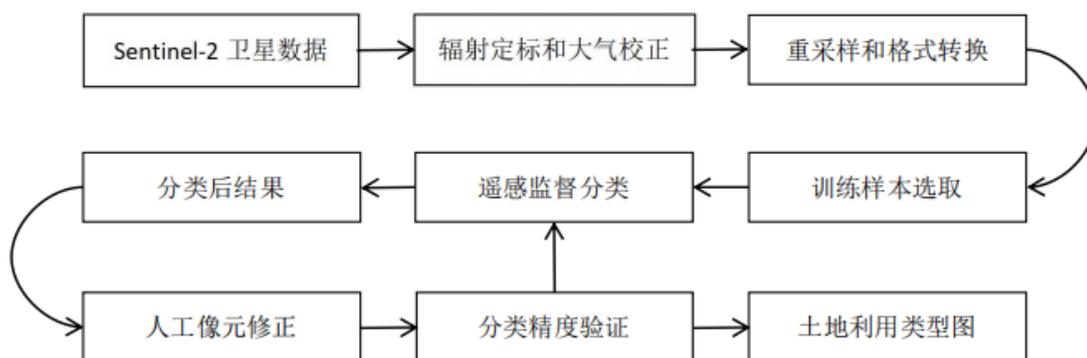


图 2 数据处理技术流程图

Figure 2 Flow chart of data processing technology

2 数据样本描述

2.1 数据集信息

本数据集包括漓江流域从 2016 年至 2020 年的土地利用空间分布数据图，总共有五期（2016、2017、2018、2019、2020）的数据图。这些数据集保存在一个压缩文件包内，各个时期的土地利用数据图都有一个日期命名，数据的存储总量为 50.3 MB。所有数据的坐标系均采用遥感影像中的 UTM-WGS84 坐标系。

2.2 数据样本

数据集采用了绿、蓝、近红外波段三个波段合成为标准假彩色图像。这三组波段合成有利于对水系、居民点和公园水体、林地等图像的判读。此次采用的分类方法是最大似然法，该方法能快速并且相对准确的依据训练样本得到图像分类结果，在提取的精度上也较于其他分类器更高，图 3 为 2016–2020 年部分漓江流域土地利用分类图。

3 数据质量控制和评估

3.1 数据精度验证的方法

本数据集精度验证采用的是混淆矩阵验证法。首先需要将处理分类好的图像转化为能进行精度

验证的 ENVI 格式，使用 ROI 工具建立一个样本验证的感兴趣区数据集，也可从外部发布的高精度分类样本数据中导入所需要的样本集。本数据集选择原始图像的随机检验样本，尽量避免与训练样本重复，然后导入至混淆矩阵分析工具中，得到关于数据集的精度评价矩阵，进而评估图像的分类精度^[13]。

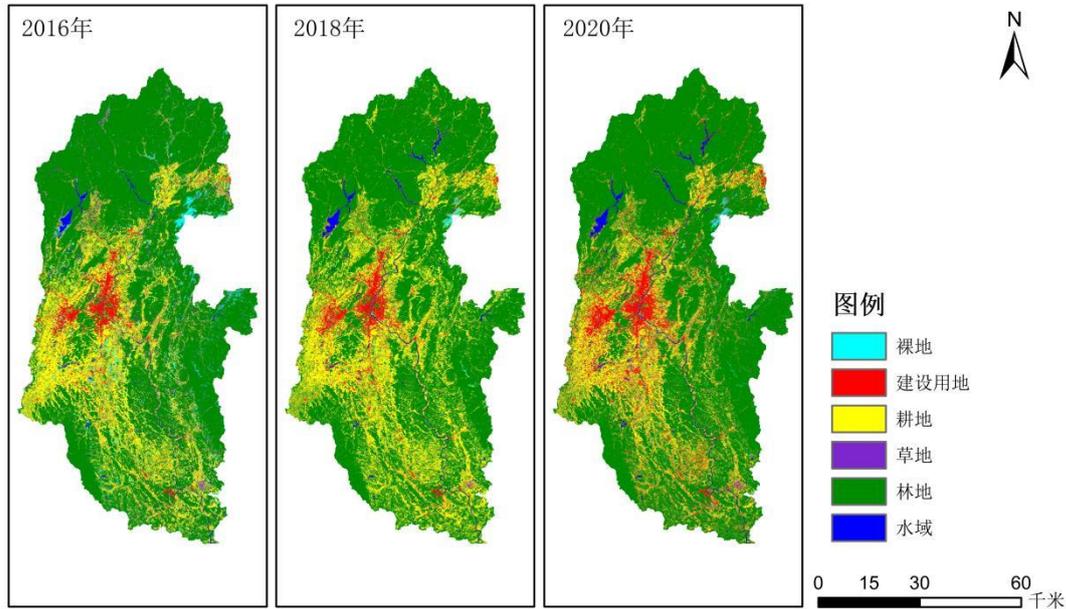


图 3 2016–2020 年部分漓江流域土地利用分类图

Figure 3 Partial land use classification map of the Lijiang River Basin during 2016–2020

3.2 数据精度验证的结果

经过混淆矩阵分析工具的计算，利用得到的总体分类精度（Overall Accuracy）和 Kappa 系数（Kappa Coefficient）对分类结果进行一个评估。分类结果的总体分类精度均大于 95%，Kappa 系数均大于 0.8。Kappa 系数大于 0.8 说明分类精度较好，Kappa 系数越高，分类精度越高，如表 3 所示。

表 3 精度验证表

Table 3 Accuracy verification table

	2016 年	2017 年	2018 年	2019 年	2020 年
Overall Accuracy	96.92%	95.30%	97.09%	95.78%	96.02%
Kappa Coefficient	0.859	0.812	0.886	0.816	0.806

4 数据使用方法和建议

本数据集为 2010–2020 年漓江流域土地覆被类型空间分布数据集，格式为 tif 格式，可以在 Envi、ArcGIS 等空间数据处理平台进行使用，可以基于该土地覆被类型数据开展相关的科学研究、理论分析、对比验证等。本数据集可为漓江流域的土地格局变化分析与预测演变提供数据支持。

作者分工职责

颜业峰（1998—），男，广西钦州市人，在读硕士研究生，研究方向为遥感监测和应用。主要承担工作：漓江流域土地利用类型的分类，精度验证和实地考察。

朱岚巍（1980—），女，内蒙古赤峰市人，副研究员，研究方向为遥感应用及微波遥感。主要承担工作：本研究实施方案设计和整体质量控制，及精度验证和实地考察。

吴美青（2000—），女，广西百色市人，在读硕士研究生，研究方向为遥感监测和应用。主要承担工作：漓江流域土地利用类型的标签定义，精度验证和实地考察。

孙希延（1973—），女，山东人，博士，教授，主要研究方向为卫星通信，卫星导航。主要承担工作：本数据集的方案设计制作，精度验证和实地考察。

邱玉宝（1978—），男，江西兴国人，博士，研究员，研究方向为被动微波遥感、冰雪及北极环境遥感及数据科学研究。主要承担工作：本数据集的精度验证和实地考察。

参考文献

- [1] 陈妮, 应丰, 王静, 等. 基于 U-Net 的高分辨率遥感图像土地利用信息提取[J]. 遥感技术与应用, 2021, 36(2): 285 - 292. [CHEN N, YING F, WANG J, et al. Research on land use information extraction based on U-net[J]. Remote Sensing Technology and Application, 2021, 36(2): 285 - 292.]
- [2] 卢辉雄, 聂振龙, 刘敏, 等. 基于 RS 和 GIS 的石羊河流域近 50 年土地覆被类型变化研究[J]. 地质与资源, 2020, 29(2): 165 - 171, 179. DOI: 10.13686/j.cnki.dzyzy.2020.02.008. [LU H X, NIE Z L, LIU M, et al. Research on land cover changes in Shiyang River Basin in recent 50 years based on rs and GIS[J]. Geology and Resources, 2020, 29(2): 165 - 171, 179. DOI: 10.13686/j.cnki.dzyzy.2020.02.008.]
- [3] 何映红. 基于多源遥感数据的漓江流域地表覆盖分类研究[D]. 西安: 长安大学, 2014. [HE Y H. Land cover classification in Lijiang River Basin based on multi-source Remote sensing data[D]. Xi'an: Chang'an University, 2014.]
- [4] 胡金龙. 漓江流域土地利用变化及生态效应研究[D]. 武汉: 华中农业大学, 2016. [HU J L. Study on Land Use Change and Ecological Effect in Lijiang River Basin[D]. Wuhan: Huazhong Agricultural University, 2016.]
- [5] 余玲, 何文, 姚月锋, 等. 基于多角度的漓江流域土地利用/覆被变化及驱动力分析[J]. 中国水土保持, 2020(1): 39 - 43, 5. DOI: 10.14123/j.cnki.swcc.2020.0038. [YU L, HE W, YAO Y F, et al. Land use/cover change and its driving forces in Lijiang River Basin based on multiple angles[J]. Soil and Water Conservation in China, 2020(1): 39 - 43, 5. DOI: 10.14123/j.cnki.swcc.2020.0038.]
- [6] 张军民, 李昱桥, 荣城. 漓江流域 2000—2020 年土地利用和覆被变化时空分异特征[J]. 水土保持通报, 2022, 42(2): 250 - 258. DOI: 10.13961/j.cnki.stbctb.2022.02.034. [ZHANG J M, LI Y Q, RONG C. Spatial and temporal characteristics of land use and land cover change in Lijiang River Basin during 2000—2020[J]. Bulletin of Soil and Water Conservation, 2022, 42(2): 250 - 258. DOI: 10.13961/j.cnki.stbctb.2022.02.034.]
- [7] 麦丽素, 乌兰吐雅. 基于 Sentinel-2 数据的水稻面积提取方法比较分析[J]. 北方农业学报, 2019, 47(5): 119 - 126. DOI: 10.3969/j.issn.2096-1197.2019.05.22. [MAI L S, WU L T Y. Contrastive analysis of

- extraction of rice area classification based on data of Sentinel-2[J]. Journal of Northern Agriculture, 2019, 47(5): 119 – 126. DOI: 10.3969/j.issn.2096-1197.2019.05.22.]
- [8] 王国重, 李中原, 张继宇, 等. 偃师市土地利用遥感分类研究[J].湖北农业科学, 2021, 60(13):60-62. DOI:10.14088/j.cnki.issn0439-8114.2021.13.011.[WANG G Z, LI Z Y, ZHANG J Y, et al. Remote sensing classification of land use in Yanshi City[J]. Hubei Agricultural Sciences, 2021, 60(13):60-62. DOI:10.14088/j.cnki.issn0439-8114.2021.13.011.]
- [9] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 土地利用现状分类: GB/T 21010—2017[S]. 北京: 中国标准出版社, 2017.[General Administration of Quality Supervision, Inspection and Quarantine of the People’s Republic of China, Standardization Administration of China. Classification of Land Use Status: GB/T 21010—2017[S]. Beijing: Standards Press of China, 2017.]
- [10] 田林, 张亭禄, 陈树果, 等. 卫星遥感融合中通量守恒重采样方法与其它常用方法的比较[J]. 中国海洋大学学报(自然科学版), 2014, 44(6): 103 – 108. DOI: 10.16441/j.cnki.hdx.2014.06.016. [TIAN L, ZHANG T L, CHEN S G, et al. Comparison of flux-conserving resampling method with commonly used resampling methods in merging of remote sensing data[J]. Periodical of Ocean University of China, 2014, 44(6): 103 – 108. DOI: 10.16441/j.cnki.hdx.2014.06.016.]
- [11] ZENG T J, WANG C J. SVM-based land use/cover classification in Shihezi area[C]//2016 Progress in Electromagnetic Research Symposium (PIERS). Shanghai, China. IEEE, 2016: 2077 – 2079. DOI: 10.1109/PIERS.2016.7734875.
- [12] CLOSE O, BENJAMIN B, PETIT S, et al. Use of sentinel-2 and LUCAS database for the inventory of land use, land use change, and forestry in Wallonia, Belgium[J]. Land, 2018, 7(4): 154. DOI: 10.3390/land7040154.
- [13] 容芳芳. 基于 RS 技术的城市土地利用变化研究[J]. 北京测绘, 2021, 35(1): 36 – 40. DOI: 10.19580/j.cnki.1007-3000.2021.01.008. [RONG F F. Research on urban land use change based on RS technology[J]. Beijing Surveying and Mapping, 2021, 35(1): 36 – 40. DOI: 10.19580/j.cnki.1007-3000.2021.01.008.]

论文引用格式

颜业峰, 朱岚巍. 漓江流域 2016-2020 年土地利用类型时空分布数据集[J/OL]. 中国科学数据, 2024, 9(2). (2023-06-25). DOI: 10.11922/11-6035.noda.2022.0020.zh.

数据引用格式

颜业峰, 朱岚巍. 漓江流域 2016–2020 年土地利用类型时空分布数据集[DS/OL]. Science Data Bank, 2022. (2023-09-20). DOI: 10.57760/sciencedb.j00001.00475.

A dataset of land use types in the Lijiang River Basin from 2016 to 2020

YAN Yefeng^{1,2,3}, ZHU Lanwei^{2,3,1*}, WU Meiqing^{1,2,3}, SUN Xiyan¹, QIU Yubao^{2,3,1}

1. School of Information and Communication Guilin University of Electronic Technology, Guilin 541004, P.R. China

2. International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, P.R. China

3. Key Laboratory of Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, P.R. China

* Email: zhulw@aircas.ac.cn (ZHU Lanwei)

Abstract: Land use monitoring is an important basis for land resource management. On the basis of Sentinel-2 remote sensing image data, we used multiple preprocessing methods such as original data calibration, Mosaic and tailoring, and established remote sensing interpretation markers as classification training samples. Then we adopted the maximum likelihood method to classify the land use situation of the Lijiang River Basin in Guilin City from 2016 to 2020. Moreover, in combined with the original random test samples, we used the confusion matrix analysis method to evaluate the accuracy of the dataset. The classification results show a relatively good classification effect of land use types in the Lijiang River Basin, with an overall classification accuracy above 95%, and the kappa coefficient above 0.8. On the whole, the land use types in the Lijiang River basin changed little, and the main land use type is woodland, accounting for about 65.2%. The spatial heterogeneity show that the construction land area expanded from 4.17% to 5.19% from 2016 to 2020. This dataset can provide data support for the study of spatio-temporal changes of land pattern and land use transfer and evolution.

Keywords: land use; remote sensing images; remote sensing interpretation; supervised classification

Dataset Profile

Title	A dataset of land use types in the Lijiang River Basin from 2016 to 2020
Data corresponding author	ZHU Lanwei (zhulw@aircas.ac.cn)
Data authors	YAN Yefeng, ZHU Lanwei
Time range	2016–2020
Geographical scope	Lijiang River Basin (24°38'03"–25°54'56"N, 110°02'26"–110°44'38"E.), covering Xiufeng District, Diecai District, Xiangshan District, Qixing District, Yanshan District, Linggui District, Yangshuo County, Lingchuan County, Xing'an County, Yongfu County, Ziyuan County and Pingde County.
Spatial resolution	10 m
Data volume	50.3 MB
Data format	*.tif
Data service system	< https://doi.org/10.57760/sciencedb.j00001.00475 >

Sources of funding	Guangxi Innovation Driven Development Project (Guike AA20302022-3-2); Chinese Academy of Sciences Strategic Pilot Science and Technology Project (Class A) (XDA19090130).
Dataset composition	The dataset is composed of a compressed package, including 5 tif type data files of the land use type dataset in the Lijiang River Basin of Guilin City from 2016 to 2020. The data files are named in the order of years.