

2013–2022 年蒙古高原逐年生长季地表水分布数据集

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14. 11-6035.csd.2022.0080.zh

文献 DOI:

10.11922/11-6035.csd.2022.0080.zh

数据 DOI:

10.57760/sciencedb.j00001.00665

文献分类: 地球科学

收稿日期: 2022-09-22

开放同评: 2022-10-14

录用日期: 2023-02-20

发表日期: 2023-03-02

李凯^{1,2}, 王卷乐^{2*}, 程文静³, 洪梦梦^{2,4}

1. 中国矿业大学(北京), 地球科学与测绘工程学院, 北京 100083

2. 中国科学院地理科学与资源研究所, 资源与环境信息系统国家重点实验室, 北京 100101

3. 中国气象科学研究院, 北京 100081

4. 山东理工大学, 建筑工程学院, 山东淄博 255049

摘要: 蒙古高原地处干旱半干旱地区, 水文水资源是其资源环境发展的重要制约条件。掌握蒙古高原的水体时空分布情况对于提示水资源和水环境时空特征及其在区域气候变化和防灾减灾方面的影响和响应具有重要意义。然而, 由于该区域辽阔且跨越中蒙两个国家, 精确自动化获取流域尺度的大范围长时序水体面临很大挑战。本研究采用本地深度学习训练和谷歌地球引擎(Google Earth Engine, GEE)分布式计算相结合的方法, 对 GEE 赋予深度学习计算能力, 使 GEE 可以快速自动化部署深度学习模型。基于此, 完成蒙古高原 2013–2022 年逐年生长季地表水分布的获取, 空间分辨率为 30 m。人工选择验证点 5000 个, 总体验证为 88.0%。数据集为 TIFF 栅格形式, 以 5°×5°×10 年的形式存储为 28 个瓦片影像, 数据量为 339 MB, 压缩后为 88.1 MB, 在原始数据格式下为 189 GB。本数据集采用的模型方法可以自动化、高效地在云端进行水体制图, 为干旱半干旱地区大范围、长时序、高效率的水体的自动化处理提供了可能, 具有应用和推广价值。

关键词: 蒙古高原; 水体分布; 水文水资源; 谷歌地球引擎; 深度学习

数据库(集)基本信息简介

数据库(集)名称	2013–2022 年蒙古高原逐年生长季地表水分布数据集
数据通信作者	王卷乐(wangjl@igsnr.ac.cn)
数据作者	李凯, 王卷乐, 程文静, 洪梦梦
数据时间范围	2013–2022 年
地理区域	蒙古高原(中国内蒙古和蒙古国)
空间分辨率	30 m
数据量	339 MB, 压缩后 88.1 MB (原始格式 189 GB)
数据格式	*.tif
数据服务系统网址	http://dx.doi.org/10.57760/sciencedb.j00001.00665
基金项目	国家自然科学基金项目(41971385、32161143025); 中国工程科技知识中心建设项目(CKCEST-2022-1-41); 资源与环境信息系统国家重点实验室自主创新项目(KPI006)

* 论文通信作者

王卷乐: wangjl@igsnr.ac.cn

数据库（集）组成	经纬度每隔5°存储为一个.tif文件，压缩包内共有28个.tif文件，每个.tif文件有10个通道（2013–2022年），文件命名方式为MP_water_ExxxNyy.tif。
----------	--

引言

水体是一切生命赖以生存的重要自然资源之一。联合国可持续发展目标 6（Sustainable Development Goals 6, SDG 6）提出为所有人提供水和环境卫生并对其进行可持续管理，其亚目标 6.6 指出到 2020 年，保护和恢复与水有关的生态系统，包括山地、森林、湿地、河流、地下含水层和湖泊。蒙古高原地处干旱半干旱地区，水文水资源是其资源环境发展的重要制约条件。掌握蒙古高原的地表水体时空变化，对于揭示本区域水资源和水环境特征及其对气候变化和防灾减灾等影响和响应具有重要的科学意义。得益于卫星对地观测技术的发展，使得年度大范围水体监测与评估成为可能。然而，由于蒙古高原地域辽阔，水体分布破碎，河流细小，且季节性河流大量存在不连续的情况，大范围跨境地理单元长时序水体自动化提取面临很大挑战。

前人的研究多以水体指数作为提取水体的主要手段。1996 年，McFeeters 利用遥感影像的绿光波段与近红外波段做归一化提出了归一化差值水体指数（Normalized Difference Water Index, NDWI），其认为该指数可以抑制土壤和植被特征用于描述水体特征并且有效估算地表水面积^[1]。徐涵秋通过改进 NDWI，将 NDWI 指数中的近红外波段替换为短波红外，构建了一种新的适用于提取建筑区水体信息的指数——改进归一化差异水体指数（Modified Normalized Difference Water Index, MNDWI）^[2]。Gudina L. Feyisa 等人构建了自动水体提取指数（Automated Water Extraction Index, AWEI），提高了通常会被误分类的阴影和暗表面地物的分类精度，总体精度优于 MNDWI 和最大似然分类方法^[3]。但是这些利用指数对水体分类的方法在不同场景中阈值难以确定，单一阈值或者单一指数往往难以在细小河流、山地阴影等区域取得较为准确的水体分类结果。步入遥感人工智能时代后，大量算法不断涌现^[4–6]。Chen 等^[7]设计了一种新型的卷积神经网络（Convolutional Neural Networks, CNN）在城市中心区从复杂的城市背景中准确地提取出水体信息。姚锦一和王卷乐^[8]利用谷歌地球引擎（Google Earth Engine, GEE）在青藏高原若尔盖地区开展水体分类试验，采用多种水体指数结合支持向量机模型获得了该地区的高精度水体分布，并随后在蒙古国色楞格河流域局部开展深度神经网络的水体提取研究并绘制了色楞格河流域 2019 年河流分布图^[9]。Li 等^[10]针对蒙古高原的图拉河流域对现有水体指数的阈值分割方法以及神经网络模型进行对比，提出了参数量少、分类精度优良的 Pixel-based CNN 水体提取模型。但由于过度依赖标签样本，水体提取依旧以小尺度、区域性的研究为主^[11–13]。因此，在不增加过多人力成本的条件下，在广袤的干旱和半干旱地区大范围、长时序、高效率地提取水体仍是一个难点。

遥感云计算平台 GEE 提供了大量影像数据以及影像处理、统计、分析的函数，其优势在于可快速、高效地进行大范围区域遥感监测和制图。遥感影像深度学习已然成为当下研究的热门，这为蒙古高原水体的长时序获取提供了技术方法基础。本数据集制作过程协同调用本地深度学习环境和 GEE 遥感数据及计算平台，采用本地深度学习训练，将训练后的模型应用于 GEE 平台，使 GEE 可以快速自动化部署深度学习模型。利用该方法生产了蒙古高原 2013–2022 年际地表水分布数据集。对比现有全球尺度的覆盖产品^[14–16]，本数据集处理具有技术方法优势，能够在较短时间内完成时间间隔为 1 年的长时序、大范围水体产品开发，极大地提高了大区域研究所需要的水体数据支撑能力。本数据集有望为蒙古高原水文水资源格局与变化、气候变化区域影响与响应、水文气象灾害防治和

以水定畜的畜牧业调控等研究提供支持。

1 数据采集和处理方法

1.1 数据生产流程

本数据集的生产主要分为数据预处理、模型训练以及模型在线部署三个部分（图 1）。第 1 部分预处理主要包括影像数据筛选、影像裁切以及数据的本地存储。第 2 部分为模型训练，包括影像标签噪声纠正、特征波段构建、模型构建等，特征波段包括可见光、近红外、线性增强波段（Linear Enhancement Band, LEB）、MNDWI、美国航天局数字高程模型（National Aeronautics and Space Administration Digital Elevation Model, NASADEM）。第 3 部分是通过模型参数解析和接口调用实现模型在线部署完成数据集制作。其中数据预处理与模型在线部署是通过本地与 GEE 云计算环境交互实现，数据的处理过程是在云端进行，模型训练则是在个人计算机中完成。

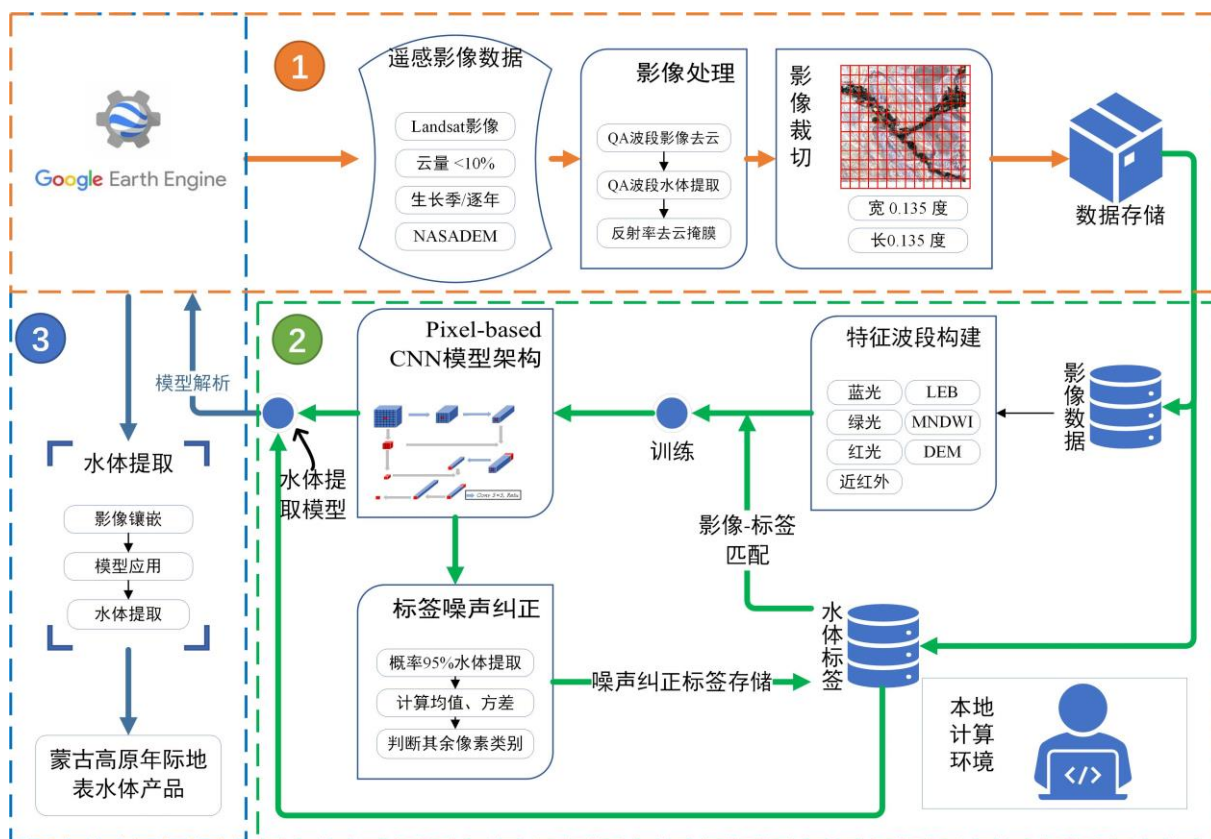


图 1 数据生产流程图

Figure 1 Data production flowchart

1.2 数据来源

影像数据源包括 Landsat 8 的地表反射率数据^[17]、NASADEM^[18]，均为 GEE 上可访问的数据集。数据的相关信息如表 1 所示，其中 Landsat 8 影像数据主要用于选取、构建对水体敏感的特征波段，DEM 主要用于减少地形造成的山地阴影的干扰，质量评估（Quality Assessment, QA）波段的参与是用于初始水体标签的构建。

表 1 数据源

Table 1 Data source

数据来源	波段	最小值	最大值	分辨率	波长
Landsat 8	B1	1	65455	30 m	0.435–0.451 μm
	B2	1	65455	30 m	0.452–0.512 μm
	B3	1	65455	30 m	0.533–0.590 μm
	B4	1	65455	30 m	0.636–0.673 μm
	B5	1	65455	30 m	0.851–0.879 μm
	B6	1	65455	30 m	1.566–1.651 μm
	B7	1	65455	30 m	2.107–2.294 μm
	QA	0	65535	30 m	/
NASA DEM	Elevation	-512	8768	30 m	/

1.3 训练数据选取

蒙古高原狭义上的范围包括蒙古国和中国的内蒙古自治区。本研究选取该范围开展研究。由于蒙古高原地域辽阔，考虑到不同场景、地形的影响，将研究区域进一步划分为北部山地区、南部荒漠区以及东部平原三部分（图 2）。对研究区的三部分单独选择训练样带，样带选择信息如图 3 所示。蒙古高原一期全覆盖影响需 Landsat 8 影像 216 景，训练样带共选择训练区域影像 71 景，覆盖不同地表、海拔，能够有效表征多种场景下的影像信息。

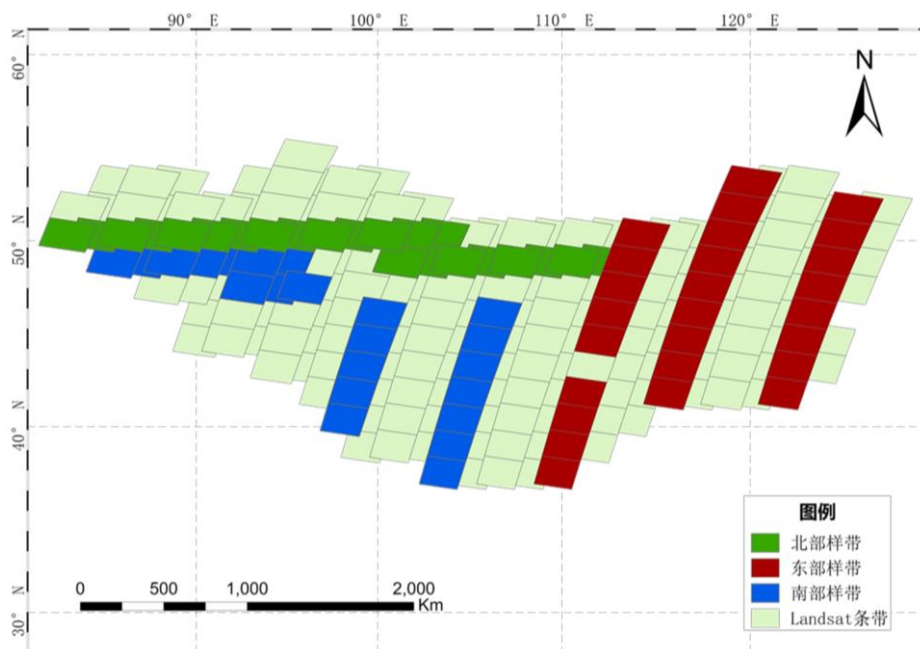


图 2 研究区域划分图

Figure 2 Zoning map of the study area

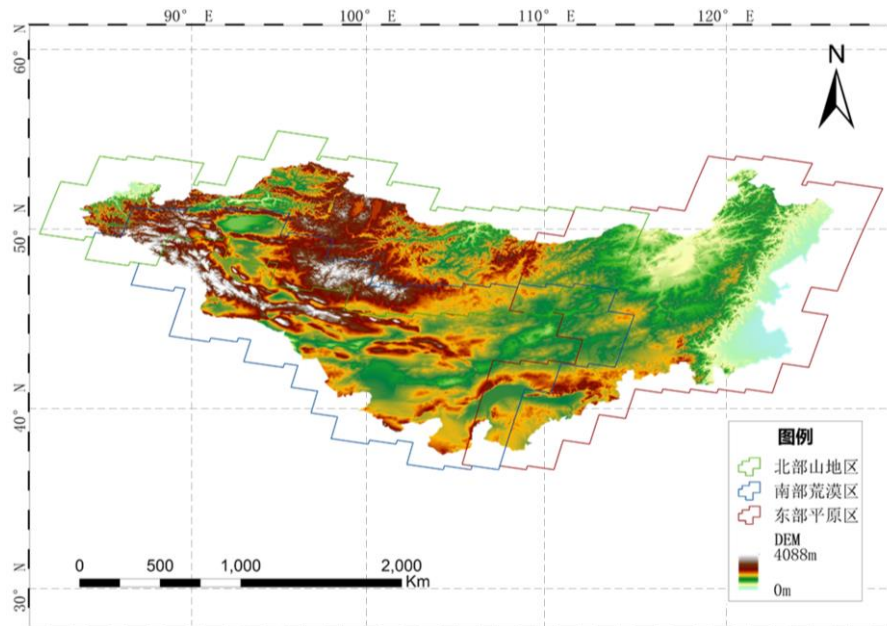


图3 训练样带分布图

Figure 3 Training transect distribution map

1.4 数据预处理

遥感影像受到云雾的干扰，短时间内难以合成大范围的无云影像。因此，本研究选取蒙古高原植被生长较为旺盛的6–8月进行无云影像的合成。依据 Landsat 8 影像中的质量评估波段中的云体信息，对影像数据进行掩膜处理，获得多景无云去云后的遥感影像。将去云后的影像在时序上做均值叠加处理最后合成6–8月的无云影像。2013–2022年每年一期无云影像共10期。在波段选择上，考虑到水体的反射率信息，我们选择了可见光、近红外、MNDWI、短波红外作为特征波段。

1.5 模型训练

模型选择李凯等人构建的 Pixel-based CNN 水体提取模型^[10]。该模型可兼顾考虑影像的像元和纹理特征，且具有参数量小易于训练的特点。提取 Landsat 影像中质量评估波段中的粗略水体信息，再利用模型进行标签噪声纠正获得参考水体标签数据^[19]。构建可见光、近红外、MNDWI、短波红外、线性增强波段、数字高程模型合成的特征数据，与参考水体标签数据联合参与模型训练。水体模型的训练根据训练样本的选择分为北部、南部、东部三个独立的模型。

1.6 模型在线部署

训练后的三个水体提取模型分别应用于蒙古高原的适用区域。通过 python 解析提取模型权重信息，利用 GEE 的接口传入权重信息，进行卷积、裁剪、拼接等操作的转译，使得 Pixel-based CNN 模型可在 GEE 云端部署。之后对三个部分的水体信息进行镶嵌，合成整个蒙古高原的水体提取数据。虽然6–8月合成一期遥感影像，但是因为云的体量较大，每年都或多或少会存在云体造成的空洞现象。对于这些空洞，采用该年 QA 波段的水体数据进行填补。完成蒙古高原2013–2022年水体数据产品的制作，数据产品的空间分辨率为30 m。原始格式数据量较大，考虑到用户计算机内存限制，将2013–2022年水体数据按照经纬度每5°进行裁剪，共获得28个5°×5°的小瓦片影像。影像有10个通道，分别对应2013–2022年该区域的水体数据。

2 数据样本描述

本数据集为栅格数据类型,地理坐标系统为 WGS84,数据保存为 tif 格式。数据存储在 ScienceDB 存储库 (<https://www.scidb.cn/>)。因为原始格式数据量较大,为便于用户使用,ScienceDB 中数据被分为 28 个瓦片存储,每个文件的大小为 $5^{\circ} \times 5^{\circ} \times 10$ 通道。数据文件的命名格式采用瓦片影像的左下角经纬度进行命名,其他字符的具体解释图 4 所示。

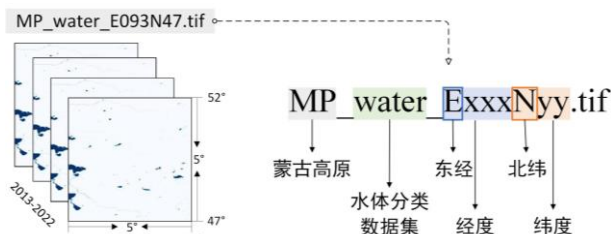


图 4 文件命名格式

Figure 4 File naming format

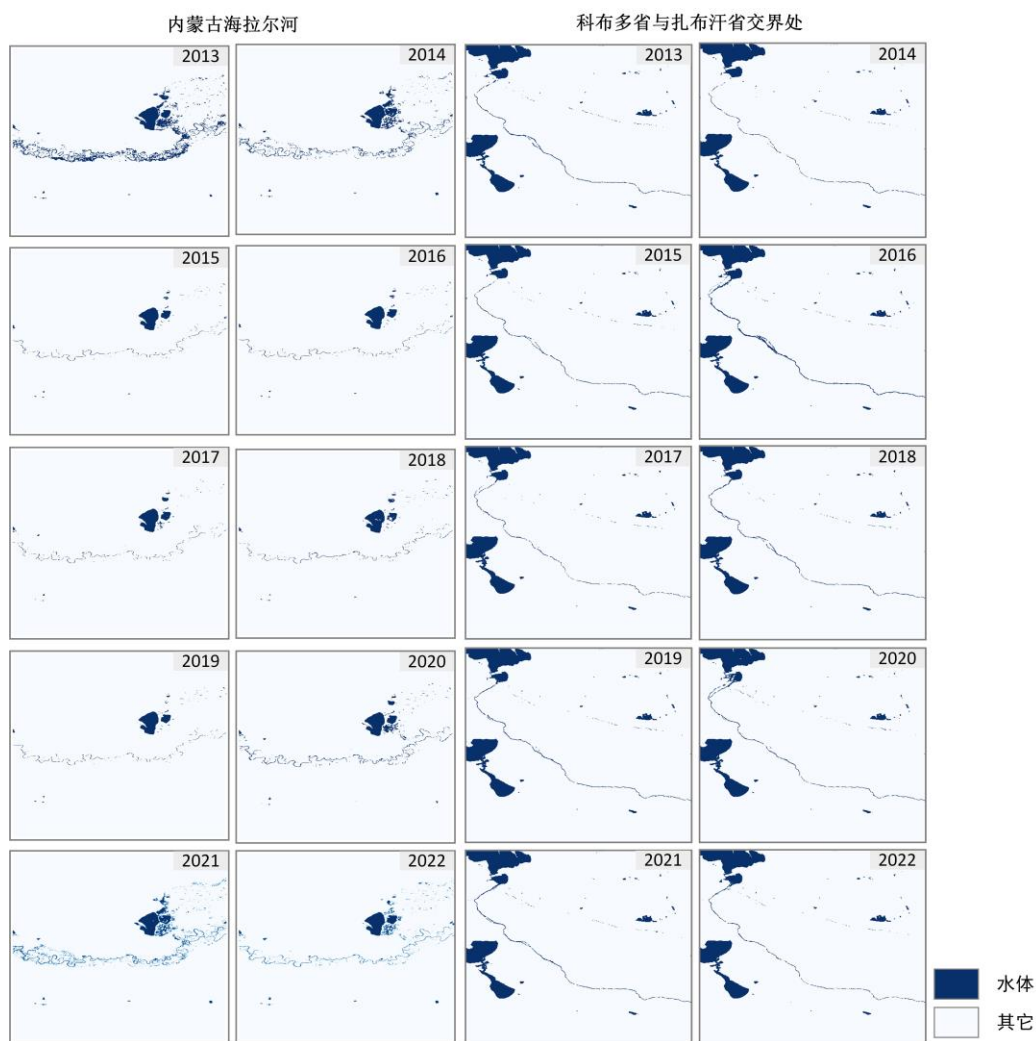


图 5 内蒙古海拉尔河以及蒙古国西北部科布多省与扎布汗省交界处水域时序变化图

Figure 5 Time series changes of Hailar River in Inner Mongolia and the water bodies at the junction of Khovd Province and Zabhan Province in the northwest of Mongolia

图 5 左右两侧分别展示了内蒙古海拉尔河以及蒙古国西北部科布多省与扎布汗省交界处水域的时序变化情况。海拉尔河全长 708.5 公里，流域面积 5.45 万平方公里，是滋养呼伦贝尔大草原的“母亲河”。海拉尔河的洪峰主要发生在 5 月融雪期以及 8 月夏雨期，2013 年夏季海拉尔河遭遇暴雨，导致水位上涨，洪水淹没面积较多。图 5 右侧的两大湖泊分别为哈尔湖和吉尔吉斯湖。该区域为蒙古国的大湖盆地，盆地内分布 300 多个湖泊，土地覆被类型以荒漠草原、沙漠以及裸地为主。

3 数据质量控制和评估

在 Google Earth 中，分别针对不同年份人工选择验证点。每年筛选 500 个验证样点，其中水体限制样点数量为 200 个，非水体为 300 个（涵盖蒙古高原的多种地类以及山体阴影区域等）。因此 10 年共有 5000 个样点，水体样点 2000 个，非水体 3000 个。将验证样点按点取值获得混淆矩阵，判读混淆矩阵如表 2 所示。经过验证，10 年分类精度均保持在 86% 以上，总体验证精度为 88.0%，10 年平均 kappa 系数为 0.75（图 6）。研究区地处干旱半干旱区域，辐射差异小、地类相似度较高，同时水资源分布及不均匀水体样本占比小，这是导致水体错分的主要原因。受制于 Landsat 影像分辨率的影响，部分宽度 ≤ 30 m 的细小河流存在漏提的问题。

表 2 混淆矩阵

Table 2 Confusion matrix

	水体	其他	水体	其他	水体	其他	水体	其他	水体	其他	总数
	2013		2014		2015		2016		2017		
水体	180	20	175	25	176	24	170	30	177	23	200
其他	33	267	35	265	35	265	40	260	38	262	300
总数	213	287	210	290	211	289	210	290	215	285	500
	2018		2019		2020		2021		2022		
水体	179	21	170	30	183	17	178	22	179	21	200
其他	39	261	34	266	33	267	41	259	38	262	300
总数	218	282	204	296	216	284	219	281	217	283	500

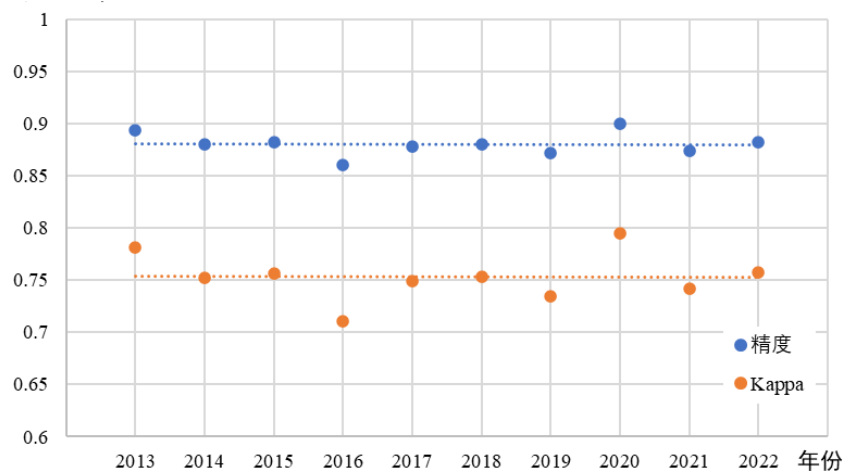


图 6 历年精度和 Kappa 系数

Figure 6 Historical precision and Kappa coefficient

4 数据使用方法和建议

本数据集可适用常规地理信息系统或遥感数据处理相关软件打开，如 QGIS、ArcGIS、ENVI 等。本数据集的空间分辨率为 30 m，可以直接用来表征蒙古高原近 10 年的水体变化情况，为蒙古高原资源、环境、生态、灾害等科学研究提供重要基础和本底数据；同时也可以结合栅格矢量转换、形态学分析等方法，为进一步揭示蒙古高原水文水资源格局与变化、气候变化区域影响与响应、与水相关的防灾减灾和畜牧业调控等研究提供支持。该模型方法也可为深度学习方法在长时序水文水资源环境监测中提供借鉴，也可为土地覆盖分类、地表参量反演等应用做支撑。本数据集也支持 Python 调用和查看，相关代码如下所示。

Python 代码：

```
import tifffile as tif
import matplotlib.pyplot as plt
water_img = tif.imread(r'XXX')#读取数据，XXX为文件路径
water2013 = water_img[0] #数据按照2013-2022每年存储为一个波段
water2014 = water_img[1]
#.....
water2022 = water_img[9]
#数据可视化
plt.imshow(water2013)
plt.show()
```

数据作者分工职责

李凯（1998—），男，江苏省南京市人，硕士，研究方向为大地测量学与测量工程。主要承担工作：模型构建与应用、数据集生产制作、文章撰写。

王卷乐（1976—），男，河南省洛阳市人，博士，研究员，研究方向为科学数据共享、地理信息系统与遥感应用研究。主要承担工作：总体设计、技术实施与成果集成、文章修改。

程文静（1998—），女，江西省景德镇市人，硕士，研究方向为中尺度气象学。主要承担工作：代码调试、模型训练。

洪梦梦（1999—），女，山东省菏泽市人，硕士，研究方向为土壤盐渍化。主要承担工作：模型部署、精度验证。

参考文献

[1] MCFEETERS S K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features[J]. International Journal of Remote Sensing, 1996, 17(7): 1425–1432. DOI:10.1080/01431169608948714.

[2] XU H Q. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery[J]. International Journal of Remote Sensing, 2006, 27(14): 3025–3033.

DOI:10.1080/01431160600589179.

- [3] FEYISA G L, MEILBY H, FENSHOLT R, et al. Automated Water Extraction Index: a new technique for surface water mapping using Landsat imagery[J]. *Remote Sensing of Environment*, 2014, 140: 23–35. DOI:10.1016/j.rse.2013.08.029.
- [4] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. IEEE, 2015: 3431–3440. DOI:10.1109/CVPR.2015.7298965.
- [5] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481–2495. DOI:10.1109/TPAMI.2016.2644615.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234–241. DOI:10.1007/978-3-319-24574-4_28.
- [7] CHEN Y, FAN R S, YANG X C, et al. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning[J]. *Water*, 2018, 10(5): 585. DOI: 10.3390/w10050585.
- [8] 姚锦一, 王卷乐. 若尔盖县弯曲河流精细提取研究[J]. *湿地科学*, 2020, 18(1): 47–53. DOI: 10.13248/j.cnki.wetlandsci.2020.01.007. [YAO J Y, WANG J L. Fine extraction of curved rivers in zoigê County[J]. *Wetland Science*, 2020, 18(1): 47–53. DOI: 10.13248/j.cnki.wetlandsci.2020.01.007.]
- [9] 姚锦一, 王卷乐, 严欣荣, 等. 基于深度神经网络的蒙古国色楞格河流域水体信息提取[J]. *地球信息科学学报*, 2022, 24(5): 1009–1017. DOI: 10.12082/dqxxkx.2022.210031. [YAO J Y, WANG J L, YAN X R, et al. Water information extraction of Selenga River Basin in Mongolia based on deep neural network[J]. *Journal of Geo-Information Science*, 2022, 24(5): 1009–1017. DOI: 10.12082/dqxxkx.2022.210031.]
- [10] LI K, WANG J L, YAO J Y. Effectiveness of machine learning methods for water segmentation with ROI as the label: a case study of the Tuul River in Mongolia[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 103: 102497. DOI:10.1016/j.jag.2021.102497.
- [11] LI L W, YAN Z, SHEN Q, et al. Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks[J]. *Remote Sensing*, 2019, 11(10): 1162. DOI:10.3390/rs11101162.
- [12] WENG L G, XU Y M, XIA M, et al. Water areas segmentation from remote sensing images using a separable residual SegNet network[J]. *ISPRS International Journal of Geo-Information*, 2020, 9(4): 256. DOI:10.3390/ijgi9040256.
- [13] LI J J, WANG C, XU L, et al. Multitemporal water extraction of Dongting Lake and Poyang Lake based on an automatic water extraction and dynamic monitoring framework[J]. *Remote Sensing*, 2021, 13(5): 865. DOI:10.3390/rs13050865.
- [14] GONG P, WANG J, YU L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data[J]. *International Journal of Remote Sensing*, 2013, 34(7): 2607–2654. DOI:10.1080/01431161.2012.748992.
- [15] GONG P, LIU H, ZHANG M N, 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

[16] ZHANG X, LIU L Y, CHEN X D, 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. DOI:10.5194/essd-13-2753-2021.

[17] Google. USGS Landsat 8 Level 2, Collection 2, Tier 2[DB/OL]. (2013–03–18) [2022–09–22]. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T2_L2.

[18] Google. NASADEM: NASA NASADEM Digital Elevation 30m [DB/OL]. (2013–02–11) [2022–09–22]. https://developers.google.com/earth-engine/datasets/catalog/NASA_NASADEM_HGT_001.

[19] LI K, WANG J L, CHENG W J, et al. Deep learning empowers the Google Earth Engine for automated water extraction in the Lake Baikal Basin[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102928. DOI:10.1016/j.jag.2022.102928.

论文引用格式

李凯, 王卷乐, 程文静, 等. 2013–2022 年蒙古高原逐年生长季地表水分布数据集[J/OL]. 中国科学数据, 2023, 8(1). (2023-03-02). DOI: 10.11922/11-6035.csd.2022.0080.zh.

数据引用格式

李凯, 王卷乐, 程文静, 等. 2013–2022 年蒙古高原逐年生长季地表水分布数据集[DS/OL]. Science Data Bank, 2022. (2022-09-23). DOI: 10.57760/sciencedb.j00001.00665.

A dataset of annual surface water distribution in the growing season on the Mongolia Plateau from 2013 to 2022

LI Kai^{1,2}, WANG Juanle^{2*}, CHENG Wenjing³, HONG Mengmeng^{2,4}

1. School of Geosciences & Surveying Engineering, China University of Mining & Technology-Beijing, Beijing 100083, P. R. China

2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, P. R. China

3. Chinese Academy of Meteorological Sciences, Beijing 100081, P. R. China

4. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo 255049, P. R. China

*Email: wangjl@igsrr.ac.cn

Abstract: Mongolia Plateau is located in arid and semi-arid areas, and hydrology and water resources are important constraints for the development of its resources and environment. Grasping the temporal and spatial distribution of water bodies on the Mongolian Plateau is of great significance for indicating the temporal and spatial characteristics of water resources and the water environment and their impacts on and responses to regional climate change as well as disaster prevention and reduction. However, as the vast Plateau spans both

China and Mongolia, it is a great challenge to accurately and automatically obtain large-scale and long time series water bodies at the basin scale. In this research, we adopted the method of combining local deep learning training and Google Earth Engine (GEE) distributed computing to endow GEE with deep learning computing capabilities so that GEE could rapidly and automatically deploy deep learning models. Based on this, we obtained the distribution of surface water in the growing season of the Mongolia Plateau from 2013 to 2022 with a spatial resolution of 30 meters. 5,000 verification points were manually selected, and the overall verification rate was 88.0%. The dataset is in the form of TIFF grid, containing 28 tile images of with $5^{\circ} \times 5^{\circ} \times 10$ years, with a data volume of 339 MB (88.1 MB compressed, 189 GB in RAW). The data volume in the raw format is 189 GB. With the method used in this dataset, users can automatically and efficiently map water bodies in the cloud platform, which makes it possible to automatically and efficiently process large-scale and long-time series water bodies in arid and semi-arid regions. This is a valuable dataset for application and promotion.

Keywords: Mongolian Plateau; surface water distribution; hydrology and water resources; Google Earth Engine; deep learning

Dataset Profile

Title	A dataset of annual surface water distribution in the growing season on the Mongolia Plateau from 2013 to 2022
Data corresponding author	WANG Juanle (wangjl@igsnr.ac.cn)
Data author(s)	LI Kai, WANG Juanle, CHENG Wenjing, HONG Mengmeng
Time range	2013–2022
Geographical scope	Mongolian Plateau (Inner Mongolia, China and Mongolia)
Spatial resolution	30 m
Data volume	399 MB, 88.1 MB compressed (189 GB in RAW)
Data format	.tif
Data service system	http://dx.doi.org/10.57760/sciencedb.j00001.00665
Source(s) of funding	National Natural Science Foundation of China (41971385, 32161143025); Construction Project of the China Knowledge Center for Engineering Sciences and Technology (CKCEST-2022-1-41); Key Project of Innovation LREIS (KPI006)
Dataset composition	The longitude and latitude are stored as a .tif file every 5 °. There are 28 .tif files in the compressed package. Each .tif file has 10 channels (2013–2022). The files are named in terms of MP_water_ExxxNyy.tif.