专题 黄河流域生态保护与高质量发展

1986–2023 年黄河流域年度地表水分布数据集

张兴旺[1]，刘戈[2]，宋辰[2]，唐朝阳[2]，靳宁[3]，夏浩铭[2,4*]

1. 河南大学生命科学学院，河南开封 475004
2. 河南大学地理与环境学院，河南开封 475004
3. 山西能源学院资源与环境工程系，山西晋中 030600
4. 黄河中下游数字地理技术教育部重点实验室（河南大学），河南开封 475004

摘要：黄河流域地表水分布的精准制图对于水资源管理、生态环境保护以及防洪减灾等方面具有重要意义。目前，黄河流域仍缺乏精准的长时间序列的地表水分布数据集。为了解决这一问题，本研究基于 Google Earth Engine（GEE）云计算平台，融合多源陆地卫星（Landsat-5/7/8/9）遥感影像，构建了黄河流域地表水提取算法，生产了 1986–2023 年黄河流域地表水分布数据集，包括地表水频率数据（WBF）和地表水类别数据（SWC）。通过对水体提取结果的精度进行评估，总体精度 OA 达到了 97.60%，Kappa 系数为 0.95。同时在黄河流域将 SWC 数据和全球地表水数据集 GSW 进行了对比，SWC 数据的表现在部分地区明显优于 GSW 数据。本研究生产的数据集可以帮助政府和公众应对黄河流域日益增长的用水需求和水安全挑战。

关键词：黄河流域；水体提取；地表水频率；地表水类别

数据库（集）基本信息简介

数据库（集）名称	1986–2023 年黄河流域年度地表水分布数据集
数据作者	夏浩铭，张兴旺
数据通信作者	夏浩铭（xiahm@vip.henu.edu.cn）
数据时间范围	1986–2023年
地理区域	黄河流域地理范围为（32°10′N–41°50′N，95°53′E–119°05′E），横跨青海、甘肃、四川、宁夏、内蒙古、陕西、山西、河南、山东九省。
空间分辨率	30 m
数据量	18.6 GB
数据格式	*.tif
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00796 http://www.ncdc.ac.cn/portal/metadata/9fb657ef-f01b-4adc-a890-1da70c75a194
基金项目	国家自然科学基金（42271118、32130066）、山西省水利科学技术研究与推广项目（2023GM44）、山西省基础研究计划自然科学研究面上项目（202203021221231）、河南省科技攻关计划项目（242102321160）、河南省科技重点项目（212102310019）。

ISSN 2096-2223
CN 11-6035/N

文献 CSTR：
32001.14.11-6035.ncdc.2024.0010.zh

文献 DOI：
10.11922/11-6035.ncdc.2024.0010.zh

数据 DOI：
10.57760/sciencedb.j00001.00796

文献分类：地球科学

收稿日期：2024-05-06
开放同评：2024-05-11
录用日期：2024-08-29
发表日期：2024-09-29

* 论文通信作者
夏浩铭：xiahm@vip.henu.edu.cn

数据库（集）组成	本数据集共包括 2 个数据，其中：（1）地表水频率数据（WBF）代表年度地表水出现的频率（波段值：0–1），共计 38 幅影像，数据量 9.07 GB；（2）年度季节性水体、永久性水体、最大水体三个地表水类别数据（SWC）：非水体（波段值：0）、水体（波段值：1），共计 114 幅影像，数据量 9.53 GB。

引　言

地表水主要由河流、湖泊、池塘、水库和沼泽组成，在维持农业、水产养殖、工业生产和陆地生态系统中发挥着关键作用[1-2]。黄河是中国第二大河流，被称为中华民族的母亲河，占全国径流量的 2.2%，为农业提供 15%的灌溉用水，满足 12%的人口用水需求[3-4]。在全球气候变化和人类活动日益频繁的背景下，黄河流域地表水体分布对人类社会和生态安全具有重要影响[5]。因此，快速、准确、及时地监测黄河流域地表水的分布至关重要。

目前，已有的黄河流域地表水数据集包括：基于多源卫星观测值建立的 1993–2007 年 0.25° 的首个全球陆地表面开放水域月度数据[6]；基于 Landsat 数据的 2000 年全球 30 m 内陆水体数据集 GIW[7]；基于 Sentinel-2 数据的近实时全球 10 m 土地利用数据[8]；基于 Landsat、中国环境卫星等数据的 2000/2010/2020 年 30 m 全球土地利用数据（GlobeLand 30）[9-10]；基于航天飞机雷达地形任务（SRTM）数据和 MODIS 数据创建的全球 250 m 水掩膜数据[11]。然而，这些数据集存在以下局限性：（1）数据集空间分辨率较低，无法识别面积较小的地表水体[12]；（2）时间序列短，即数据集被限制在一年或几年的时间内，使得捕捉地表水的长期变化具有挑战性。欧盟委员会可持续资源理事会联合研究中心（JRC）发布了全球地表水数据集（GSWv1.4）[13]，该数据集制作了 1984–2021 年 30 m 的地表水数据，将地表水分类为永久和季节性地表水，但其分类结果在黄河流域的表现存在较大的不确定性。武汉大学黄昕等[14]制作了 1990–2019 年中国 30 m 年度土地覆盖数据集，包括耕地、草地、森林、水体等 9 个类别；但是，水类别没有进一步划分，在识别季节性水体方面的表现低于 GSW，并且其总体精度（79%）比其他水体数据集的总体精度低。总之，长时间序列且精准的黄河流域高分辨率地表水分布数据集较为缺乏。

为了解决这一问题，本研究基于 GEE 云计算平台，利用 HAND 地形指数改进了 Zou 等人提出的基于决策树的水体提取算法，使用 Landsat-5/7/8/9 数据构建了黄河流域 1986–2023 年的 30 m 地表水分布数据集。本数据集将指导黄河流域水资源利用，支持水资源分配政策，促进水资源的可持续发展和有效管理。

1　数据采集和处理方法

1.1　研究区概况

黄河是中国的第二长河，源头位于青海省的巴颜喀拉山脉，流经青海、甘肃、四川、宁夏、内蒙古、陕西、山西、河南和山东九省，最终在山东省东营市注入渤海。黄河全长约 5,464 km，流域面积约达 752,443 km^2，占中国国土面积的 8%，海拔范围为 0-6300 m[15]（图 1）。黄河流域的地表水情况复杂多样，从干旱半干旱的西北地区到湿润的东部地区，地表水的分布和变化特征具有显著

的地域差异。在上游的青藏高原地区，冰雪融水是主要的水源，而中游和下游地区的地表水体分布受季节性降水和人类活动的影响较大[16]。本文选择了扎陵湖、龙门镇河段、小浪底水库三个区域展示数据样本。

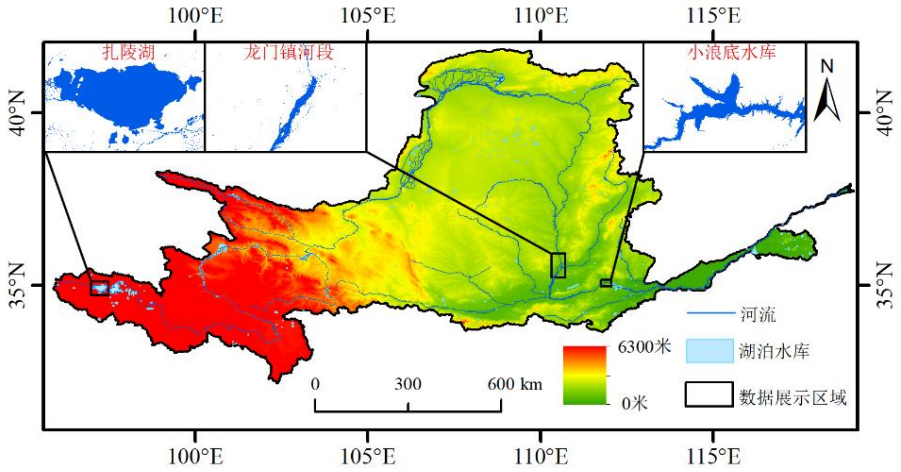


图 1　研究区位置及高程

Figure 1　Location and elevation of the study area

1.2　数据源

1.2.1　Landsat-5/7/8/9 数据

本文使用 1986–2023 年 Landsat-5/7/8/9 的影像数据提取研究区地表水，共计 78298 幅影像被用于地表水提取。图 2 展示了每年的影像数量，由于加入了 Landsat-9 影像数据，在 2022 年和 2023 年超过 4000 幅影像被用于地表水提取，极大地提高了地表水监测的频率。图 3 展示了每年不同月份影像数量百分比，除了 1986 年，其他年份影像数量分布均匀。图 4 展示了影像数量的空间分布，陆地卫星在影像范围重叠区域有超过 800 次观测，在影像范围未重叠区域观测次数在 400–800 次之间。使用的 Landsat-5/7/8/9 数据集在 GEE 平台中的名称分别为"LANDSAT/LT05/C02/T1_L2"、"LANDSAT/LE07/C02/T1_L2"、"LANDSAT/LC08/C02/T1_L2"、LANDSAT/LC09/C02/T1_L2。

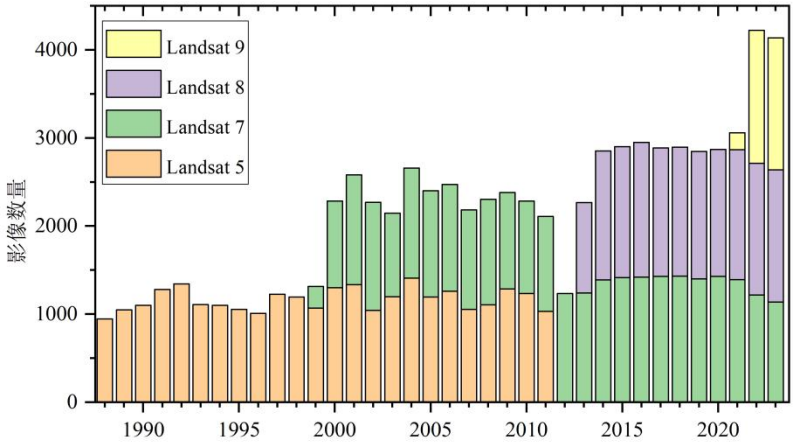


图 2　研究区每年的陆地卫星影像数量

Figure 2　The number of Landsat images in the study area per year

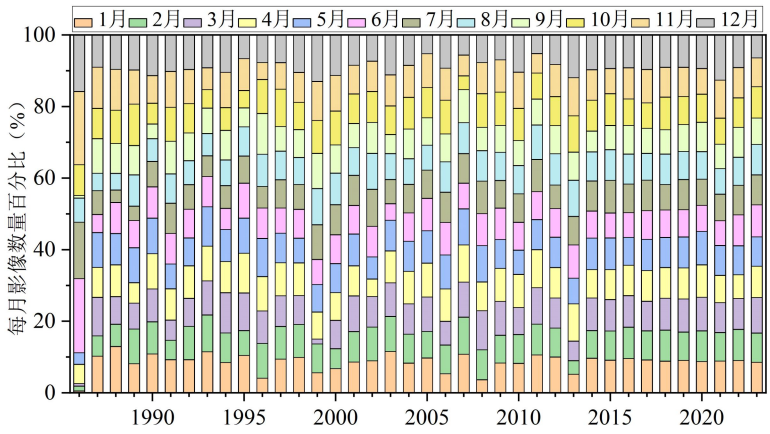


图 3　研究区每年的陆地卫星影像数量

Figure 3　Percentage of the image number in different months of the year in the study area

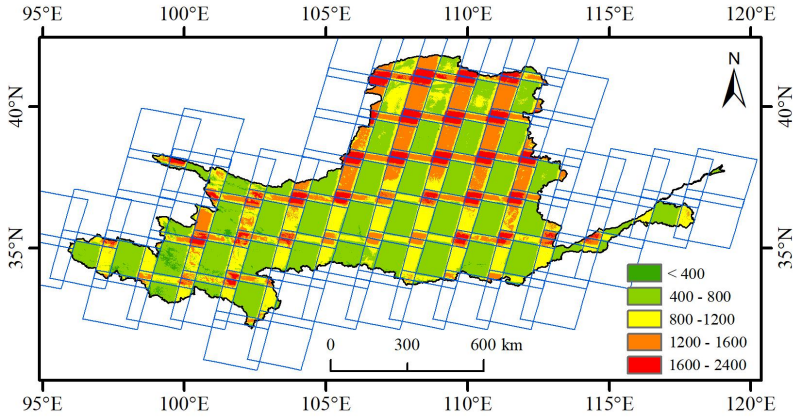


图 4　研究区影像数量空间分布

Figure 4　Spatial distribution of the number of images in the study area

1.2.2　验证数据

本文在整个黄河流域，利用多时相的谷歌地球影像和 Sentinal-2 影像，采用分层采样法，生成 4249 个地面真实样本点，其中包括 2027 个地表水样本点和 2222 个非地表水样本点（图 5）。

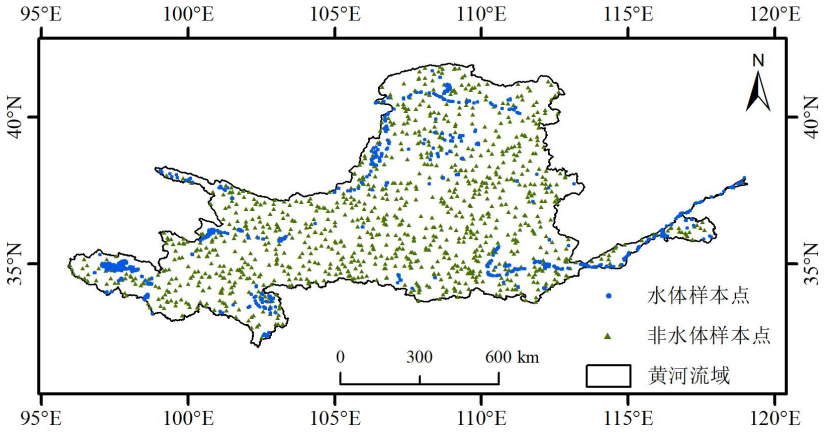


图 5　验证样本点空间分布

Figure 5　Spatial distribution validation of sample points

1.2.3 GSW 数据集

本文使用了全球地表水数据集（GSW）的年度地表水分类数据（JRC Yearly Water Classification History, v1.4），该数据集基于 1984 年 3 月 16 日至 2021 年 12 月 31 日期间获取的 Landsat-5、7 和 8 卫星的 4,716,475 幅影像生成的。年度地表水分类数据将每个像素分类为季节性水体或永久水体或非水体。在 GEE 平台中该数据的名称为"JRC/GSW1_4/YearlyHistory"。

1.2.4 HAND 数据

HAND 指数（最近水系网络上方高度）是根据航天飞机雷达地形任务（SRTM）的 DEM 数据和水系网络信息开发的一种新地形指数，代表相对于最近水系网络的高程值[17-18]。HAND 指数用于消除地表水检测中出现的山体阴影等潜在错误[19]。

1.3 数据处理方法

1.3.1 地表水提取方法

水体指数（mNDWI，公式（1）被广泛用于地表水检测，但在像元是水体与其他土地覆盖类别混合时，尤其是当像元包含植被和水体时，存在一定的误差。为了减少植被对水体识别的影响，可以利用水体指数（mNDWI）与归一化植被指数（NDVI，公式（2）和增强植被指数（EVI，公式（3）之间的关系来提取地表水。以往的研究主要是基于水体和植被指数的算法进行水体提取[1,5,20-21]。本文采用 Zou 等人提出的基于决策树的水体提取方法[16-22]，当像元的光谱指数满足（mNDWI-NDVI>0 且 EVI<0.1）或（mNDWI-EVI>0 且 EVI<0.1）时，该像元为水体，否则为非水体公式（4）。

$$mNDWI = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}} \tag{1}$$

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{2}$$

$$EVI = 2.5 \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + 6 \times \rho_{red} - 7.5 \times \rho_{blue} + 1} \tag{3}$$

$$Water = \begin{cases} 1, EVI < 0.1 \ and \ (mNDWI > EVI \ or \ mNDWI > NDVI) \\ 0, \text{非水体} \end{cases} \tag{4}$$

其中，ρ_{green}是绿波段地表反射率，ρ_{swir1}代表短波红外 1 波段，ρ_{nir}代表近红外波段，ρ_{red}代表红波段，ρ_{blue}代表蓝波段。

1.3.2 水体频率计算方法

年度水体频率（WBF）计算公式如下[1,5,23]：

$$WBF = \frac{\sum N_{Water}}{\sum N_{Good}} \tag{5}$$

其中，$\sum N_{Water}$是水体像元的年总观测次数，$\sum N_{Good}$是高质量像元（影像去云后的像元）的年总观测次数，WBF 代表水体观测次数在高质量观测次数中的比例，其值在 0–1 范围内。

地表水类别数据（SWC）的分类标准为季节性水体（0.25≤WBF<0.75）和永久性水体（WBF≥0.75），公式（6）所示。最大水体定义为（WBF≥0.25），即季节性水体和永久性水体的和。

$$SWC = \begin{cases} 0(\text{非水体}), & WBF < 0.25 \\ 1(\text{季节性水体}), & 0.25 \le \text{WBF} < 0.75 \\ 2(\text{永久性水体}), & \text{WBF} \ge 0.75 \end{cases} \quad (6)$$

1.3.3 水体频率数据矫正方法

HAND 是基于水系网络的地形指数。为了减少地形对水体数据集的影响，本研究只保留距离最近水系相对高度 20 m 范围内的数据。这种方法有助于避免高海拔地区的错误分类，并最大限度地减少地形对地表水监测的影响。

1.3.4 精度评价方法

使用混淆矩阵对水体提取结果进行精度验证。计算了生产者精度（Production Accuracy，PA）、用户精度（User Accuracy，UA）、总体精度（Overall Accuracy，OA）和 Kappa 系数（Kappa）四个指标。

1.3.5 数据生产流程概述

在 GEE 平台中基于 JavaScript 语言对 Landsat 5/7/8/9 数据进行预处理，提取地表水后通过 HAND 地形指数对地表水范围进行了矫正。最后计算了年度地表水频率，得到了 1986–2023 年的年度地表水频率数据 WBF。通过采集的验证样本点对地表水提取结果进行了精度验证。另外，对年度地表水频率数据 WBF 进行分类，得到了年度地表水分类数据 SWC；并且将该数据和 JRC 的年度地表水分类数据进行了对比。图 6 展示了本文的数据预处理、方法构建、精度验证、数据集生成等生产流程。

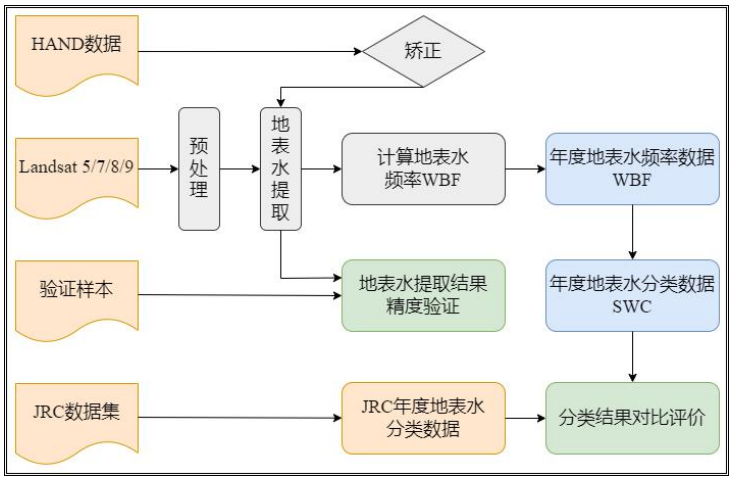


图 6　数据生产流程图

Figure 6　Data production flowchart

2　数据样本描述

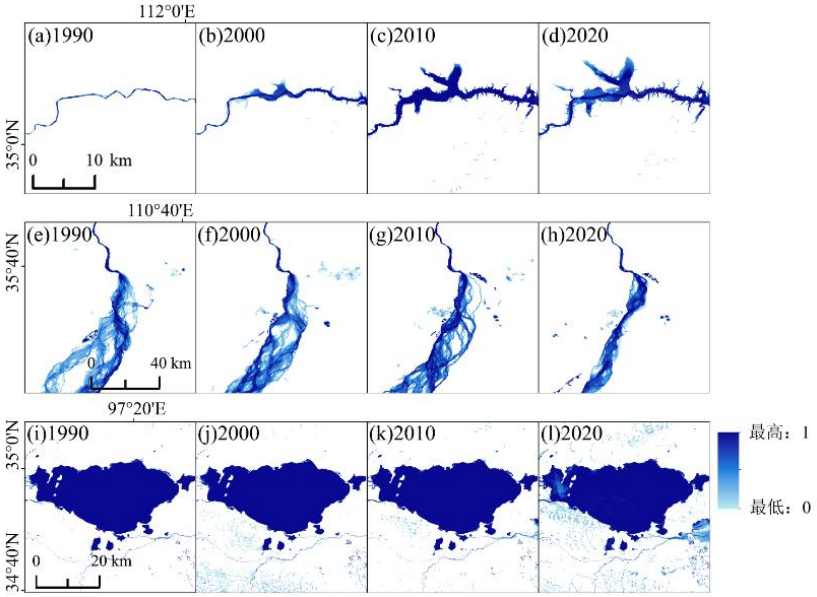
2.1　数据组成

本文生产的数据集包括 1986–2023 年黄河流域年度地表水频率数据（WBF）和地表水类别（永久性、季节性、最大水体）数据（SWC）。两套数据均是栅格数据，采用 WGS-84 地理坐标系，空间

分辨率是 30 m，文件类型均为*.tif 文件，文件命名格式为 YYYY.tif （YYYY 为对应年份）。其中地表水频率数据波段值为 0–1（浮点型），地表水类别数据波段值为 0（非水体）、1（水体）。

2.2 数据样本

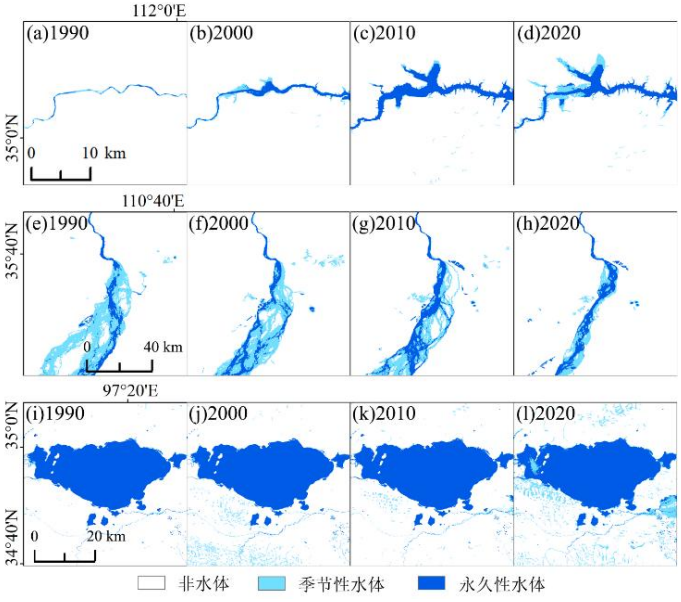
以小浪底水库（图 7、8a–d）、龙门镇河段（图 7、8e–h）、扎陵湖区域（图 7、8i–l）为例，分别对 1990 年、2000 年、2010 年、2020 年的地表水频率数据和地表水类别数据进行了展示。图 7 是年度地表水频率数据，图 8 是年度地表水类别数据（分为非水体、季节性水体、永久性水体）。



a–d 是小浪底水库部分区域，e–h 是龙门镇河段，i–l 是扎陵湖区域

图 7 地表水频率数据

Figure 7 Surface water body frequency data



a–d 是小浪底水库部分区域，e–h 是龙门镇河段，i–l 是扎陵湖区域

图 8 地表水类别数据

Figure 8 Surface water category data

3 数据质量控制和评估

本研究利用在谷歌地球影像和 Sentinel-2 影像上采集的 2023 年地面真实样本点，包括 2027 个水体样本点和 2222 个非水体样本点，对水体数据集精度进行了直接评价（表 1），其中，非水体和水体提取结果的用户精度分别是 96.40%和 98.91%，生产者精度分别是 98.98%和 96.16%，总体精度 OA 为 97.60%，Kappa 系数为 0.95。因此，本研究使用的水体提取方法在黄河流域表现出优异的性能，利用该方法生产的数据集具有高度的可靠性。为了进一步评估本研究数据集的质量，搜集了现有的几种水体数据集和本研究数据集进行对比分析，对比的信息包括地理范围、时间范围、空间分辨率、总体精度、数据来源（表 2）。本研究生产的地表水数据总体精度可达到 98%。

表 1　水体提取结果和地面真实样本的混淆矩阵

Table 1　Confusion matrix between water body extraction results and ground truth samples

		地面真实样本/个			
		非水体	水体	总计	用户精度（%）
提取结果	非水体	2142	80	2222	96.40%
	水体	22	2005	2027	98.91%
	总计	2164	2085	4249	OA=97.60%
	生产者精度（%）	98.98%	96.16%		Kappa = 0.95

表 2　本研究生产的数据集与现有水体数据集的对比结果

Table 2　Results of comparison between the dataset produced in this study and existing water body datasets

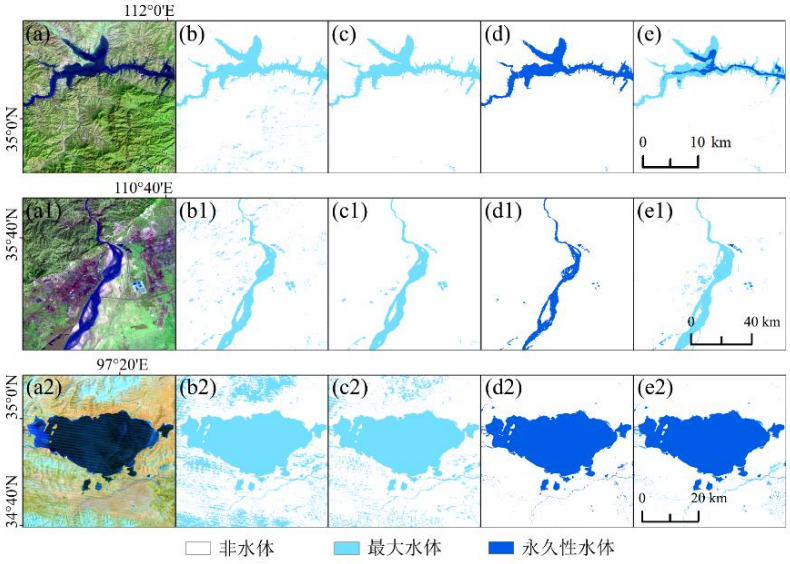
数据集名称	地理范围	时间范围	空间分辨率	总体精度%	数据来源
陆地表面开放水域月度数据[6]	全球	1993–2007 年	0.25°	无	AVHRR、SSM/I、ERS
GIW 内陆水体数据集[7]	全球	2000 年	30m	95%	Landsat
Dynamic World 近实时土地利用数据集[8]	全球	2015 年至今	10m	99%	Sentinel-2
Global Land 30 土地利用数据[9-10]	全球	2000 年、2010 年、2020 年	30m	96%	Landsat、中国环境一号
水掩膜栅格数据[11]	全球	2005 年	250m	无	MODIS
GSW v1.4 地表水数据[13]	全球	1984–2021 年	30m	99%	Landsat
CLCD 土地利用数据[14]	中国	1990–2019 年	30m	79%	Landsat
本研究生产的地表水数据	黄河流域	1986–2023 年	30m	98%	Landsat

4 数据价值

现有数据集中只有 GSW v1.4 地表水数据和 CLCD 土地利用数据同时具备长时间序列和高空间分辨率。Dynamic World 近实时土地利用数据集在精度和空间分辨率方面都具备极高的标准，但是

该数据集只存在 2015 年之后的数据，难以满足长时间序列研究的需求。精度方面，本研究生产的地表水数据总体精度（98%）略低于 GSW v1.4 地表水数据（99%），明显高于 CLCD 土地利用数据（79%）。在黄河流域，我们将本研究的数据集和 GSW 进行了对比，发现 GSW 在部分区域有明显的错误分类情况，本研究数据集的分类结果更符合实际水体类别范围（图 9）。总体来说，在综合考虑所有数据集的时间序列长度、空间分辨率、总体精度、在黄河流域的适用性几种需求时，本研究生产的数据集是最优的。

图 9 为 2021 年本研究生产的地表水类别数据 SWC 和 GSW 数据集的地表水类别数据的对比结果。图 9（a）和图 9（b）对比可以看到 HAND 数据集在去除山体阴影方面展现出良好的性能。图 9（c、d）和图 9（e）对比可以看到，GSW 数据集在某些区域会大幅度低估永久性水体，尤其是图 9（e1）河道中没有永久性水体这是明显错误的。在最大水体范围的提取结果中，本研究的水体类别数据集和 GSW 数据集范围几乎一致，具有高度一致性。



a–e 是小浪底水库部分区域，a1–e1 是龙门镇河段，a2–e2 是扎陵湖区域，a、a1、a2 为 Landsat 影像的短波红外 2、近红外、绿波段假彩色组合结果；b、b1、b2 为未使用 HAND 数据掩膜的最大水体结果；c、c1、c2 为使用 HAND 数据掩膜后的最大水体结果；d、d1、d2 为本研究的永久性水体结果；e、e1、e2 为 GSW 数据集的最大水体和永久性水体结果。

图 9　地表水类别数据 SWC 和 GSW 数据集对比

Figure 9　Comparison between the Surface Water Category (SWC) dataset and the Global Surface Water (GSW) dataset

5　数据使用方法和建议

本研究生产的 1986–2023 年黄河流域年度地表水分布数据集是公开发布的，包括地表水频率数据和地表水类别（永久性和季节性水体）数据，具有 30 m 的空间分辨率。数据集可在 ENVI、Arc GIS、QGIS 等遥感软件中进行查看、读取、统计分析。黄河流域地表水分布数据集可帮助政府和学者深入

理解该地区水文地质特征和水文过程，促进相关学科领域的发展。同时对于保障黄河流域的水资源安全、生态环境健康和社会经济可持续发展提供决策支持。

作者分工职责

张兴旺（1998—），男，甘肃省庆阳市人，硕士，研究方向为水体遥感。主要承担工作：算法构建、数据处理、数据质量控制和论文撰写。

刘戈（2000—），男，河南省洛阳市人，硕士，研究方向为全球变化生态学。主要承担工作：数据采集、部分数据处理。

宋辰（2001—），男，云南省玉溪市人，硕士，研究方向为植被遥感。主要承担工作：数据采集、部分数据处理。

唐朝阳（2001—），男，河南省濮阳市人，硕士，研究方向为植被遥感。主要承担工作：数据采集、部分数据处理。

靳宁（1983—），男，山西襄汾人，博士，研究方向为生态环境信息化。主要承担工作：部分数据处理。

夏浩铭（1983—），男，河南省永城市人，博士，副教授，博士生导师，研究方向为定量遥感及其综合应用。主要承担工作：总体工作部署、整体思路设计、算法构建、数据处理和论文修改。

参考文献

[1] WANG X, XIAO X, ZOU Z, et al. Gainers and losers of surface and terrestrial water resources in China during 1989-2016 [J]. Nature Communications, 2020, 11(1): 3471. DOI:10.1038/s41467-020-17103-w.

[2] 李欢, 万玮, 冀锐, 等. 中国卫星遥感地表水资源监测能力分析与展望[J]. 遥感学报,2023,27(07):1554-73. [LI H, WAN W, JI R, et al. Inspects and prospects of satellite remote sensing monitoring ability for land surface water in China[J]. Journal of Remote Sensing, 2023, 27(07): 1554-73.]

[3] MIAO C, NI J, BORTHWICK A G L. Recent changes of water discharge and sediment load in the Yellow River basin, China [J]. Progress in Physical Geography: Earth and Environment, 2010, 34(4): 541-61. DOI:10.1177/0309133310369434.

[4] CHEN Y-P, FU B-J, ZHAO Y, et al. Sustainable development in the Yellow River Basin: Issues and strategies [J]. Journal of Cleaner Production, 2020, 263: 121223. DOI:10.1016/j.jclepro.2020.121223.

[5] ZOU Z, XIAO X, DONG J, et al. Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016 [J]. Proceedings of the National Academy of Sciences, 2018, 115(15): 3810-5. DOI:10.1073/pnas.1719275115.

[6] PRIGENT C, PAPA F, AIRES F, et al. Changes in land surface water dynamics since the 1990s and relation to population pressure [J]. Geophysical Research Letters, 2012, 39(8). DOI:10.1029/2012GL051276.

[7] FENG M, SEXTON J O, CHANNAN S, et al. A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic–spectral classification algorithm [J]. International Journal of Digital Earth, 2016, 9(2): 113-33. DOI:10.1080/17538947.2015.1026420.

[8] 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). 10.1038/s41597-022-01307-4.

[9] LIAO A, CHEN L, CHEN J, et al. High-resolution remote sensing mapping of global land water [J]. Science China Earth Sciences, 2014, 57: 2305-16. DOI:10.1007/s11430-014-4918-0.

[10] CHEN J, CHEN J. GlobeLand30: Operational global land cover mapping and big-data analysis [J]. Science China Earth Sciences, 2018, 61(10): 1533-4. DOI:10.1007/s11430-018-9255-3.

[11] CARROLL M L, TOWNSHEND J R, DIMICELI C M, et al. A new global raster water mask at 250 m resolution [J]. International Journal of Digital Earth, 2009, 2(4): 291-308. DOI:10.1080/17538940902951401.

[12] GAO H, BIRKETT C, LETTENMAIER D P. Global monitoring of large reservoir storage from satellite remote sensing [J]. Water Resources Research, 2012, 48, W09504. DOI:10.1029/2012WR012063.

[13] PEKEL J F, COTTAM A, GORELICK N, et al. High-resolution mapping of global surface water and its long-term changes [J]. Nature, 2016, 540(7633): 418-422. DOI:10.1038/nature20584.

[14] YANG J, HUANG X. 30 m annual land cover and its dynamics in China from 1990 to 2019 [J]. Earth System Science Data Discussions, 2021, 2021: 1-29. DOI:10.5194/essd-13-3907-2021.

[15] 鲍怡, 胡春春, 王喆, 等. 近十年黄河流域地表水面积时空变化与影响因素分析[J]. 水文, 2024, 27(07): 1-12. [BAO Y, HU C C, WANG Z, et al. Spatial-Temporal Changes and Influencing Factors of Surface Water in the Yellow River Basin in Recent Ten Years[J]. Journal of China Hydrology, 2024, 27(07): 1-12.]

[16] 王瑞萌. 黄河流域及典型区域地表水时空变化及驱动因素研究 [D]. 开封: 河南大学, 2022. [WANG R M. Temporal and Spatial Variation and Driving Factors of Surface Water in the Yellow River Basin and Typical Regions[D]. Kaifeng: Henan University, 2022.]

[17] RENNó C D, NOBRE A D, CUARTAS L A, et al. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia [J]. Remote Sensing of Environment, 2008, 112(9): 3469-81. DOI:10.1016/j.rse.2008.03.018.

[18] NOBRE A D, CUARTAS L A, HODNETT M, et al. Height Above the Nearest Drainage–a hydrologically relevant new terrain model[J]. Journal of Hydrology, 2011, 404(1/2): 13–29. DOI: 10.1016/j.jhydrol.2011.03.051.

[19] BIORESITA F, PUISSANT A, STUMPF A, et al. A method for automatic and rapid mapping of water surfaces from Sentinel-1 imagery [J]. Remote Sensing, 2018, 10(2): 217. DOI:10.3390/rs10020217.

[20] ZHOU Y, DONG J, XIAO X, et al. Continuous monitoring of lake dynamics on the Mongolian Plateau using all available Landsat imagery and Google Earth Engine [J]. Science of the Total Environment, 2019, 689: 366-80. DOI:10.1016/j.scitotenv.2019.06.341.

[21] ZOU Z, DONG J, MENARGUEZ M A, et al. Continued decrease of open surface water body area in Oklahoma during 1984-2015 [J]. Science of the Total Environment, 2017, 595: 451-60. DOI:10.1016/j.scitotenv.2017.03.259.

[22] ZHOU H, LIU S, HU S, et al. Retrieving dynamics of the surface water extent in the upper reach of Yellow River [J]. Science of the Total Environment, 2021, 800: 149348. DOI:10.1016/j.scitotenv.2021.149348.

[23] XIA H, ZHAO J, QIN Y, et al. Changes in water surface area during 1989–2017 in the Huai River Basin using Landsat data and Google earth engine [J]. Remote Sensing, 2019, 11(15): 1824. DOI:10.3390/rs11151824.

论文引用格式

张兴旺, 刘戈, 宋辰. 等. 1986–2023 年黄河流域年度地表水分布数据集[J/OL]. 中国科学数据, 2024, 9(3). (2024-09-29). DOI: 10.11922/11-6035.ncdc.2024.0010.zh.

数据引用格式

夏浩铭, 张兴旺. 1986–2023 年黄河流域年度地表水分布数据集[DS/OL]. V1. Science Data Bank, 2024. (2024-09-11). DO: 10.57760/sciencedb.j00001.00796.

A dataset of annual surface water distribution in the Yellow River Basin (1986–2023)

ZHANG Xingwang[1], LIU Ge[2], SONG Chen[2], TANG Zhaoyang[2], JIN Ning[3], XIA Haoming[2,4*]

1. The College of Life Sciences, Henan University, Kaifeng 475004, P.R. China
2. The College of Geography and Environmental Science, Henan University, Kaifeng 475004, P.R. China
3. The Department of Resources and Environmental Engineering, Shanxi University of Energy, Jinzhong 030600, P.R. China
4. The Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education, Kaifeng 475004, P.R. China

*Email: xiahm@vip.henu.edu.cn

Abstract: Accurate mapping of surface water distribution in the Yellow River Basin is of great significance for effective water resource management, ecological environment protection, and flood prevention. However, there is a lack of precise long-term dataset of surface water distribution currently in the Yellow River Basin. To address this issue, using the Google Earth Engine (GEE) cloud computing platform, we integrated multi-source Landsat satellite (Landsat-5/7/8/9) remote sensing imagery to develop a surface water extraction algorithm for the Yellow River Basin in this study. Based on this algorithm, we

further produced a dataset of annual surface water distribution in the Yellow River Basin from 1986 to 2023, including Surface Water Frequency (WBF) data and Surface Water Category (SWC) data. The accuracy assessment of the water extraction results showed an overall accuracy (OA) of 97.60% and a Kappa coefficient of 0.95. Additionally, when compared with the Global Surface Water dataset (GSW) in the Yellow River Basin, the SWC data demonstrated superior performance in certain areas. The dataset produced in this study can assist the government and the public in addressing the growing water demand and water security challenges in the Yellow River Basin.

Keywords: Yellow River Basin; water body extraction; surface water frequency; surface water class

Dataset Profile

Title	A dataset of annual surface water distribution in the Yellow River Basin (1986–2023)
Data authors	XIA Haoming, ZHANG Xingwang
Data corresponding author	XIA Haoming (xiahm@vip.henu.edu.cn)
Time range	1986–2023
Geographical scope	The Yellow River Basin (32°10′N–41°50′N and from 95°53′E–119°05′E), encompassing nine provinces (i.e. Qinghai, Gansu, Sichuan, Ningxia, Inner Mongolia, Shaanxi, Shanxi, Henan, and Shandong)
Spatial resolution	30 m
Data volume	18.6 GB
Data format	*.tif
Data service system	<https://doi.org/10.57760/sciencedb.j00001.00796> <http://www.ncdc.ac.cn/portal/metadata/9fb657ef-f01b-4adc-a890-1da70c75a194>
Sources of funding	The National Natural Science Foundation of China (42271118,32130066), Shanxi Province Water Science and Technology Research and Promotion Project (2023GM44), Natural Science Research General Program of Shanxi Province Basic Research Project (202203021221231), Henan Provincial Science and Technology Research Project (242102321160), Henan Province Science and Technology Key Project (212102310019).
Dataset composition	The dataset comprises two types of data: (1) Surface Water Body Frequency (WBF) Data, which indicates the annual frequency of surface water occurrence (band values: 0-1), totaling 38 images with a data volume of 9.07 GB; (2) three categories of surface water (SWC) Data (i.e. seasonal water bodies, permanent water bodies, and maximum extent water bodies), including two labels: non-water (band value: 0) and water (band value: 1), totaling 114 images with a total data volume of 9.53 GB.