

1990–2021 年蒙古高原年度丰水期水体数据集

ISSN 2096-2223

CN 11-6035/N

王妍¹, 刘帅¹, 邱丽莎¹, 单炜^{1,2,3*}

1. 东北林业大学, 寒区科学与工程研究院, 哈尔滨 150040
2. 东北多年冻土区地质环境系统教育部野外科学观测研究站, 哈尔滨 150040
3. 东北多年冻土区环境、道路建设与养护协同创新中心, 哈尔滨 150040

摘要: 蒙古高原生态系统脆弱, 土地荒漠化严重, 是亚洲地区主要的气候敏感区域之一, 其水资源的时空变化有重要的研究价值。本数据集利用 Google Earth Engine(GEE)平台, 对近 32 年 (1990–2021 年) 可用的高质量 Landsat 系列卫星影像进行处理。利用最小云量合成算法, 得到每年的丰水期 (6–9 月) 最小云量影像。计算 NDWI 后采用大津算法进行阈值分割, 提取得到 32 年蒙古高原逐年丰水期 30 m 分辨率水体数据, 最终结果以 GeoTIFF 格式保存。通过对比, 数据集与 JRC 年度水体数据的永久水体和最大水体的一致性平均值分别为 93.0% 和 90.9%, 具有较高的可信度。数据集可以为蒙古高原水资源变化、生态建设规划、环境保护等提供数据支撑。

关键词: 蒙古高原; 水体; 大津算法; 归一化水体指数

文献 CSTR:

32001.14. 11-6035.csd.2022.0061.zh

文献 DOI:

10.11922/11-6035.csd.2022.0061.zh

数据 DOI:

10.57760/sciencedb.01943

文献分类: 地球科学

收稿日期: 2022-07-17

开放同评: 2022-07-26

录用日期: 2022-11-21

发表日期: 2023-02-02

数据库 (集) 基本信息简介

数据库 (集) 名称	1990–2021 年蒙古高原年度丰水期水体数据集
数据通信作者	单炜 (shanwei456@163.com)
数据作者	王妍、刘帅、单炜
数据时间范围	1990–2021 年
地理区域	地理范围为北纬 37°46'N–53°08'N, 东经 87°40'E–122°15'E, 其中包括内蒙古自治区和北部毗邻的蒙古国。地理区域范围涉及到的国家有中国, 蒙古。
空间分辨率	30 m
数据量	1.77 GB
数据格式	*.tif
数据服务系统网址	http://doi.org/10.57760/sciencedb.01943
基金项目	黑龙江省交通运输厅科技项目, 多年冻土退化区路堑边坡植物固坡机理与适生性研究
数据集组成	数据集包括 1990–2021 年每年 1 期、共计 32 期的蒙古高原丰水期水体空间分布数据, 每期包含 3 幅影像, 共计 96 个子文件, 数据存贮为 GeoTIFF 数据文件, 命名方法为 “yyyyMongolianWater-1/2/3”, 如 1990MongolianWater-1。影像空间分辨率为 30 m, 投影系统为 WGS-84。由于数据范围过大, 每一期数据被分割为 3 幅子数据保存。全部数据保存为 1 个文件 “1990–2021 年蒙古高原丰水期水体数据集.rar”, 总数据量为 1.77GB。

* 论文通信作者

单炜: shanwei456@163.com

引言

蒙古高原主体由我国内蒙古自治区和北部毗邻我国的蒙古国共同组成，属于温带大陆性气候，是典型的干旱和半干旱地区。蒙古高原地域广袤（37°46'N–53°08'N，87°40'E–122°15'E），是海拔较高的内陆高原^[1]。作为深刻影响东亚乃至全球气候和碳循环的重要地区，蒙古高原的水系变化对其生态系统的稳定性、区域经济发展等诸多方面有着深刻的影响^[2]。随着全球气候变暖和人为因素的干扰，蒙古高原自然灾害频发^[3]，植被分布面积不断减少^[4]，冻土分布面积也在不断缩减^[5]，土地荒漠化日趋严重^[6]，水系面积不断萎缩^[7]，且退化总体速度远超世界平均水平^[8]。

水体的面积变化可以在一定程度上揭示该地区的环境变化和生态响应趋势，其时空变换和分配直接反映出区域水资源的分布情况和变化特征^[9]，因而水体面积基础数据对于地区生态环境、气候变化、农业植被以及经济工程建设等多个方面具有重大的意义和价值^[10]。自 1980 年卫星遥感技术开始应用于水体提取，至今已发展出基于雷达和光学遥感数据源进行水体信息提取的阈值法^[11]、目标分类法^[12]等许多方法^[13]。随着遥感信息云计算处理水平的提高和技术的发展，GEE 作为专业的海量遥感数据处理云平台被广泛应用于数据集加工和生产。利用 GEE 平台可以对多源遥感数据和卫星影像进行批量处理与分析，进而将结果用于科学研究与利用，加快遥感信息处理和应用的进程^[14]。

欧盟联合研究中心（JRC）提供的全球地表水数据集、美国航天局（NASA）提供的地表水体掩膜数据产品 MOD44W.006 和内陆地表水体数据集 GLCF 是目前较为常用的水体数据集。在精度上，JRC 的全球地表水体数据集和 NASA 的内陆地表水体数据集 GLCF 数据集都为 30 m 分辨率，尽管地表水体掩膜数据产品 MOD44W.006 相比内陆地表水体数据集 GLCF 其空间覆盖范围更广，但精度大为降低，为 250 m 分辨率。在时间跨度上，MOD44W.006 仅有 2000 年至 2014 年，共计 15 年的数据可用性^[15]，GLCF 虽然以高分辨率很好地显示了内陆地表水体，包括淡水湖和咸水湖、河流和水库等，但其可用时间跨度仅 2000 年 1 年，很难反映出区域水体在时空上的变化情况。JRC 的水体数据集包含 1984 年至 2020 年地表水的位置和时间分布图，描绘了过去 35 年地表水时空分布的不同方面，并提供这些水面的范围和变化的统计数据。数据集将每个像素单独分为水/非水，不考虑从未探测到水的区域。其结果整理成整个时间段的月历史记录和两个时期（1984–1999 年，2000–2020 年）的变化检测，具有长时序变化检测的优点^[16]。JRC 数据集在 1984–1999 年间使用了来自 Landsat5(TM) 的数据影像，但由于 Landsat5(TM) 没有装备机载数据记录器，故其与数据中继卫星的连接随着时间的推移而失效，因此其收集的数据影像覆盖范围通常仅限于接收站的视线范围，而且无法提供完整的全球覆盖范围。从 1984 年到 1999 年，全球各洲各地随着时间推移和技术进步依次成像，直到 1999 年 4 月具有机载数据记录能力的 Landsat7 开始运行，覆盖率和成像的问题才得到很好的改善，故 JRC 数据集在 2000 年以前的数据中包含很多无像素数据。

本研究在 1990–1999 年间的数据采集和选取上，通过人工逐年检查筛选，对在 GEE 平台收集的蒙古高原地区丰水期（6–9 月）影像进行质量把控（如图 1）。对于丰水期高质量影像较少的年份，筛选 3 月到 10 月的原始影像，采用像元级最小云量合成算法合成最小云量影像，从而解决部分地区在时空上像素缺失的问题，同时保证本数据集在 2000 年前的数据质量完整性。对于水库、湖泊等水体，当人为进行放水、蓄水等季节性调控时，JRC 的监测分类也存在不准确性。本研究采用大津算法实现局部自适应阈值水体信息的提取^[18]。该方法可以进行大范围水体信息提取，精度相较 JRC 和 NASA 的内陆地表水体数据集在丰水期的水体数据有所提高，并在一定程度上消除积雪、山体阴影

等的影响，降低对于农田等细小水体的误分和漏提。

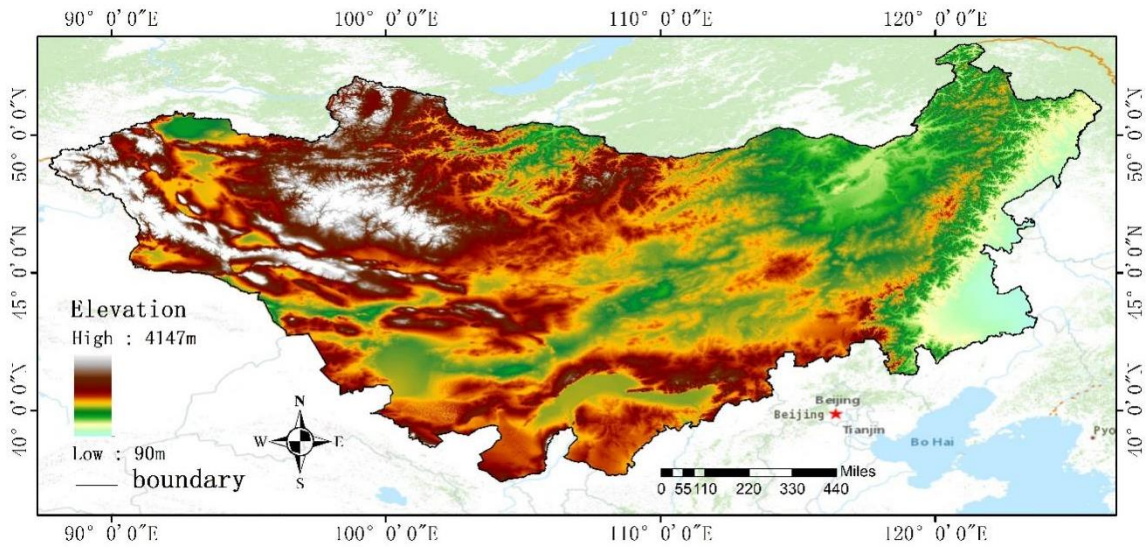


图 1 研究区示意图

Figure 1 Schematic diagram of study areas

1 数据采集和处理方法

1.1 数据采集方法

在本研究中，使用了蒙古高原地区 1990–2021 年的丰水期（6–9 月）的所有可用的 Landsat5/7/8 的影像数据。具体为 1990–1998 年采用 Landsat5 数据；1999–2011 年采用 Landsat5 和 Landsat7 的数据；2012 年采用 Landsat7 的数据；2013–2021 年采用 Landsat7 和 Landsat8 的数据。将上述数据通过像元级最小云量合成算法得到每年一幅的蒙古高原丰水期最小云量影像，进行水体提取。值得注意的是，由于 1995 年丰水期高质量影像过少，这一年我们筛选了 3 月到 10 月的原始影像来合成最小云量影像。上述影像取自 GEE 平台提供的开放数据集^[19]，如表 1 所示。

表 1 数据源基本信息

Table 1 Basic information of data source

序号	数据源卫星	数据年份	数据集
1	Landsat 5 TM	1990–2011	USGS Landsat 5 TM Collection 2 Tier 1–Raw Scenes
2	Landsat 7 ETM+	1999–2021	USGS Landsat 7 Collection 2 Tier 1 and Real–Time data Raw Scenes
3	Landsat 8 OLI/TIRS	2013–2021	USGS Landsat 8 Collection 2 Tier 1–Raw Scenes

对于数据质量的把控，本数据集使用了 GEE 平台开放数据集所有可用的 Landsat5/7/8 的影像数据。作为目前世界上功能较为先进的全球性云计算平台之一，其数据质量经过全球各专业机构的修正和检验，具有一定的可信度。在数据处理上通过严格对影像质量逐年筛选进行把控，对于因卫星原因造成像素数据缺失的年份，通过扩大临近的筛选时段后采用像元级最小云量合成算法进行处理，最终合成每年一幅的影像，确保了数据影像质量的完整性。研究基于 RS 技术，参考了多篇相关前人研究文献，以及得出的目前最新的研究成果，以因地制宜选取的大津算法为主，以 GEE 平台针对蒙古高原水体提取阈值的调控为辅，来尽可能确保数据产品的精度和准确性有着良好的提升效果。

1.2 水体提取方法

本研究基于 GEE 平台，使用了蒙古高原地区 1990–2021 年的丰水期（6–9 月）的所有可用的 Landsat5/7/8 的影像数据，通过像元级最小云量合成算法得到每年一幅的蒙古高原丰水期最小云量影像，进行水体提取。提取用到的参数为归一化水体指数(NDWI)，NDWI 是基于绿波段与近红外波段的归一化比值指数（公式 1），最早由 Mcfeeters 在 1996 年提出^[17]。

$$NDWI = \frac{Band_{green} - Band_{NIR}}{Band_{green} + Band_{NIR}} \quad (1)$$

式中 $Band_{green}$ 为绿光波段， $Band_{NIR}$ 为近红外波段。与其他地物相比，水对绿光和近红外光有特殊的吸收和反射性质，因此 NDWI 在水体提取中有着广泛的应用^[20]。另外，由于植被一般在近红外波段的反射率最强，因此采用绿光波段与近红外波段的比值可以最大程度地抑制植被的信息，从而达到突出水体的目的。但土壤和建筑物在绿光和近红外波段的波谱特征与水体几乎一致，容易在成像中与水体混淆形成噪音^[17]。由于蒙古高原植被覆盖较大而城市群占地较小，故 NDWI 指数在蒙古高原具有适用性，且采用大津算法对 NDWI 数据进行处理，尽可能地消除了因城市土壤覆盖对提取结果造成的干扰，从而保证了数据的提取精度和质量。在基于 NDWI 的水体提取研究中，许多学者将 $NDWI > 0$ 作为判断水体的标准^[21]。这样的分割标准有着良好的效果^[22-23]，但是由于不同时期的影像数据有差异，往往需要设置不同的阈值来提取水体。

大津算法(OSTU)是一种自适应阈值的图像分割算法，该算法通过最小化类内强度方差，或者等效地通过最大化类间方差确定一个单一的强度阈值，将像素分成两类，以完成图像的二值化分割^[24]。

本文采用基于大津算法处理 NDWI 数据的水体提取方法（Water body extraction method based on Otsu algorithm to process NDWI data，以下简称 BONM），先将每年蒙古高原丰水期最小云量影像通过波段运算得到每年蒙古高原的 NDWI 数据，然后利用大津算法计算 NDWI 数据的自适应强度阈值 T ，以 $NDWI > T$ 为提取条件提取从 1990 年到 2021 年每年 1 期的蒙古高原水体，共得到 32 期丰水期蒙古高原水体数据，技术路线如图 2。

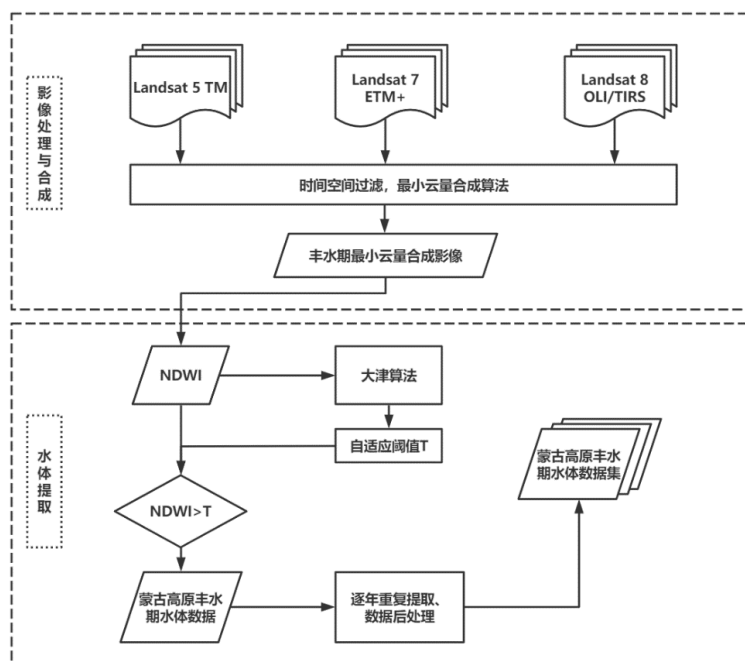


图 2 BONM 水体提取技术路线

Figure 2 Technical roadmap of BONM water extraction

将 $NDWI > 0$ 和 BONM 的提取结果进行对比实验。按照影像数据来源的不同, 将实验分为大尺度多影像源 (A 组) 和小尺度单幅影像 (B 组) 两组。A 组对蒙古高原地区, 利用 2014 年丰水期合成最小云量影像进行水体提取; B 组对呼伦湖地区, 利用 2014 年 9 月 11 日的 Landsat 8 单幅影像 (WRS_PATH: 124, WRS_ROW: 26) 进行水体提取, 提取结果的细节对比如图 3 所示。

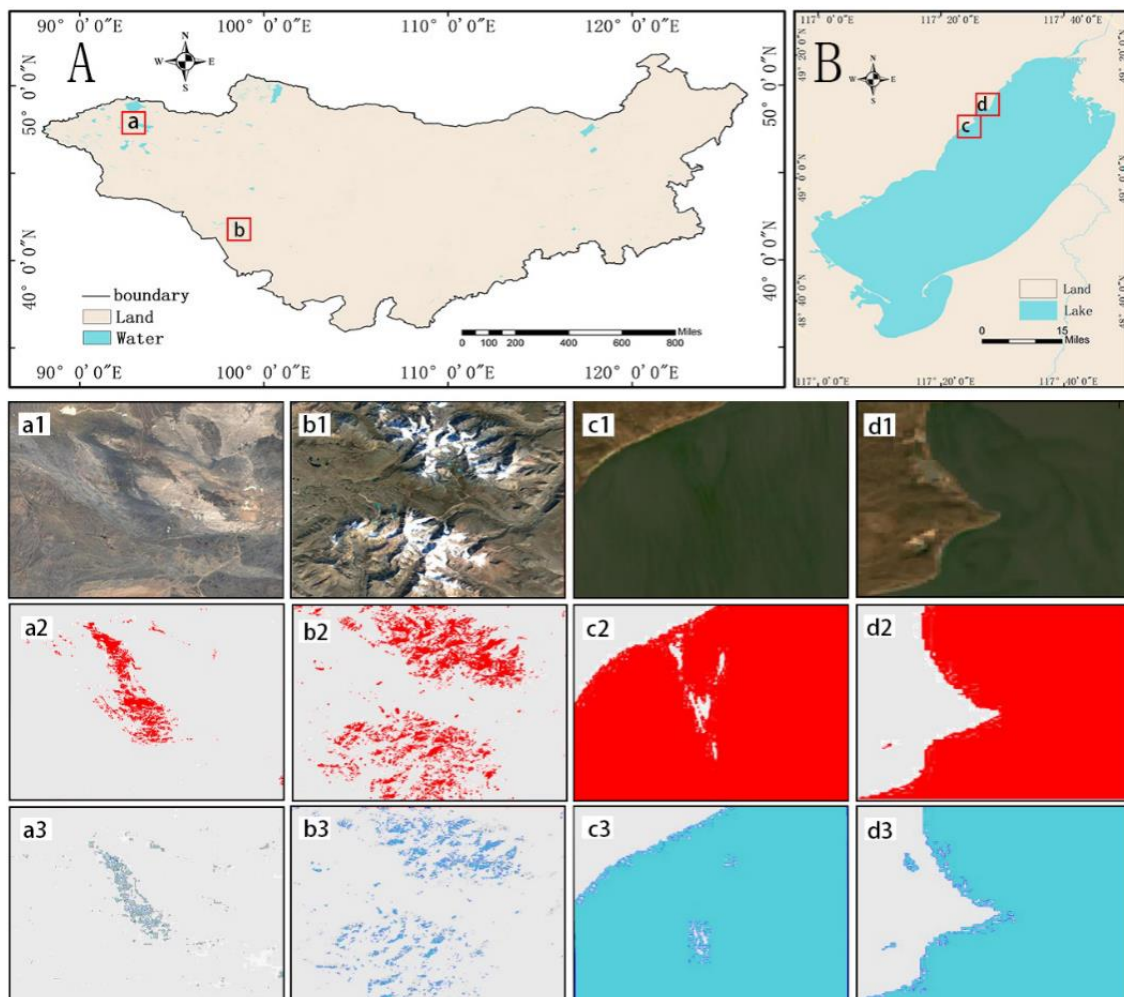


图 3 $NDWI > 0$ 和 BONM 水体提取差异对比图。包含 A 组利用合成影像提取水体和 B 组利用单幅影像提取水体的对比。A 组选取 a (湿地) 和 b (雪山) 展示, B 组选取 c (湖面波纹) 和 d (细小水体) 展示。在每处提取结果对比中, 1 为卫星影像, 2 (红色) 为 $NDWI > 0$ 提取结果, 3 (蓝色) 为 BONM 提取结果

Figure 3 Comparison diagram of differences in water extraction between $NDWI > 0$ and BONM. It includes the comparison between group A using synthetic image to extract water and group B using single image to extract water. Group A selected a (wetland) and b (snow mountain) for display, and group B selected c (ripple) and d (small water body) for display. In the comparison of extraction results at each place, 1, 2 and 3 stand for satellite image, $NDWI > 0$ extraction result and BONM extraction result respectively.

湿地和雪山是水体提取中两种较常见的误差来源。如图 3 所示, 在 A 组实验中, 位于 b 处的雪山积雪被误分为水体, 但 BONM 的误分量更小; 对于只有少量地表水的湿地 (a), 相对于 $NDWI > 0$ 的提取结果, BONM 也有更好的鲁棒性。湖面的波纹的消除和细小水体提取也是水体提取中较难解决的问题, 在 B 组实验中发现, 对于 d 处的细小水体, BONM 的识别效果更好; 同时, BONM 对湖

面波纹（c）干扰的处理更为理想。

对比结果表明，无论大尺度多影像源合成影像，还是小尺度单幅影像，在基于 NDWI 提取水体时，采用基于大津算法的阈值提取水体都有较好的适用性。因此，采用 BONM 方法提取 1990–2021 年的蒙古高原丰水期水体数据。

2 数据样本描述

本数据集共包括 1990–2021 年每年 1 期，共计 32 期的蒙古高原丰水期水体空间分布数据，数据存贮为 GeoTIFF 数据文件。由于数据范围过大，每一期数据被分割为 3 幅子数据保存，子文件命名方法为“yyyyMongolianWater-1/2/3”（如 2014MongolianWater-1.tif），共计 96 个子文件，全部数据保存为 1 个压缩文件“1990–2021 年蒙古高原丰水期水体数据集.rar”。数据集的空间分辨率为 30 m，投影系统为 WGS-84，总数据量为 1.77 GB。以 2014 年为例，图 4 展示了采用 BONM 方法提取的蒙古高原丰水期水体数据局部放大示意图。

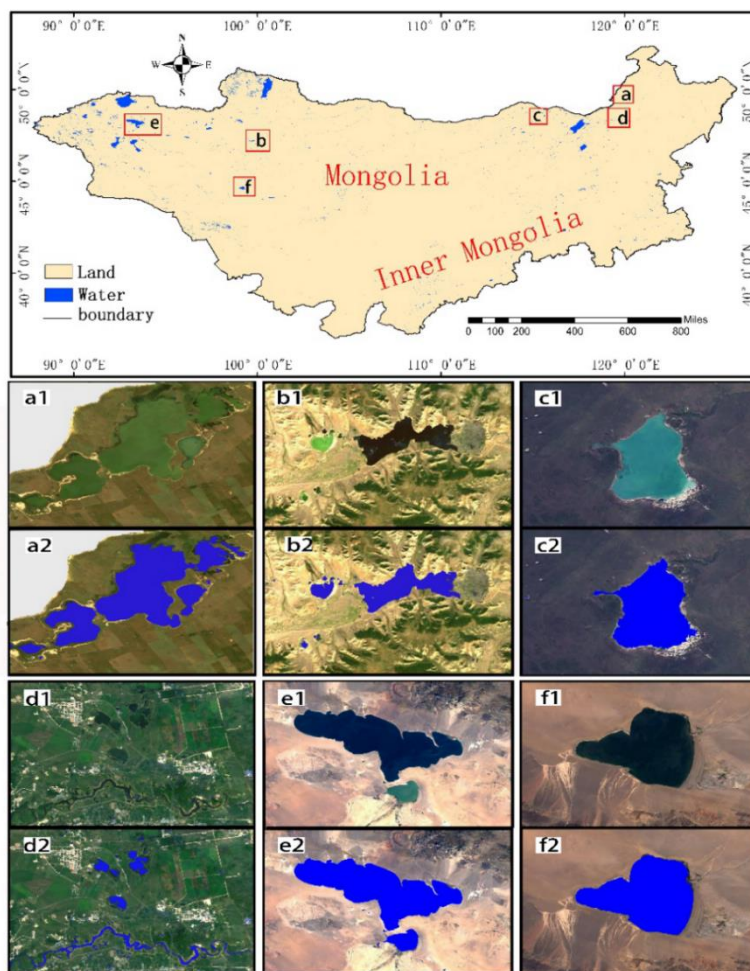


图 4 BONM 提取蒙古高原丰水期水体效果局部放大图（2014 年）。选取蒙古高原不同地区（a–f）的水体样本进行展示，在每一处的提取结果对比中，1 为卫星影像，2 为 BONM 提取结果

Figure 4 Partial enlarged picture of BONM extraction of water body effect in wet seasons of Mongolian Plateau (2014). Water samples from different areas (a–f) of the Mongolian Plateau are selected for display. In the comparison of extraction results at each place, 1 is satellite image and 2 is BONM extraction result

3 数据质量控制和评估

3.1 空间评估

将 BONM 提取结果与 JRC 年度水体数据进行空间对比分析。JRC 年度水体数据提供了在 1984–2020 年期间的每一年全球地表水的信息，根据全年水的季节性分类，将地表水分为季节性地表水和永久地表水^[16]。

分别计算每年 BONM 结果与 JRC 年度水体数据中的永久性水体(以下简称 JRC_P)、季节性水体(以下简称 JRC_S)和最大水体(以下简称 JRC_M)的一致性比例。我们定义最大水体为季节性水体与永久性水体相加，即：

$$\text{JRC_M} = \text{JRC_P} + \text{JRC_S} \quad (2)$$

定义一致性比例的计算方法为：

$$\text{POV}_i = \frac{\text{PO}_i}{\text{O}_i} \times 100\% \quad (3)$$

$$\text{SOV}_i = \frac{\text{SO}_i}{\text{O}_i} \times 100\% \quad (4)$$

$$\text{MOV}_i = \frac{\text{MO}_i}{\text{O}_i} \times 100\% \quad (5)$$

其中 i 表示年份， $i \in [2000, 2020]$ 且 i 为整数。 POV_i 表示 i 年 JRC_P 与 BONM 结果的一致性比例； PO_i 表示 i 年 JRC_P 与 BONM 结果重合部分的面积； O_i 表示 i 年 BONM 结果的面积。以此类推， SOV_i 表示 i 年 JRC_S 与 BONM 结果的一致性比例； SO_i 表示 i 年 JRC_S 与 BONM 结果重合部分的面积； MOV_i 表示 i 年 JRC_M 与 BONM 结果的一致性比例； MO_i 表示 i 年 JRC_M 与 BONM 结果重合部分的面积。

由于研究区 2000 年以前的 JRC 数据包含太多的无数据像素，所以选取 2000–2020 年的 JRC 年度水体数据与水体提取结果进行比较。计算了从 2000 年到 2020 年 BONM 提取结果与 JRC_P、JRC_S 和 JRC_M 的一致性比例，结果如图 5 所示。

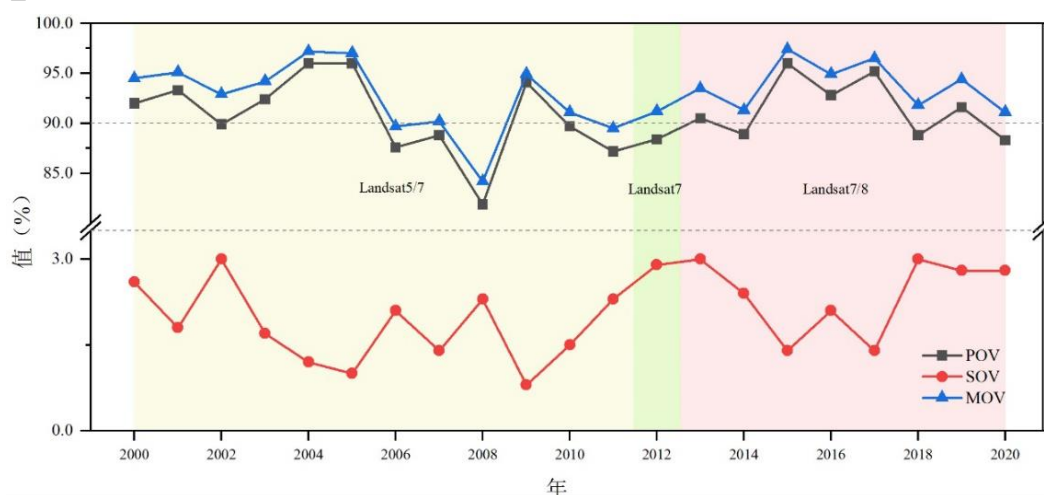


图 5 2000–2020 年 BONM 提取结果与 JRC_P、JRC_S 和 JRC_M 的一致性比例

Figure 5 Consistency ratio of BONM extraction results compared with JRC_P, JRC_S and JRC_M from 2000 to 2020

由图 5 可以看出,在 2000–2020 年间,BONM 提取结果与 JRC_P 和 JRC_M 具有较高的一致性。一致性比例 POV 和 MOV 的平均值分别为 93.0%和 90.9%,最高出现在 2015 年,达到 POV=96.0%,MOV=97.4%;最低出现在 2008 年,POV=81.9%,MOV=84.2%。由于季节性水体的出现在空间和时间上具有很强的可变性,在降水较多和干旱年份的数据差异较大,卫星在季节性水体的像素捕捉上具有一定难度,需要保证无云的卫星观测影像和季节性水体出现的时间一致,在季节性水体出现时间较短的年份,这一过程更具有挑战性。且季节性水体水量相较永久性水体和最大水体来说相对较小,其水体提取结果在 JRC 数据和 BONM 提取的逐年总水体面积 O_i 中所占比例在 1%–3.8%之间,所以展现出的一致性经过计算最终能够控制在 O_i 的 3%以内波动,可以认为一致性较高。

进一步探讨 BONM 提取结果与 JRC 年度水体数据在空间分布上的差异。由于 2000 年到 2011 年的原始影像来自 Landsat5/7 卫星,2013 年到 2020 年的原始影像来自 Landsat7/8 卫星,2012 年的原始影像来自 Landsat7 卫星。按照使用卫星数据来源的不同,分别取出一致性比例最低的两年,即 2008 年和 2020 年,将提取的水体数据与 JRC 年度水体数据进行对比,并在对比结果中筛选出具有代表性的样本,来对提取结果的差异进行展示与说明,如图 6 所示。

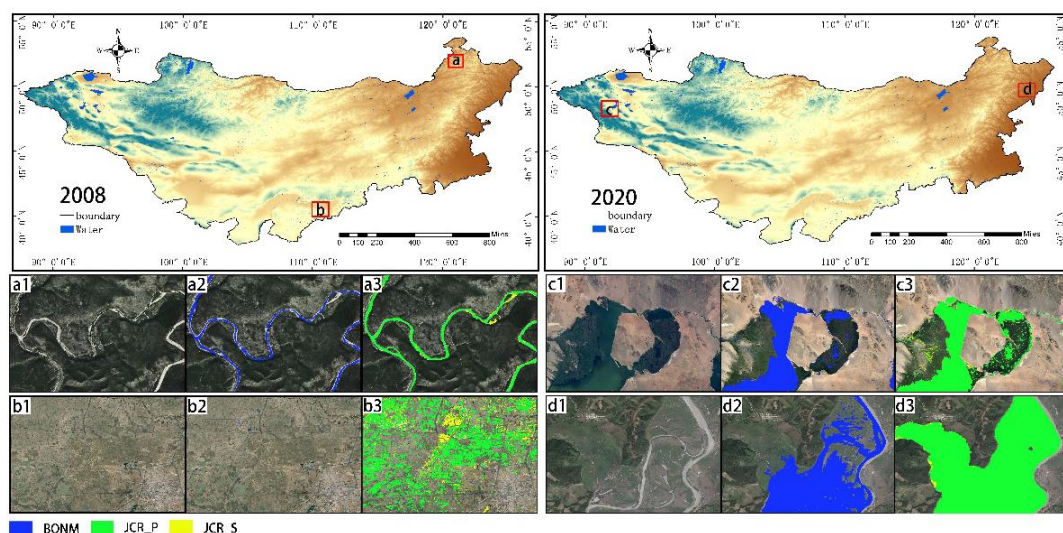


图 6 BONM 提取结果与 JRC 年度水体数据 (JRC_P 和 JRC_S) 差异图。放大对比 2008 年 (a 和 b) 和 2020 年 (c 和 d) BONM 提取结果与 JRC 年度水体数据的主要差异。在每一处对比中, 1 为卫星影像, 2 为 BONM 提取结果, 3 为 JRC 年度水体数据

Figure 6 Difference diagram between BONM extraction results and JRC annual water body data (JRC_P and JRC_S). The main differences between the BONM extraction results in 2008 (a and b) and 2020 (c and d) and the annual water body data of JRC are amplified and compared. In each comparison, 1, 2 and 3 stand for satellite image, BONM extraction result and JRC annual water body data respectively.

从图 6 所示 2008 年对河道 (a) 和 2020 年对滩地 (c) 的水体提取结果对比中可看出, BONM 提取结果对于变化的水体刻画效果要略逊于 JRC 年度数据。最主要原因是, BONM 的水体数据的提取基础是本数据集采用像元级最小云量合成算法处理所得的蒙古高原丰水期卫星影像数据, 最终合成每年一幅无像素缺失的丰水期高质量影像。在全年流动水体的动态变化方面, BONM 相较于以全年影像为提取基础而不是对丰水期数据有所侧重的 JRC 数据集, 因其提取对象侧重在时间跨度上的选取而导致不如 JRC 数据集提取对象的时间跨度广, 所以在时空捕捉上对于全年都有所赋存的河流

等动态水体较为逊色。但是, 对于一些地区农田 (b), JRC 年度水体数据存在将大片农田误分为水体的现象, 这与 JRC 数据的生成算法有关, 而 BONM 在处理农田干扰时的表现更好。在一些水库地区 (d), BONM 提取结果和 JRC 年度数据存在明显差异, 其原因是丰水期水库放水, 而 JRC 数据通过水体出现频率判定整个水库为永久水体。

3.2 时间序列变化评估

计算 BONM 提取结果和 JRC 年水体数据的水体总面积, 进行时间上的数值和趋势变化分析。这里也选取 2000–2020 年的数据进行比较, 如图 7 所示。在 2000–2020 年, BONM 提取结果均值为 17 781 km², 最大值出现在 2008 年, 为 19 657 km², 最小值出现在 2000 年, 为 16 819 km²; JRC_P 均值为 18 551 km², 最大值出现在 2000 年, 为 19 728 km², 最小值出现在 2007 年, 为 17 862 km²。

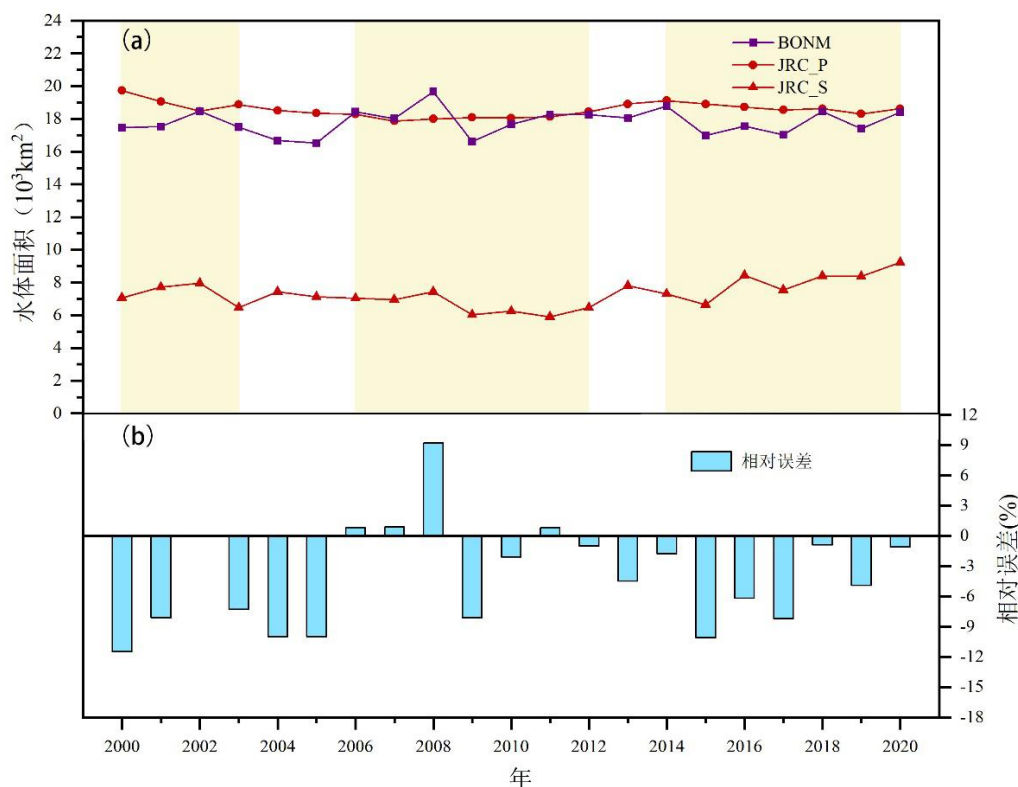


图 7 2000–2020 年 BONM 提取结果与 JRC 年水体数据的水体面积的数值和趋势变化分析。(a)为水体总面积数值变化折线图,(b)为每年 BONM 水体面积与 JRC_P 水体面积的相对误差

Figure 7 The numerical value and trend change analysis of the water area of BONM extraction results and JRC water data from 2000 to 2020. (a) is the line chart of the numerical change of the total water area; (b) is the annual BONM water area and JRC_P relative error of water area

我们发现, 在数值上 BONM 提取结果与 JRC_P 比较接近, 二者的相对误差仅在 2000 年达到 14.7%, 此外的 20 年均保持在 $\pm 10\%$ 以内。此外, 在变化趋势上, BONM 提取结果与 JRC_S 存在一定程度的相似性。在 2000–2003 年、2006–2012 年和 2014–2020 年间, BONM 提取结果与 JRC_S 的变化趋势几乎相同。

综上所述, BONM 提取结果在蒙古高原有很好的适用性, 在时间和空间分布上都具有较好的可信度。

4 数据价值

蒙古高原是中国重要的生态屏障，也是亚洲乃至世界生态系统的重要组成部分，其水体的赋存状态和时空变化对于延缓和阻止该地区及周边地区的荒漠化进程、改善生态、乃至全球碳循环都有着重要的意义。本研究利用 GEE 平台，基于 Landsat 系列影像，提供近 32 年（1990–2021 年）蒙古高原丰水期 30 m 分辨率水体数据，可以为蒙古高原水系变化动态监测、生态建设规划、环境保护与研究提供数据信息和依据。

本数据集可作为蒙古高原湖泊水体赋存状态，如水量、湖泊面积和数量等方面的时空变化监测等研究的基础数据。但对于蒙古高原部分山区存在的细小水体，由于所占面积较小，故本数据集未剔除因山体阴影造成的水体提取结果的误分，进行相关研究和使用时需要采用坡度筛选等手段进行再处理。本数据集在展现蒙古高原近 30 年水体时空演变规律的同时，也可以为蒙古高原气候、降水、生态变化等方面的研究提供数据参考。

5 数据使用方法和建议

由于研究区的部分山区存在细小水体，本文没有对由于山体阴影导致的少量误分进行剔除。如有需要，可通过坡度筛选等手段进行数据再加工。

本数据集包含 1990–2021 共 32 年逐年蒙古高原水体数据，储存为 TIF 格式文件，可以利用 ArcGIS、ENVI、GEE 等地理信息处理软件 and 平台对本数据集中数据进行查看、编辑以及进行相关研究分析。

数据作者分工职责

王妍（1999—），女，河南省濮阳市人，硕士研究生，研究方向为寒区交通基础设施工程。主要承担工作：数据校核、论文撰写。

刘帅（1998—），男，安徽省淮北市人，博士研究生，研究方向为冻土区碳生态调控与工程减排。主要承担工作：数据生产、论文撰写。

邱丽莎（1993—），女，甘肃高台县人，博士研究生，研究方向为冻土地质与生态遥感。主要承担工作：论文图表绘制、质量控制。

单炜（1965—），男，黑龙江省哈尔滨市人，博士，教授，研究方向为寒区工程、基础与边坡工程。主要承担工作：技术指导、质量控制。

参考文献

- [1] 包刚, 包玉海, 覃志豪, 等. 近 10 年蒙古高原植被覆盖变化及其对气候的季节响应[J]. 地理科学, 2013, 33(5): 613–621. DOI:10.13249/j.cnki.sgs.2013.05.016. [BAO G, BAO Y H, QIN Z H, et al. Vegetation cover changes in Mongolian Plateau and its response to seasonal climate changes in recent 10 years[J]. Scientia Geographica Sinica, 2013, 33(5): 613–621. DOI:10.13249/j.cnki.sgs.2013.05.016.]
- [2] 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/dqxkx.2022.210031.]
- [3] 丹丹. 蒙古高原近 35 年气候变化[D]. 内蒙古师范大学, 2014. [DAN D. Climate Changes In Mongolia Plateau During Last 35 Years[D]. Inner Mongolia Normal University, 2014.]
- [4] 缪丽娟, 蒋冲, 何斌, 等. 近 10 年来蒙古高原植被覆盖变化对气候的响应[J]. 生态学报, 2014, 34(5): 1295–1301. DOI:10.5846/stxb201304100659. [MIAO L J, JIANG C, HE B, et al. Response of vegetation coverage to climate change in Mongolian Plateau during recent 10 years[J]. Acta Ecologica Sinica, 2014, 34(5): 1295–1301. DOI:10.5846/stxb201304100659.]
- [5] 高红豆, 萨楚拉, 孟凡浩, 等. 2003—2019 年蒙古高原多年冻土时空动态变化及其影响因素分析[J]. 干旱区资源与环境, 2022, 36(3): 99–106. DOI:10.13448/j.cnki.jalre.2022.070. [GAO H D, SA C L, MENG F H, et al. Temporal and spatial characteristics of permafrost on Mongolian Plateau from 2003 to 2019[J]. Journal of Arid Land Resources and Environment, 2022, 36(3): 99–106. DOI:10.13448/j.cnki.jalre.2022.070.]
- [6] 孟翔冲. 蒙古国沙质荒漠化对中国北方沙质荒漠化影响研究[D]. 吉林大学, 2012. [MENG X C. A Study on the Influences of Mongolia Desertification on the Desertification in Northern China[D]. Jilin University, 2012.]
- [7] 周岩. 蒙古高原 2000–2015 年湖泊变化及其成因分析[D]. 中国地质大学(北京), 2018. [ZHOU Y. Lake changes and their causes analysis on the Mongolia Plateau from 2000 to 2015[D]. China University of Geosciences(Beijing), 2018]
- [8] 恩和. 蒙古高原草原荒漠化的文化学思考[J]. 内蒙古社会科学(汉文版), 2005, 26(3): 136–141. DOI:10.3969/j.issn.1003-5281.2005.03.031. [EN H. Cultural thinking on grassland desertification in Mongolian Plateau[J]. Inner Mongolia Social Sciences, 2005, 26(3): 136–141. DOI:10.3969/j.issn.1003-5281.2005.03.031.]
- [9] 郝美玉, 罗泽. 1987—2017 年青海湖水体边界数据集[J]. 中国科学数据, 2021(1): 193–199. [HAO M Y, LUO Z. A dataset of Qinghai Lake water body periphery during 1987–2017[J]. China Scientific Data, 2021(1): 193–199.]
- [10] 朱刚, 高会军, 曾光. 近 35a 来新疆干旱区湖泊变化及原因分析[J]. 干旱区地理, 2015, 38(1): 103–110. DOI:10.13826/j.cnki.cn65-1103/x.2015.01.014. [ZHU G, GAO H J, ZENG G. Lake change research and reasons analysis in Xinjiang arid regions during the past 35 years[J]. Arid Land Geography, 2015, 38(1): 103–110. DOI:10.13826/j.cnki.cn65-1103/x.2015.01.014.]
- [11] 毕海芸, 王思远, 曾江源, 等. 基于 TM 影像的几种常用水体提取方法的比较和分析[J]. 遥感信息, 2012, 27(5): 77–82. DOI:10.3969/j.issn.1000-3177.2012.05.014. [BI H Y, WANG S Y, ZENG J Y, et al. Comparison and analysis of several common water extraction methods based on TM image[J]. Remote Sensing Information, 2012, 27(5): 77–82. DOI:10.3969/j.issn.1000-3177.2012.05.014.]
- [12] 都金康, 黄永胜, 冯学智, 王周龙. SPOT 卫星影像的水体提取方法及分类研究[J]. 遥感学报, 2001, 5(3): 214–219. DOI:10.3321/j.issn: 1007-4619.2001.03.009. [DU J K, HUANG Y S, FENG X Z, et al. Study on water bodies extraction and classification from SPOT image[J]. Journal of Remote Sensing, 2001, 5(3): 214–219. DOI:10.3321/j.issn: 1007-4619.2001.03.009.]
- [13] 李丹, 吴保生, 陈博伟, 等. 基于卫星遥感的水体信息提取研究进展与展望[J]. 清华大学学报

- (自然科学版), 2020, 60(2): 147–161. DOI:10.16511/j.cnki.qhdxxb.2019.22.038. [LI D, WU B S, CHEN B W, et al. Review of water body information extraction based on satellite remote sensing[J]. Journal of Tsinghua University (Science and Technology), 2020, 60(2): 147–161. DOI:10.16511/j.cnki.qhdxxb.2019.22.038.]
- [14] ZHOU Y, DONG J W, XIAO X M, 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: 366–380. DOI:10.1016/j.scitotenv.2019.06.341.
- [15] 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): 113–133. DOI:10.1080/17538947.2015.1026420.
- [16] 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.
- [17] 徐涵秋. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究[J]. 遥感学报, 2005, 9(5): 589–595. DOI:10.3321/j.issn: 1007-4619.2005.05.012. [XU H Q. A study on information extraction of water body with the modified normalized difference water index (MNDWI)[J]. Journal of Remote Sensing, 2005, 9(5): 589–595. DOI:10.3321/j.issn: 1007-4619.2005.05.012.]
- [18] 彭燕, 何国金, 王桂周, 等. 澜湄流域 30 m 分辨率陆表水体产品[J]. 中国科学数据, 2020, 5(4):97-105. DOI: 10.11922/csdata.2020.0029.zh. [Peng Yan, He Guojin, Zhang Zhaoming, Yin Ranyu. Inland surface water products with 30 m spatial resolution in Lanmei Watershed [J]. China Scientific Data, 2020, 5(4) :97-105. DOI: 10.11922/csdata.2020.0029.zh.]
- [19] Landsat Levels of Processing[EB/OL]. [2022-07-01]. <https://www.usgs.gov/land-resources/nli/landsat/landsat-levels-processing>.
- [20] 孙芳蒂, 熊立. 2000 – 2015 年咸海边界数据集[J]. 中国科学数据, 2019, 4(4) : 56-61. DOI: 10.11922/csdata.2019.0033.zh. [Sun Fangdi, Xiong Li. A dataset of the Aral Sea periphery during 2000 – 2015 [J]. China Scientific Data,2019, 4(4) : 56-61. DOI: 10.11922/csdata.2019.0033.zh.]
- [21] 王大钊, 王思梦, 黄昌. Sentinel-2 和 Landsat8 影像的四种常用水体指数地表水体提取对比[J]. 国土资源遥感, 2019, 31(3): 157–165. DOI:10.6046/gtzyyg.2019.03.20. [WANG D Z, WANG S M, HUANG C. Comparison of Sentinel-2 imagery with Landsat8 imagery for surface water extraction using four common water indexes[J]. Remote Sensing for Land & Resources, 2019, 31(3): 157–165. DOI:10.6046/gtzyyg.2019.03.20.]
- [22] 徐蓉, 张增祥, 赵春哲. 湖泊水体遥感提取方法比较研究[J]. 遥感信息, 2015, 30(1): 111–118. DOI:10.3969/j.issn.1000-3177.2015.01.019. [XU R, ZHANG Z X, ZHAO C Z. Different models used in extraction of lake water body based on MODIS data[J]. Remote Sensing Information, 2015, 30(1): 111–118. DOI:10.3969/j.issn.1000-3177.2015.01.019.]
- [23] 席晓燕, 沈楠, 李小娟. ETM+影像水体提取方法研究[J]. 计算机工程与设计, 2009, 30(4): 993–996. DOI:10.16208/j.issn1000-7024.2009.04.035. [XI X Y, SHEN N, LI X J. Research of automatic extraction of water Bodies on ETM+images[J]. Computer Engineering and Design, 2009, 30(4): 993–996. DOI:10.16208/j.issn1000-7024.2009.04.035.]
- [24] 吴冰, 秦志远. 自动确定图像二值化最佳阈值的新方法[J]. 测绘学院学报, 2001, 18(4): 283–286.

DOI:10.3969/j.issn.1673-6338.2001.04.014. [WU B, QIN Z Y. New approaches for the automatic selection of the optimal threshold in image binarization[J]. Journal of the PLA Institute of Surveying and Mapping, 2001, 18(4): 283–286. DOI:10.3969/j.issn.1673-6338.2001.04.014.]

论文引用格式

王妍, 刘帅, 邱丽莎, 等. 1990–2021 蒙古高原年度丰水期水体数据集[J/OL]. 中国科学数据, 2023, 8(1). (2023–02–02). DOI: 10.11922/11–6035.csd.2022.0061.zh.

数据引用格式

王妍, 刘帅, 单炜. 1990–2021 蒙古高原年度丰水期水体数据集[DS/OL]. Science Data Bank, 2022. (2022–07–26). DOI: 10.57760/sciencedb.01943.

A dataset of the annual water bodies in wet seasons on the Mongolian Plateau during 1990–2021

WANG Yan¹, LIU Shuai¹, QIU Lisha¹, SHAN Wei^{1,2,3*}

1. Institute of Cold Regions Science and Engineering of Northeast Forestry University, Harbin 150040, P. R. China
2. Ministry of Education Observation and Research Station of Permafrost Geo–environment System in Northeast China (MEORS–PGSNEC), Harbin 150040, P. R. China
3. Collaborative Innovation Centre for Permafrost Environment and Road Construction and Maintenance in Northeast China (CIC–PERCM), Harbin 150040, P. R. China

*Email: shanwei456@163.com

Abstract: The Mongolian Plateau is characteristic of fragile ecosystem and serious land desertification. As one of the main climate-varied sensitive regions in Asia, it is of great value for the study on temporal and spatial changes of water resources. For the production of this dataset, we used Google Earth engine (GEE) platform to process the high–quality Landsat series satellite images available in the past 32 years (1990–2021). Using the minimum cloud amount synthesis algorithm, we obtained the minimum cloud amount images in the wet season (from June to September) of each year. After calculating NDWI, we further used the OTSU algorithm for threshold segmentation, and extracted the yearly water body data with 30m resolution on the Mongolian Plateau in the wet season for 32 years. The final data results are saved in GeoTIFF format. Through comparison, the average consistency between this dataset and the JRC annual water body data in both permanent water body and maximum water body is 93.0% and 90.9% respectively, which indicates the high reliability of this dataset. The dataset can provide data support for water resource changes, ecological construction planning, environmental protection, etc. on the Mongolian Plateau.

Keywords: Mongolian Plateau; water body; OTSU algorithm; normalized water body index

Dataset Profile

Title	A dataset of the annual water bodies in wet seasons on the Mongolian Plateau during 1990–2021
Data corresponding author	SHAN Wei (shanwei456@163.com)
Data author(s)	WANG Yan, LIU Shuai, SHAN Wei
Time range	1990–2021
Geographical scope	Inner Mongolia Autonomous Region of China and neighboring Mongolia in the north (37°46′–53°08′N, 87°40′–122°15′E)
Spatial resolution	30 m
Data volume	1.77 GB
Data format	*.tif
Data service system	http://doi.org/10.57760/sciencedb.01943
Source(s) of funding	Heilongjiang Provincial Department of Transportation Science and Technology Project, Study on the Mechanism and Suitability of Plant Slope Reinforcement of Cutting Slope in Permafrost Degraded Areas
Dataset composition	The dataset is composed of 32 periods of spatial distribution data of water bodies in the wet season of the Mongolian Plateau from 1990 to 2021. Each period contains 3 images, a total of 96 sub files. The data are stored in GeoTIFF files named "yyyymongolianwater–1/2/3" (e.g. 1990mongolianwater–1). The spatial resolution of the images is 30m, and the projection system is WGS–84. As the data range is too large, each phase of data is divided into three pieces of sub data for saving. All the data are saved as a file ("1990–2021 Mongolian Plateau wet season water body data set.rar"), with a total data volume of 1.77GB.