

蒙古高原资源环境科学数据产品及验证

王卷乐^{1,2*}, 李 凯^{1,2}, Ochir Altansukh³, 徐书兴^{1,2}, 魏海硕⁴

1. 中国科学院地理科学与资源研究所资源环境与信息系统国家重点实验室, 北京 100101; 2. 中国科学院大学资源与环境学院, 北京 101408; 3. 蒙古国立大学工程与应用科学学院, 乌兰巴托 14201, 蒙古; 4. 北京师范大学地理科学学部, 北京 100875;

摘 要: 蒙古高原是中蒙俄经济走廊绿色发展的关键区域, 对亚洲生态安全具有重要意义。本研究针对该区域面临的土地退化、水资源短缺和沙尘暴频发等资源生态问题, 采用遥感和地理信息系统 (GIS) 技术, 研发了一套蒙古高原资源环境科学数据产品。该数据集包括土地覆盖、春季沙尘暴分布、产草量估算和地表水分布等子数据集, 旨在为区域可持续发展提供科学支撑。数据结果反映了蒙古高原的生态环境。数据产品在精度和可靠性方面均达到了较高标准, 并能够有效支持蒙古高原生态屏障的监测和管理。

关键词: 蒙古高原; 土地覆盖; 沙尘暴; 产草量; 地表水

DOI: <https://doi.org/10.3974/geodp.2024.04.11>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2024.04.11>

1 前言

蒙古高原是“一带一路”中蒙俄经济走廊绿色发展的核心区域, 处于西伯利亚针叶林与亚洲荒漠草原之间的过渡区域, 扮演着中国北方生态屏障的重要角色。这一区域对亚洲文明和地缘格局有着重要影响。然而, 全球气候变暖和人类活动的影响导致该地区面临土地退化、荒漠化、沙尘暴频发、畜牧业无序扩张和水资源短缺等严峻的生态与环境挑战^[1]。蒙古国是全球荒漠化研究的热点之一, 其国土有超过四分之三的面积正遭受不同程度的荒漠化, 并且这一现象正向东蔓延至东方省和肯特省等优质草原地区, 严重阻碍了联合国可持续发展目标 SDG15.3.1 土地退化零增长目标的实现。近来, 中蒙两国遭遇了近十年来最强烈的沙尘暴, 给当地人民的生命财产安全带来了巨大损失。

习近平总书记近年多次考察内蒙古, 强调指出筑牢北方重要的生态安全屏障。2022 年上合峰会, 中蒙俄三国元首表示将中蒙俄经济走廊发展规划建设延长五年。2023 年蒙古国总理访华期间进一步加强了中蒙跨境生态治理的合作意向。然而由于历史原因, 蒙古高原区域长期未受重视且高时空分辨的科学数据积累严重匮乏。随着对地观测和处理技术的快

收稿日期: 2024-09-10; 修订日期: 2024-11-02; 出版日期: 2024-12-24

基金项目: 中华人民共和国科学技术部 (2022YFE0119200); 内蒙古自治区 (2023 KJHZ0027); 资源与环境信息系统国家重点实验室 (KPI006); 中国工程科技知识中心 (CKCEST-2023-1-5)

*通讯作者: 王卷乐, 中国科学院地理科学与资源研究所, wangjl@igsnrr.ac.cn

引用方式: 王卷乐, 李凯, Altansukh, O.等. 蒙古高原资源环境科学数据产品及验证[J]. 全球变化数据学报, 2024, 8(4): 442–448. <https://doi.org/10.3974/geodp.2024.04.11>. <https://cstr.escience.org.cn/CSTR:20146.14.2024.04.11>.

速发展,开展精细化的蒙古高原资源环境科学数据的研发和生产成为可能。

本研究面向大数据驱动的地理科学研究范式变革,以蒙古高原生态屏障智能计算的需求为牵引,打通数据-模型-产品-场景之间的链条,解决支撑大区域地理单元智能计算的关键数据产品反演算法、多源数据融合、知识发现场景应用问题。实现蒙古高原高分辨率地表特征参量数据产品反演算法、蒙古高原生态安全相关产品的研发方法,提高我国自主数据产品的保障能力,发挥其对蒙古高原生态屏障建设的支撑作用。

2 数据研发方法

针对蒙古高原生态屏障的干旱半干旱特点,从植被、水土条件、生态环境等方面,筛选土地覆被、地表水、植被覆盖度、叶面积指数、地表温度、草地生物量、植被供水指数、土壤湿度等关键指标,开展特征参量计算的源数据指标一致性、规范性分析与设计。设计关键特征参量智能计算数据产品共享、精度验证与质量评价的协同技术方案,构建一个精准和可共享的蒙古高原关键地表参量数据产品体系(图1)。

对比分析各关键参数现有监测算法特点,基于遥感和GIS空间分析算法,构建基于图像分割算法的土地覆盖生产与变化监测模型、优化验证地表温度反演劈窗算法。建立各特征参量智能分类和计算的土地覆被类型、草地生物量和叶面积指数(LAI)、土壤水分的训练和验证样本集。构建改进的基于卷积神经网络的地表水体提取模型、基于植被指数和神经网络的LAI反演模型、基于随机森林的植被覆盖度重构模型、基于机器学习的草地生物量估算模型。

结合国际和国内卫星数据源的筛选和可调用资源支持,实施长时间序列、高分辨率的关键特征参量数据产品计算。开展数据产品协同验证,完成关键特征参量产品的智能生成,并开放发布与共享,允许其他人员或机器调用本数据产品。

3 数据结果与验证

3.1 数据集组成

蒙古高原资源环境科学数据产品由五个子数据集组成,包括基于多源数据和丰富样本标记的蒙古高原长时序土地覆盖数据集、揭示蒙古国土地覆盖变化的可更新数据集、蒙古高原春季沙尘暴分布数据集(2000–2021年)、蒙古国30 m分辨率产草量估算数据集(2017–2021年)以及2013–2022年蒙古高原逐年生长季地表水分布数据集。

3.2 数据结果

针对蒙古高原的森林、灌木、草甸草地、典型草地、荒漠草地、湿地、水体、耕地、建筑、裸地、沙漠、沙地以及冰雪13个主要类别,作者团队构建覆盖蒙古高原的等经纬度渔网文件,参考遥感影像,在每个渔网网格中目视解译标记标签数据43,223个。2020、2015、2010、2005、2000、1995、1990年分别采集了11,295、4,521、4,887、5,794、4,459、9,807以及2,460个样点,其中90%的样点用于模型训练,10%的样点用于产品的精度验证,采用Google Earth Engine作为数据采集和模型训练的平台。集成历年的样本点数据,将1990至2020年7个时期的影像和数据进行统一。构建包含蓝、绿、红、近红外、短波红外1、

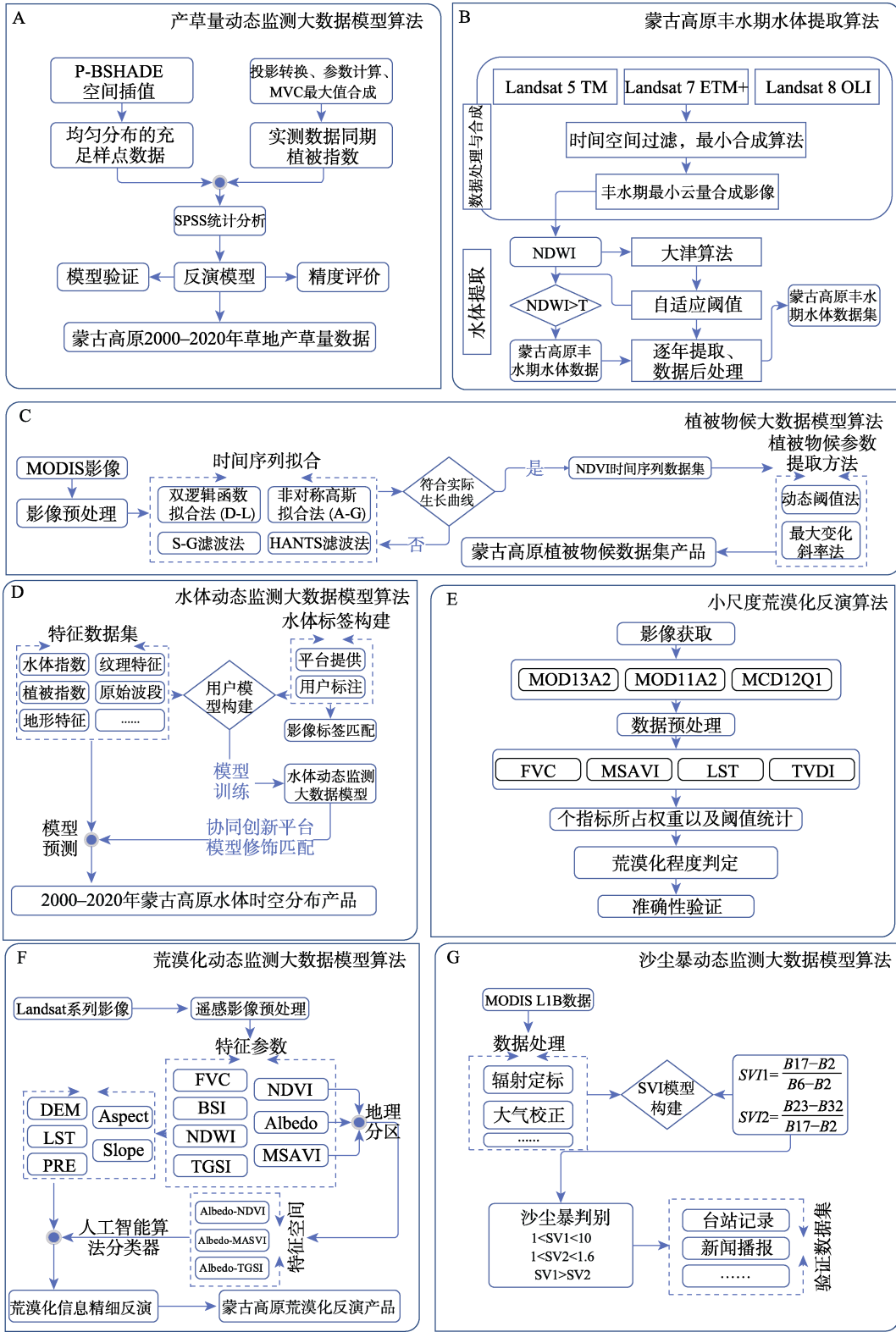


图 1 蒙古高原资源环境科学数据产品研发技术路线图

短波红外 2、NDVI、NDWI、高程、坡度以及夜间灯光数据的特征数据集，将特征像元值与标签值一一对应。采用随机森林算法对训练点集进行训练，其中子树的数量设置为 100 个。将训练好的随机森林分类器应用到预测影像集，实现蒙古高原土地覆盖数据产品的生产（图 2）。

作者使用 eCognition 进行面向对象的遥感解释。将森林、草甸草地、典型草地、荒漠草地按照归一化植被指数，裸地按照归一化土壤指数，沙地按照各波段总和，耕地按照紧凑度，水体按照归一化水体指数进行参考阈值的设定，再结合人工目视判别调整阈值识别沙漠、建筑和冰雪^[2]。最后构建了 11 个类别的蒙古国土地覆盖数据（图 3）。

利用 MODIS L1B 数据构建归一化尘埃指数、热红外沙尘指数、亮温差指数算法以及无阈值法沙尘指数（DSDI）^[3]。发现 DSDI 指数可有效提取沙尘像元，且对于所有的沙尘事件，DSDI 值均大于 0，有效规避了阈值差异问题，适用于蒙古高原研究时空尺度。将该指数应用于 2000–2021 年春季影像，获得蒙古高原近 20 多年春季沙尘分布（图 4）。

基于数据源获得所需的遥感图像、实测土地等数据，结合上章制作的蒙古高原土地覆盖数据提取蒙古国草地范围，并使用 GEE 平台进行预处理，生成所需的训练数据集。对比了四种模型，分别是多元线性回归、随机森林、K 近邻和人工神经网络，并选择表现最佳的产草量估算模型，将其应用于产草量的估算^[4]（图 5）。

构建包含红、绿、蓝、近红外、归一化水体指数、短波红外、水体指数增强波段、数字高程模型等多通道合成的特征数据。将标签噪声纠正方法应用于质量评估波段的水体信息得到校正的水体标签，并与特征数据结合构建训练数据集。在本地训练得到基于深度学

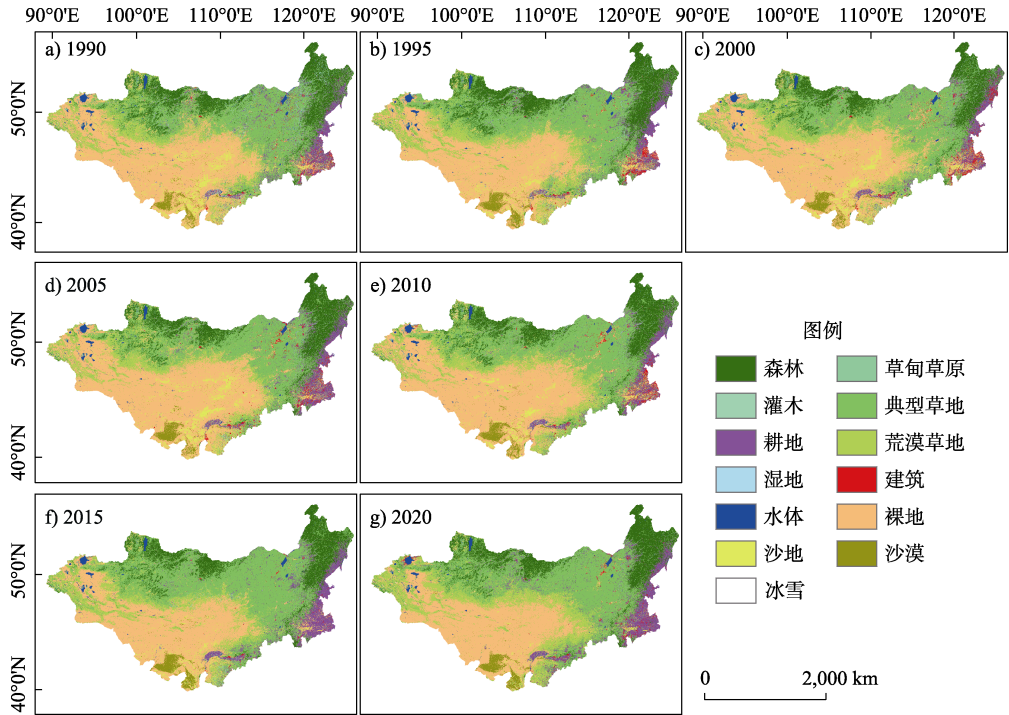


图 2 蒙古高原土地覆盖图（1990–2020）

（该图基于国家自然资源部地图技术审查中心标准地图（审图号: GS（2024）0650 号）绘制，底图无修改）

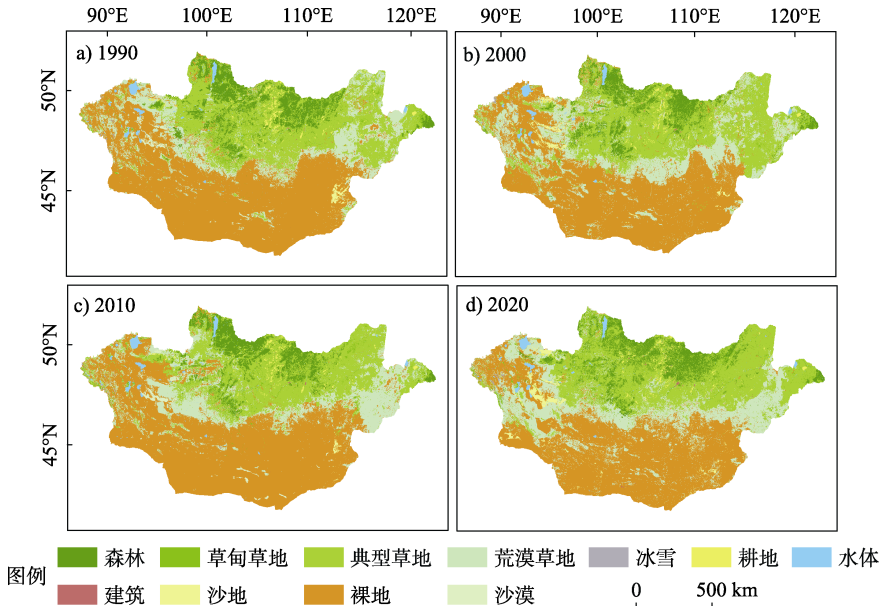


图 3 蒙古国土地覆盖图（1990–2020）

（该图基于国家自然资源部地图技术审查中心标准地图（审图号：GS（2024）0650 号）绘制，底图无修改）

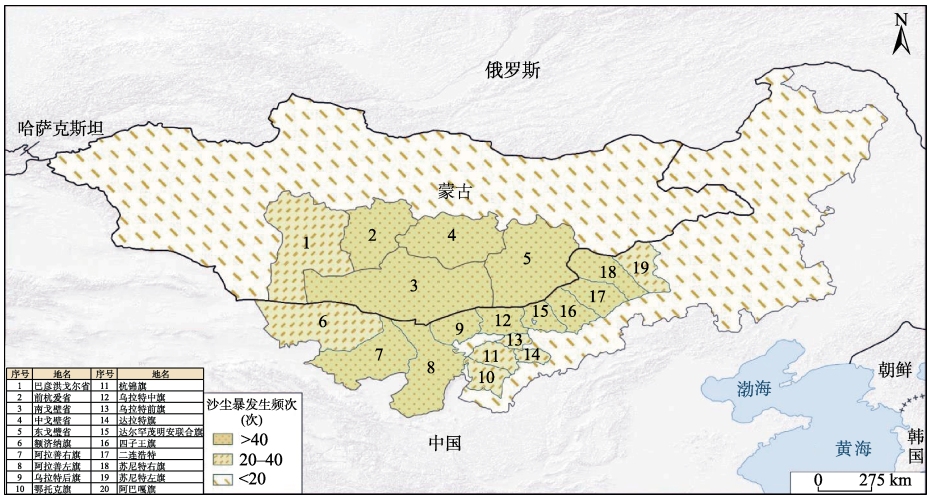


图 4 蒙古高原 2000–2021 年春季沙尘暴频次分布图

（该图基于国家自然资源部地图技术审查中心标准地图（审图号：GS（2024）0650 号）绘制，底图无修改）

习的水体分类模型^[5]。采用本地深度学习训练和 Google Earth Engine 分布式计算相结合的方法，通过解析深度学习模型结构和 GEE 的函数接口，实现深度学习对 GEE 的赋能^[6]，将水体分类模型自动化部署在 GEE 中在线计算，应用于 2013 至 2022 年逐年的特征影像，完成蒙古高原 2013–2022 年逐年生长季地表水分布的获取（图 6）。

3.3 数据结果验证

蒙古高原土地覆盖：1990 至 2020 年 7 期数据共采集验证样点 4,383 个。经计算，总体

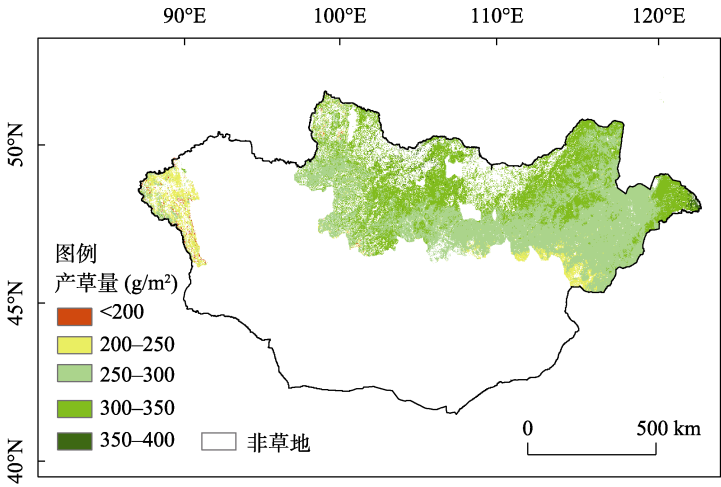


图 5 蒙古国产草量时空分布图（2021）

（该图基于国家自然资源部地图技术审查中心标准地图（审图号：GS（2024）0650 号）绘制，底图无修改）

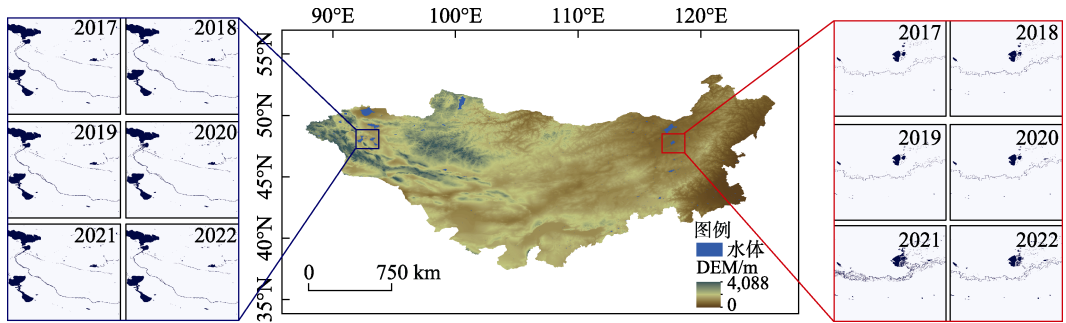


图 6 部分年份蒙古高原地表水时空变化分布图（2017–2022）

（该图基于国家自然资源部地图技术审查中心标准地图（审图号：GS（2024）0650 号）绘制，底图无修改）

验证精度为 83.9%，Kappa 系数为 0.817，精确度均值为 86.1%，召回率均值为 81.4%，加权 F1 得分为 84.0%。从不同年份来看，1990、1995、2000、2005、2010、2015 以及 2020 的总体精度分别为 80.9%、73.4%、78.2%、85.5%、67.6%、94.5%以及 97.7%。Kappa 系数分别为 0.78、0.70、0.75、0.83、0.62、0.94 以及 0.97。

蒙古国土地覆盖：1990 年、2000 年、2010 年和 2020 年土地覆被数据的总体分类精度分别为 84.19%、82.12%、81.84%和 81.84%。1990 年、2000 年、2010 年和 2020 年土地覆被数据的 Kappa 系数分别为 0.805,2、0.765,6、0.798,5 和 0.799,1^[2]。

蒙古高原沙尘暴分布：在确定训练样本点的真实值后，通过误差矩阵评估沙尘探测指数的提取精度。DSDI 沙尘暴探测指数提取的蒙古高原近春季沙尘分布的总体分类精度可达 85.24%，Kappa 系数为 0.763,6^[7]。

蒙古国产草量：将人工神经网络、随机森林、K-最邻近和多源线性回归模型四种模型方法进行比对。其中人工神经网络模型（ $R^2=0.78$ ， $RMSE=48.7\text{ g/m}^2$ ）和随机森林模型（ $R^2=0.72$ ， $RMSE=55.28\text{ g/m}^2$ ）的准确性明显高于其他两种模型，并且两种模型均可用于蒙古国的产草量估算。K-最近邻模型的准确性略低于上述两种模型，而多元线性回归模型

只能表征 40% 的方差, 略高于单植被指数的统计模型。因此选择使用人工神经网络来估算蒙古国的产草量^[8]。

蒙古高原地表水: 在 Google Earth 中, 对 2013–2022 年进行逐年验证点的人工筛选, 10 年共 5,000 个验证样点 (限制每年水体样点数量为 200 个, 非水体 300 个)。水体和非水体样点总计分别为 2,000 和 3,000 个。经计算, 总体精度优于 86%, 平均 Kappa 系数为 0.75^[9]。

4 总结

本研究针对蒙古高原这一“一带一路”中蒙俄经济走廊的核心区域, 针对区域生态环境问题, 提出了以地球大数据技术为核心的解决方案。研究利用遥感技术和地理信息系统, 开发了包括土地覆盖、春季沙尘暴分布、产草量估算和地表水分布等在内的多种数据集。这些数据集涵盖了蒙古高原的主要生态特征, 并采用了先进的算法和技术, 如随机森林算法、深度学习模型, 以及面向对象的遥感解译, 以确保数据的高精度和高可靠性。最终形成的数据集不仅能为蒙古高原的生态环境监测提供科学依据, 还为中蒙俄三国的跨境生态治理合作提供了重要的支撑。成果在 UNESCO “防灾减灾知识服务系统” “地球大数据支撑联合国可持续发展目标” 案例报告、国家遥感中心 “地球观测优秀应用案例 100 例” 等得到应用。

作者分工: 王卷乐对数据集的开发做了总体设计并撰写了论文; 李凯获取了蒙古高原的地表水数据, 共同撰写了论文; Ochir Altansukh 支撑了野外考察和验证数据的采集; 徐书兴、魏海硕获取了蒙古高原的土地覆盖数据。

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] 王卷乐. 《蒙古高原资源与环境数据专题》卷首语[J]. 中国科学数据, 2023, 8(1): 7.
- [2] Wang, J. L., Wei, H. S., Cheng, K., *et al.* Updatable dataset revealing decade changes in land cover types in Mongolia [J]. *Geoscience Data Journal*, 2022, 9(2): 341–345. DOI: 10.1002/gdj3.149.
- [3] Zhang, Y., Wang, J. L., Altansukh, O., *et al.* Dynamic evolution of spring sand and dust storms and cross-border response in Mongolian plateau from 2000 to 2021 [J]. *International Journal of Digital Earth*, 2023, 16(1): 2341–2355.
- [4] Li, M. H., Wang, J. L., Li, K., *et al.* Spatial-temporal pattern analysis of grassland yield in Mongolian Plateau based on artificial neural network [J]. *Remote Sensing*, 2023, 15(16): 19. DOI: 10.3390/rs15163968.
- [5] 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(7): 102497. DOI: 10.1016/j.jag.2021.102497.
- [6] 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.
- [7] 张煜, 王卷乐. 蒙古高原春季沙尘暴分布数据集 (2000–2021 年) [J/OL]. 中国科学数据, 2023, 8(1): 123–133. DOI: 10.11922/11-6035.csd.2023.0032.zh.
- [8] 李梦晗, 王卷乐, 李凯. 蒙古国 30 米分辨率产草量估算数据集 (2017–2021 年) [J/OL]. 中国科学数据, 2023, 8(1): 14–22. DOI: 10.11922/11-6035.csd.2023.0006.zh.
- [9] 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. DOI: 10.57760/sciencedb.j00001.00665.