

基于 Landsat 的胶州湾悬浮物浓度反演数据集 (1984–2020)

黄 珏¹, 张 祥¹, 赵永芳^{2,*}

1. 山东科技大学测绘与空间信息学院, 青岛 266590; 2. 中国科学院海洋研究所胶州湾生态站, 青岛 266071

摘 要: 悬浮物 (Total suspended solids, TSS) 浓度 (mg/L) 是近岸和内陆湖泊的重要水环境参数。基于 1984–2020 年 Landsat TM/ETM /OLI 影像和“留一法交叉验证”方法使用 22 组遥感反射率和 TSS 浓度数据进行建模, 剩余 1 组数据进行验证, 上述过程依次进行, 共得到 23 个模型, 并计算决定系数 (R^2) 和均方根误差 (RMSE), 最终选择精度最高的模型作为 TSS 浓度的定量反演模型, 生成了胶州湾时间序列的 TSS 产品数据集。数据集内容包括: (1) 研究区边界和胶州湾 1984、2020 年边界数据; (2) 318 幅 TSS 数据, 空间分辨率为 30 m。数据集存储为.shp 和.tif 格式, 由 341 个数据文件组成, 数据量为 236 MB (压缩为 1 个文件 232 MB)。

关键词: 胶州湾; Landsat; 悬浮物浓度; 定量反演; 时间序列

DOI: <https://doi.org/10.3974/geodp.2022.03.14>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2022.03.14>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2022.02.09.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2022.02.09.V1>.

1 前言

胶州湾 (JiaoZhou Bay, JZB) 位于中国黄海中部、山东半岛南岸、青岛市市区与黄岛区之间, 大约在 120°04'E 到 120°23'E, 35°38'N 到 36°18'N 之间 (图 1), 为半封闭海湾, 近似喇叭状, 出口朝东。胶州湾面积近 370 km², 且入口窄、航道深、浊度低、潮汐弱、海浪低, 作为青岛市的天然港口促进了青岛市交通运输产业的发展^[1]。湾口以团岛南端、薛家岛北端为界, 宽约 2.5 km。湾内中部东西宽约 27.8 km, 南北向最大长度约 33.3 km, 海岸线长约为 163 km, 滩涂面积约为 125 km²。然而, 与大多数毗邻城市化地区的海湾一样, 为了解决土地供需矛盾, 胶州湾的沿海地区也发生了很大的变化。根据以往的研究, 胶州湾的面积比 1935 年缩小了 36.6%, 海岸线长度缩短了 30.4%^[2]。

收稿日期: 2022-01-20; 修订日期: 2022-06-20; 出版日期: 2022-09-25

基金项目: 国家自然科学基金 (42076185, 41706194); 山东科技大学科研创新团队支持计划 (2019TDJH103)。

*通讯作者: 赵永芳, 中国科学院海洋研究所胶州湾生态站, yfzhao@qdio.ac.cn

数据引用方式: [1] 黄珏, 张祥, 赵永芳. 基于 Landsat 的胶州湾悬浮物浓度反演数据集 (1984–2020) [J]. 全球变化数据学报, 2022, 6(3): 433–439. <https://doi.org/10.3974/geodp.2022.03.14>. <https://cstr.escience.org.cn/CSTR:20146.14.2022.03.14>.

[2] 黄珏, 张祥, 赵永芳. 基于 Landsat 的胶州湾悬浮物浓度反演数据集 (1984–2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.02.09.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.02.09.V1>.

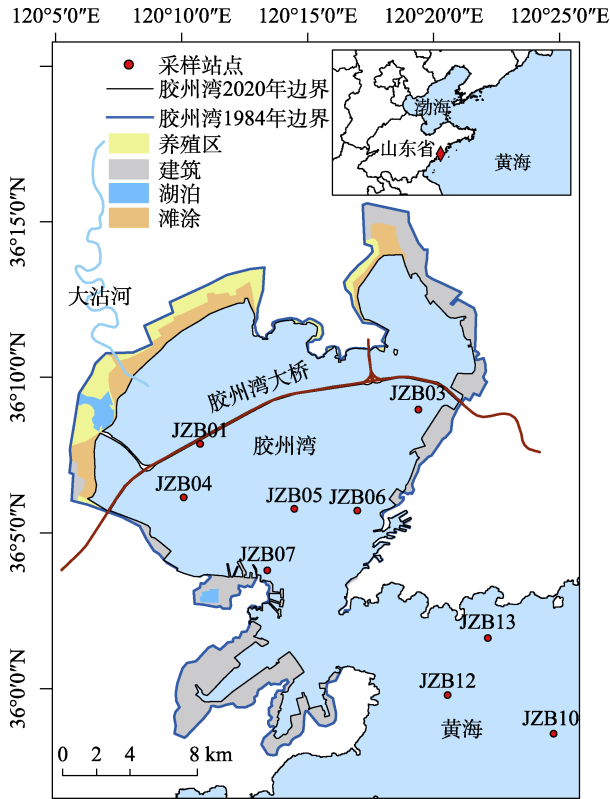


图 1 胶州湾位置图

作为近岸和内陆湖泊水环境质量检测的重要参数之一，悬浮物（total suspended solids, TSS）与水体中的许多物理化学过程有着重要的关系。一方面，TSS 浓度作为常见的水质参数，对于污染程度的指示以及水土流失的准确估算有重要作用；另一方面，悬浮物减少了进入水体的光能量，影响了水生植物的生长，制约着浮游植物的初级生产力^[3]。

近年来，随着人类生产的不断进步和生活水平的不断提高，海洋河流湖泊等水资源受到严重破坏，TSS 浓度的监测成为当前环境改善的一个重要环节。传统测量 TSS 浓度的方法是采用室内仪器对实测站点采集的水样进行测量^[4]，这种方法不仅费时费力，而且测站数量有限，对于人和船无法到达的区域无能为力，无法实现区域性和周期性的测量。而遥感手段具有省时省力、成本低、周期短的优势，可以同步获取大面积水域的遥感信息，实现实时、宏观、动态的水环境监测^[5]。

虽然胶州湾及附近海域的 TSS 变化一直以来是科学家和管理者关注的问题，但胶州湾 TSS 长时间序列数据依然缺乏。因此，本研究基于实测数据和 Landsat 遥感数据建立了 TSS 浓度的定量反演模型，并得到了 1984–2020 年胶州湾的 TSS 数据集，为水色遥感和水环境监测等方面提供数据支撑。

2 数据集元数据简介

《基于 Landsat 的胶州湾悬浮物浓度反演数据集（1984–2020）》^[6]的名称、作者、地理区域、数据年代、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等

信息见表 1。

表 1 《基于 Landsat 胶州湾悬浮物浓度反演数据集（1984-2020）》元数据简表

条目	描述
数据集名称	基于 Landsat 的胶州湾悬浮物浓度反演数据集（1984-2020）
数据集短名	JZBay_TSS_1984-2020
作者信息	黄珏, 山东科技大学测绘与空间信息学院, huangjue@sdust.edu.cn 张祥 AAG-5860-2022, 山东科技大学测绘与空间信息学院, zhangxiang0112@sdust.edu.cn 赵永芳, 中国科学院海洋研究所胶州湾生态站, yfzhao@qdio.ac.cn
地理区域	青岛市胶州湾区域（35°55'N–36°18'N, 120°04'E–120°22'E）
数据年代	1984–2020 年
空间分辨率	30 m
数据格式	.shp、.tif
数据量	232 MB（压缩后）
数据集组成	包括胶州湾 1984 和 2020 年的边界矢量数据, 以及 Landsat 的 3 个传感器反演得到的 1984–2020 年 318 幅胶州湾悬浮物浓度数据。“LandsatTMTSS”包含 144 幅数据, “LandsatETM+ TSS”包含 120 幅数据, “LandsatOLITSS”包含 54 幅数据
基金项目	国家自然科学基金（42076185, 41706194）；中华人民共和国科学技术部 CNERN 项目
数据计算环境	Google Earth Engine
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《全球变化数据学报（中英文）》发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议, 获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源 ^[7]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

3 数据研发方法

3.1 胶州湾矢量数据

本研究做出了研究区边界和胶州湾 1984、2020 年的边界矢量数据。由于近年来填海造陆的影响, 胶州湾的面积一直在减小, 边界一直在变化, 因此采用 2020 年边界, 保证研究区范围内全是海水。胶州湾 2020 年边界在北部距离大陆有一定的距离, 是因为胶州湾北部多为滩涂, 且水深较浅, 而底质泥沙的高反射率会严重影响 TSS 反演的结果, 因此剔除了受泥沙严重影响的区域, 最终确定了此边界作为“反演边界”, 尽量减少滩涂的影响。此外,

经过测试，胶州湾大桥与海水有明显的反射率差异，导致 TSS 反演结果不正确。为了减少胶州湾大桥对 TSS 反演的影响，本研究也对胶州湾大桥进行了掩膜处理。

3.2 实测数据

本研究从胶州湾海洋生态系统研究站 (<http://jzw.qdio.cas.cn/>) 共获得 2011–2015 年间的 23 组实地测量数据。其 TSS 浓度根据重量法测量得到。水样首先用干燥和预称重的 0.45 μm 滤膜过滤。接下来，将滤膜在 45 $^{\circ}\text{C}$ 下干燥 24 小时，然后称重。重复干燥过程，直到连续称重的 TSS 值之差小于 0.01 mg/L。采样点位置信息见图 1。

3.3 Landsat 数据及预处理

本研究使用了来自 Google Earth Engine(GEE)平台的 Landsat 反射率数据,包括 Theme Mapper (TM)、Enhanced Theme Mapper Plus (ETM+) 和 Operating Land Imager (OLI)。为了质量控制,排除了云量大于 10%的图像。GEE 平台上的 Landsat TM 和 ETM+反射率数据的大气校正方法为 LEDAPS(Landsat Ecosystem Disturbance Adaptive Processing System), Landsat OLI 反射率数据的大气校正方法为 LaSRC(Landsat Land Surface Reflectance Code)。本研究利用质量评估 (Quality Assessment, QA) 波段对 Landsat 卫星影像上的云和云阴影进行检测,同时生成云掩模文件,用于云的去除。本研究使用的 Landsat 卫星数据来自三个不同的传感器,尽管 TM/ETM+和 OLI 具有相似的光谱波段,但这三个传感器获取的影像反射率之间仍然存在差异。

因此,利用经验线方法^[8]建立了 OLI 与 TM/ETM+的反射率之间的线性关系,将 LandsatTM/ETM+影像的反射率调整到与 OLI 反射率相同的水平。2003 年 5 月 31 日, LandsatETM+机载扫描行校正器 (scan line corrector, SLC) 故障,导致此后获取的影像出现了数据条带丢失,严重影响了 Landsat ETM+遥感影像的使用。因此,本文采用焦点均值方法对其进行了去条带处理。

3.4 TSS 遥感定量反演模型的构建

本文首先进行了波段敏感性分析,发现红光波段与 TSS 浓度相关性最高,因此,本研究将红光波段与其他波段组合作为遥感因子,与 TSS 浓度进行了相关性分析。表 2 展示了 8 个与 TSS 浓度相关性最好的波段组合,可以看出,相关性最高的波段组合为(G+R)/(G/R) (G 为绿光波段遥感反射率, R 为红光波段遥感反射率)。因此,本研究将上述波段组合用

表 2 不同波段组合与 TSS 相关性分析

波段组合*	相关性系数
(G+R)/(G/R)	0.82
G+R	0.80
NIR/R	0.79
(G+R)/(B/R)	0.79
(R - B)/(R+B)	0.77
B+R	0.76
B+NIR	0.73
R/G	0.68

*: B, G, R 和 NIR 分别表示蓝光波段, 绿光波段, 红光波段和近红外波段的反射率。

于 TSS 浓度定量反演模型的建立。

本研究采用了一种可靠的“留一法交叉验证”方法对模型进行了校准和验证^[9]。使用 22 组遥感反射率和 TSS 浓度数据进行建模，剩余 1 组数据进行验证。上述过程依次进行，共得到 23 个模型，并计算决定系数（ R^2 ）和均方根误差（RMSE），最终选择精度最高的模型作为 TSS 浓度的反演模型。本研究提出的模型获得了最佳性能，反演模型如下所示：

$$Y=0.71\times\exp(21.31X) \tag{1}$$

式中，对于 Landsat 5 和 Landsat 7，式中 X 为波段组合 $(B2+B3)/(B2/B3)$ ；对于 Landsat 8，式中 X 为波段组合 $(B3+B4)/(B3/B4)$ ； Y 为 TSS 浓度。B2，B3 和 B4 是 Landsat 对应波段的反射率。

4 数据结果与验证

4.1 数据集组成

经过上述数据处理，将反演模型应用到 Landsat 影像上，最终得到了 1984–2020 年 318 幅胶州湾悬浮物浓度反演数据集，数据格式为.tif，空间分辨率为 30 m，单位为 mg/L。其中，TM 有 144 个文件，时间范围为 1984–2011 年；ETM+有 120 个文件，时间范围为 1999–2020 年；OLI 有 54 个文件，时间范围为 2013–2020 年。命名规则为 XXXX_120035_YYYYMMDD.tif，其中 XXXX 代表 Landsat 不同传感器（LT05、LE07 和 LC08），120035 代表胶州湾所在影像的行列号，YYYYMMDD 代表年月日。不同传感器反演得到的 TSS 可以混合使用。除此之外，本数据集还包括了研究区边界和胶州湾 1984、2020 年的边界矢量数据（.shp 格式）。

4.2 数据结果

以胶州湾 2013 年 8 月 19 日的 TSS 分布图为例，如图 2 所示。图中所缺像元为去云导致，一般不会影响使用。胶州湾 TSS 浓度主要在 0–150 mg/L 范围内变化，为低浊度海域，呈现从西北到东南逐渐下降的分布规律。河口区域及近岸区域 TSS 浓度多高于 30 mg/L，中心海域 TSS 浓度基本低于 30 mg/L。

图 3 显示了 1984–2020 年 Landsat 数据反演的年平均悬浮物浓度变化。年平均悬浮物浓度在 1984–2018 年内变化起伏不定，1993 年达到最大值（26.94 mg/L），2016 年降到最小值（10.69 mg/L）。但整体上呈现下降的趋势，与 Gao 等^[10]的研究一致。

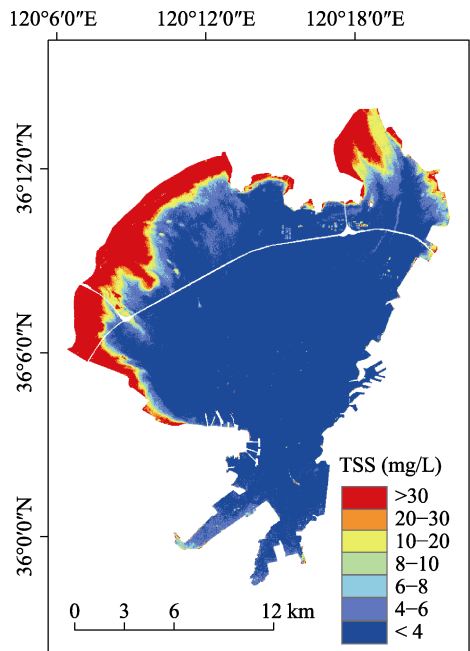


图 2 胶州湾 2013 年 8 月 19 日 TSS 分布图

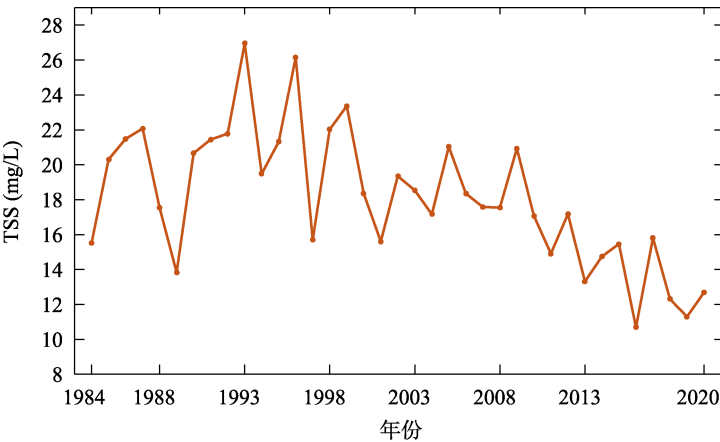


图3 1984–2020 年胶州湾年平均悬浮物浓度的变化

4.3 数据结果验证

在 TSS 遥感定量模型的构建中，采用“留一法交叉验证”最大限度地减少因实测数据数量较少而产生的随机因素影响。图 4 显示了 TSS 实测值与预测值之间的相关性。该模型具有较高的解释能力和预测能力, R^2 为 0.77, RMSE 为 1.82 mg/L,趋势线斜率为 1.07,与 1 : 1 趋势线基本一致。散点均匀分布在趋势线两侧，所有预测的 TSS 值均在实测数据范围内。

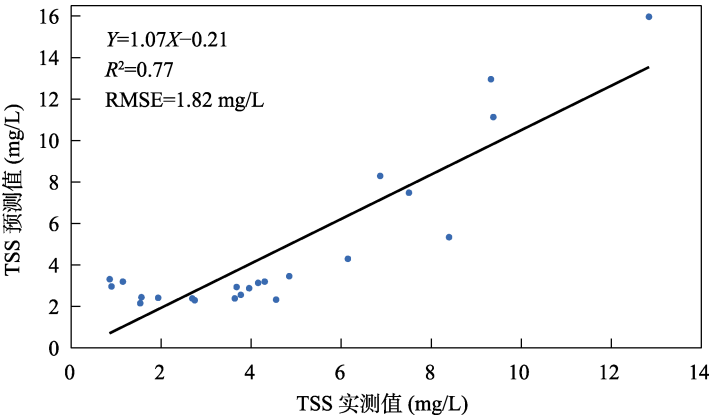


图4 TSS 实测值与预测值之间的关系

5 讨论和总结

本数据集为基于 Landsat 的 1984–2020 年胶州湾悬浮物浓度反演数据集，空间分辨率为 30 m，单位为 mg/L。数据集资料来源可靠，主要基于 GEE 等平台生产完成，数据格式为 TIF，用户可以借助 ArcGIS 或 ENVI 等图像处理软件打开使用。若需要其他格式，可以在上述软件中进行格式转换。

值得说明的是，本数据集可能存在一定的不确定性：由于现场实测数据有限，导致现

场实测数据与卫星数据的采集时间存在差异, 不可避免地给反演模型带来了不确定性。其次, 由于缺乏近岸的观测站点, 实测数据的最高浓度为 13 mg/L, 但反演结果中小部分 TSS 浓度高于 30 mg/L。基于较窄的浓度动态范围而开发的反演模型也会引入一定的不确定性。然后, 由于受胶州湾夏季多云天气的影响, 夏季可用的影像数据相对其他季节较少。最后, 胶州湾北部沿岸因海水较浅, 受底质泥沙高反射率的影响较大, 多显示出较高的 TSS 浓度值, 这可能也与实际情况会有一定的差距。

长期动态的 TSS 数据集可以有效地为管理部门提供数据支持, 加快改善水质。本数据集提供了 1984–2020 年胶州湾 TSS 浓度产品, 数据量大, 时间范围广, 为研究胶州湾 TSS 时空分布变化规律及影响因素等水色遥感问题提供了重要的数据支撑, 也为水环境监测和环境保护政策的制定提供支持。

作者分工: 黄珏对数据集的开发做了总体设计, 参与模型设计和论文的撰写与修改; 赵永芳采集实测数据并对研究做了总体指导; 张祥处理了基础数据, 参与模型开发, 做了数据验证以及撰写了数据论文等。

致谢: 感谢胶州湾海洋生态系统国家野外科学观测研究站提供的胶州湾悬浮物浓度实测数据。

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] 杨广普, 江涛, 赵永芳等. 基于长时序遥感影像的胶州湾秋季叶绿素 a 浓度变化及其影响因素研究[J]. 海洋学报, 2019, 41(1): 183–190.
- [2] Yuan, Y., Jalón-Rojas, I., Wang, X. H., *et al.* Design, construction and application of a regional ocean database: a case study in Jiaozhou Bay, China [J]. *Limnology and Oceanography*, 2019, 17(3): 210–222.
- [3] Zhang, Y., Wu, Z., Liu, M. Thermal structure and response to long-term climatic changes in Lake Qian-daohu, a deep subtropical reservoir in China [J]. *Limnology and Oceanography*, 2014, 59(4): 1193–1202.
- [4] 张铭汉. 胶州湾海水中悬浮体的分布及其季节变化[J]. 海洋科学集刊, 2000(00): 49–54.
- [5] Zhang, X., Song, Y., Chen, J. J., *et al.* Landsat image-based retrieval and analysis of spatiotemporal variation of total suspended solid concentration in Jiaozhou Bay, China [J]. *Remote Sensing*, 2021, 13(23): 4796–4796.
- [6] 黄珏, 张祥, 赵永芳. 基于 Landsat 的胶州湾悬浮物浓度反演数据集 (1984–2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2022. <https://doi.org/10.3974/geodb.2022.02.09.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2022.02.09.V1>.
- [7] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策 [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [8] Han, X., X., Chen, X. L., Feng, L., *et al.* Four decades of winter wetland changes in Poyang Lake based on Landsat observations between 1973 and 2013 [J]. *Remote Sensing of Environment*, 2015, 156:426–437.
- [9] Huang, J., Wu, M., Cui, T. W., *et al.* Quantifying DOC and its controlling factors in major Arctic rivers during ice-free conditions using Sentinel-2 data [J]. *Remote Sensing*, 2019, 11(24): 2904.
- [10] Gao, G., Wang, X., Bao, X., *et al.* The impacts of land reclamation on suspended-sediment dynamics in Jiaozhou Bay, Qingdao, China [J]. *Estuarine Coastal and Shelf Science*, 2018, 206: 61–75.