

HY-1C/1D 卫星 COCTS 数据叶绿素 a 浓度反演神经网络算法研究

韦美怡¹, 吴璟瑜¹, 郑鹭飞¹, 王道生¹, 李忠平¹, 商少凌¹,
叶小敏², 林供¹, 余小龙¹

1. 厦门大学 海洋生物地球化学全国重点实验室, 厦门 361102;

2. 国家卫星海洋应用中心, 北京 100081

摘要: 水体叶绿素 a 浓度 (Chl-a) 作为核心气候参数, 其精准的遥感反演对于全球碳循环研究、海洋环境监测和治理具有重要意义。中国海洋一号系列卫星 (HY-1C/1D) 搭载的海洋水色水温扫描仪 (COCTS), 可实现公里级全球海洋水色监测。研发针对 COCTS 数据的高精度 Chl-a 遥感算法, 是实现自主卫星数据支撑海洋监测与气候变化研究的核心前提。本研究基于覆盖全球海域的 2165 组实测数据, 构建了基于 HY-1C/1D COCTS 多波段遥感反射比 (R_{rs}) 与环境参数的多层感知器神经网络 (MLP-NN) 模型, 实现全球水体 Chl-a 浓度高精度反演。通过输入因子敏感性分析, 最终确定以 412 nm、443 nm、490 nm、520 nm、565 nm、670 nm 波段的 R_{rs} 、纬度、月份、前月月平均海表面温度和前月气候态光合有效辐射参数构成最优输入组合。验证结果表明, MLP-NN 模型估算的 Chl-a 浓度 log 转换后计算的均方根偏差 RMSD (Root Mean Square Difference) 为 0.22, 中位数绝对百分偏差 MAPD (Median Absolute Percentage Difference) 为 29.1%, 较 NASA 业务化标准的水色指数算法 (OCI) 分别降低了 0.1 和 16.9%。进一步的卫星与实测数据的匹配验证结果显示, MLP-NN 模型估算的 Chl-a 的 RMSD 和 MAPD 分别较 OCI 算法降低了 0.09 和 9.8%, 展现出优异的算法稳健性。本研究为国产卫星在海洋生态监测领域提供了一种创新方法, 显著增强了利用 HY-1C/1D 卫星观测全球 Chl-a 浓度的能力。

关键词: 叶绿素 a (Chl-a), 遥感反射比, 反演算法, HY-1C/1D 卫星, 神经网络, COCTS, 海洋水色

中图分类号: P2

引用格式: 韦美怡, 吴璟瑜, 郑鹭飞, 王道生, 李忠平, 商少凌, 叶小敏, 林供, 余小龙. 2025. HY-1C/1D 卫星 COCTS 数据叶绿素 a 浓度反演神经网络算法研究. 遥感学报, 29(12): 3459-3474

Wei M Y, Wu J Y, Zheng L F, Wang D S, Lee Z P, Shang S L, Ye X M, Lin G and Yu X L. 2025. Retrieving Chlorophyll-a concentration by using a neural network for COCTS sensors aboard HY-1C/1D satellites. National Remote Sensing Bulletin, 29(12):3459-3474 [DOI:10.11834/jrs.20255079]

1 引言

叶绿素 a (Chlorophyll-a, Chl-a) 浓度是估算浮游植物生物量和海洋初级生产力的关键指标 (Huot 等, 2007), 对海洋碳循环、生态系统乃至全球气候变化都具有深远的影响 (Behrenfeld 等, 2005; Carvalho 等, 2013; Henson 等, 2010)。准确获取 Chl-a 时空分布特征, 有助于海洋生态系统健康评估, 并为气候政策制定提供科学依据。近期, 中国自主研制的海洋一号 C 星 (HY-1C) 和

海洋一号 D 星 (HY-1D) 搭载的 COCTS (Chinese Ocean Color and Temperature Scanner) 能够获取可靠的海洋遥感数据 (Chen 等, 2021; 刘建强 等, 2023), 为 Chl-a 浓度和海表面温度等关键参数的提取提供了重要支持 (蒋兴伟 等, 2019; 史鑫皓 等, 2023)。

传统的 Chl-a 浓度测量依赖于现场采样和实验室分析, 通常采用分光光度法、荧光法和高效液相色谱法 (Ritchie, 2006; Holm-Hansen 等, 1965; Holm-Hansen 和 Riemann, 1978; Wright 等,

收稿日期: 2025-03-06; 预印本: 2025-06-24

基金项目: 国家重点研发计划 (编号: 2022YFC3104900, 2022YFC3104903)

第一作者简介: 韦美怡, 研究方向为海洋水色遥感。E-mail: mywei@stu.xmu.edu.cn

通信作者简介: 余小龙, 研究方向为水色遥感机理与应用。E-mail: xlyu@xmu.edu.cn

1991)。尽管这些实测方法测量精度高,但在空间覆盖范围和时效性方面存在一定局限性。相比之下,星载水色遥感能够大范围、准实时监测海洋生态环境,可有效弥补现场采样的不足(宋庆君等, 2025)。卫星反演 Chl-a 浓度主要方法包括经验算法和半解析模型。其中,经验算法依赖于光谱波段组合与 Chl-a 浓度之间的统计关系,常见方法有波段比值算法、颜色指数算法以及针对特定区域的经验算法等(O'Reilly 等, 1998; O'Reilly 和 Werdell, 2019; Hu 等, 2012, 2019; Volpe 等, 2007)。半解析模型基于水体光学辐射传输理论,通过分解水体的固有光学特性(如吸收和散射系数)反演 Chl-a 浓度,与经验算法相比,该方法具有更强的物理约束性和更广泛的适用性(Carder 等, 1999; Lee 等, 2002)。

当前针对 HY-1C/1D 卫星的 Chl-a 浓度反演算法研究多聚焦于 CZI (Coastal Zone Imager) 传感器(Chen 等, 2019; 滕越 等, 2022)。尽管已有学者尝试基于 COCTS 数据开展 Chl-a 浓度反演研究,但方法体系主要沿用传统经验算法框架。例如, Ye 等(2021)采用 OCI 算法从 COCTS 传感器获得全球 Chl-a 浓度,并验证了算法的良好性能。此外,李芝凤等(2022)针对中国南海北部海域提出了基于 4 种多波段比值的 Chl-a 浓度反演算法,并发现五波段比值算法的表现较好。尽管传统算法在开阔海域能够实现较精确的 Chl-a 浓度反演结果,但在河口和近海等光学复杂水域存在较大误差,这主要由归因于高浓度的悬浮物和溶解有机物对水体光学信号的复杂影响(Dierssen, 2010; Morel 和 Prieur, 1977)。因此,亟需突破传统经验算法的局限,开发新的适用于 COCTS 数据的不同水体类型下 Chl-a 浓度的反演方法,以提升国产卫星在全球范围水环境的监测能力。

随着人工智能技术的发展,机器学习在 Chl-a 浓度反演的应用受到广泛关注。通过对大量光谱数据和 Chl-a 实测数据的训练,机器学习模型能够更有效捕捉 Chl-a 浓度与水体光学特性之间的非线性关系,尤其在悬浮物浓度较高或水体成分复杂的区域,展现出了更强的适应性和稳健性(Dransfeld 等, 2004; Zhang 等, 2003)。目前,针对不同遥感观测数据,已经有多种基于机器学习的 Chl-a 浓度反演算法被提出,并在初步验证中展现了较高的精度。例如, Pahlevan 等(2020)通过融合

Sentinel-2 与 Sentinel-3 卫星数据,提出了一种混合密度网络模型,成功实现了在多样化生物光学条件下的内陆与沿海水域 Chl-a 浓度的高精度反演。Cao 等(2020)结合 Landsat 8 OLI 数据与极端梯度提升树模型,开发了一种专门针对浑浊湖泊水体的 Chl-a 浓度反演算法,验证结果显示该算法显著优于传统比值算法。李修竹等(2019)针对长江口及其邻近海域,构建了基于支持向量机(SVM)的 Chl-a 浓度估算模型,展现了 SVM 在复杂的水体环境中的反演能力。朱云芳等(2017)利用 GF-1 WFV 影像,在太湖区域建立了 Chl-a 浓度 BP 神经网络反演模型,该算法在光学复杂水体中具有良好适用性与精度。

针对 HY-1C/1D 卫星的海洋水色监测, Yang 等(2023)尝试将基于残差网络模型的机器学习方法应用于 COCTS 波段的 Chl-a 浓度反演。一方面,该研究中算法仅采用遥感反射比数据进行 Chl-a 浓度反演,未充分考虑环境参数对 Chl-a 分布的潜在影响;另一方面,其算法训练数据主要采集自北半球的近岸区域,限制了模型在全球不同类型水体的适应性(Hu 等, 2023; Li 等, 2023; Zhang 等, 2024)。针对上述不足,本研究基于 HY-1C/1D 卫星 COCTS 中心波段遥感反射比,以全球尺度实测数据为训练样本,结合地理坐标、时间信息以及海表温度与光合有效辐射等环境参数,采用多层感知器神经网络,开发了针对 COCTS 数据的 Chl-a 浓度反演算法,为海洋生态环境的持续监测提供了新的算法支持。

2 数据与方法

2.1 数据来源

(1) 实测 R_{rs} 光谱数据处理: 本研究用于 Chl-a 浓度算法开发与验证的数据集包括 Valente 数据集(Valente 等, 2022)、Nechad 数据集(Nechad 等, 2015)和厦门大学光学海洋学实验室(OOL)采集的实测遥感反射比(R_{rs} ; 单位: sr^{-1})和 Chl-a 浓度(mg/m^3)数据集。其中, Valente 数据集是国际水色研究广泛认可的全球现场水体生物光学数据集,汇集了全球多个航次观测计划的实测数据,但是由于数据来源广泛,该数据集的 R_{rs} 是离散的多光谱数据; Nechad 数据集中仅在少数波段有 R_{rs} 观测数据(412.5 nm、442.5 nm、490 nm、510 nm、

560 nm、620 nm、665 nm、681.25 nm 和 708.75 nm), 主要采集自欧洲近海和南非西部海域, 用于评估海洋水色算法在沿海水域水质参数 (例如, 悬浮物和 Chl-a 浓度) 反演中的准确性; OOL 数据集的样本主要来自中国近海, 包括黄海、台湾海峡、南中国海、西太平洋 4 个海域的观测结果, 其中 R_{rs} 数据采用 FOBY 漂浮式光谱仪测量 (Lee 等, 2013; Yu 等, 2021), 光谱范围涵盖 360—1000 nm, 分辨率为 5 nm, Chl-a 浓度基于荧光法测量 (Catherine 等, 2012)。

(2) 卫星数据: 本研究构建 Chl-a 反演模型时引入了 COCTS 传感器采集的 2023 年 8 月 25 日以来的 Level-2A R_{rs} 数据产品, 空间分辨率为 1.1 km, 数据通过中国海洋卫星数据服务系统 (<https://osdds.nsoas.org.cn/DataRetrieval> [2025-03-06]) 获取。同时, 本研究考虑到水温和光照条件对海洋浮游植物的物候特性 (对应 Chl-a 浓度表征) 的驱动机制, 特别选择 MODIS 传感器反演的海表温度 SST (Sea Surface Temperature) 和光合有效辐射 PAR

(Photosynthetically Active Radiation) 数据作为关键环境变量。其中, MODIS 提供了 SST 和 PAR 的 Level 3 级日均值、月均值以及气候态月均值产品。相关产品的空间分辨率为 4 km, 并可通过 NASA Ocean Color 数据平台 (<https://oceancolor.gsfc.nasa.gov/> [2025-03-06]) 获取。鉴于目前尚无法直接获取 COCTS 传感器的 SST 气候态月均值产品和 PAR 产品, 本研究采用 MODIS 的 PAR 和 SST 数据产品, 以保障模型训练与分析中对时空连续性和参数完备性的要求, 后续开发的 Chl-a 反演模型可根据 HY 系列卫星提供的 SST 和 PAR 数据进行更新。

2.2 数据预处理

表 1 展示了本研究使用的现场观测数据集 (具体空间分布详见 <https://zenodo.org/records/15683396> [2025-03-06] 中图 A1 所示), 涵盖了全球范围内不同类型的水体, 划分为训练集、测试集和独立验证集。具体预处理步骤包括 (1) 光谱数据匹配, (2) 卫星数据与实测数据匹配和 (3) 训练和测试划分。

表 1 本研究使用的现场观测数据集概况
Table 1 Summary of the in-situ datasets utilized in this study

数据源	时间范围	站位分布	数据用途	样本数
Valente	1997 年—2019 年	全球海域	85% 训练、15% 测试	1278
Nechad	2002 年—2010 年	欧洲近海和南非西部海域	85% 训练、15% 测试	278
OOL	2003 年—2018 年	台湾海峡	85% 训练、15% 测试	424
OOL	2003 年—2017 年	黄海、南中国海、西太平洋	独立验证	185

(1) 实测光谱数据处理: 由于不同实测数据集中 R_{rs} 的波长差异, 我们采用了不同的策略获取与 COCTS 传感器的波长等效的实测 R_{rs} 数据集。Valente 数据集提供离散的多光谱 R_{rs} , 我们根据其论文推荐的标准方法通过在 ± 6 nm 范围内取 R_{rs} 数据的平均值 (Valente 等, 2022) 以匹配 HY-1C/1D 卫星中心波长 (包括 412 nm、443 nm、490 nm、520 nm、565 nm 和 670 nm)。对于 Nechad 数据集, 其光谱仅包含几个波段 (见 2.1 节), 与 COCTS 的中心波长较接近, 因此采用线性插值 (linear) 来获取 COCST 对应波段数据。OOL 数据集中提供的是高光谱 R_{rs} 测量数据, 则通过光谱响应函数卷积获得 COCTS 相应波段的数据。

(2) 卫星观测与实测数据匹配: 本研究中, 与 COCTS R_{rs} 匹配的实测数据来源于 OOL 数据集, 时间范围为 2019 年—2024 年。为获取更多有效匹

配样本, 每个实测站点的卫星 R_{rs} 数据匹配采用以站点为中心的 5×5 像元窗口。若窗口内有效像元数量达到设定阈值 (≥ 5), 且 R_{rs} (565) 的变异系数 CV (Coefficient of Variation) 小于 30%, 则选取窗口内数据的中位数作为该站点的匹配值。对于 SST 和 PAR 参数, 由于其空间分辨率较大, 匹配时采用 3×3 像元窗口, 其他条件与 R_{rs} 的匹配方法保持一致。

(3) 训练和测试划分: 本研究中共获取到 2165 组实测数据用于模型的训练、测试和验证。训练集和测试集由 Valente、Nechad 和 OOL (台湾海峡) 3 个数据来源混合而成, 共 1980 组数据, 其中 85% 用于模型的训练过程, 15% 的数据用于测试模型的性能。独立验证集选用了 OOL 数据集的黄海、西太平洋和南中国海 3 个海域的观测数据, 共 185 组, 专用于评估模型的泛化性能。选择

该3个海域观测数据主要是因其观测的 Chl-a 浓度基本覆盖了训练数据集 Chl-a 浓度的范围, 因此有较好的代表性。图1(a)展示了本数据集中包含多种水体类型特性的 R_{rs} 的光谱及其对应的

Chl-浓度。图1(b)则呈现了 Chl-a 浓度的分布情况, Chl-a 浓度范围为 $0.02\text{--}95\text{ mg/m}^3$, 其中有146个站位的 Chl-a 浓度大于 20 mg/m^3 , 充分反映了不同水体条件下的生物光学特征。

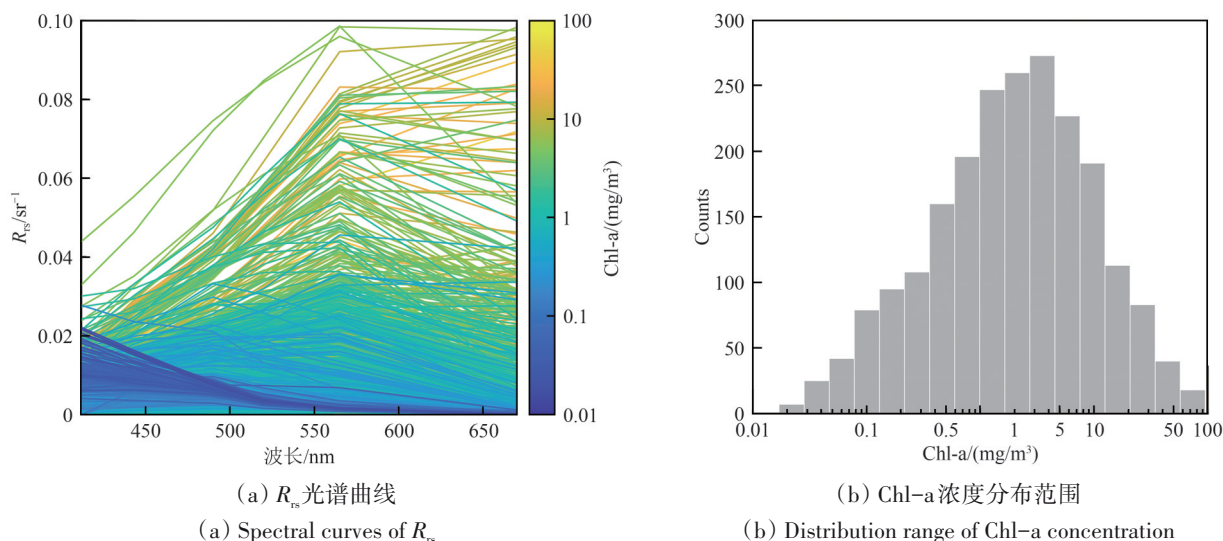


图1 实测数据集的 R_{rs} 光谱特征及 Chl-a 浓度分布统计

Fig. 1 Spectral R_{rs} and the distribution of Chl-a concentration in the in-situ datasets

2.3 神经网络模型

(1) 模型选择: 为了实现更佳的模型性能和稳健性, 本研究对多种被广泛应用的机器学习模型进行了比较分析, 最终选定基于多层感知器 MLP (Multi-Layer Perceptron) 的神经网络 NN (Neural Network) 模型 (简称 MLP-NN) 用于 Chl-a 浓度反演。因为 MLP-NN 模型能够通过调整网络层数和节点数适配多种特性的数据集 (即高灵活性), 从而能够有效捕捉环境数据中的非线性关系 (Hecht-Nielsen, 1992; Kim 和 Adali, 2002), 这对提高 Chl-a 浓度估算精度起到关键作用。其他参与候选的模型包括: 支持向量回归 (SVR)、长短记忆网络 (LSTM)、随机森林 (RF) 和通过遗传算法优化的反向传播神经网络 (GA-BP), 模型对比结果见 3.1 节 (表 3)。

(2) 模型设置: 为了提高模型的泛化能力, 所有输入数据在训练前进行了标准化处理 (Sola 和 Sevilla, 1997), 以消除不同特征之间的量纲差异, 确保各特征对模型的影响处于相同的尺度, 从而加速模型的收敛并提高模型训练的稳定性。定义的网络结构包括一个输入层、两个隐藏层 (包含 6 个和 3 个神经元) 以及一个输出层。由于数据集

较小, 本研究选择 LM (Levenberg-Marquardt) 算法作为优化方法, 该算法结合了梯度下降法和牛顿法的优点, 能够高效地优化网络权重并加速模型的收敛速度 (Bilski 等, 2020; Hagan 和 Menhaj, 1994)。LM 算法具备可变和自适应特性, 能在有限样本情况下避免过拟合, 确保训练稳定性与快速收敛 (Mukherjee 和 Routroy, 2012)。在模型训练过程中, 本研究采用 K 折交叉验证法 (Arlot 和 Celisse, 2010) 以增强模型稳健性。交叉验证通过多次训练和验证, 保证模型在不同数据子集上的表现一致性, 进而提升其泛化能力, 并避免因数据划分不均导致的性能波动。此外, 训练过程中引入了正则化和早停机制 (Srivastava 等, 2014; Zheng 等, 2018) 以提升模型性能。正则化通过对过大权重进行惩罚, 简化模型结构并提升其泛化能力; 而早停机制则在验证集误差连续迭代 5 次不再改善时终止训练, 从而防止模型过度拟合训练集。

(3) 模型输入: 本研究构建的模型在输入参数设计上采用多维度特征融合的策略: 除 COCTS 6 个波段的 R_{rs} 以外, 引入 R_{rs} 波段比值以增强水体光学特征区分度, 同时整合地理坐标 (经纬度)、

时间信息（月份）及不同时间尺度的 SST 与 PAR 数据，各因子的定义和计算公式如表 2 所示。其中， R_{rs} 波段比只选取 3 个波段的比值，主要考虑了水体中 Chl-a 浓度对蓝光波段（412、443、490）和绿光波段（565）的敏感性（Tran 等，2023）。针对经纬度与月份数据的特殊属性，引入正弦—余弦编码方式（Rußwurm 等，2024；Yang 等，2022），实现周期性信号的连续性表征，避免不同输入特征存在数量级差异，有效平滑周期数据，确保相邻周期值在数值上更为一致，从而提高模型的学习效率。考虑环境变量的时序依赖性规律，本模型引入了 SST 与 PAR 的 5 种时间尺度数据：包括天平均（当天）、月平均（当月）、气候态月平均（当月）、月平均（前月）和气候态月平均（前月）。其中，天平均数据有助于模型捕捉温度的实时变化；月平均和气候态月平均分别表示某一月内所有有效观测数据的平均值，以及基于 2002 年 7 月到 2024 年 6 月期间所有月度观测数据计算得到的长期平均值，分别提供了区域短期和长期的气候背景；前月月平均数据可用于应对 Chl-a 受 SST、PAR 的影响可能存在滞后效应（Dunstan 等，2018；Trombetta 等，2019）。由于 Chl-a 浓度的量级差异较大，因此在模型输出前对其进行对数转换，以缩小数据范围，提升模型训练的稳定性及反演精度。为明晰各参数对模型输出的贡献度，本研究分析了各参数对输出数据的敏感度（见 3.2 节）。

2.4 评价指标

为全面衡量模型的算法性能，本研究采用了多个评价指标，包括相关决定系数 R^2 （Coefficient of Determination）、均方根偏差 RMSD（Root Mean Square Difference）、中位数绝对百分比偏差 MAPD（Median Absolute Percentage Difference）和无偏百分比偏差 UPD（Unbiased Percentage Difference）。所选评价指标已广泛应用于 Chl-a 浓度的遥感反演与评估工作中（张亭禄和贺明霞，2002；Ye 等，2021；Shahvaran 等，2024）。具体计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (\log_{10}(\text{est}_i) - \log_{10}(\text{obs}_i))^2}{\sum_{i=1}^n (\text{Mean}(\log_{10}(\text{obs}_i)) - \log_{10}(\text{obs}_i))^2}, \quad (1)$$

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (\log_{10}(\text{est}_i) - \log_{10}(\text{obs}_i))^2}{n}} \quad (2)$$

$$\text{MAPD} = \text{Median} \left(\left[\left| \frac{\text{est}_i - \text{obs}_i}{\text{obs}_i} \right| \times 100\% \right]_{i=1,2,3,\dots,n} \right) \quad (3)$$

$$\text{UPD}_i = \frac{(\text{Chla}_{\text{NN}})_i - (\text{Chla}_{\text{OCI}})_i}{0.5 \times ((\text{Chla}_{\text{NN}})_i + (\text{Chla}_{\text{OCI}})_i)} \times 100\% \quad (4)$$

式中， est_i 表示反演得到的 Chl-a 浓度， obs_i 表示实测的 Chl-a 浓度； Chla_{NN} 和 Chla_{OCI} 分别表示基于 MLP-NN 模型和 NASA 业务化运行 OCI 算法（<https://www.earthdata.nasa.gov/apr/documents/chlor-a/> [2025-03-06]）反演得到的 Chl-a 浓度， UPD_i 则衡量任意给定网格单元（ i ），本研究方法与 OCI 算法在 Chl-a 浓度估算结果上的差异。

表 2 各输入参数的定义

Table 2 Definitions of the input parameters

参数	描述
R1	$R_{rs} (412)/R_{rs} (565)$
R2	$R_{rs} (443)/R_{rs} (565)$
R3	$R_{rs} (490)/R_{rs} (565)$
Lon(lon1, lon2)	经度的正弦值 $\text{lon1} = \sin(2\pi \frac{\text{经度}}{360})$ 经度的余弦值 $\text{lon2} = \cos(2\pi \frac{\text{经度}}{360})$
Lat(lat1, lat2)	纬度的正弦值 $\text{lat1} = \sin(2\pi \frac{\text{纬度}}{180})$ 纬度的余弦值 $\text{lat2} = \cos(2\pi \frac{\text{纬度}}{180})$
Month(m1, m2)	月份的正弦值 $\text{m1} = \sin(2\pi \frac{\text{月份}}{12})$ 月份的余弦值 $\text{m2} = \cos(2\pi \frac{\text{月份}}{12})$
SSTd	天平均 SST
PARd	天平均 PAR
SST0	月平均 SST(当月)
PAR0	月平均 PAR(当月)
SST0 _{clim}	气候态月平均 SST(当月)
PAR0 _{clim}	气候态月平均 PAR(当月)
SST1	月平均 SST(前月)
PAR1	月平均 PAR(前月)
SST1 _{clim}	气候态月平均 SST(前月)
PAR1 _{clim}	气候态月平均 PAR(前月)

注：参数选取依据：COCTS 波段比值的选取参考了 Kim 等(2022)的研究；空间信息（经纬度）依据了 Zhang 等(2024)的方法；季节信息的选取参照了 Hu 等(2023)的研究；环境变量(SST 和 PAR)的使用依赖于 Hu 等(2023)和 Li 等(2023)的分析。

3 结果和讨论

3.1 不同机器学习模型对比

本研究通过系统性的模型筛选实验，评估了包括 SVR、LSTM、RF、GA-BP 和 MLP-NN 等在内

的5种机器学习算法在Chl-a浓度反演任务中的适用性(表3)。研究中所有模型均基于统一划分的训练集与测试集进行训练和验证。结果表明,MLP-NN模型在 R^2 、RMSD和MAPD等评估指标上均显著优于其他模型,这突出展现了MLP-NN模型在处理非线性关系和小样本数据方面的优越性。其他对照模型虽然也展现出可比的性能,但存在模型解释性不足和对样本数据要求高等问题。例如,LSTM模型虽然擅长处理时间序列数据,但在小样本数据上训练时容易受到长时间依赖问题的影响,从而限制了其对数据中非线性特征的捕捉能力,影响了模型的表现。基于以上分析,本研究选择基于MLP-NN模型作为COCTS传感器Chl-a浓度反演算法的核心架构。

表3 不同机器学习模型对于Chl-a浓度反演的性能比较,使用表2中所有特征参数作为模型输入

Table 3 Comparison of different machine learning models for Chl-a concentration retrieval. All parameters in Table 2 were used as model inputs

机器学习模型	R^2	RMSD	MAPD
SVR	0.85	0.27	34.2%
LSTM	0.85	0.27	40.0%
RF	0.88	0.23	33.4%
GA-BP	0.86	0.25	33.3%
MLP-NN	0.89	0.22	30.8%

3.2 输入因子分析

本研究采用控制变量法定量评估输入参数Chl-a浓度反演精度的贡献,以选择最优的神经网络模型输入因子,避免冗余信息对模型性能的干扰。以6个波段(412 nm、443 nm、490 nm、520 nm、565 nm和670 nm)的 R_{rs} 为基准模型,逐一引入候选因子(表2)。选择因子的依据为引入候选因子后的MAPD的变化:MAPD差值越大,表示该因子对模型性能的正向贡献越显著;反之,表示该因子对模型带来负向干扰特性。

从图2可以看出,不同输入因子对模型性能的影响存在显著差异。引入Lat、SST1或 $PAR1_{clim}$ 后,模型性能显著提升,表明这些因子对模型贡献较大。其中,Lat在周期性编码后有效捕捉到了空间变化特征,并且作为地理空间参数,其本质上反映了不同纬度带太阳辐射的角度和强度。SST1反映了前月海表温度的动态变化及其对Chl-a浓度的滞后效应,这一影响可能揭示了海洋生态系统对

温度变化的响应延迟,为模型训练提供了更稳定的环境参考。同时,作为气候变化的重要指标之一,有助于提高模型估算Chl-a浓度的准确性(Hu等,2023)。 $PAR1_{clim}$ 通过前月的气候态光照条件,提供了浮游植物光合作用能力的解释,并利用气候态背景所呈现的空间基本模式,有效消除年际异常波动,从而稳定地描述了光学特性对Chl-a浓度的调控作用。Month作为反映季节性变化的重要时间因子,尽管对性能提升有限,但能够有效表征和捕捉模型中浮游植物生长与消退的物候特征。

对于其他因子, $PAR0_{clim}$ 的引入导致模型性能下降,可能是由于当月气候态光照的短期波动干扰了对Chl-a浓度的估算。短期水温变化(如SSTd和SST0)对模型性能产生负面影响,可能是由于其波动性较大,无法稳定描述Chl-a浓度的变化; R_{rs} 波段比未能显著改善模型性能,说明 R_{rs} 数据已包含波段相关信息,因此未纳入作为输入因子,避免潜在冗余; $SST0_{clim}$ 虽具有一定贡献,但考虑SST1已包含关键的水温信息,因此也未被选用。需要指出的是,引入环境参数虽然能够提升模型的反演精度,但也可能带来副作用。例如,多源参数的耦合可能增加模型的复杂性,使其对训练数据过于敏感,尤其是在数据量较小的情况下,容易导致过拟合风险的增加。此外,月平均的SST和PAR具有较为粗糙的空间分辨率,其输入误差可能通过模型传递至Chl-a反演结果,导致误差的传递效应。因此,在引入环境参数时,必须综合评估其对模型性能的正负面影响,以确保模型的准确性和适用性。本研究通过评估不同环境参数对MLP-NN模型的敏感性(图2),选出最优的输入参数,确保其有助于Chl-a遥感估算的精度提升。综合以上分析,本研究的MLP-NN模型选择了以下因子作为输入参数:412 nm、443 nm、490 nm、520 nm、565 nm和670 nm波段的 R_{rs} 、SST1、Lat、 $PAR1_{clim}$ 和Month。

3.3 模型性能评估

本研究通过占总数据集的15%的测试数据(2.2节)分别对MLP-NN模型与OCI算法进行测试,评估了两种反演模型在Chl-a浓度估算中的精度和稳健性。图3为MLP-NN模型与OCI算法估算的Chl-a浓度与实测浓度之间的散点图,可见MLP-NN模型在各评价指标均优于OCI算法,表现出与实测数据更高的一致性。

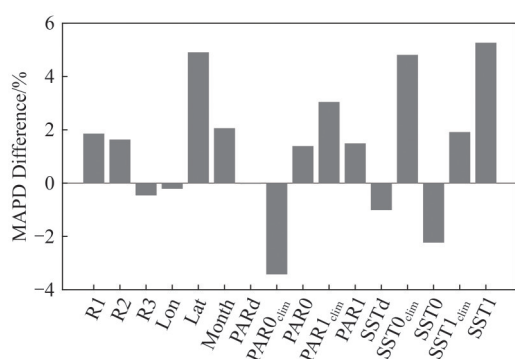
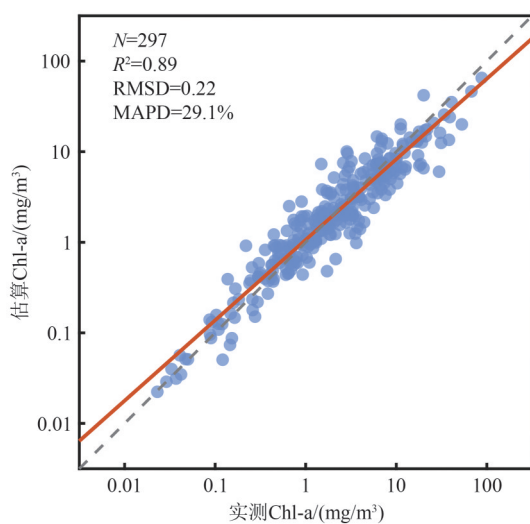
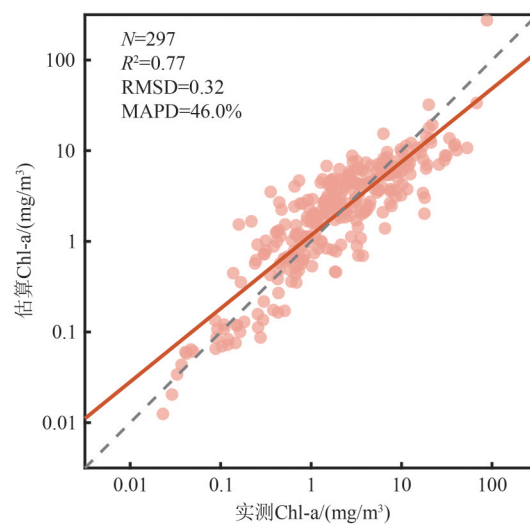


图2 不同输入因子对模型性能的影响(MAPD的相对变化)

Fig. 2 Impacts of different input parameters on the performance of MLP-NN model (Indicated by the relative change in MAPD)



(a) MLP-NN 反演结果与实测值对比散点图
(a) MLP-NN retrievals vs. in situ observations



(b) OCI 反演结果与实测值对比散点图
(b) OCI retrievals vs. in situ observations

图3 不同 Chl-a 浓度反演模型的性能评估: MLP-NN, OCI

Fig. 3 Evaluation of the retrieved Chl-a concentration by MLP-NN and OCI

本研究从测试集中筛选出包含悬浮物 TSM (Total Suspended Matter) 浓度数据的样本 (共45个), 评估模型在不同 TSM 浓度下的模型的性能表现 (图4 (a)、(b))。图4 (c) 展示了测试集中 TSM 浓度的分布范围 (0—100 mg/L), 平均浓度为 20.26 mg/L。根据图4 (a) 和 4 (b) 的结果, MLP 模型和 OCI 算法在不同 TSM 浓度下均表现出了较为良好的拟合性能。然而, 与 OCI 算法相比, MLP-NN 模型在不同 TSM 浓度条件下对 Chl-a 浓度的估算结果更为稳定, 其 MAPD 值比 OCI 算法低 10.5%。由于数据集中高浊度 (TSM > 100 mg/L) 样本较少, 而神经网络的性能依赖于训练数据中各参数的范围, 因此在高浑浊水体中模型的性能

如图3 (a) 所示, MLP-NN 模型的反演结果大多数数据点紧密分布在 1:1 线附近, R^2 为 0.89, 表明该模型能够较好地再现实测 Chl-a 浓度的分布。相比之下, OCI 算法的 R^2 为 0.77 (相差 0.12), 在高值区域出现低估、低值区域出现高估的现象较明显 (图3 (b))。在精度指标方面, MLP-NN 模型误差较低, RMSD 降低了 0.1, MAPD 降低了 16.9%。与之相比, OCI 算法的散点分布较为离散, 不同浓度范围内误差波动较大, 精度低于 NN 模型。以上结果进一步表明, MLP-NN 模型在不同浓度水平下的稳定性较强, 能够较好地适应不同范围的 Chl-a 浓度变化。

仍需进一步验证。

为进一步评估模型的泛化能力, 本研究通过独立数据集 ($N=185$) 对比分析了 MLP-NN 模型与 OCI 算法的泛化表现 (图4)。对于图4 (a)—(b), 散点的颜色表示 440 nm 处有色溶解有机物 CDOM (Colored Dissolved Organic Matter) 的吸收系数 ($a_g(440)$)。图4 (f) 则显示了 $a_g(440)$ 的直方图分布 (0—0.6 m^{-1}), 其平均值 (\pm 标准差) 为 0.06 (\pm 0.07) m^{-1} , 高 CDOM ($a_g(440) \geq 0.1 m^{-1}$) 的样本仅 27 个, 因此模型在估算高溶解有机物吸收系数下水体的 Chl-a 浓度仍存在较大不确定性。图4 (a)、(b) 显示, MLP-NN 模型不仅在训练数据和测试数据上具有较好的拟合能力, 在独立数据集上同样展现了在

不同CDOM范围下模型的稳健性。然而,OCI算法虽然与独立验证数据集的相关性与MLP-NN模型相当,但误差方面明显更高(尤其是MAPD相比MLP-NN高出了19.8%),表明MLP-NN模型对复杂水体表层Chl-a浓度估算的适应性更强。值得注意的是,对于部分样品,MLP-NN和OCI的部分反演结

果均表现明显高估,遥感估算Chl-a是实测值的5—10倍,这些数据主要采集自于长江口和黄海沿岸海域相对浑浊和高CDOM水体。究其原因,在于两个算法的训练数据集相对缺乏代表这类型水体的样本,导致反演Chl-a不确定性增加。

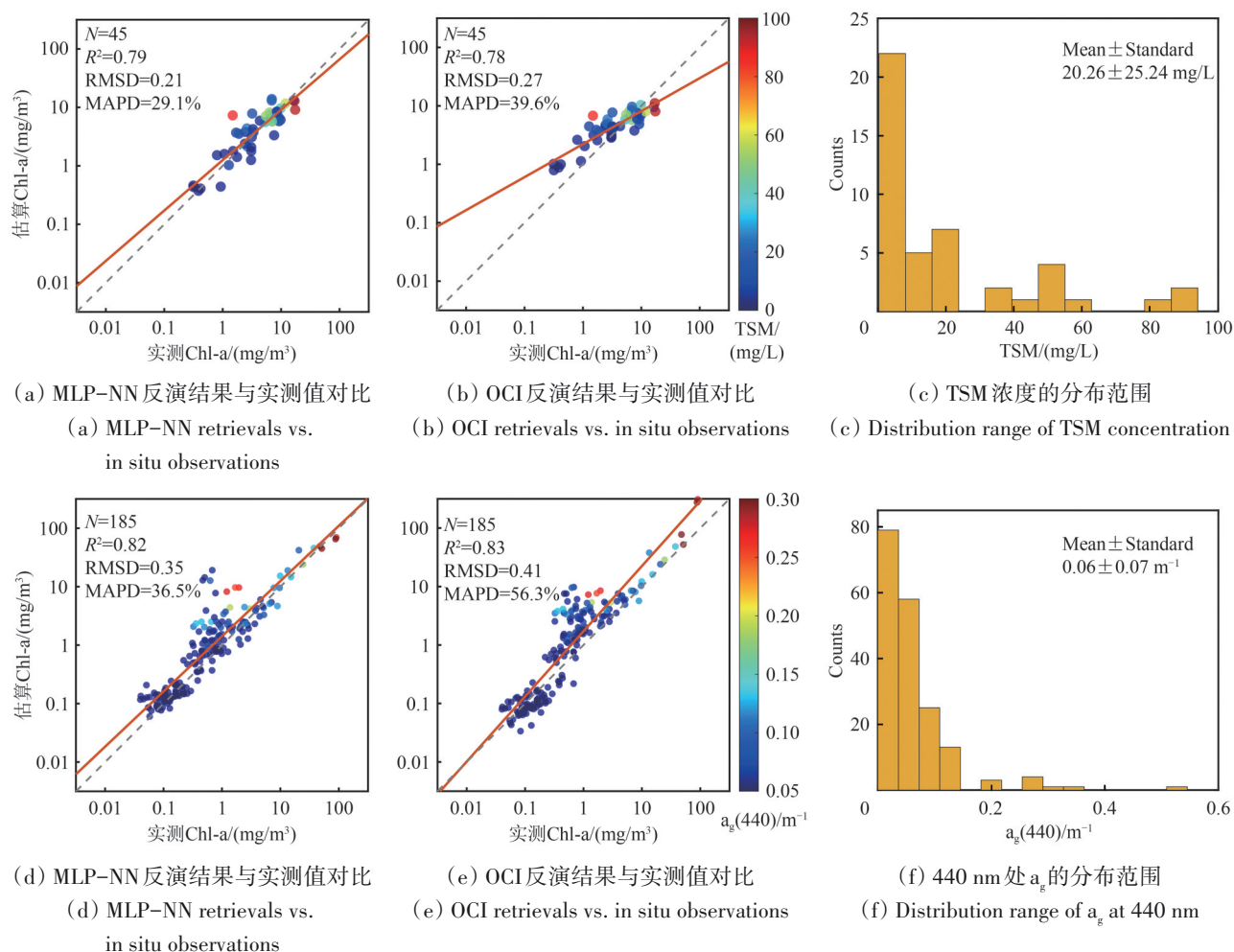


图4 不同TSM和CDOM浓度下MLP-NN和OCI模型反演Chl-a的独立验证对比(其中(a)和(b)中散点颜色表示TSM浓度,(d)和(e)中散点颜色表示440 nm处CDOM的吸收系数($a_g(440)$))

Fig. 4 Independent validation of the retrieved Chl-a concentration by MLP-NN and OCI for samples with different TSM and CDOM concentrations (The color of scatter points in (a) and (b) represents the TSM concentration, while represents the absorption coefficient of CDOM at 440 nm ($a_g(440)$) in (d) and (e))

3.4 基于COCTS匹配数据的验证

考虑到COCTS标准 R_{rs} 产品在近海可能存在较大不确定性,本研究通过构建站点观测的 R_{rs} 数据、Chl-a浓度与COCTS标准 R_{rs} 产品的匹配数据集,评估COCTS R_{rs} 数据的精度(图5),并进一步验证MLP-NN模型在COCTS上估算Chl-a浓度的能力(图6)。最终共获得21个实测—卫星匹配样本,

由于部分实测站点含有Chl-a浓度,但是 R_{rs} 数据缺失,因此二者的样本数量存在一定差异。

根据实测 R_{rs} 对COCTS的验证结果显示,总体而言,6个波段(412 nm、443 nm、490 nm、520 nm、565 nm和670 nm)的 R_{rs} 与实测数据相比存在较大差异,整体拟合效果欠佳,与韩冰等(2023)研究结论一致。尤其是在蓝光波段(如412 nm和

443 nm), MAPD 值分别高达 119.9%、76.7% 和 46.1%, 表明 COCTS R_{rs} 产品在这些波段与实测值之间存在显著偏差。相对而言, COCTS 在绿光波段的 565 nm 处误差较小, MAPD 值为 12.9%。图 6 显示了 MLP-NN 模型与 OCI 算法基于 COCTS R_{rs} 估算 Chl-a 浓度与实测值的对比, 验证结果表明, 两

种算法估算的 Chl-a 浓度与实测数据具有较好的相关性 ($R^2 \geq 0.76$), 但误差表现出显著差异。相较于 OCI 算法, MLP-NN 模型在整体反演精度上更优, 其 MAPD 值降低了 9.8%, 提供了更为准确的 Chl-a 浓度估算结果。

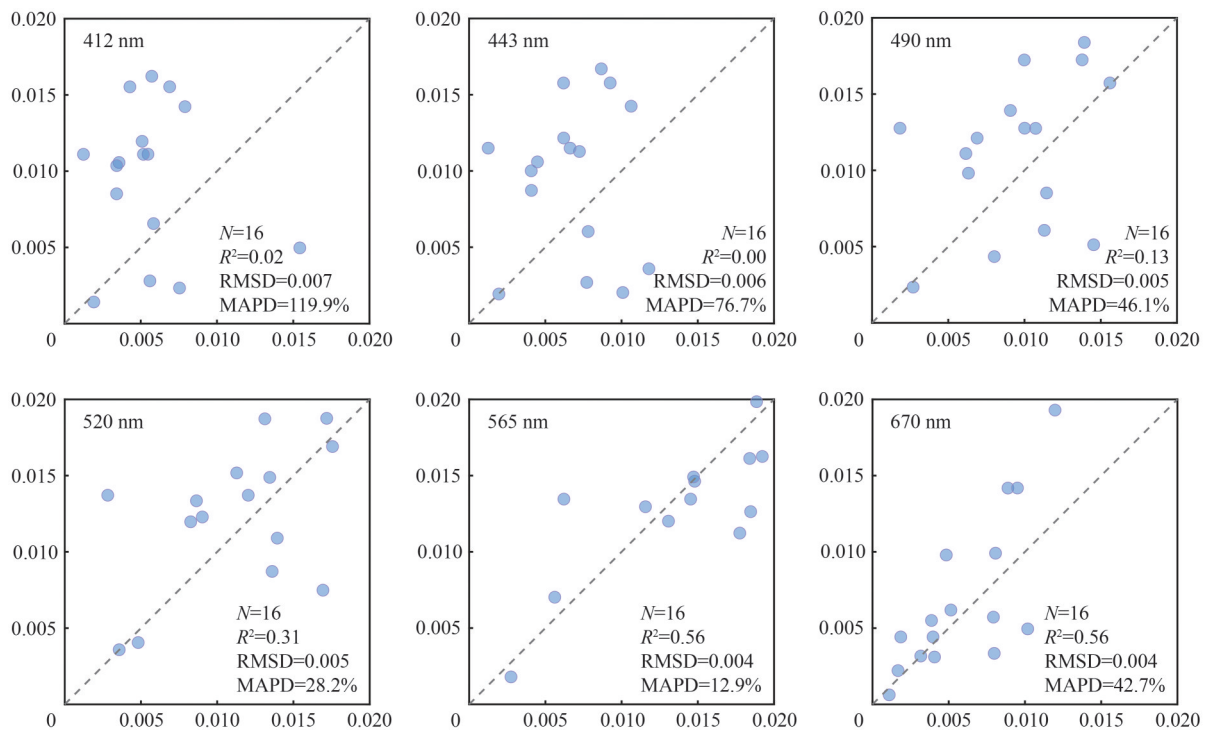
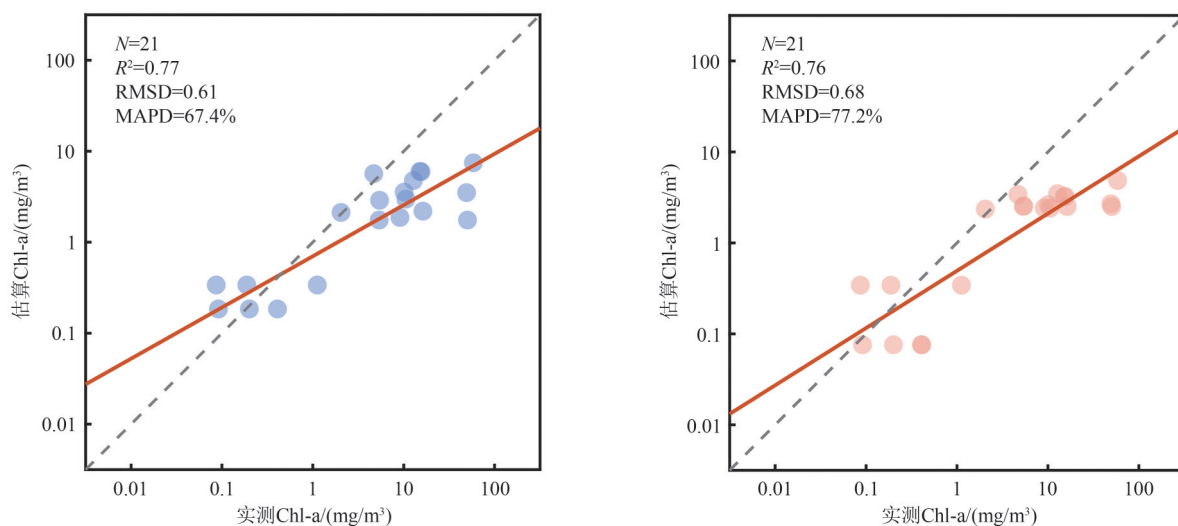


图 5 6 个波段(412 nm、443 nm、490 nm、520 nm、565 nm、670 nm)COCTS 的 R_{rs} 与实测 R_{rs} 对比

Fig. 5 Comparison of R_{rs} from COCTS and field-measured R_{rs} at 412 nm, 443 nm, 490 nm, 520 nm, 565 nm and 670 nm



(a) MLP-NN 反演结果与实测值对比散点图

(a) MLP-NN retrievals vs. in situ observations

(b) OCI 反演结果与实测值对比散点图

(b) OCI retrievals vs. in situ observations

图 6 基于 COCTS R_{rs} 估算的 Chl-a 浓度与实测值对比

Fig. 6 Comparison of Chl-a concentrations estimated based on COCTS R_{rs} with field measurements

尽管 MLP-NN 模型在 Chl-a 浓度估算中优于 OCI 算法,但在高浓度近岸区域,模型估算结果与实测数据仍存在差异,主要源于近海水色遥感观测的固有不确定性及卫星与站点数据匹配时的时空错位。一方面,本研究使用的卫星 R_{rs} 数据经过 COCTS 标准大气校正算法处理,但在近海区域该算法的校正精度可能受到大气条件的显著影响(例如,较为复杂的气溶胶类型),进而导致 Chl-a 浓度的估算值与实测数据之间出现系统性误差(Nazeer 等, 2021; Shen 等, 2020; 王道生 等, 2023)。另一方面,本研究设定时间窗口为一天进行卫星与实测数据的匹配,空间窗口设置为 5×5 像元,旨在尽可能提高样本匹配数量。如果时间窗口设置过短,匹配数据量会显著减少,进而影响统计分析的有效性和结果的可靠性。譬如,若采用 ± 3 h、 ± 6 h 时间窗口进行卫星 R_{rs} 和现场测量 Chl-a 数据匹配,仅能匹配到 0 个、8 个匹配样品。而若采用更小的空间匹配窗口,如 3×3 窗口,有效匹配数据仅剩余 9 个。

此外,必须强调的是近岸海域水体受到潮汐、风速等多种环境因素的共同作用,表现出明显的时空异质性(Lin 等, 2022; Laignel 等, 2023),尤其是在较短的时间尺度内,水体状态的快速变化可能难以被遥感观测数据完全捕捉。本研究中, SST、PAR 数据和实测点的空间匹配窗口与 R_{rs} 数据匹配窗口在空间尺度上的差异不可避免地对 Chl-a 浓度的遥感估算结果产生一定影响。为降低这种空间尺度差异带来的不确定性,本研究对匹配窗口内的水体特征进行了像元均匀性质量控制(即窗口内 R_{rs} (565) 有效数据的变异系数小于 30%),在一定程度上缓解了窗口内水体异质性对遥感估算结果的干扰。

针对目前应用于 COCTS 传感器 R_{rs} 数据的大气校正算法所带来的不确定性,未来需要进一步开展与 HY-1C/1D 卫星过境同步的现场观测,以系统性地扩充实测 R_{rs} 数据样本。同时可通过设计敏感性实验,系统评估匹配时空窗口与样本数量之间的权衡关系,以优化卫星观测与地面监测数据的匹配和验证策略。值得参考的是,已有研究利用卫星传感器获取的 R_{rs} 数据反演 Chl-a 浓度(Smith 等, 2021),在一定程度上能够将大气校正误差隐含于模型之中,减少其对反演精度的直接干扰。然而,受限于当前配对样本数量的严重不足(本

文仅 21 个卫星—实测 Chl-a 匹配样本),尚无法支撑具备良好泛化能力的神经网络模型构建,因此该方法在本研究中尚不具备实际应用条件。随着 HY 系列卫星的持续观测以及实测样本数量的积累,可进一步探索以 COCTS 获取的 R_{rs} 为输入、实测 Chl-a 浓度为输出的神经网络模型。此外,直接使用卫星测量的瑞利校正反射率和大气顶层反射率等参数(Cao 等, 2020; Soomets 等, 2022)也是实现高精度 Chl-a 浓度反演的一种可行路径。未来研究应进一步优化大气校正算法,特别是在近海复杂水体环境下,同时结合高分辨率气象数据,以提高大气校正精度。

3.5 中国渤海 Chl-a 浓度空间分布

为比较本研究开发的 MLP-NN 模型与现有 OCI 算法反演 Chl-a 浓度在空间分布上的差异,本研究以 2019 年 8 月 30 日 HY-1C/1D 卫星数据为例,对渤海 Chl-a 浓度空间分布开展对比分析(黄渤海区域的完整空间分布对比图见 <https://zenodo.org/records/15683396>[2025-03-06]中的图 A2)。该日期的选择综合考虑云量和数据覆盖率等因素,有助于确保图像清晰度以及提高 Chl-a 浓度反演的稳健性。具体来说,2019 年 8 月 30 日的 HY-1C/D 卫星影像在渤海区域云量较少,一方面有效观测比较多,另一方面受云和杂散光影响的像元相对更少,卫星数据质量更高,基于卫星数据不同 Chl-a 浓度反演结果的对比更直观。

结果表明, MLP-NN 模型和 OCI 算法估算的 Chl-a 浓度在空间分布特征上具有较高的一致性(图 7 (b)、7 (d))。二者均显示出渤海区域 Chl-a 浓度整体较高,并呈现由内海向外海递减的梯度,与现场观测数据的空间格局一致(朱碧泓 等, 2021)。尽管 MLP-NN 和 OCI 算法估算的 Chl-a 浓度整体空间分布模式趋同,但在不同区域数值上仍存在显著差异。空间差异分析表明,在高 Chl-a 浓度的浑浊水体区域(渤海湾和莱州湾沿岸)和低 Chl-a 浓度的渤海海峡, MLP-NN 模型的估算值明显高于 OCI 算法,其 UPD 的大致范围为 80%—100% (图 7 (c)),表明 MLP-NN 模型对 Chl-a 浓度的高值和低值的捕捉能力更强。相反,在渤海大部分区域, MLP-NN 模型的估算值普遍低于 OCI 算法,对应 UPD 的大致范围为 -20 至 -60% (图 7 (c))。该差异可能源于两种算法对水体光学特性的敏感性不同

以及与 MLP-NN 训练数据中浑浊水域样本占比有关。需强调的是, 通过不同日期的验证结果证实, 上述差异性特征具有普遍性, 尽管在不同日期和

不同区域该差异的幅度存在波动性, 表明该现象与算法之间的差异存在关联。

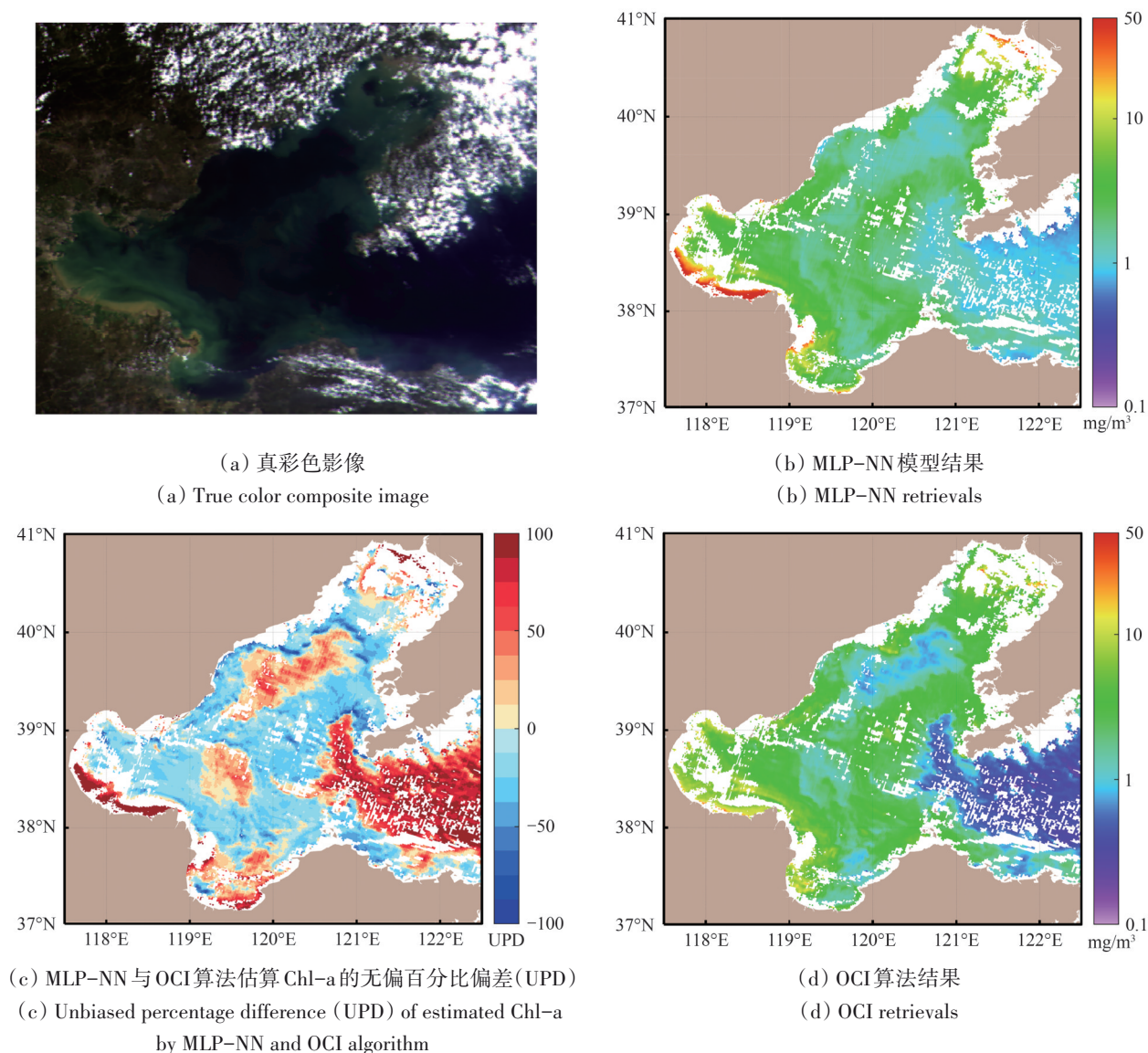


图7 基于 HY-1C/1D 卫星 COCTS 观测的 2019 年 8 月 30 日渤海 Chl-a 浓度空间分布对比

Fig. 7 Comparison of Chl-a concentration distributions from MLP-NN and OCI for Bohai, China using the COCTS image acquired on August 30, 2019

4 结论和展望

本研究通过评估多类机器学习模型的性能表现, 成功为 HY-1C/1D 卫星构建了基于多层感知器神经网络 (MLP-NN) 的 Chl-a 浓度反演方法, 并与常规 OCI 算法的反演结果进行了对比评估。通过输入参数特征筛选, 最终确定最佳模型输入组合为: 412 nm、443 nm、490 nm、520 nm、565 nm 和 670 nm 波段的 R_{rs} 、纬度、月份、前月月均 SST 和

前月气候态月均 PAR, 该结论为 Chl-a 浓度反演的机理研究提供了一定参考。独立验证结果表明, MLP-NN 模型的估算精度整体优于 OCI 算法, 其中对于高 Chl-a 浓度区域的反演优势尤为明显, MAPD 显著降低了 19.8%。卫星匹配数据集验证结果表明, 相较于 OCI 算法, MLP-NN 模型的 RMSD 和 MAPD 实现了 0.09、9.8% 的精度提升。以中国渤海 Chl-a 浓度空间分布为例, 两个算法均可有效反映中国渤海 Chl-a 浓度的分布格局, 但前者呈现出

在极高值和低值区均偏低的差异特征,表明MLP-NN模型对极端浓度值的捕捉能力更优。

现有数据规模的限制可能影响MLP-NN模型的外推适用性和空间泛化能力。建议未来研究重点拓展以下几个方向:(1)加强训练数据集规模,拓宽实测数据的采样范围和多样性,尤其需要扩大高浓度的悬浮物和高有色溶解有机物数据样本,以进一步提高模型的全球水体适用性;(2)训练方案的设计上考虑引入多源卫星传感器、辐射传输模型和海洋模式的Chl-a浓度产品等数据;(3)纳入风速和营养盐等辅助环境变量作为模型的协同输入;(4)构建深度学习模型,采用更高级的神经网络架构(如深度前馈神经网络),能够进一步提升反演精度。值得关注的是,2023年11月16日成功发射的海洋一号E星(HY-1E)搭载的第二代水色水温扫描仪(COCTS2)在紫外-可见光波段实现了光谱分辨率的升级,这不仅印证了中国自主海洋水色卫星的持续发展与观测能力,更为本研究成果业务化应用创造了优越的条件,进一步显示了本研究在水色遥感领域的实践潜力与发展前景。

志 谢 特别感谢国家海洋卫星中心提供的HY-1C/1D原始数据,以及Valente等、Nechad等整理和发布全球海洋的实测生物光学数据。同时,感谢所有航次参与人员在现场数据采集过程中所付出的辛勤努力。本文研发的Chl-a算法程序可从ZENODO公开获取:<https://zenodo.org/records/15683396>[2025-06-18]。

参考文献(References)

- Arlot S and Celisse A. 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4: 40-79 [DOI: 10.1214/09-SS054]
- Behrenfeld M J, Boss E, Siegel D A and Shea D M. 2005. Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, 19(1): GB1006 [DOI: 10.1029/2004GB002299]
- Bilski J, Kowalczyk B, Marchlewska A and Zurada J M. 2020. Local Levenberg-Marquardt algorithm for learning feedforward neural networks. *Journal of Artificial Intelligence and Soft Computing Research*, 10(4): 299-316 [DOI: 10.2478/jaiscr-2020-0020]
- Cao Z G, Ma R H, Duan H T, Pahlevan N, Melack J, Shen M and Xue K. 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sensing of Environment*, 248: 111974 [DOI: 10.1016/j.rse.2020.111974]
- Carder K L, Chen F R, Lee Z P, Hawes S K and Kamykowski D. 1999. Semianalytic Moderate-Resolution Imaging Spectrometer algorithms for chlorophyll *a* and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research: Oceans*, 104(C3): 5403-5421 [DOI: 10.1029/1998JC900082]
- Carvalho L, McDonald C, de Hoyos C, Mischke U, Phillips G, Borics G, Poikane S, Skjelbred B, Solheim A L, Van Wichelen J and Cardoso A C. 2013. Sustaining recreational quality of European lakes: minimizing the health risks from algal blooms through phosphorus control. *Journal of Applied Ecology*, 50(2): 315-323 [DOI: 10.1111/1365-2664.12059]
- Catherine A, Escoffier N, Belhocine A, Nasri A B, Hamlaoui S, Yéprémian C, Bernard C and Troussellier M. 2012. On the use of the FluoroProbe®, a phytoplankton quantification method based on fluorescence excitation spectra for large-scale surveys of lakes and reservoirs. *Water Research*, 46(6): 1771-1784 [DOI: 10.1016/j.watres.2011.12.056]
- Chen S G, Du K P, Lee Z P, Liu J Q, Song Q J, Xue C, Wang D S, Lin M S, Tang J W and Ma C F. 2021. Performance of COCTS in global ocean color remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2): 1634-1644 [DOI: 10.1109/TGRS.2020.3002460]
- Chen X Y, Zhang J, Tong C, Liu R J, Mu B and Ding J. 2019. Retrieval algorithm of chlorophyll-a concentration in turbid waters from satellite HY-1C coastal zone imager data. *Journal of Coastal Research*, 90(SI): 146-155 [DOI: 10.2112/SI90-018.1]
- Dierssen H M. 2010. Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40): 17073-17078 [DOI: 10.1073/pnas.0913800107]
- Dransfeld S, Tatnall A R, Robinson I S and Mobley C D. 2004. A comparison of Multi-layer Perceptron and multilinear regression algorithms for the inversion of synthetic ocean colour spectra. *International Journal of Remote Sensing*, 25(21): 4829-4834 [DOI: 10.1080/01431160412331269661]
- Dunstan P K, Foster S D, King E, Risbey J, O' Kane T J, Monselesan D, Hobday A J, Hartog J R and Thompson P A. 2018. Global patterns of change and variation in sea surface temperature and chlorophyll *a*. *Scientific Reports*, 8: 14624 [DOI: 10.1038/s41598-018-33057-y]
- Hagan M T and Menhaj M B. 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6): 989-993 [DOI: 10.1109/72.329697]
- Han B, Jia D, Gao F, Guo K, Zhu J H, Li T J, Ma C F, Liu J Q and Zibordi G. 2023. Validation of L2A operational products of COCTS onboard HY-1C satellite across coastal waters in China and Europe. *National Remote Sensing Bulletin*, 27(1): 26-42 (韩冰, 贾

- 迪, 高飞, 郭凯, 朱建华, 李铜基, 马超飞, 刘建强, Zibordi G. 2023. HY-1C 卫星 COCTS L2A 级产品在中国和欧洲近海的真实性检验. *遥感学报*, 27(1): 26-42 [DOI: 10.11834/jrs.20235007]
- Hecht-Nielsen R. 1992. Theory of the backpropagation neural network//Wechsler H, ed. *Neural Networks for Perception*. San Diego: Academic Press: 65-93 [DOI: 10.1016/B978-0-12-741252-8.50010-8]
- Henson S A, Sarmiento J L, Dunne J P, Bopp L, Lima I, Doney S C, John J and Beaulieu C. 2010. Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences*, 7(2): 621-640 [DOI: 10.5194/bg-7-621-2010]
- Holm-Hansen O, Lorenzen C J, Holmes R W and Strickland J D H. 1965. Fluorometric determination of chlorophyll. *ICES Journal of Marine Science*, 30(1): 3-15 [DOI: 10.1093/icesjms/30.1.3]
- Holm-Hansen O and Riemann B. 1978. Chlorophyll a determination: improvements in methodology. *Oikos*, 30(3): 438-447 [DOI: 10.2307/3543338]
- Hu C M, Feng L, Lee Z P, Franz B A, Bailey S W, Werdell P J and Proctor C W. 2019. Improving satellite global chlorophyll a data products through algorithm refinement and data recovery. *Journal of Geophysical Research: Oceans*, 124(3): 1524-1543 [DOI: 10.1029/2019JC014941]
- Hu C M, Lee Z P and Franz B. 2012. Chlorophyll a algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans*, 117 (C1): C01011 [DOI: 10.1029/2011JC007395]
- Hu Q W, Chen X Y, Bai Y, He X Q, Li T and Pan D L. 2023. Reconstruction of 3-D ocean chlorophyll a structure in the Northern Indian Ocean using satellite and BGC-Argo data. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4200513 [DOI: 10.1109/TGRS.2022.3233385]
- Huot Y, Babin M, Bruyant F, Grob C, Twardowski M S and Claustre H. 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies?. *Biogeosciences Discussions*, 4: 707-745
- Jiang X W, He X Q, Lin M S, Gong F, Ye X M and Pan D L. 2019. Progresses on ocean satellite remote sensing application in China. *Haiyang Xuebao*, 41(10): 113-124 [蒋兴伟, 何贤强, 林明森, 龚芳, 叶小敏, 潘德炉. 2019. 中国海洋卫星遥感应用进展. *海洋学报*, 41(10): 113-124] [DOI: 10.3969/j.issn.0253-4193.2019.10.007]
- Kim T and Adali T. 2002. Fully complex multi-layer perceptron network for nonlinear signal processing. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, 32(1): 29-43 [DOI: 10.1023/A:1016359216961]
- Kim Y W, Kim T H, Shin J, Lee D S, Park Y S, Kim Y and Cha Y K. 2022. Validity evaluation of a machine-learning model for chlorophyll a retrieval using Sentinel-2 from inland and coastal waters. *Ecological Indicators*, 137: 108737 [DOI: 10.1016/j.ecolind.2022.108737]
- Laignel B, Vignudelli S, Almar R, Becker M, Bentamy A, Benveniste J, Birol F, Frappart F, Idier D, Salameh E, Passaro M, Menende M, Simard M, Turki E I and Verpoorter C. 2023. Observation of the coastal areas, estuaries and deltas from space. *Surveys in Geophysics*, 44(5): 1309-1356 [DOI: 10.1007/s10712-022-09757-6]
- Lee Z P, Carder K L and Arnone R A. 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41(27): 5755-5772 [DOI: 10.1364/AO.41.005755]
- Lee Z P, Pahlevan N, Ahn Y H, Greb S and O'Donnell D. 2013. Robust approach to directly measuring water-leaving radiance in the field. *Applied Optics*, 52(8): 1693-1701 [DOI: 10.1364/AO.52.001693]
- Li X L, Yang Y, Ishizaka J and Li X F. 2023. Global estimation of phytoplankton pigment concentrations from satellite data using a deep-learning-based model. *Remote Sensing of Environment*, 294: 113628 [DOI: 10.1016/j.rse.2023.113628]
- Li X Z, Su R G, Zhang C S and Shi X Y. 2019. A Chl-a prediction model based on support vector machine in Yangtze River estuaries and its adjacent sea areas. *Periodical of Ocean University of China*, 49(1): 69-76 (李修竹, 苏荣国, 张传松, 石晓勇. 2019. 基于支持向量机的长江口及其邻近海域叶绿素 a 浓度预测模型. *中国海洋大学学报(自然科学版)*, 49(01): 69-76) [DOI: 10.16441/j.cnki.hdx.20170428]
- Li Z F, Han B, Jia D and Zhu J H. 2022. Evaluation of remote sensing inversion algorithm and COCTS Inversion Algorithm for chlorophyll a concentration in the north South China Sea. *Journal of Ocean Technology*, 41(6): 1-7 (李芝凤, 韩冰, 贾迪, 朱建华. 2022. 南海北部叶绿素 a 浓度遥感反演算法评估与 COCTS 反演算法构建. *海洋技术学报*, 41(6): 1-7) [DOI: 10.3969/j.issn.1003-2029.2022.06.001]
- Lin L, Liu D Y, Fu Q J, Guo X Y, Liu G L, Liu H and Wang S L. 2022. Seasonal variability of water residence time in the Subei Coastal Water, Yellow Sea: the joint role of tide and wind. *Ocean Modelling*, 180: 102137 [DOI: 10.1016/j.ocemod.2022.102137]
- Liu J Q, Ye X M, Song Q J, Ding J and Zou B. 2023. Products of HY-1C/D ocean color satellites and their typical applications. *National Remote Sensing Bulletin*, 27(1): 1-13 (刘建强, 叶小敏, 宋庆君, 丁静, 邹斌. 2023. HY-1C/D 海洋水色卫星产品体系及其典型应用. *遥感学报*, 27(1): 1-13) [DOI: 10.11834/jrs.20235002]
- Morel A and Prieur L. 1977. Analysis of variations in ocean color1. *Limnology and Oceanography*, 22(4): 709-722 [DOI: 10.4319/lo.1977.22.4.0709]
- Mukherjee I and Routroy S. 2012. Comparing the performance of neural networks developed by using Levenberg-Marquardt and Quasi-Newton with the gradient descent algorithm for modelling a multiple response grinding process. *Expert Systems with Applications*, 39(3): 2397-2407 [DOI: 10.1016/j.eswa.2011.08.087]
- Nazeer M, Ilori C O, Bilal M, Nichol J E, Wu W C, Qiu Z F and Gayene B K. 2021. Evaluation of atmospheric correction methods for low to high resolutions satellite remote sensing data. *Atmospheric*

- Research, 249: 105308 [DOI: 10.1016/j.atmosres.2020.105308]
- Nechad B, Ruddick K, Schroeder T, Oubelkheir K, Blondeau-Patissier D, Cherukuru N, Brando V, Dekker A, Clementson L, Banks A C, Maritorena S, Werdell P J, Sá C, Brotas V, Caballero de Frutos I, Ahn Y H, Salama S, Tilstone G, Martinez-Vicente V, Foley D, McKibben M, Nahorniak J, Peterson T, Siliò-Calzada A, Röttgers R, Lee Z, Peters M and Brockmann C. 2015. CoastColour Round Robin data sets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. *Earth System Science Data*, 7(2): 319-348 [DOI: 10.5194/essd-7-319-2015]
- O'Reilly J E, Maritorena S, Mitchell B G, Siegel D A, Carder K L, Garver S A, Kahru M and McClain C. 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans*, 103(C11): 24937-24953 [DOI: 10.1029/98JC02160]
- O'Reilly J E and Werdell P J. 2019. Chlorophyll algorithms for ocean color sensors-OC4, OC5 and OC6. *Remote Sensing of Environment*, 229: 32-47 [DOI: 10.1016/j.rse.2019.04.021]
- Pahlevan N, Smith B, Schalles J, Binding C, Cao Z G, Ma R H, Alikas K, Kangro K, Gurlin D, Hà N, Matsushita B, Moses W, Greb S, Lehmann M K, Ondrusek M, Oppelt N and Stumpf R. 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. *Remote Sensing of Environment*, 240: 111604 [DOI: 10.1016/j.rse.2019.111604]
- Ritchie R J. 2006. Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynthesis Research*, 89(1): 27-41 [DOI: 10.1007/s11120-006-9065-9]
- Rußwurm M, Klemmer K, Rolf E, Zbinden R and Tuia D. 2024. Geographic location encoding with spherical harmonics and sinusoidal representation networks. *arXiv preprint arXiv: 2310.06743* [DOI: 10.48550/arXiv.2310.06743]
- Shahvaran A R, Kheyrollah Pour H and Van Cappellen P. 2024. Comparative evaluation of semi-empirical approaches to retrieve satellite-derived chlorophyll-a concentrations from nearshore and offshore waters of a large lake (Lake Ontario). *Remote Sensing*, 16(9): 1595 [DOI: 10.3390/rs16091595]
- Shen C Y, Zhao H, Chen F J and Xiao H W. 2020. The distribution of aerosols and their impacts on chlorophyll-a distribution in the South China Sea. *Journal of Geophysical Research: Biogeosciences*, 125(6): e2019JG005490 [DOI: 10.1029/2019JG005490]
- Shi X H, Chen S G, Lin M S, Liu J Q, Ma C F, Song Q J, Xue C and Hu L B. 2023. Preliminary performance of the COCTS onboard HY-1D satellite in the global ocean. *National Remote Sensing Bulletin*, 27(4): 943-952 (史鑫皓, 陈树果, 林明森, 刘建强, 马超飞, 宋庆君, 薛程, 胡连波. 2023. 中国海洋水色卫星传感器 COCTS HY-1D 产品初步评价. *遥感学报*, 27(4): 943-952) [DOI: 10.11834/jrs.20221666]
- Smith B, Pahlevan N, Schalles J, Ruberg S, Errera R, Ma R H, Giardinò C, Bresciani M, Barbosa C, Moore T, Fernandez V, Alikas K and Kangro K. 2021. A chlorophyll-a algorithm for Landsat-8 based on mixture density networks. *Frontiers in Remote Sensing*, 1: 623678 [DOI: 10.3389/frsen.2020.623678]
- Sola J and Sevilla J. 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3): 1464-1468 [DOI: 10.1109/23.589532]
- Song Q J, Ma C F, Lin M S, Jiang X W, Wang L L, Xu P M, Tang J W, Chen P, Lu Y C, Wei J and Zhang K L. 2025. Advances in ocean color observation satellite technology. *National Remote Sensing Bulletin*, 29(6): 399-1425 (宋庆君, 马超飞, 林明森, 蒋兴伟, 王丽丽, 徐彭梅, 唐军武, 陈鹏, 陆应诚, 危峻, 张可立. 2025. 海洋水色卫星技术进展. *遥感学报*, 29(6): 399-1425) [DOI: 10.11834/jrs.20255013]
- Soomets T, Toming K, Paavel B and Kutser T. 2022. Evaluation of remote sensing and modeled chlorophyll-a products of the Baltic Sea. *Journal of Applied Remote Sensing*, 16(4): 046516 [DOI: 10.1117/1.JRS.16.046516]
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1): 1929-1958
- Teng Y, Zou B and Ye X M. 2022. Study on the chlorophyll a concentration retrieved from HY-1C satellite coastal zone imager data. *Haiyang Xuebao*, 44(5): 25-34 (滕越, 邹斌, 叶小敏. 2022. HY-1C 卫星海岸带成像仪叶绿素 a 浓度反演研究. *海洋学报*, 44(5): 25-34) [DOI: 10.12284/hyxb2022052]
- Tran M D, Vantrepotte V, Loisel H, Oliveira E N, Tran K T, Jorge D, Mériaux X and Paranhos R. 2023. Band ratios combination for estimating chlorophyll-a from Sentinel-2 and Sentinel-3 in coastal waters. *Remote Sensing*, 15(6): 1653 [DOI: 10.3390/rs15061653]
- Trombetta T, Vidussi F, Mas S, Parin D, Simier M and Mostajir B. 2019. Water temperature drives phytoplankton blooms in coastal waters. *PLoS One*, 14(4): e0214933 [DOI: 10.1371/journal.pone.0214933]
- Valente A, Sathyendranath S, Brotas V, Groom S, Grant M, Jackson T, Chuprin A, Taberner M, Airs R, Antoine D, Arnone R, Balch W M, Barker K, Barlow R, Bélanger S, Berthon J F, Beşiktepe Ş, Borsheim Y, Bracher A, Brando V, Brewin R J W, Canuti E, Chavez F P, Cianca A, Claustre H, Clementson L, Crout R, Ferreira A, Freeman S, Frouin R, García-Soto C, Gibb S W, Goericke R, Gould R, Guillocheau N, Hooker S B, Hu C, Kahru M, Kämpel M, Klein H, Kratzer S, Kudela R, Ledesma J, Lohrenz S, Loisel H, Mannino A, Martinez-Vicente V, Matrai P, McKee D, Mitchell B G, Moisan T, Montes E, Muller-Karger F, Neeley A, Novak M, O'Dowd L, Ondrusek M, Platt T, Poulton A J, Repecaud M, Röttgers R, Schroeder T, Smyth T, Smythe-Wright D, Sosik H M, Thomas C, Thomas R, Tilstone G, Tracana A, Twardowski M, Vellucci V, Voss K, Werdell J, Wernand M, Wojtasiewicz B, Wright S and Zibordi G. 2022. A compilation of global

- bio-optical in situ data for ocean colour satellite applications – version three. *Earth System Science Data*, 14(12): 5737-5770 [DOI: 10.5194/essd-14-5737-2022]
- Volpe G, Santoleri R, Vellucci V, d'Alcalà M R, Marullo S and d'Ortenzio F. 2007. The colour of the Mediterranean Sea: global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sensing of Environment*, 107(4): 625-638 [DOI: 10.1016/j.rse.2006.10.017]
- Wang D S, Du K P, Chen S G, Xue C, Ye X M and Lee Z P. 2023. Construction of ocean color remote sensing data processing system based on open source code: taking HY-1C/D as an example. *National Remote Sensing Bulletin*, 27(1): 68-78 (王道生, 杜克平, 陈树果, 薛程, 叶小敏, 李忠平. 2023. 基于开源代码构建水色遥感数据处理系统——以 HY-1C/D 为例. *遥感学报*, 27(1): 68-78) [DOI: 10.11834/jrs.20235008]
- Wright S W, Jeffrey S W, Mantoura R F C, Llewellyn C A, Bjørnland T, Repeta D and Welschmeyer N. 1991. Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. *Marine Ecology Progress Series*, 77(2/3): 183-196 [DOI: 10.3354/meps077183]
- Yang G Y, Ye X M, Xu Q, Yin X B and Xu S Y. 2023. Sea surface chlorophyll-a concentration retrieval from HY-1C satellite data based on residual network. *Remote Sensing*, 15(14): 3696 [DOI: 10.3390/rs15143696]
- Yang N S, Shi H Z, Tang H and Yang X. 2022. Geographical and temporal encoding for improving the estimation of PM_{2.5} concentrations in China using end-to-end gradient boosting. *Remote Sensing of Environment*, 269: 112828 [DOI: 10.1016/j.rse.2021.112828]
- Ye X M, Liu J Q, Lin M S, Ding J, Zou B and Song Q J. 2021. Global ocean chlorophyll-a concentrations derived from COCTS onboard the HY-1C satellite and their preliminary evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 9914-9926 [DOI: 10.1109/TGRS.2020.3036963]
- Yu X L, Lee Z P, Shang Z H, Lin H and Lin G. 2021. A simple and robust shade correction scheme for remote sensing reflectance obtained by the skylight-blocked approach. *Optics Express*, 29(1): 470-486 [DOI: 10.1364/OE.412887]
- Zhang T L, Fell F, Liu Z S, Preusker R, Fischer J and He M X. 2003. Evaluating the performance of artificial neural network techniques for pigment retrieval from ocean color in Case I waters. *Journal of Geophysical Research: Oceans*, 108(C9): 3286 [DOI: 10.1029/2002JC001638]
- Zhang T L and He M X. 2002. A method to retrieve the oceanic chlorophyll-a concentrations in case I water based on artificial neural network. *Journal of Remote Sensing (in Chinese)*, 6(1): 40-44 (张亭禄, 贺明霞. 2002. 基于人工神经网络的一类水域叶绿素-a 浓度反演方法. *遥感学报*, 6(1): 40-44) [DOI: 10.11834/jrs.20020108]
- Zhang Y, Shen F, Li R H, Li M Y, Li Z X, Chen S Y and Sun X R. 2024. AIGD-PFT: the first AI-driven global daily gap-free 4 km phytoplankton functional type data product from 1998 to 2023. *Earth System Science Data*, 16(10): 4793-4816 [DOI: 10.5194/essd-16-4793-2024]
- Zheng Q H, Yang M Q, Yang J J, Zhang Q R and Zhang X X. 2018. Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process. *IEEE Access*, 6: 15844-15869 [DOI: 10.1109/ACCESS.2018.2810849]
- Zhu B H, Yuan Y B and Hu J B. 2021. Spatial and temporal distribution characteristics of chlorophyll-a in the Eastern China seas based on field data. *Jiangsu Science and Technology Information*, 38(34): 74-80 (朱碧泓, 袁逸博, 胡剑波. 2021. 基于现场数据的东中国海叶绿素时空分布特征. *江苏科技信息*, 38(34): 74-80) [DOI: 10.3969/j.issn.1004-7530.2021.34.021]
- Zhu Y F, Zhu L, Li J G, Chen Y J, Zhang Y H, Hou H Q, Ju X and Zhang Y Z. 2017. The study of inversion of chlorophyll a in Taihu based on GF-1 WFV image and BP neural network. *Acta Scientiae Circumstantiae*, 37(1): 130-137 (朱云芳, 朱利, 李家国, 陈宜金, 张永红, 侯海倩, 鞠星, 张雅洲. 2017. 基于 GF-1 WFV 影像和 BP 神经网络的太湖叶绿素 a 反演. *环境科学学报*, 37(1): 130-137) [DOI: 10.13671/j.hjkxxb.2016.0275]

Retrieving Chlorophyll-a concentration by using a neural network for COCTS sensors aboard HY-1C/1D satellites

WEI Meiyi¹, WU Jingyu¹, ZHENG Lufei¹, WANG Daosheng¹, LEE Zhongping¹, SHANG Shaoling¹,
YE Xiaomin², LIN Gong¹, YU Xiaolong¹

1. National Key Laboratory of Marine Biogeochemistry, Xiamen University, Xiamen 361102, China;

2. National Satellite Ocean Application Service, Beijing 100081, China

Abstract: Chlorophyll-a (Chl-a) concentration is an essential climate variable and fundamental to global carbon cycle studies and ocean environmental monitoring. HY-1C/1D satellites, equipped with the Chinese Coastal Ocean Color and Temperature Scanner (COCTS), enable global ocean color monitoring at kilometer-scale resolution. High-accuracy remote sensing algorithms for Chl-a concentration on the basis

of COCTS data need to be developed to fully leverage these Chinese autonomous satellites for ocean monitoring and climate research.

This study developed a Chl-a retrieval algorithm on the basis of a multilayer perceptron neural network (MLP-NN) for the COCTS sensor. The model inputs included remote sensing reflectance (R_{rs}) at COCTS center bands and environmental variables, such as geolocation, Sea Surface Temperature (SST), and Photosynthetically Active Radiation (PAR). The model was trained using 2,165 in-situ measurements collected from the global ocean. After a comparative analysis of mainstream machine learning models, MLP was selected as the core architecture for the NN framework. A multidimensional feature fusion strategy was implemented to construct the MLP-NN model. Given that multidimensional inputs could introduce redundancy, sensitivity analysis was conducted to quantify the contribution of each input, identify the optimal input set, and improve the model's efficiency and generalization.

The sensitivity analysis identified the following optimal combination for MLP-NN: R_{rs} at 412, 443, 490, 520, 565, and 670 nm; latitude; month; average SST from the previous month; and climatological PAR from the previous month. Validation indicated that Chl-a estimated by MLP-NN achieved a Root Mean Square Difference (RMSD) of 0.22 and a Median Absolute Percentage Difference (MAPD) of 29.1% for log-transformed Chl-a, which are 0.1 and 16.9% lower than those estimated by the NASA operational Ocean Color Index (OCI) algorithm, respectively. Further validation using satellite and in-situ matchups confirmed that MLP-NN outperformed OCI, reducing RMSD and MAPD by 0.09 and 9.8%, respectively, highlighting its improved robustness. In China's Bohai Sea, both algorithms effectively captured the spatial distribution patterns of Chl-a. However, OCI exhibited systematic bias, underestimating Chl-a concentrations at high and low extremes. By contrast, the MLP-NN model demonstrated high accuracy in retrieving extreme Chl-a values.

Overall, the MLP-NN model developed in this study substantially improves the estimation of Chl-a concentrations from HY-1C/1D satellite observations. It offers valuable algorithmic support for leveraging domestic satellites in ocean ecological monitoring.

Key words: Chlorophyll-a (Chl-a), remote sensing reflectance, retrieval algorithm, HY-1C/1D satellites, neural network, COCTS, ocean color

Supported by National Key Research and Development Program of China (No. 2022YFC3104900, 2022YFC3104903)