J. Lake Sci.(湖泊科学), 2026, 38(3):000-000
DOI 10.18307/2026.0316

基于便携式高光谱近感快速监测内陆水体叶绿素 a 浓度*

骆夏杨[1, 2, 3], 李娜[1, 2**], 张运林[1, 2, 3**], 郭宇龙[4], 马宗伟[5]
(1:中国科学院南京地理与湖泊研究所湖泊与流域水安全重点实验室, 南京 211135)
(2:中国科学院大学南京学院, 南京 211135)
(3:中国科学院大学, 北京 100049)
(4:南京中科深瞳科技研究院有限公司, 南京 211899)
(5:江苏省无锡谱视界科技有限公司, 无锡 214131)

摘　要：浮游植物叶绿素 a（Chl-a）浓度是评价水体富营养化程度的重要指标。然而传统实验室分析法流程繁琐、费时费力；水下原位传感器易受生物附着、精度低、运维成本高；传统的卫星和无人机遥感受限于大气校正、专业化程度高、时效性差，皆难以满足高精度的即时监测需求。高光谱近感技术的出现和应用，解决了传统监测方法的不足，提升了内陆水体 Chl-a 浓度的监测效率。因此，本研究利用一种新型便携式高光谱近感水质监测设备，结合 2021-2024 年 8 个湖泊、水库和河流获取的 533 个同步实测 Chl-a 浓度数据，引入线性回归算法与机器学习算法，构建和比选了高精度 Chl-a 浓度反演模型，实现了 Chl-a 浓度的即时监测。对比基于线性回归算法、随机森林（Random Forest，RF）算法、极度梯度提升树算法（Extreme Gradient Boosting，XGBoost）和支持向量机算法（Support Vector Machine，SVM）构建的 Chl-a 浓度估算模型发现，基于 XGBoost 的 Chl-a 浓度估算模型性能最好（R^2=0.87，*RMSE*=6.02 μg/L，*MAE*=3.98 μg/L）。该监测方法实现了光谱采集与 Chl-a 浓度估算的同步性，简化了流程，降低了专业化门槛，显著提升了野外监测效率。
关键词：便携式高光谱；近感监测；叶绿素 a；内陆水体；机器学习。

Rapid monitoring of chlorophyll-a concentration in inland water bodies using a portable proximal sensing technology*

Luo Xiayang[1, 2, 3], Li Na[1, 2**], Zhang Yunlin[1, 2, 3**], Guo Yulong[4] & Ma Zongwei[5]
(1: Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 211135, P.R.China)
(2: College of Nanjing, University of Chinese Academy of Sciences, Nanjing 211135, P.R.China)
(3: University of Chinese Academy of Science, Beijing 100049, P.R.China)
(4: Nanjing Zhongke Deep Insight Technology Research Institute Co., Ltd., Nanjing 211899, P.R.China)
(5: Wuxi Specvision Technology Co., Ltd., Wuxi 214131, P.R.China)

Abstract: Phytoplankton chlorophyll-a (Chl-a) concentration is a key indicator for assessing water eutrophication status. Conventional monitoring approaches face significant limitations: laboratory analyses are time-consuming and labor-intensive, while in situ sensors are susceptible to biofouling, low accuracy, and high maintenance costs. Traditional satellite remote sensing is also unsuitable for high-precision, real-time monitoring due to challenges such as atmospheric correction errors, technical

* 2025-03-21 收稿; 2025-09-25 收修改稿。
江苏省生态环境科研项目（2023003）、江苏省重点研发计划（产业前瞻与关键核心技术）项目（BE2022152）、国家重点研发项目（2022YFC3204100）、江苏省卓越博士后计划和国家自然科学基金项目（42401479）联合资助。
** 通信作者; E-mail: nli@niglas.ac.cn，ylzhang@niglas.ac.cn。

complexity, and low temporal resolution. The emergence of hyperspectral proximal sensing technology offers a promising alternative for improving Chl-a monitoring efficiency. In this study, we utilized a novel portable hyperspectral proximal sensing device to collect 533 synchronized in situ Chl-a measurements across eight lakes, reservoirs, and rivers between 2021 and 2024. We developed and compared high-accuracy Chl-a inversion models using both linear regression and machine learning methods. Among the evaluated algorithms—linear regression, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM)—the XGBoost-based model exhibited the best performance (R^2 = 0.87, *RMSE* = 6.02 μg/L, *MAE* = 3.98 μg/L). This approach enables simultaneous spectral acquisition and Chl-a estimation, streamlining field monitoring workflows, lowering technical barriers, and significantly improving operational efficiency.
Keywords: Portable hyperspectral; proximal sensing; Chlorophyll-a; Inland waters; Machine learning.

淡水资源是支撑人类生存与可持续发展的基础性资源[1,2]。然而，随着工业化与城市化的快速推进，大量污染物持续汇入湖泊、水库等内陆水体，导致营养盐浓度显著升高[3]，加剧了水体富营养化进程，严重削弱了水体的自净能力和生态系统稳定性[4,5]。在全球气候变化背景下，蓝藻水华在我国频发，对生态安全构成严峻挑战[6,7]。水质持续恶化不仅威胁饮用水安全，也对粮食生产与公共健康产生潜在影响[8]。因此，建立科学完善的水环境管理方法和高效的监测技术体系已成为当务之急。浮游植物叶绿素 a（Chl-a）作为藻类普遍存在的重要色素分子，是指示蓝藻水华的关键指标[9]。Chl-a 浓度不仅能直接反映水体中藻类生物量，还可间接表征水体营养状态，因而在水质监测与生态评价中具有重要应用价值[10,11]。然而，湖库中的 Chl-a 浓度常呈现显著时空不确定性，是光照强度、温度、营养盐浓度及水体扰动等多重环境因子的综合作用的结果[12,13]。这一特性对数据获取的时效性与精度提出了更高要求。因此，建立一种高效、精准、快速的 Chl-a 浓度监测方法，对于水质管理和内陆水体生态系统保护具有重要的技术支撑意义。

目前，Chl-a 浓度的监测主要依赖原位采样与实验室分析方法，该方法虽然精度较高，但操作流程繁琐，数据获取周期长，难以满足动态变化的快速响应需求。多参数水质分析仪如 EXO 2（美国 YSI 公司）、Hydrolab DS 5（美国 HACH 公司）等原位自动监测设备，尽管其时效性更强，但传感器易受藻类等生物附着干扰，长期运行的可靠性与精度难以保障[14]。此外调制叶绿素荧光仪 Phyto-PAM（德国 Walz 公司）和野外藻类分析仪 BBE（德国 BBE Moldaenk 公司）等高端荧光设备虽具备较强的藻类识别与分类能力，但高昂的采购与维护成本限制了其在常规水质监测中的推广应用[15]。遥感技术可实现中大尺度水体的 Chl-a 浓度反演，但常规卫星遥感受限于重访周期长、波段设置不精细以及大气干扰显著等问题，难以对小型湖泊、城市水体等异质性强的水域实现高精度、实时监测[16,17]。无人机与地面高光谱设备（如 ASD、SPECIM）虽在高分辨率近感监测中具备一定优势，但其设备体积大、续航能力弱、操作复杂，不适合开展高频次的连续观测[18]。高光谱近感技术凭借其光谱分辨率高、实时监测能力强和不受大气干扰等优势，已在内陆水体水质监测领域得到了广泛的应用[19,20]。便携式高光谱近感监测设备作为一种新型的集成化仪器，进一步发挥了该技术快速监测的优势，并兼具便携特性，在水质参数原位监测领域展现出巨大潜力。例如，张立福等[21]基于一种便携式高光谱水质监测仪对于宁波某地水体的总氮、总磷进行监测，与传统实验室水质分析相比，实现了快速、实时、智能的技术突破。然而，目前该设备尚未建立针对 Chl-a 浓度的有效反演方法。因此，本研究旨在开发适用于该设备的 Chl-a 浓度反演算法。

目前，Chl-a 浓度的遥感反演方法主要可分为三类：基于统计关系的经验模型、基于生物光学原理的半解析模型和解析模型，以及数据驱动的机器学习方法[18,22]。在具体应用方面，徐逸等[23]采用 HJ-1B 卫星 CCD 多光谱影像数据，结合深度学习算法构建了太湖 Chl-a 浓度反演模型，模型验证结果显示其预测精度为：决定系数（R^2）=0.91，均方根误差（*RMSE*，Root Mean Square Error）=3.46 μg/L。包颖等[24]分别基于 GOCI 影像的 b5、b6、b7 波段数据，运用二波段半分析模型成功实现了太湖 Chl-a 浓度的分类反演，该研究实现了时间分辨率达 1 h 的遥感监测，精度最高达 R^2=0.82，为水体动态监测提供了新思路。梁文广等[25]利用 Sentinal-2 卫星影像的 B5 和 B9 波段和 FA-SVR 模型对于骆马湖的 Chl-a 浓度进行反演，模型精度达到 R^2=0.86，*RMSE*=2.77 μg/L，平均绝对误差（*MAE*，Mean Absolute Error）=2.10 μg/L，分析了骆马湖

的水质环境情况及其成因。其中，机器学习方法凭借其强大的非线性建模能力和环境适应性优势，已成为当前水体 Chl-a 浓度高精度反演研究中的重要技术手段，尤其在处理高维非线性回归问题方面表现出显著优越性[26]。虽然轻量级深度学习模型在遥感反演 Chl-a 浓度中展现出潜力，但其训练过程通常对初始权重、学习率等超参数敏感，可能导致结果波动[27]。相比之下，在小样本条件下，传统机器学习算法如随机森林（Random Forest，RF）、极度梯度提升树（Extreme Gradient Boosting，XGBoost）和支持向量机（Support Vector Machine，SVM）通常具有更稳定的性能表现。

因此，本研究基于便携式高光谱近感设备获取的光谱数据，结合同步实测 Chl-a 浓度数据集，采用多种机器学习算法构建其高精度反演模型，并通过系统验证，为水体 Chl-a 浓度的实时快速监测提供新的技术方案。

1 材料与方法

1.1 采样点概况

本研究选取了中国东部和西南地区具有代表性的内陆水体作为研究对象，包括太湖流域、滇池、沙河水库、大溪水库、千岛湖、洪泽湖、白马湖以及南京市内河道。这些水体在类型、营养状态和生态环境特征等方面具有显著差异，能够较好地反映中国内陆水体的空间异质性和环境梯度特征。太湖（30°55'N~31°32'N，120°08'E~120°43'E）是我国第三大淡水湖，平均水深约 2.0 m，属于典型的浅水富营养化湖泊，夏季蓝藻水华频发[28]。滇池（24°40'N~25°02'N，102°36'E~102°47'E）位于云贵高原，是该区域面积最大的淡水湖泊，平均水深约 5.3 m，长期处于富营养化状态，具有典型的高原深水湖泊特征。在研究的水库样本中，千岛湖（29°22'N~29°50'N，118°34'E~119°15'E）作为深水水库的代表，最大水深超过 90 m，整体处于中营养水平；沙河水库（31.25°N~31.40°N，119.35°E~119.50°E）和大溪水库（31.25°N~31.35°N，119.30°E~119.40°E）均位于江苏省，属于人工调控型水库，水力滞留时间较长，水体流动性较低，受周边人类活动影响显著[29]。洪泽湖（33°06'N~33°40'N，118°10'E~118°52'E）与白马湖（33°15'N~33°25'N，118°55'E~119°05'E）为淮河流域重要的浅水通江湖泊，兼具渔业资源开发与生态保护功能，其水文情势受流域来水和人为调度共同影响。南京市内河道（31°59'N~32°07'N，118°42'E~118°50'E）作为高度城市化区域的水体代表，水质受城市排污、地表径流和人类活动干扰较为强烈，表现出典型的城市水体污染特征[30]。这些研究区域涵盖了从贫营养到富营养、从深水到浅水、从自然湖泊到人工水库等不同类型的内陆水体，其多样化的环境特征为水质参数反演模型的建立和验证提供了理想的研究平台。

1.2 Chl-a 与总悬浮物采样与浓度测定分析方法

水样采集工作于 2022 年 11 月 10 日至 2024 年 1 月 24 日期间开展，所有采样均在日间（08:00—17:00）进行。采样涵盖春、夏、秋、冬四季，其中夏季样本占比 10.2%以覆盖蓝藻高发期。采样过程中，使用专业采水器采集水体表层（水面 0~20 cm）水样，采集后立即将样品置于 4℃冷藏箱中保存，并尽快运送至实验室进行后续水质参数分析。采集样本统计描述如表 1 所示。

由于内陆水体 Chl-a 浓度通常指浮游植物 Chl-a 浓度，本文所涉及的 Chl-a 浓度全部指浮游植物 Chl-a 浓度。本研究严格采用“热乙醇-分光光度法”测定水体中 Chl-a 浓度[31]。具体步骤如下：首先，使用孔径为 0.7 μm 的 Whatman GF/F 玻纤滤膜进行过滤；随后，将滤膜置于-80°C 冷冻 48 h，以充分破碎蓝藻细胞；然后，采用 90%热乙醇（80°C）对滤膜中的 Chl-a 进行提取。最后，利用岛津 Shimadzu UV-2550 分光光度计分别测量加入 1 mol/L 稀盐酸酸化前后 665 nm 和 750 nm 的吸光度，对比计算得到水体中 Chl-a 浓度。

总悬浮物（TSM）浓度采用烘干称重法进行测定[32]，将表层水样本用经过 550℃预灼烧 4 h 的 0.7 μm 孔径 Whatman GF/F 玻纤滤膜过滤，随后将滤膜置于 105℃烘干，通过比较滤膜前后重量差值计算 TSM 浓度。

1.3 水体遥感反射率采集与计算

利用美国 ASD 公司生产的光谱分辨率为 1 nm、测量范围 350~2500 nm 的 ASD 地物光谱仪（FieldSpec spectroradiometer，ASD FieldSpec）进行水体遥感反射率（$Rrs(\lambda)$）光谱采集。具体步骤如下：首先使用 ASD

地物光谱仪垂直测量标准灰板的辐射亮度（L_{p1}），然后在太阳光入射方位角90°~135°、与水面法线呈30°~45°夹角的方向上测量水体总辐射亮度（L_{sw}），接着在同一观测平面内将探头向上旋转 90°~120°测量天空漫散射辐射亮度（L_{sky}），最后再次测量标准灰板的辐射亮度（L_{p2}）并计算两次灰板测量的均值（L_p）。基于测量数据，先计算水面下行辐照度（$E_d(0+)$），再结合风速条件选取水-气界面反射率计算离水辐亮度（L_w），最终得到遥感反射率。每次采集 10 条光谱取平均值以减少光场变化的影响。计算公式如下所示。

表 1 2022-2024 年不同内陆水体野外原位水质样品采集统计表

Table 1 Statistics for in situ water samplings in different inland water during 2022-2024

类型	研究区	县市	样品数量	采样时间
湖库	白马湖	宝应县、淮安市	4	2023/2/16
	洪泽湖	淮安市	5	2023/2/15
	千岛湖	淳安县	9	2023/1/31-2/2
	太湖	无锡市、苏州市、湖州市	183	2022/11/8-11/17、2023/5/29-5/31、2023/6/1-6/3、2023/6/7-6/9
	沙河水库	溧阳市	63	2023/2/13、2024/1/22-1/24
	大溪水库	溧阳市	17	2022/11/7
	滇池	昆明市	19	2022/12/26-12/28
河流	环太湖河道	无锡市、苏州市、湖州市	149	2022/11/18、2023/2/8、2023/2/10-2/11、2023/2/13-2/14、2023/2/17、2023/2/20
	南京市内河道	南京市	78	2022/11/22-11/24、2022/11/30

$$L_p = \frac{L_{p1} + L_{p2}}{2} \quad (1)$$

$$L_{sw} = L_w + r_{sky} \times L_{sky} \quad (2)$$

$$E_d(0+) = L_p \times \pi \div P_\rho \quad (3)$$

$$Rrs = L_w \div E_d(0+) \quad (3)$$

式中，r_{sky}为水-气界面对天空光漫散射辐射的反射率，其大小与太阳高度、风速和水面粗糙度有关。其中，当静风条件下 r_{sky}取值为 2.2%；当风速介于 5 m/s 左右时，r_{sky}取值为 2.5%；当风速为 10 m/s 左右时 r_{sky}取值介于 2.6~2.8%。P_ρ为标准灰板的反射率，取值为 30%。

1.4 便携式高光谱近感光谱采集与处理

便携式高光谱近感设备由无锡谱视界研发，是一款集成高效的近感光谱监测系统。该设备采用分层架构，融合了传感采集、无线通信与智能终端处理技术。由滑动镜头盖和校准板、偏振选择镜头、微型光谱模块、内部电路和磁吸外壳等组成，可通过磁吸支架吸附在智能手机背面，通过蓝牙连接手机上的 APP，执行光谱校准、光谱测量、数据传输、参数反演等一系列功能（图 1）。前端感知模块基于微型光谱单元实现水体反射率的实时采集，并通过低功耗蓝牙协议稳定、高效地传输至智能手机端。终端应用集成光谱校准、反演模型计算、结果可视化及数据管理等功能模块，构成完整的数据闭环处理链，实现了从原始信号获取到水质参数估算与展示的全自动化处理，显著降低了操作复杂性与人为误差，提升了现场监测的即时性与效率。便携式高光谱近感设备的镜头具有 5°视场角，较小的视场角可以减小测量结果受到漂浮物等的干扰，内置的偏振元件可选择性过滤水面反射光，使光谱测量更准确、更稳定。仪器内部采用微型光栅进行分光，用线阵探测器记录 400~850 nm 范围的光谱数据，光谱分辨率约 10 nm，光谱采样间隔为 1 nm。

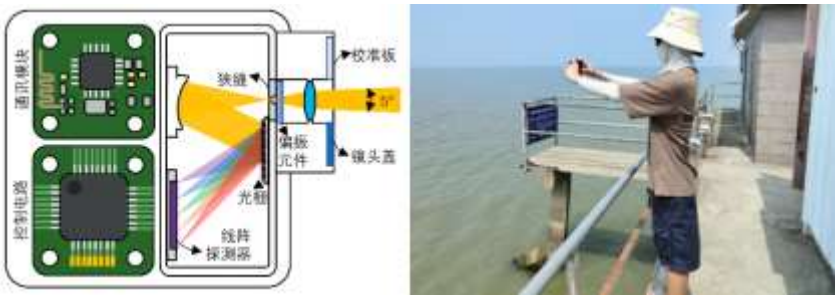

图 1 仪器结构与实际操作示意图

Fig.1 Instrument structure and practical operation schematic diagram

使用便携式高光谱近感设备进行研究区光谱采集时，首先需要进行光谱校准，即分别用乳白色校准板和金属镜头盖盖住镜头采集光源和暗背景光谱，并根据内置的反射率校准系数计算目标物的实际光谱反射率。光谱校正完成后，可以进行水体光谱采集，采集过程中，保证便携式近感设备竖直倾角约为 53°，水平倾角约为 0°。拍摄时背对太阳，拍摄方向保证为太阳照射方向，向左或者向右倾斜 45°，保证水面不能有太阳直射光的镜面反射或阴影遮挡。随即进行连续拍摄 10 条光谱并保存其均值高光谱数据。在拍摄过程中应避免水生植物等干扰。为保证光谱数据与水质参数的时空一致性，光谱采集与水质采样需同步进行，两者采集时间间隔不得超过 2 min。研究采用 Savitzky-Golay 卷积平滑算法对原始光谱数据进行预处理，以有效抑制噪声并保留光谱形态特征。该方法基于局部多项式最小二乘拟合原理，选取二阶多项式进行拟合，并采用窗口宽度为 11 个采样点（对应约 10 nm 波长范围）的移动窗口进行卷积计算。基于对光谱信噪比的评估，位于 400~449 nm 和 721~850 nm 的波段反射率数据因偏差百分比大于 7%而被剔除，最终研究采用 450~720 nm 的光谱范围进行后续分析。

1.5 Chl-a 浓度反演模型建立方法

本研究中采用几种成熟的机器学习算法，包括 RF、XGBoost、SVM。并将机器学习算法与传统的线性回归模型精度进行比较。

RF 是一种利用集成学习思想构建多棵决策树的集成算法，其基本单元是一棵决策树，每棵决策树都是一个预测器[33,34]。通过投票(分类问题)或平均(回归问题)对所有决策树的建模结果进行集成[35]，因此，整个模型的结果具有较高的准确性和泛化性能。随机森林回归算法可以在大数据集上高效运行，可以处理具有高维特征而不降维的输入样本，可以评估每个输入特征对模型的重要性。XGBoost 是 boosting 算法的一种实现方式，其基本组成元素是决策树；我们将这些决策树成为“弱学习器”，这些“弱学习器”共同组成了 XGboost[36]。组成 XGBoost 的决策树之间是有先后顺序的；后一棵决策树的生成会考虑前一棵决策树的预测结果，即将前一棵决策树的偏差考虑在内（在目标函数中有体现）。生成每棵决策树使用的数据集，

是整个数据集。所以可以将每棵决策树的生成都看作是一个完整的决策树生成过程[37]。SVM 是一类按监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面，SVM 使用铰链损失函数计算经验风险并在求解系统中加入了正则化项以优化结构风险，是一个具有稀疏性和稳健性的分类器[38]。SVM 可以通过核方法进行非线性分类，是常见的核学习方法之一。

建模前首先根据 *Z-score* 标准分数进行异常值筛选[39]，将绝对值大于 3 的进行清洗，清洗后的数据集共 532 条。

$$Z-score\ =\frac{X-\mu}{\sigma} \tag{5}$$

其中 X 代表原始数据，μ代表所有数据的平均值，σ代表所有数据的标准差。

采用 Python 3.11 编程环境下的机器学习工具包实现模型构建与优化。在建模前，首先对所有输入波段进行标准归一化处理（StandardScaler）。其中 SVM 和 RF 模型使用 scikit-learn 实现，XGBoost 模型使用 XGBoost 库实现。建模时采用 7：3 的比例随机将总数据集划分为训练集和验证集。通过采用格网搜索方法，结合五折交叉验证，优化各机器学习模型的超参数组合，以有效降低模型的过拟合风险。最终采用的参数组合如表 2 所示。此外，本研究引入 SHAP（SHapley Additive exPlanations）分析，该分析基于 Python 3.11 编程环境的 shap 包实现，用于解释机器学习模型的预测结果并量化各输入特征对输出结果的影响贡献。

表 2 各机器学习模型的超参数组合

Tab.2 Hyperparameter combinations of each machine learning model

模型	参数			
SVM	C	cache_size	epsilon	gamma
	22	200	0.2	auto
	kernel	shrinking	tol	
	rbf	True	0.001	
RF	bootstrap	max_ depth	max_ features	min_samples_leaf
	True	20	sqrt	1
	min_samples_ split	n_estimators		
	5	100		
XGBoost	colsample_ bytree	gamma	learning_ rate	max_ depth
	0.8	0	0.01	10
	min_child_weight	n_ estimators	ubsample	
	10	300	0.8	

1.6 数据处理与统计方法

利用 SPSS（Statistical Product and Service Solutions）22.0 进行最大值、最小值、平均值及标准差的计算，皮尔逊相关性的检验，线性拟合检验，光谱角测度（Spectral angle metric，SAM）与模型评价指标的计算，规定 $p \leq 0.05$ 时具有统计学意义上的显著。为了评价模型性能，本研究选用 3 个指标评价，包括 R^2、*RMSE* 和 *MAE*。引入误差贡献度 P_K 对各水体对于模型总误差的贡献度进行评估。引入 SAM 对光谱的一致性进行评价，SAM 越小，则待测光谱与参考光谱的形状相似性越高。采用 OriginLab 2019 和 ArcGIS 10.2 实现制图。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Meas_i - Esti_i)^2}{\sum_{i=1}^{n}(Meas_i - \overline{Esti_i})^2} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Meas_i - Esti_i)^2}{n}} \tag{7}$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} |Meas_i - Esti_i| \tag{8}$$

$$P_K = \frac{\sum_{i=1}^{n_k}\left(Meas_{i_i}^{\ k} - Esti_{i_i}^{\ k}\right)^2}{\sum_{j=1}^{K}\sum_{i=1}^{n_j}\left(Meas_{i_i}^{\ j} - Esti_{i_i}^{\ j}\right)^2} \times 100\% \tag{9}$$

其中，$Meas_i$ 代表实验室测量的真实值，$Esti_i$ 代表预测量，n 代表样本数，n_k 表示第 k 个水体的样本数，K 代表总水体数。

$$\mathrm{SAM}\,(x, y) = \cos^{-1}\frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}} \tag{10}$$

其中 x_i 代表待测光谱第 i 波段反射率；y_i 代表参考光谱第 i 波段反射率

2 结果与分析

2.1 光谱准确性分析

在不同环境条件下，采用 ASD 地物光谱仪采集水面实测高光谱与便携式高光谱近感监测设备所采集的高光谱数据同步进行比对（图 2）。结果表明，两者的线性拟合斜率为 0.97，表明便携式高光谱近感监测设备采集的光谱存在系统性偏低；但 R^2为 0.98，*RMSE* 为 0.003 sr^{-1}，*MAE* 为 0.002 sr^{-1}，表明两套设备的光谱数据整体表现出较好的一致性。此外，通过计算 SAM 评估两套设备光谱数据的光谱相似性，其中，SAM 最大值为 4.11°，最小值为 1.70°，平均值为 3.29±0.83°，表明 ASD 地物光谱仪与便携式高光谱近感设备采集的光谱在形状上高度相似。综上所述，便携式高光谱近感监测设备获取的光谱数据与标准 ASD 光谱数据高度一致，具备较高的可靠性和精度。

2.2 实测水质分析

在建模前，对各研究水体的实测 Chl-a 浓度和 TSM 浓度进行了统计分析（表 3）。总体来说，不同水体的水质特征差异较大，Chl-a 浓度变化范围为 0.54~75.30 μg/L，TSM 浓度变化范围为 0.81~127.10 mg/L。具体来说，实测 Chl-a 浓度最高值出现在太湖，为 75.30 μg/L，最低值出现在环太湖河道，为 0.54 μg/L。实测 TSM 浓度最高值同样出现在太湖，为 127.10 mg/L，最低值出现在千岛湖，为 0.81 mg/L。太湖 Chl-a 平均浓度（36.82 μg/L）和 TSM 平均浓度（35.50 mg/L）均为所有水体中最高，表明太湖富营养化水平最高；其次是洪泽湖和南京市内河道，Chl-a 平均浓度分别为 20.81 μg/L 和 13.11 μg/L。相比之下，千岛湖的 Chl-a 和 TSM 平均浓度最低，分别为 2.31 μg/L 和 5.34 mg/L，表明其水质最优。

2.3 Chl-a 浓度反演模型构建与验证

分别对训练集与验证集的 Chl-a 浓度进行了基本统计（表 4），在建模数据集中，训练集 Chl-a 浓度为

0.54~75.30 μg/L，平均值为 19.49 μg/L，样本量为 372；验证集浓度范围为 1.00~73.71 μg/L，平均值为 19.87 μg/L，样本量为 160。训练集与验证集在浓度分布和均值上保持一致性，确保了模型训练的代表性合结果的可代表性。

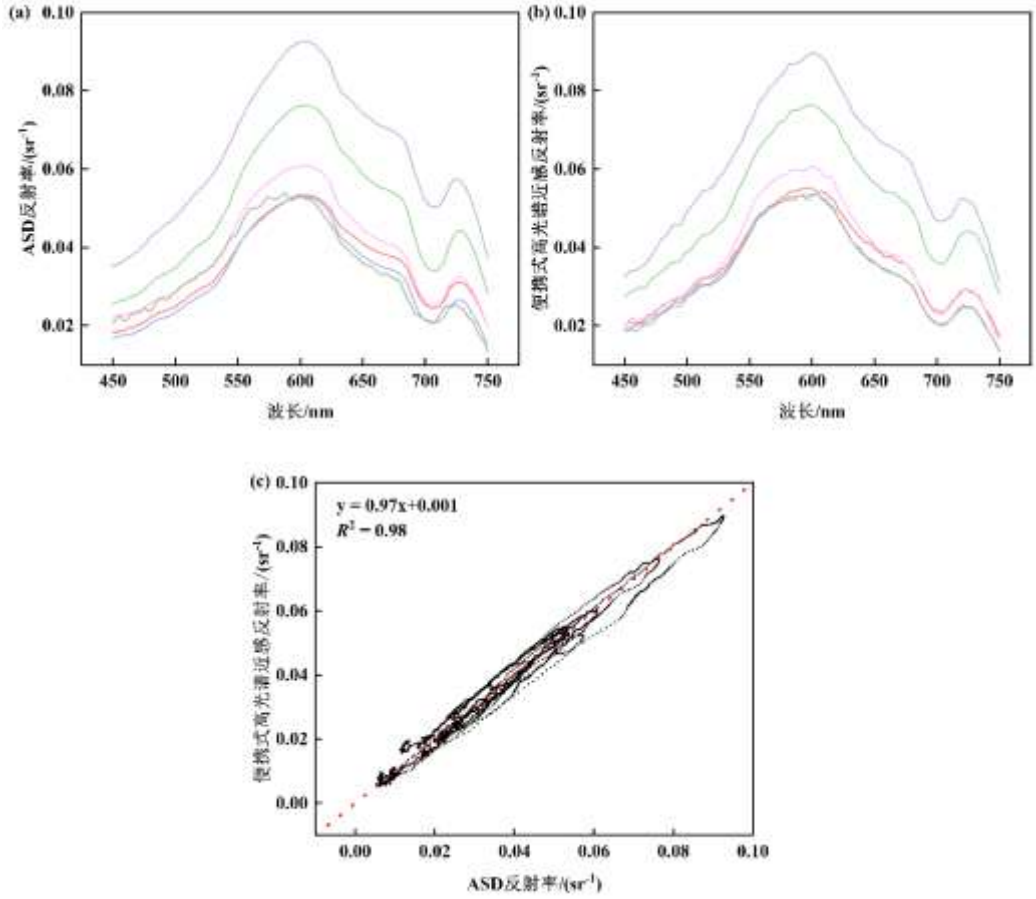

图 2 ASD 地物光谱仪与便携式高光谱近感监测设备实测高光谱比对（a，b）与拟合图（c）

（a 图与 b 图中同颜色曲线为同一水面目标的光谱数据，颜色一一对应）

Fig.2 Comparison of the measured hyperspectra of the ASD geophysical spectrometer and the portable proximal monitoring equipment (a, b) and the fitted plot (c) (Curves with the same color in (a) and (b) represent spectral data from the same water surface target, with one-to-one color correspondence)

在构建 Chl-a 浓度的线性回归反演模型前，首先基于便携式高光谱近感监测设备采集的 450–720 nm 波段范围内的高光谱反射率数据，开展与实测 Chl-a 浓度的皮尔逊相关性分析（图 3），以识别对 Chl-a 浓度变化最为敏感的波段区域。结果表明，单波段 r 值最高在 707 nm，为 0.47（p≤0.05），其次为 708 nm，r 值为 0.469（p≤0.05）。因 675 nm 处 Chl-a 强吸收作用会形成反射谷，选取波段 675 nm 和波段 707 nm 构建归一化模型（图 4），所用模型如式（10）所示。

$$\mathrm{C(Chl-a)} = 9.17 \cdot \frac{R_{rs}(707) - R_{rs}(675)}{R_{rs}(707) + R_{rs}(675)} + 19.47 \tag{11}$$

结果表明，反演结果较差，具体表现为：训练集：R^2=0.29，*RMSE*=14.59 μg/L，*MAE*=11.11 μg/L；验证集：R^2=0.34，*RMSE*=13.85 μg/L，*MAE*=10.46 μg/L。难以满足精度需求。

机器学习能够深入挖掘数据中的潜在模式和关联，进而实现对复杂问题的精准建模。因此分别采用 SVM、RF 和 XGBoost 对 Chl-a 浓度进行反演，为了更好的进行潜在联系的学习，采用 450~720 nm 的全

波段作为输入进行反演，并对反演精度进行比较（图 5）

表 3 实测 Chl-a 和 TSM 浓度的描述性统计

Tab.3 Descriptive statistics and sample size of measured concentration of Chl-a and TSM

点位	Chl-a 浓度/(μg/L) 最小值~最大值	Chl-a 浓度/(μg/L) 平均值	TSM 浓度 /(mg/L) 最小值~最大值	TSM 浓度 /(mg/L) 平均值
太湖	4.77~75.30	36.82	9.17~127.10	35.50
环太湖河道	0.54~40.73	9.36	3.05~80.20	19.69
沙河水库	5.03~18.60	10.76	9.92~63.10	24.82
大溪水库	8.45~11.44	9.91	5.63~10.80	7.71
千岛湖	1.00~5.60	2.31	0.81~11.50	5.34
洪泽湖	11.60~33.97	20.81	4.23~5.54	4.76
白马湖	4.30~9.74	8.00	3.00~6.81	5.29
南京市内河道	1.89~74.11	13.11	7.67~87.73	28.50
滇池	6.00~52.00	11.94	10.00~18.00	13.64

表 4 训练集与验证集实测 Chl-a 浓度描述性统计

Tab.4 Descriptive statistics of measured Chl-a for training and validation sets

	训练集			验证集		
	最小值~最大值	平均值	样本量	最小值~最大值	平均值	样本量
Chl-a 浓度 /(μg/L)	0.54~75.30	19.49	372	1.00~73.71	19.87	160

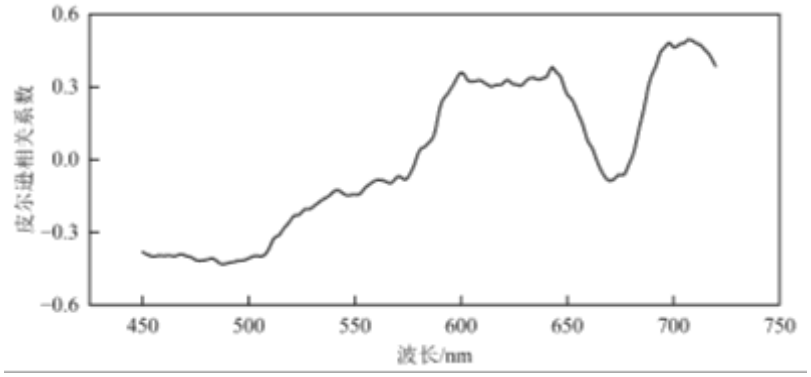

图 3450~720 nm 处的高光谱与实测 Chl-a 浓度的皮尔逊相关系数

Fig.3 Pearson's correlation coefficient between hyperspectra at 450-720 nm and measured Chl-a concentration

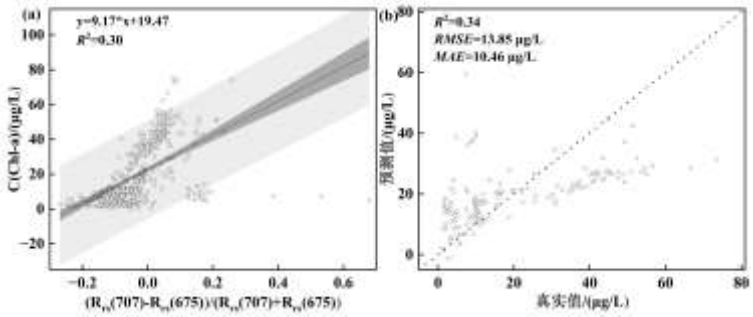

图 4 基于 Chl-a 浓度的归一化反演模型的线性拟合（a，深灰色为 95%的预测区间，浅灰色为 95%的置信区间）和模型验证集精度图（b）

Fig.4 Linear fit of the normalized inversion model based on Chl-a concentration (a, 95% prediction intervals in dark gray, 95% confidence intervals in light gray) and model validation set accuracy plots (b)。

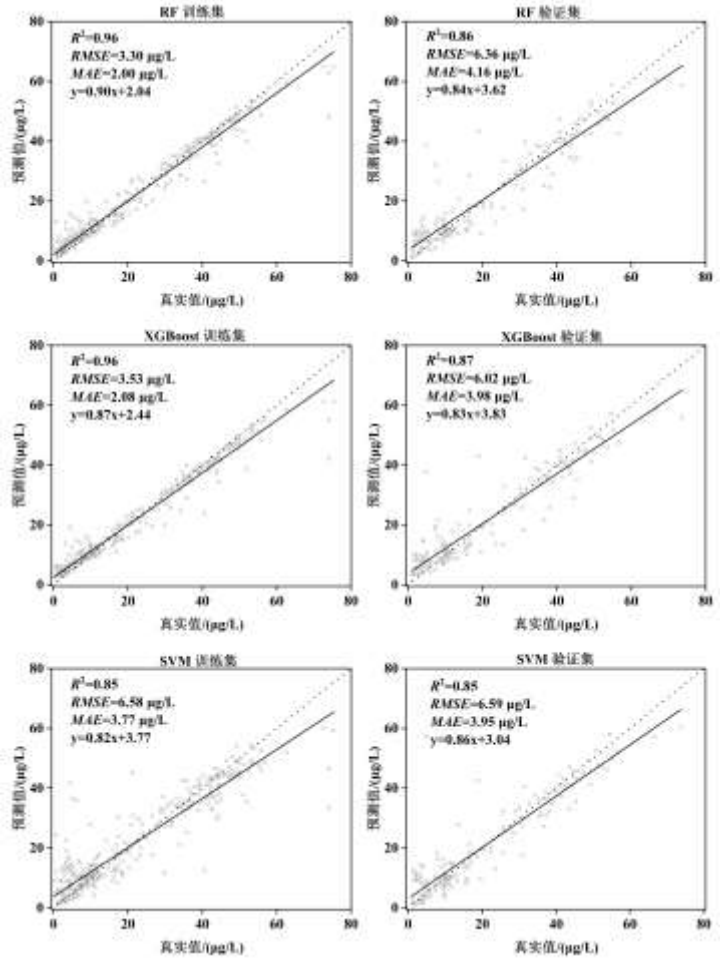

图 5 基于 RF、XGBoost 和 SVM 的 Chl-a 浓度反演模型训练集和验证集真实值与预测值比对图（图中虚线表示理论直线 y=x，实线为模型预测值与真实值的线性拟合直线）

Fig.5Plot of true versus predicted values of Chl-a concentration inversion model based on RF, XGBoost and SVM (The dashed line represents the y=x reference line, while the solid line shows the linear regression fit between predicted and true values)

总体而言，XGBoost 模型在 Chl-a 浓度反演任务中表现最佳。在训练集上，该模型的拟合精度为 R^2=0.96，*RMSE*=3.53 μg/L，*MAE*=2.08 μg/L；在验证集上，R^2=0.87，*RMSE*=6.02 μg/L，*MAE*=3.98 μg/L。相比之下，RF 模型在训练集上的性能与 XGBoost 相近（R^2=0.96，*RMSE*=3.30 μg/L，*MAE*=2.00 μg/L），在验证集上的 R^2=0.86，*RMSE*=6.36 μg/L，*MAE*=4.16 μg/L。SVM 模型的整体表现相对较低，以验证集为例，其 R^2值较 XGBoost 模型低 2.3%，*RMSE* 高出 8.6%。进一步将 XGBoost 模型应用于完整数据集，并对其在不同水体类型中的反演精度进行评估，结果显示各水体的 *RMSE* 分别为：大溪水库（1.14 μg/L）、千岛湖（1.84 μg/L）、沙河水库（2.14 μg/L）、白马湖（3.78 μg/L）、南京市内河道（4.72 μg/L）、太湖流域（4.78 μg/L）、滇池（5.65 μg/L）和洪泽湖（5.83 μg/L），平均 *RMSE* 为 3.73 μg/L。

从各水体对模型总误差的贡献来看，样本量较大且 *RMSE* 较高的水体对整体误差贡献显著。其中，太湖样本量最多，为 330 个，其误差贡献占比高达 72.4%，是模型总误差的主要来源；南京市内河道样本量次之，为 78 个，贡献占比为 16.6%；滇池样本量为 17 个，贡献占比为 5.2%；其余水体如沙河水库、千岛湖、大溪水库、洪泽湖和白马湖的样本量较少，其误差贡献占比均低于 3.0%，对总误差影响较小。这一结果表明，样本量与局部误差共同决定了不同水体对模型总体误差的贡献程度。

总体而言，XGBoost 模型在所有水体中表现出较高的反演精度，具有良好的鲁棒性与泛化能力，适用于基于便携式高光谱近感技术的水体 Chl-a 浓度动态监测。

2.4 滇池 Chl-a 浓度空间分布

基于便携式近感水质监测设备反演的 Chl-a 浓度数据，进一步分析了 2022 年 12 月 26 日至 28 日期间云南滇池 17 个采样点位的监测结果（图 6）。结果表明，滇池的 Chl-a 浓度呈现较大的空间异质性，变化范围在 10.06~38.64 μg/L，全湖平均浓度为 14.44±6.89 μg/L。具体来说，滇池 Chl-a 浓度整体呈现"北高南低"的分布格局。以纬度 24.80°N 为界，将滇池划分为北部湖区和南部湖区，其中北部湖区 9 个点位的 Chl-a 浓度最高值为 38.64 μg/L，最低值为 10.06 μg/L，平均值为 14.91±9.09 μg/L；南部湖区 8 个点位的 Chl-a 浓度最高值为 21.64 μg/L，最低值为 10.44 μg/L，平均值为 13.91±3.64 μg/L。北湖湖区的 Chl-a 浓度最高值为南部湖区的 1.79 倍，而北部湖区的标准差比南部湖区高出了 5.45 μg/L，表明北部湖区 Chl-a 浓度空间异质性更为明显。本研究所采用的便携式高光谱近感监测方法，有效揭示了滇池 Chl-a 浓度的空间分布格局及其区域差异特征。

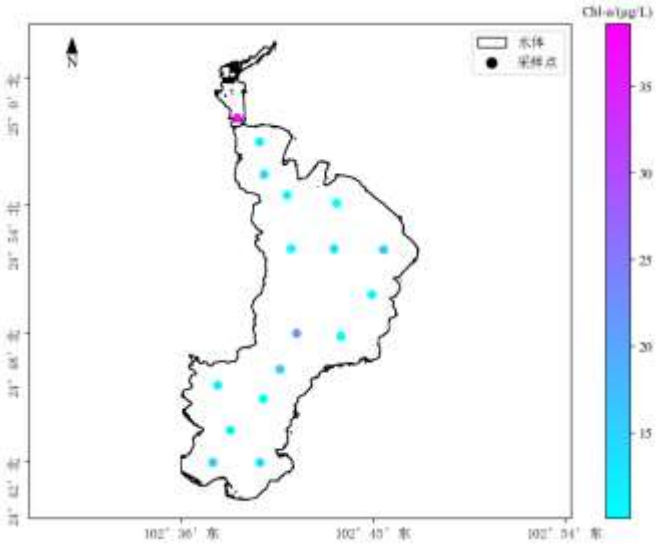

图 6 基于便携式近感水质监测设备 2022 年 12 月 26 日至 28 日滇池 Chl-a 浓度分布图

Fig.6 Spatial distribution map of Chl-a concentration in Lake Dianchi from December 26 to 28, 2022, based on portable proximal water quality monitoring equipment

3 讨论

3.1 模型的优缺点

本研究构建的 XGBoost 模型在多个典型淡水水体中表现出较好的反演精度与鲁棒性，原因有三点。首先，训练数据涵盖了如太湖、千岛湖、南京市内河道等水体，包含了湖泊、水库和河流，代表了一定范围内的水色类型与营养状况。其次，模型所选 450~720 nm 波段覆盖了 Chl-a 浓度的主要光谱响应区域，基于 SHAP 分析的结果进一步揭示了不同光谱波段对模型预测的重要性（图 7）：重要性较高的波段主要分布在蓝光区域（488 nm）、红光吸收区（670~680 nm）及红边区域（690~705 nm），其中，蓝光波段 488 nm 的重要性最高（0.048），其次为 678 nm（0.044）和 673 nm（0.041），这些波段分别位于 Chl-a 的蓝光吸收区和红光吸收区，与 Chl-a 的光学特性高度一致[40,41]。SHAP 分析从特征贡献角度验证了模型所采用波段的物理合理性，说明模型有效捕捉了 Chl-a 浓度的敏感光谱信号。最后，XGBoost 模型通过其独特的正则化机制和集成学习框架，在处理同类问题时在拟合能力、鲁棒性和解释性之间实现了更优平衡[37]，其集成学习结构能够充分挖掘多波段间复杂的非线性关系，有效提升了表层 Chl-a 浓度反演的精度与适应性，表现出良好的综合性能。

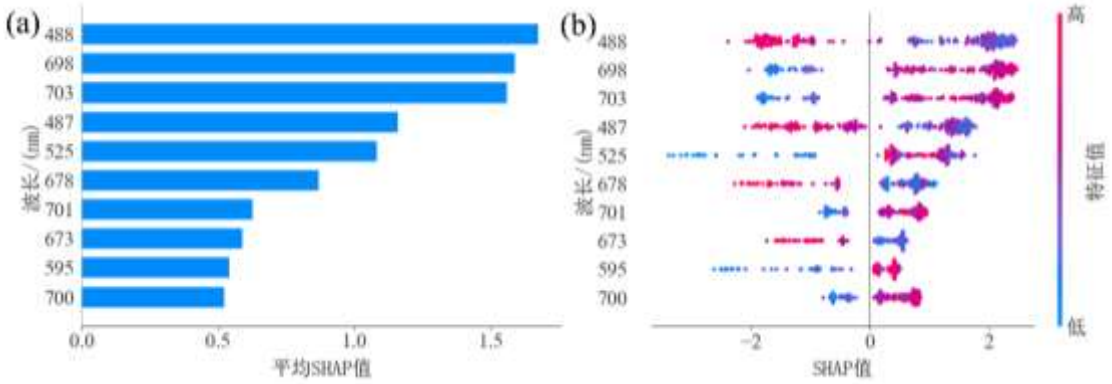

图 7 基于 SHAP 值的前 10 特征重要性排序

Fig.7 SHAP value summary of top 10 features

然而，该模型的适用性仍受限于训练数据所代表的光学条件。当应用对象超出已有水体类型时，其预测精度可能下降。例如，在盐湖等高矿化度水体中，水体电导率与离子成分改变了反射率结构。此外，采集与反演结果也主要反映表层 Chl-a 浓度分布，难以表征垂向分层条件下的 Chl-a 浓度结构。因此，在模型推广至光学背景差异较大的水体前，仍需结合目标水体特性开展适用性评估与必要的本地化校准，以确保反演结果的科学性与可靠性。此外，模型出现了少数低值高估的情况，这主要源于训练数据中低 Chl-a 浓度样本的代表性不足，其中 Chl-a 浓度小于 5 μg/L 的样本仅有 106 条，需补充此类样本，以优化模型在贫营养化水体中的预测性能。未来还需要结合迁移学习等方法优化模型结构，同时针对不同光学背景水体开发自适应校准模型，以提升模型在复杂条件下的适用性。

3.2 便携式高光谱近感设备在水质监测中的优缺点

近年来，高光谱遥感技术，尤其是搭载于无人机平台的高光谱成像系统，在近岸水域和中小型水体的 Chl-a 浓度监测中逐渐受到关注[42,43]。尽管其具备较高的光谱和空间分辨率，能在一定程度上提高局部区域水质参数的反演精度，但该技术在实际应用中仍存在诸多限制。首先，高光谱设备成本高昂、体积和重量较大，增加了无人机平台的载荷压力和系统集成难度，不利于日常监测任务的低成本推广。其次，无人机飞行时间受限，续航能力较弱，难以实现对大面积水体的连续覆盖，限制了其在流域尺度或区域尺度的广泛应用。此外，无人机遥感在多种天气条件下的稳定性较差，受风速、降水和光照变化影响显著，不利于开展长期、连续、可比性强的时序监测。而高光谱卫星虽然具备大范围覆盖能力，但受限于空间分辨率、复访周期和数据获取成本，在内陆水体的高精度动态监测中同样存在实际应用障碍[41]。总体而言，高光谱

遥感技术在广域、水质多变区域的稳定性应用仍有待进一步提升。

此外，气候变化与人类活动的共同影响，使内陆水体藻华的暴发呈现出更加复杂和不可预测的趋势[44]。为了应对这一挑战，自动化监测设备的应用逐渐普及，但在此基础上，人工监测验证的频率需求也随之提高。这种情况下，便携式高光谱近感监测设备的引入，为人工监测提供了一种高效便捷的解决方案，具备诸多优势。首先，采样时具备高光谱分辨率（1 nm）和 450~720 nm 波长范围，从而为内陆水域水质反演提供了更多光谱通道。其次，便携式高光谱近感设备设置简单、便于携带，具备快速采集和高效作业的优点。其效率优势体现在：单点测量时间由传统方法的小时级（4~6 h/点）缩短至分钟级（约 2 min/点），日监测能力大大提升；同时，与传统 ASD FieldSpec 设备相比，其从光谱采集到结果展示无需经过繁复的反射率计算与反演分析等步骤，显著缩短了数据处理时间。便携设备通过内置算法实现结果的即时输出，真正实现了从采样到参数结果的实时一体化完成，为水质管理提供了高效、实用的技术方案，为快速响应与精准决策提供有力支持。设备可与手机 APP 蓝牙连接进行实时水体监测，也可配合云台辅助拍摄，确保测量角度精准。相较于在设备内置无线传输模块，采用手机吸附和蓝牙连接的方式虽然传输速率较低，但省去独立的无线通信模块，降低了硬件成本和设备重量，更适合野外小规模采样便携使用，且直接利用手机算力实现实时数据显示、参数反演和 GPS 定位，且 APP 可远程升级算法，无需更换硬件。第三，卫星遥感数据在获取水体反射率时，常受到大气校正等因素的影响，导致反射率精度下降。而便携式高光谱近感设备采用 1~3 m 离水距离进行近水观测，因此无需复杂的大气校正过程，能有效避免大气成分（如水汽、臭氧、气溶胶等）的干扰，获得更为精准的光谱反射率数据。相比传统光谱测量设备（如 ASD FieldSpec 和 TriOS RAMSES），本研究所采用的便携式设备基于近水测量模式，并通过智能手机 APP 实现“一键式”控制操作，显著降低了使用门槛。传统设备需分别对天空、白板和水体进行多次测量，对光照环境稳定性要求高，操作繁琐且易引入误差。该便携式设备集成测量与控制功能，减少了校准的繁复操作，从源头降低了数据获取过程中的不一致性与人为误差，更适用于高频次、动态连续的监测场景。第四，本研究不同于以往的内陆水质参数估算方法，采样工作涵盖了八个主要内陆水体，且在同一采样区域，光谱采集与水质采样的时间差控制在 2 min 以内，这一操作确保了水质参数与地表水体反射率数据的高度同步，为后续模型训练提供了可靠数据支持，也为开展时间序列上的动态分析与预警奠定了坚实基础。

最后，便携式高光谱近感设备能够将训练好的水质参数估算算法嵌入仪器中，在精度、效率和实用性方面达到了理想平衡，为内陆水体水质监测提供了新的技术思路。

3.3 不足与展望

虽然该设备与配套算法适用于不同类型的水体，已经具有较高的普适性和泛化能力。然而，本研究仍存在一定局限性。高光谱反射率数据易受到环境光照、水面扰动及操作角度等因素影响，导致反演误差。例如，阳光直射、波浪干扰或漂浮物存在可能引起光谱曲线异常，影响建模稳定性[45]；不同气象条件下采集的数据在光谱一致性上也存在不确定性[46]；不同水体类型在水质组成和光学特性方面存在差异[47,48]，可能导致相同波段下的反射率响应存在系统性偏移[49]。因此仍需在未来加强误差源识别与不确定性评估。最后，本研究采用的数据驱动型机器学习算法，其可移植性可能受限于所使用的训练数据集。

未来的研究应进一步扩大样本覆盖范围，采集不同生态特征的湖泊、河流和水库的更多样本，以丰富内陆水体反射率和水质参数数据库。同时，探索更多适用的机器学习和深度学习算法，优化水质参数反演模型，以期获得更为精确的反演结果。

4 结论

本研究基于一种新型的便携式高光谱近感水质监测设备，结合光谱数据和同步水质参数，利用机器学习算法建立了 Chl-a 浓度的反演模型，并与传统线性回归模型进行了比较。结果表明，采用 XGBoost 算法的模型在精度和鲁棒性方面表现最佳，验证集表现为 R^2=0.87，*RMSE*=6.02 μg/L，*MAE*=3.98 μg/L，能够有效估算内陆水体中的 Chl-a 浓度。

便携式高光谱近感设备具备数据精度高、易于携带、操作便捷等优势，能够在现场实现对水体 Chl-a

浓度的快速、连续监测，显著提升了水质监测的时效性。本研究提出的方法有效提高了内陆水体水质监测的响应速度和精度，为水环境管理提供了一种可行、灵活且具有推广价值的技术路径，具有广泛的应用前景。

致谢：感谢牛永康、钱海铭、杨华音、顾锦程提供采样与测样帮助。

5 参考文献

[1] Zhang YL, Qin BQ, Zhu GW, *et al.* Importance and main ecological and environmental problems oflakes in China. *Chinese Science Bulletin*, 2022, **67**(30): 3503-3519. DOI: 10.1360/TB-2022-0178. [张运林, 秦伯强, 朱广伟等. 论湖泊重要性及我国湖泊面临的主要生态环境问题. 科学通报, 2022, **67**(30): 3503-3519.]

[2] Zhang YL, Cai YJ, Peng K, *et al.* Challenges and protection strategies of ecological environmenta of lakes along the Eastern Route of the South to North Water Diversion Project. *Journal of Lake Sciences*, 2024, **36**(5): 1289-1302. DOI: 10.18307/2024.0500. [张运林, 蔡永久, 彭凯等. 南水北调东线沿线湖泊生态环境面临的挑战与保护对策. 湖泊科学, 2024, **36**(5): 1289-1302.]

[3] Qiu GY, Zhang XN. China's urbanization and its ecological environment challengesin the 21s century. *Advances in Earth Science*, 2019, **34**(6): 640-649. DOI: 10.11867/j.issn.1001-8166.2019.06.0640. [邱国玉, 张晓楠. 21 世纪中国的城市化特点及其生态环境挑战. 地球科学进展, 2019, **34**(6): 640-649.]

[4] Liu D, Zhang C, Yan N, *et al.* Eutrophication exacerbated organic pollution in lakes across China during the 1980s–2010s. *Water Research*, 2025, **268**: 122782.

[5] Yang LY, Yang XY, Ren LM, *et al.* Mechanism and control strategy of cyanobacterial bloom in Lake Taihu. *Journal of Lake Sciences*, 2019, **31**(1): 18-27. DOI: 10.18307/2019.0102. [杨柳燕, 杨欣妍, 任丽曼等. 太湖蓝藻水华暴发机制与控制对策. 湖泊科学, 2019, **31**(1): 18-27.]

[6] Zhang M, Duan H, Shi X, *et al.* Contributions of meteorology to the phenology of cyanobacterial blooms: Implications for future climate change. *Water Research*, 2012, **46**(2): 442-452.

[7] Huisman J, Codd GA, Paerl HW, *et al.* Cyanobacterial blooms. *Nature Reviews Microbiology*, 2018, **16**(8): 471-483.

[8] Plessis D, Anja. Persistent degradation: Global water quality challenges and required actions. *One Earth*, 2022, **5**(2): 129-131.

[9] Li YM, Li MT, Ai W, *et al.* Distribution,chlorophyll-a and relationship and significanceof phytoplankton，environment variables in spring season of the Zhelin Reservoir, Jiangxi Province. *Journal of Lake Sciences*, 2017, **29**(3): 625-636. DOI: 10.18307/2017.0311. [李懿淼, 李茂田, 艾威等. 江西柘林水库春季浮游藻类、叶绿素 a 与环境因子的分布、关系及意义. 湖泊科学, 2017, **29**(3): 625-636.]

[10] Jin XC, Zhu X. Tropic characteristics and changes of water bodies of the main lakesand reservoirs in China. *Research of Environmental Sciences*, 1991, **4**(1): 11-20. DOI: 10.13198/j.res.1991.01.13.jinxc.003. [金相灿, 朱萱. 我国主要湖泊和水库水体的营养特征及其变化. 环境科学研究, 1991, **4**(1): 11-20.]

[11] Huang J, Li ZM, Zhang K, Jiang T. Spatio-temporal dynamic analysis of phytoplankton biomassin Chinese lakes based on Google Earth. *Acta Geographica Sinica*, 2021, **76**(7): 1693-1707. DOI: 10.11821/dlxb202107009. [黄珏, 李正茂, 张珂等. 基于 GEE 的中国湖泊浮游植物生物量时空动态分析. 地理学报, 2021, **76**(7): 1693-1707.]

[12] Qi LY, Huang JC, Gao JF, *et al.* Spatial-temporal variation characteristics of chlorophyll-a concentration in Lake Hongze. *Journal of Lake Sciences*, 2016, **28**(3): 583-591. DOI: 10.18307/2016.0314. [齐凌艳, 黄佳聪, 高俊峰等. 洪泽湖叶绿素 a 浓度的时空变化特征. 湖泊科学, 2016, **28**(3): 583-591.]

[13] Shi X, Yu H, Zhao S, *et al.* Impacts of environmental factors on Chlorophyll-a in lakes in cold and arid regions: A 10-year study of Wuliangsuhai Lake, China. *Ecological Indicators*, 2023, **148**: 110133.

[14] Chawla I, Karthikeyan L, Mishra AK. A review of remote sensing applications for water security: Quantity, quality, and extremes. *Journal of Hydrology*, 2020, **585**: 124826.

[15] Sun Y, Wang D, Li L, *et al.* Application of remote sensing technology in water quality monitoring: From traditional approaches to artificial intelligence. *Water Research*, 2024, **267**: 122546.

[16] Sun X, Zhang YL, Shi K, *et al.* Monitoring water quality using proximal remote sensing technology. *Science of the Total Environment*,

2022, **803**: 149805.

[17] Duan HT, Cao ZG, Shen M*, et al.* Review of lake remote sensing research. *National Remote Sensing Bulletin*, 2022, **26**(1): 3-18. DOI: 10.11834/jrs.20221301. [段洪涛, 曹志刚, 沈明等. 湖泊遥感研究进展与展望. 遥感学报, 2022, **26**(1): 3-18.]

[18] Wang SM, Qin BQ. Research progress on remote sensing monitoring of lake water quality parameters. *Environmental Science*, 2023, **44**(3): 1228-1243. DOI: 10.13227/j.hjkx.202203285. [王思梦, 秦伯强. 湖泊水质参数遥感监测研究进展. 环境科学, 2023, **44**(3): 1228-1243.]

[19] The unprecedented 2022 extreme summer heatwaves increased harmful cyanobacteria blooms. *Science of the Total Environment*, 2023, **896**: 165312.

[20] Monitoring water transparency, total suspended matter and the beam attenuation coefficient in inland water using innovative ground-based proximal sensing technology. *Journal of Environmental Management*, 2022, **306**: 114477.

[21] Zhang LF, Zhang LS, Sun XJ*, et al.* Spectral Monitoring Online System for Water Quality Assessment Based on Satellite–Ground Data Integration . *Journal of Global Change Data & Discovery*, 2021, **5**(01): 1-10. DOI: 10.3974/geodp.2021.01.01. [张立福, 张琳姗, 孙雪剑等. 星地协同水质光谱在线监测系统及应用研究. 全球变化数据学报, 2021, **5**(01): 1-10.]

[22] Luo JCY, Qin LJ, Mao P*, et al.* Research Progress in the Retrieval Algorithms for Chlorophyll-a, a Key Element of Water Quality Monitoring by Remote Sensing. *Remote Sensing Technology and Application*, 2021, **36**(03): 473-488. DOI: 10.11873/j.issn.1004-0323.2021.3.0473. [罗婕纯一, 秦龙君, 毛鹏等. 水质遥感监测的关键要素叶绿素 a 的反演算法研究进展. 遥感技术与应用, 2021, **36**(03): 473-488.]

[23] Xu Y, Dong XY, Wang JJ. Use of Remote Multispectral Imaging to Monitor Chlorophyll-a in
Taihu Lake: A Comparison of Four Machine Learning Models. *Journal of Hydroecology*, 2019, **40**(04): 48-57. DOI: 10.15928/j.1674-3075.2019.04.007. [徐逸, 董轩妍, 王俊杰. 4 种机器学习模型反演太湖叶绿素 a 浓度的比较. 水生态学杂志, 2019, **40**(04): 48-57.]

[24] Bao Y, Tian QJ, Chen M, Lv CG. Analysis on Diurnal Variation of Chlorophyll-a Concentration of Taihu Lake Based on Optical Classification with GOCI Data. *Spectroscopy and Spectral Analysis*, 2016, **36**(08): 2562-2567. DOI: 10.3964/j.issn.1000-0593(2016)08-2562-06. [包颖, 田庆久, 陈旻等. 基于 GOCI 影像分类的太湖水体叶绿素 a 浓度日变化分析. 光谱学与光谱分析, 2016, **36**(08): 2562-2567.]

[25] Liang WG, Chen W, Wang JD*, et al.* Inversion of Chlorophyll-a concentration in Luoma Lake using Sentinel-2 satellite imagery. *Bulletin of Surveying and Mapping*, 2024, **0**(10): 1-6. DOI: 10.13474/j.cnki.11-2246.2024.1001. [梁文广, 陈伟, 王金东等. 利用 Sentinel-2 卫星影像反演骆马湖叶绿素 a 浓度. 测绘通报, 2024, **0**(10): 1-6.]

[26] Zhou LN, Pan SM, Wang JW, Vasilakos AV. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 2017, **237**: 350-361.

[27] Wei Chen, Han Z. Study on application of deep learning models to forecast chlorophyll-a concentration in the Estuary of Changjiang River. *Journal of Dalian Fisheries University*, 2025, **40**(02): 330-337. DOI: 10.16535/j.cnki.dlhyxb.2024-168. [陈炜, 韩震. 深度学习模型在长江口叶绿素 a 浓度预测中的应用研究. 大连海洋大学学报, 2025, **40**(02): 330-337.]

[28] Qin BQ, Wang XD, Tang XM*, et al.* Drinking Water Crisis Caused by Eutrophication and Cyanobacterial Bloom in Lake Taihu: Cause and Measurement. *Advances in Earth Science*, 2007, **22**(09): 896-906. DOI: 10.11867/j.issn.1001-8166.2007.09.0896. [秦伯强, 王小冬, 汤祥明等. 太湖富营养化与蓝藻水华引起的饮用水危机——原因与对策. 地球科学进展, 2007, **22**(09): 896-906.]

[29] Yang WB, Duan WX, Cui Y*, et al.* Long-term Changes and Drivers of Ecological Security in Shahe Reservoir, China. *Environmental Science*, 2021, **42**(10): 4739-4752. DOI: 10.13227/j.hjkx.202101125. [杨文斌, 段文秀, 崔扬等. 天目湖沙河水库水生态安全状况长期变化及影响因素. 环境科学, 2021, **42**(10): 4739-4752.]

[30] Ma XX, Gong C, Guo JX*, et al.* Water Pollution Characteristics and Source Apportionment in Rapid Urbanization Region of the Lower Yangtze River: Considering the Qinhuai River Catchment. *Environmental Science*, 2021, **42**(07): 3291-3303. DOI: 10.13227/j.hjkx.202011184. [马小雪, 龚畅, 郭加汛等. 长江下游快速城市化地区水污染特征及源解析:以秦淮河流域为例. 环境科学, 2021, **42**(07): 3291-3303.]

[31] Chen YW, Chen KN, Hu YD. Discussion on possible error for phytoplankton chlorophyll-a concentration analysis usinghot-ethanol

extraction method. *Journal of lake Sciences*, 2006, **18**(5): 550-552. DOI: 10.18307/2006.0519. [陈宇炜, 陈开宁, 胡耀辉. 浮游植物叶绿素 a 测定的“热乙醇法”及其测定误差的探讨. 湖泊科学, 2006, **18**(5): 550-552.]

[32] Hu Y, Zhang Y, Yang B, Zhang Y. Short-term dynamics and driving factors of total suspended matter concentration in Lake Taihu using high frequent geostationary ocean color imager data. *Journal of Lake Sciences*, 2018, **30**(04): 992-1003. DOI: 10.18307/2018.0412. [胡耀躲, 张运林, 杨波, 张毅博. 基于高频次 GOCI 数据的太湖悬浮物浓度短期动态和驱动力分析. 湖泊科学, 2018, **30**(04): 992-1003.]

[33] Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2014, **2**(1): 602-609.

[34] Fang KN, Wu JB, Zhu JP, Xie BC. A review of technologies on Random Forests. *Statistics & Information Forum*, 2011, **26**(3): 32-38. DOI: 10.3969/j.issn.1007-3116.2011.03.006. [方匡南, 吴见彬, 朱建平等. 随机森林方法研究综述. 统计与信息论坛, 2011, **26**(3): 32-38.]

[35] Mcclarren RG: Decision Trees and Random Forests for Regression and Classification, Mcclarren RG, editor, Machine Learning for Engineers: Using data to solve problems for physical systems, Cham: Springer International Publishing, 2021: 55-82.

[36] Zhu HL, Liu HZ, Zhou QM, Cui AH. A XGBoost-Based Downscaling-Calibration Scheme for Extreme Precipitation Events. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**.

[37] Bentéjac C, Csörgo A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 2021, **54**(3): 1937-1967.

[38] Chang CC, Lin CJ. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, **2**(3): 1-27.

[39] O'leary B, Reiners JJ, Xu XH, Lemke LD. Identification and influence of spatio-temporal outliers in urban air quality measurements. *Science of the Total Environment*, 2016, **573**: 55-65.

[40] Remote estimation of chl-a concentration in turbid productive waters — Return to a simple two-band NIR-red model?. *Remote Sensing of Environment*, 2011, **115**(12): 3479-3490.

[41] Li FT, Liang S, Tian LQ*, et al.* Remote Sensing Retrieval of Chlorophyll-a from Hyperspectral Images of ZY-1E Satellite Based on Neural Network:A Case Study of East Lake in Wuhan City. *Geospatial Information*, 2025, **23**(03): 61-65. DOI: 10.3969/j.issn.1672-4623.2025.03.015. [厉芳婷, 梁思, 田礼乔等. 基于神经网络的 ZY-1E 卫星高光谱影像叶绿素 a 遥感反演——以武汉市东湖为例. 地理空间信息, 2025, **23**(03): 61-65.]

[42] Zhao F, Tan K, Pan Z. Water quality monitoring in the demonstration zone of integrated development of the Yangtze River delta using airborne hyperspectral imagery. *Bulletin of surveying and mapping*, 2024, **2024**(10): 7-12. DOI: 10.13474/j.cnki.11-2246.2024.1002. [赵峰, 谭琨, 潘琛. 基于机载高光谱遥感的长三角一体化示范区水质监测. 测绘通报, 2024, **2024**(10): 7-12.]

[43] Feng CJ, Fang CQ, Yuan GY*, et al.* Water pollution monitoring based on unmanned aerial vehicle(UAV)hyperspectral and BP neural network. *Chinese Journal of Environmental Engineering*, 2023, **17**(12): 3996-4006. DOI: 10.12030/j.cjee.202308120. [冯翠杰, 方晨琦, 袁亘宇等. 基于无人机高光谱和 BP 神经网络的城市水体污染监测. 环境工程学报, 2023, **17**(12): 3996-4006.]

[44] Qin BQ, Yang GJ, Ma JR*, et al.* Dynamics of variability and mechanism of harmful cyanobacteriabloom in Lake Taihu, China. *Chinese Science Bulletin*, 2016, **61**(7): 759-770. DOI: 10.1360/N972015-00400. [秦伯强, 杨桂军, 马健荣等. 太湖蓝藻水华“暴发”的动态特征及其机制. 科学通报, 2016, **61**(7): 759-770.]

[45] Werther M, Odermatt D, Simis SGH*, et al.* Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, **190**: 279-300.

[46] Zhong P, Guo X, Jiang X*, et al.* Study on the Effect of Sky Scattered Light on the Reflectance of UAV Hyperspectral about Remote Sensing of Water Color. *Remote Sensing Technology and Application*, 2024, **39**(05): 1128-1140. DOI: 10.11873/j.issn.1004-0323.2024.5.1128. [钟朴, 郭兴健, 蒋听桐等. 面向水色遥感的无人机高光谱天空散射光影响研究. 遥感技术与应用, 2024, **39**(05): 1128-1140.]

[47] Zhu GW, Xu H, Zhu MY*, et al.* Changing characteristics and driving factors of trophic state of lakes in the middle andlower reaches

of Yangtze River in the past 30 years. *Journal of Lake Sciences*, 2019, **31**(6): 1510-1524. DOI: 10.18307/2019.0622. [朱广伟, 许海, 朱梦圆等. 三十年来长江中下游湖泊富营养化状况变迁及其影响因素. 湖泊科学, 2019, **31**(6): 1510-1524.]

[48] Gao F, Chen XL, Yang WF*, et al.* Study on the absorption characteristies of different types of water particlesand CDOM in summer in Taiyuan. *Ecology and Environmental Sciences*, 2021, **30**(7): 1455-1469. DOI: 10.16258/j.cnki.1674-5906.2021.07.015. [高峰, 陈晓玲, 杨文府等. 太原市不同类型夏季水体颗粒物与 CDOM 吸收特性研究. 生态环境学报, 2021, **30**(7): 1455-1469.]

[49] Guo Y, Li H, Huai X*, et al.* Uncertainty Evaluation of Water Color Parameters Inversion in Optically Complex Lakes. *Environmental Science and Technology*, 2023, **46**(08): 15-26. DOI: 10.19672/j.cnki.1003-6504.0453.23.338. [郭羽羽, 李含含, 隗晓琪等. 光学复杂水体中水色参数反演的不确定性评价. 环境科学与技术, 2023, **46**(08): 15-26.]