

DOI: 10.20098/j.cnki.1001-5132.2025.0522

基于无人机高光谱影像与机器学习算法的汤逊湖 水质参数反演研究

郭亚会^{1,2}, 周伟^{1,2}, 李松泽^{1,2}, 马瑞晨^{1,2}, 高敬丹^{1,2}, 张璇³, 郝芳华^{1,2}

(1.华中师范大学 地理过程分析与模拟湖北省重点实验室, 湖北 武汉 430079; 2.华中师范大学
城市与环境科学学院, 湖北 武汉 430079; 3.北京师范大学 水科学研究院, 北京 100875)

摘要: 及时准确地开展城市湖泊水质监测, 对于保护城市生态环境、保障居民健康和促进可持续发展具有重要意义。结合无人机高光谱影像与机器学习算法, 开展武汉市汤逊湖的水质参数反演研究, 分别使用误差反向传播神经网络 (Backpropagation Neural Network, BP)、随机森林 (Random Forest, RF)、极限梯度提升 (eXtreme Gradient Boosting, XGBoost) 和支持向量机 (Support Vector Machine, SVM) 构建叶绿素 a (Chla)、总磷 (TP)、总氮 (TN) 和悬浮物 (SS) 的反演模型。结果表明: (1) Chla、TP、TN 和 SS 反演的最优机器学习算法分别为 BP ($R^2=0.949$, RMSE=9.337 $\mu\text{g/L}$)、RF ($R^2=0.948$, RMSE=0.013 mg/L)、XGBoost ($R^2=0.928$, RMSE=0.095 mg/L)、XGBoost ($R^2=0.875$, RMSE=4.052 mg/L); (2) Chla 最佳反演模型为 BP 神经网络结合高光谱全波段, TP 最优反演模型为 RF 结合高光谱敏感波段, TN 和 SS 的最优反演模型为 XGBoost 结合高光谱敏感波段。研究结果可为城市湖泊水质动态监测和管理提供技术支撑, 有助于推动城市水环境的精细化管理和保护。

关键词: 无人机遥感; 高光谱影像; 水质参数反演; 机器学习算法

中图分类号: X824; TP79

文献标志码: A

文章编号: 1001-5132 (2025) 05-0020-11

城市湖泊的健康对城市水环境安全、生态平衡及可持续发展具有关键意义^[1]。湖泊作为城市生态系统的核心组成部分, 不仅可以为水生生物提供栖息地^[2], 还可以保障城市用水安全, 为城市可持续发展奠定基础。因此, 及时准确地开展湖泊水质监测, 开展水质参数反演工作, 对于科学评估湖泊生态状况、实施精准治理具有重要的现实意义^[3]。

传统水质参数的获取方式主要为人工采样及实验室分析^[4], 此类方式虽能提供高精度的水质参数测量结果, 但同时也存在诸多局限性。人工采样不仅耗时费力、成本高昂, 而且只能反映样本点的局部情况, 难以全面反映湖泊水质的整体状况。此外, 传统水质参数获取方法还存在主观性强、时效性差等问题。相比之下, 遥感技术具有监测范围广、

速度快、成本低的优势, 能够突破时空限制, 实现湖泊水质的长期、快速和动态监测^[5]。此外, 遥感技术凭借其独特的时空覆盖能力和快速监测优势, 通过经验、半经验以及相关分析方法, 利用特定遥感波段构建水质参数反演模型, 能够高效、精准地获取湖泊的水质参数信息^[6]。该技术不仅可以有效弥补传统水质参数监测在时空连续性、监测效率等方面的不足, 还能够揭示水质参数的迁移特征和时空分布变化, 显著提升湖泊水质参数获取的精度与效率。

当前, 卫星遥感已广泛应用于典型水体水质参数的反演工作^[7-8], 其中水质参数主要包括叶绿素 a (Chlorophyll-a, Chla)、总磷 (Total Phosphorus, TP)、总氮 (Total Nitrogen, TN)、悬浮物 (Suspended

收稿日期: 2025-05-23.

宁波大学学报(理工版)网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家社科基金重大项目 (23&ZD105)

第一作者: 郭亚会, 硕士/副教授, 主要研究方向为无人机遥感。E-mail: guoyh@ccnu.edu.cn

Sediment, SS)等。Chla是浮游植物光合作用的主要色素,其浓度可反映水体藻类生物量和富营养化程度^[9];TP涵盖溶解态和颗粒态磷,TN包括有机氮和无机氮(如硝酸盐、铵盐),两者过量均易引发藻类暴发^[10];SS指悬浮在水中的颗粒物,如泥沙、有机碎屑等,其浓度影响着水体透光度和生态健康^[11]。近年来,基于卫星遥感的水质参数反演技术取得了显著进展,为湖泊水质监测提供了新的技术手段。刘曼等^[12]以漓江为研究对象,综合利用Landsat 8 OLI、GF-1、Sentinel-2A多光谱影像以及Sentinel-1A SAR影像等多种遥感数据源,构建针对叶绿素a和悬浮物浓度的水质参数反演模型,研究表明,Chla浓度反演模型的均方根误差(RMSE)范围为0.18~7.88 mg/m³,SS浓度反演模型RMSE范围为0.17~12.55 g/m³。Guo等^[13]利用经过大气校正的Landsat遥感反射率数据以及同步采集的水质测量数据,对加拿大Simcoe湖的总磷和总氮浓度进行了估算,研究表明反演参数TN平均绝对误差(MAE)为35.05%。然而,卫星遥感水质参数反演技术在应用中仍面临诸多挑战,空间分辨率较低和重返周期较长是其主要局限性因素,一定程度上制约了卫星遥感在水质监测中的应用效率,尤其是在需要高时空分辨率数据捕捉水质动态变化的方面^[14]。因此,为更进一步提升水质参数监测的精度和时效性,急需更高时空分辨率的遥感影像作为数据支撑。

近年来,无人机遥感技术凭借其高分辨率、机动灵活、受大气干扰小等优势,为水质监测和水质参数反演提供了有力的数据支持^[15]。无人机搭载的高分辨率传感器可在短时间内进行成像,空间分辨率可达厘米级,甚至更高。此外,无人机遥感技术还具备成本低、操作简便、数据采集和处理速度快等特点,可显著提高水质参数反演精度和效率。这些优点使无人机遥感技术在水质监测领域具备更广阔的应用前景,能够为湖泊水质的动态监测和管理提供更为精准的数据支撑。当前,基于无人机遥感的水质参数反演也取得了一些进展。刘彦君^[16]使用线性回归分析分别构建了研究水域TP、SS、浊度的反演模型,水质参数反演 R^2 值分别为0.74、0.89、0.67。赵琳^[17]使用多光谱数据的敏感波段组合构建了TN、TP等水质参数的一元线性回归模

型,水质反演精度可达75%。

目前,多数研究利用无人机多光谱影像开展了水质监测与分析。然而,多光谱影像光谱分辨率有限,难以满足复杂环境下水质参数反演需求^[18]。相比之下,高光谱影像能提供更全面的光谱信息^[19],为水质参数反演提供更强的数据支撑。此外,多数水质参数反演研究主要依赖于传统回归分析方法,而这些方法在处理复杂遥感影像时,往往难以充分提取其中的有效信息,从而限制了水质参数反演精度的进一步提升。机器学习算法具备自动学习特征和处理高维数据能力,在构建遥感影像与水质参数间关系时,能够更充分地挖掘和提取信息,并能有效利用这些信息以提升模型性能^[20]。由此可见,机器学习算法为水质参数反演提供了有力技术支撑,能够显著提升水质参数反演的准确性和可靠性。

本文以汤逊湖典型湖区的水质参数反演为研究对象,将无人机高光谱影像与机器学习算法相结合,开展水质参数反演研究,主要研究工作包括:

(1)开展高光谱遥感波段与各水质参数之间的敏感性分析;(2)利用4种不同的机器学习算法构建水质参数Chla、TP、TN和SS的反演模型,并针对每种水质参数筛选出最优的机器学习反演算法;(3)利用优选的机器学习模型结合无人机高光谱影像,开展研究区各水质参数的反演并进行空间分布图绘制。

1 材料与方法

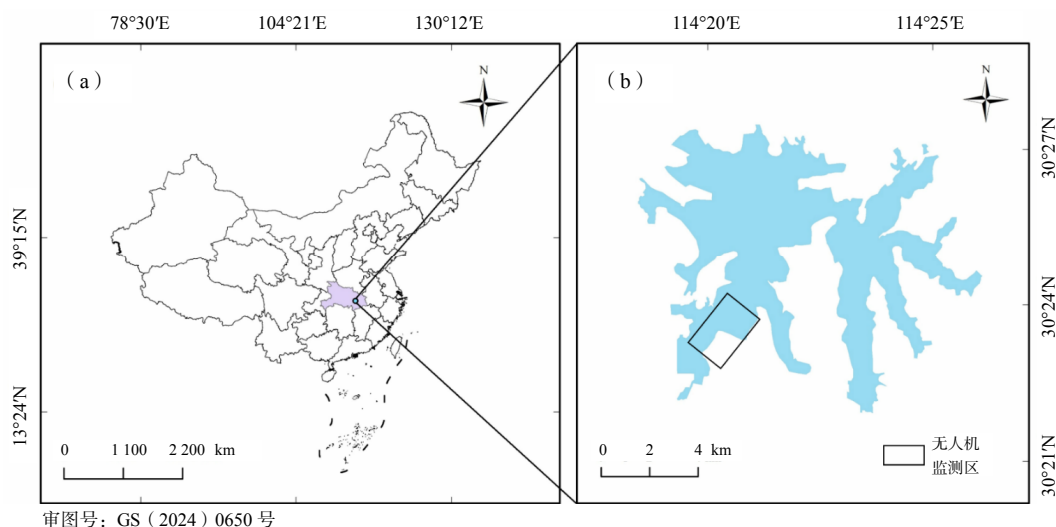
1.1 研究区概况

汤逊湖(东经114°18'~114°26',北纬30°22'~30°28')位于湖北省武汉市,是亚洲最大的城中湖。汤逊湖水体面积达47.6 km²,湖泊平均水深1.85 m,蓄水量1.17亿 m³。汤逊湖所在地气候为亚热带季风气候,降水主要集中在每年的6—8月。本文选取汤逊湖的一处典型区域作为研究区,开展水质参数反演研究,地理位置如图1所示。

1.2 数据采集及方法

1.2.1 水体无人机高光谱影像获取

使用大疆M600 Pro无人机搭载X20P高光谱相机(<https://www.cubert-hyperspectral.com/products/ultris-x20-plus>)分别于2023年6月2日、9月18日和12月27日的上午11:00—12:00采集研究区的



注: 本文的研究区为无人机监测区。

图1 研究区概况

水体高光谱影像。影像采集日均为晴朗天气,光照强度高,成像效果好。无人机飞行高度设为300 m,航向重叠率为40%,旁向重叠率为20%,飞行速度15 m/s。在影像采集前,预先使用相机获取标准参考白板(100%)的实际反射率,以校准影像反射率,确保能准确反映水体的实际光谱。X20P高光谱相机覆盖光谱波长范围为350~1 000 nm,光谱采样间隔为4 nm,可提供连续164个光谱波段。为减少噪声干扰,提高水质参数反演的精度,选取400~900 nm范围内的光谱数据用于水质参数反演研究。

1.2.2 水质参数数据

水样数据同样采集于2023年6月2日、9月18日和12月27日上午11:00—12:00。在研究区内

7个不同位置的采样点进行水体的样本采集,采集每个位置表层0.3~0.5 m深度处的水样,并记录具体地理空间坐标。将采集水样低温保存后送至实验室进行理化检验。水质参数Chla、TP、TN和SS的统计结果如图2所示。通过水质采样点的精确坐标确定经过该位置的无人机高光谱影像,在每个采样点对应的影像位置选取一个3×3像素的窗口区域。通过计算该窗口内各波段光谱反射率的平均值,获取该采样点处的高光谱全波段反射率数据用于水质参数反演建模,并以此构建采样点水质数据和高光谱无人机影像之间的关系。

1.2.3 水质参数的光谱敏感性分析

高光谱影像蕴含丰富的光谱信息,可为开展水质参数反演提供数据支撑。但高光谱影像波段众

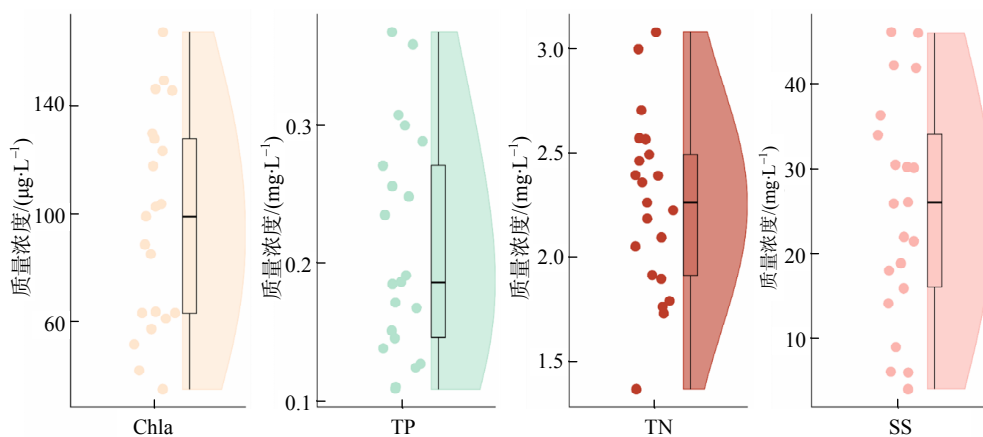


图2 水质参数实测数据

多, 各波段间具有较强相关性, 在进行水质参数反演之前, 需要针对各水质参数开展遥感波段的光谱敏感性分析。依次从高光谱反射率数据中任选两个不同波段, 构建双波段光谱指数 (Two-dimensional Spectral Index, 2D-SI)。常用双波段指数包括改进型归一化水体指数 (Refined Normalized Difference Water Index, RNDWI)、改进型比值水体指数 (Refined Ratio Water Index, RRWI) 和改进型差值水体指数 (Refined Difference Water Index, RDWI) [21], 具体公式如下:

$$R_{\text{RNDWI}} = \frac{R_{\lambda_1} - R_{\lambda_2}}{R_{\lambda_1} + R_{\lambda_2}} \quad (1)$$

$$R_{\text{RRWI}} = \frac{R_{\lambda_1}}{R_{\lambda_2}} \quad (2)$$

$$R_{\text{RDWI}} = R_{\lambda_1} - R_{\lambda_2} \quad (3)$$

式中, R_{λ_1} 和 R_{λ_2} 分别为高光谱反射率数据中任选的两个不同的遥感光谱波段数值。采用线性回归分析方法计算所构建的 2D-SI 光谱指数与目标水质参数之间的相关性, 具体是分别计算 2D-SI 光谱指数与 Chla、TP、TN、SS 之间的决定系数 R^2 , 并基于此值衡量 2D-SI 光谱指数对不同水质参数的反演能力。通过比较不同 2D-SI 光谱指数的 R^2 值, 筛选出与特定水质参数具有较强相关性的光谱指数, 进而构建了一套敏感光谱指数的样本集。

1.3 基于机器学习的水质参数反演模型构建

本文采用 4 种不同的机器学习算法构建水质参数反演模型, 包括误差反向传播神经网络 (Back-propagation Neural Network, BP)、极限梯度提升 (eXtreme Gradient Boosting, XGBoost)、随机森林 (Random Forest, RF) 和支持向量机 (Support Vector Machine, SVM)。BP 神经网络是一种多层前馈神经网络, 其结构由输入层、一个或多个隐藏层以及输出层构成。该算法核心机制是误差反向传播, 通过计算输出误差的梯度, 并借助链式法将误差从输出层逐层反向传播至输入层, 进而逐层调整权重和偏置参数, 最终实现预测误差的最小化 [22]。RF 算法是一种基于决策树的集成学习方法, 通过构建多个决策树, 并将这些决策树的输出结果进行综合, 从而实现对目标变量的分类或回归预测 [23]。XGBoost 算法是一种基于梯度提升决策树的优化算法, 以其高效性和高性能而闻名 [15]。该算法通过

逐步添加新决策树来修正前一阶段模型的残差, 每棵新树的构建目标均为减少前个模型的预测误差。此外, XGBoost 还引入正则化项, 对模型复杂度进行惩罚, 防止模型出现过拟合。SVM 算法是一种经典的二分类模型, 其核心思想是寻找一个能够最大化分类间隔的超平面, 实现对不同类别数据的有效分离 [24]。

本文分别以高光谱全波段反射率和筛选出的敏感光谱指数作为自变量, 以各水质参数 (Chla、TP、TN、SS 的质量浓度) 作为因变量, 利用机器学习算法构建水质参数反演模型。数据集被随机分成训练数据集 (70%) 和测试数据集 (30%) [25], 为了更可靠地进行模型参数的选择, 采用五折交叉验证法进行模型参数调优。在每轮建模时, 训练数据集被划分为 5 个大小相等且分布均匀的子集, 每个子集在 5 轮训练中依次作为独立的验证集, 剩余 4 个子集用作训练集, 对模型超参数进行调优。在模型训练和参数优化完成后, 使用测试集对模型进行检验, 以衡量其预测性能。通过比较不同模型在各水质参数反演中的表现, 评估各机器学习算法对于水质参数反演建模的有效性, 最后针对每个水质参数确定最优的机器学习反演算法。

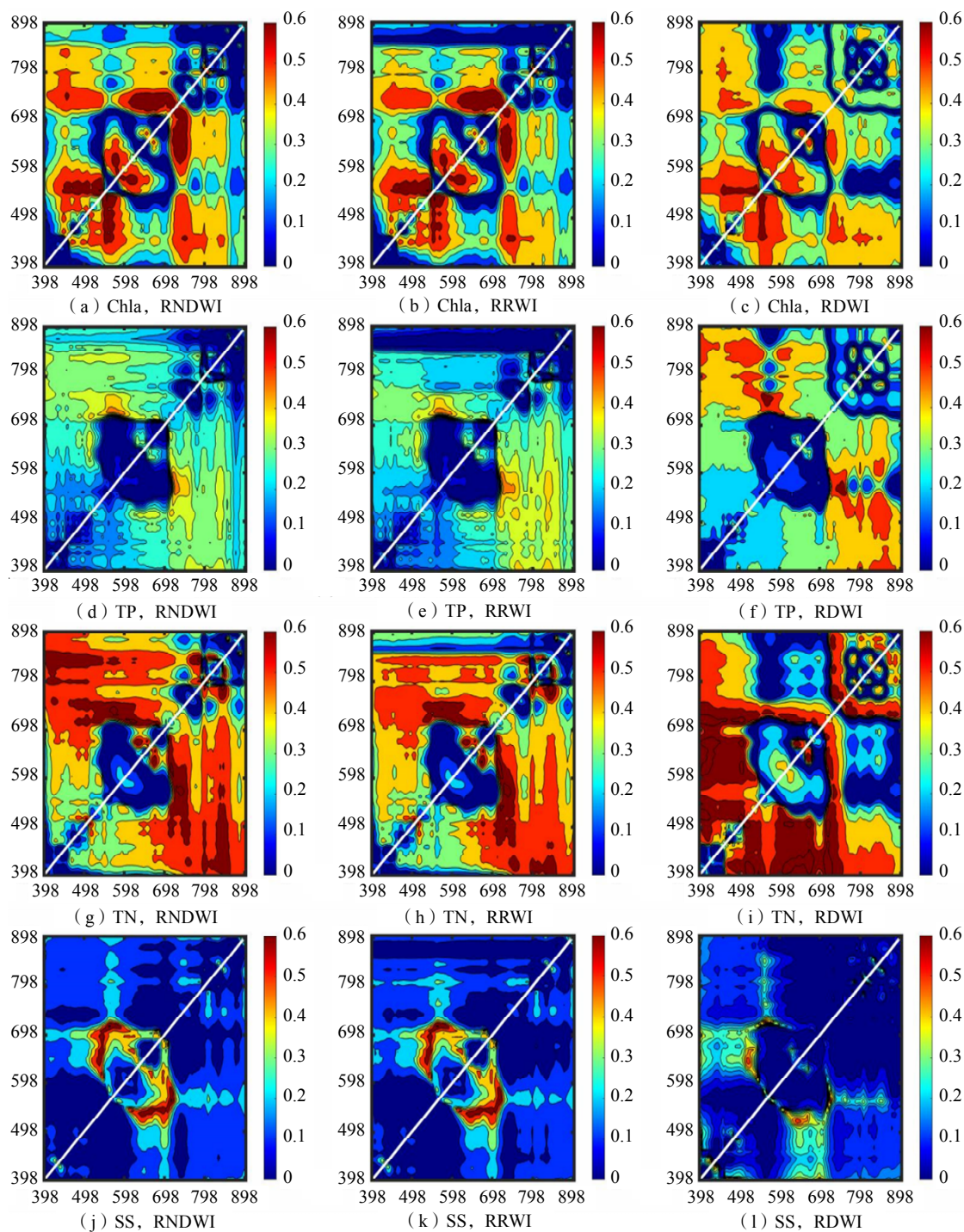
1.4 水质参数空间分布图绘制

针对每个水质参数, 利用研究区域的水体高光谱影像, 结合最优的机器学习算法, 完成整个研究区的 Chla、TP、TN 和 SS 反演工作。利用每幅高光谱影像中心对应的经纬度坐标, 确定其地理位置, 将反演得到的水质参数进行空间展示, 得到各水质参数的空间分布图。再对时间序列上的水质参数进行对比与分析, 明确水质参数的变化趋势并解释其驱动因素。

2 结果与分析

2.1 水质参数的光谱敏感性分析

高光谱全波段影像构建的 2D-SI 与各水质参数间的敏感性分析结果如图 3 所示。由图可见, Chla 与使用 500~600 nm 和 700~800 nm 波段区间构建的 3 种 2D-SI 具有较强相关性; TP 与 RDWI 之间的相关性高于 RNDWI 和 RRWI, 敏感波长主要位于 700~900 nm; TN 与 3 种 2D-SI 相关性均较强, 其中 RNDWI 和 RRWI 的光谱敏感波长位于 700~900



注：纵横坐标均为波长，单位 nm。

图 3 水质参数遥感敏感波段分析结果

nm, RDWI 的光谱敏感波长位于 500~750 nm; SS 与 RNDWI 和 RRWI 间的相关性明显高于 RDWI, 且敏感波长主要位于 600~700 nm。通过决定系数 R^2 大小可以看出, 用于 Chla、TP、TN、SS 水质参数反演建模的敏感波段范围分别为 RNDWI (500~600 nm, 700~800 nm)、RDWI (700~900 nm)、

RNDWI (700~900 nm) 和 RRWI (600~700 nm)。

2.2 基于机器学习的水质参数反演模型构建与检验

2.2.1 基于高光谱全波段数据的水质参数反演模型构建

对所有时相的高光谱全波段数据以及对应的水质参数实测数据, 分别使用 BP、XGBoost、RF

和 SVM 4 种机器学习算法进行水质参数反演建模。通过表 1 可得, 对于水质参数 Chla、TP 和 SS, 基于 BP 神经网络构建的水质参数反演模型优于其他机器学习模型, 取得了最高的反演精度, R^2 分别为 0.949、0.946、0.939, RMSE 分别为 9.337 $\mu\text{g/L}$ 、

0.013 mg/L、2.830 mg/L。对于 TN 反演, XGBoost 构建的水质参数反演模型取得了最高的反演精度, R^2 和 RMSE 分别为 0.923 和 0.098 mg/L。而 RF 和 SVM 在水质参数反演上的表现明显弱于 BP 神经网络和 XGBoost, 详细结果如图 4 所示。

表 1 结合高光谱全波段和机器学习模型的水质参数反演模型精度对比

水质参数	指标	BP	XGBoost	RF	SVM
Chla	R^2	0.949	0.807	0.884	0.836
	RMSE/ $(\mu\text{g}\cdot\text{L}^{-1})$	9.337	18.139	14.030	16.708
TP	R^2	0.946	0.911	0.942	0.826
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	0.013	0.017	0.014	0.024
TN	R^2	0.903	0.923	0.763	0.798
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	0.110	0.098	0.172	0.159
SS	R^2	0.939	0.834	0.829	0.796
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	2.830	4.670	4.737	5.174

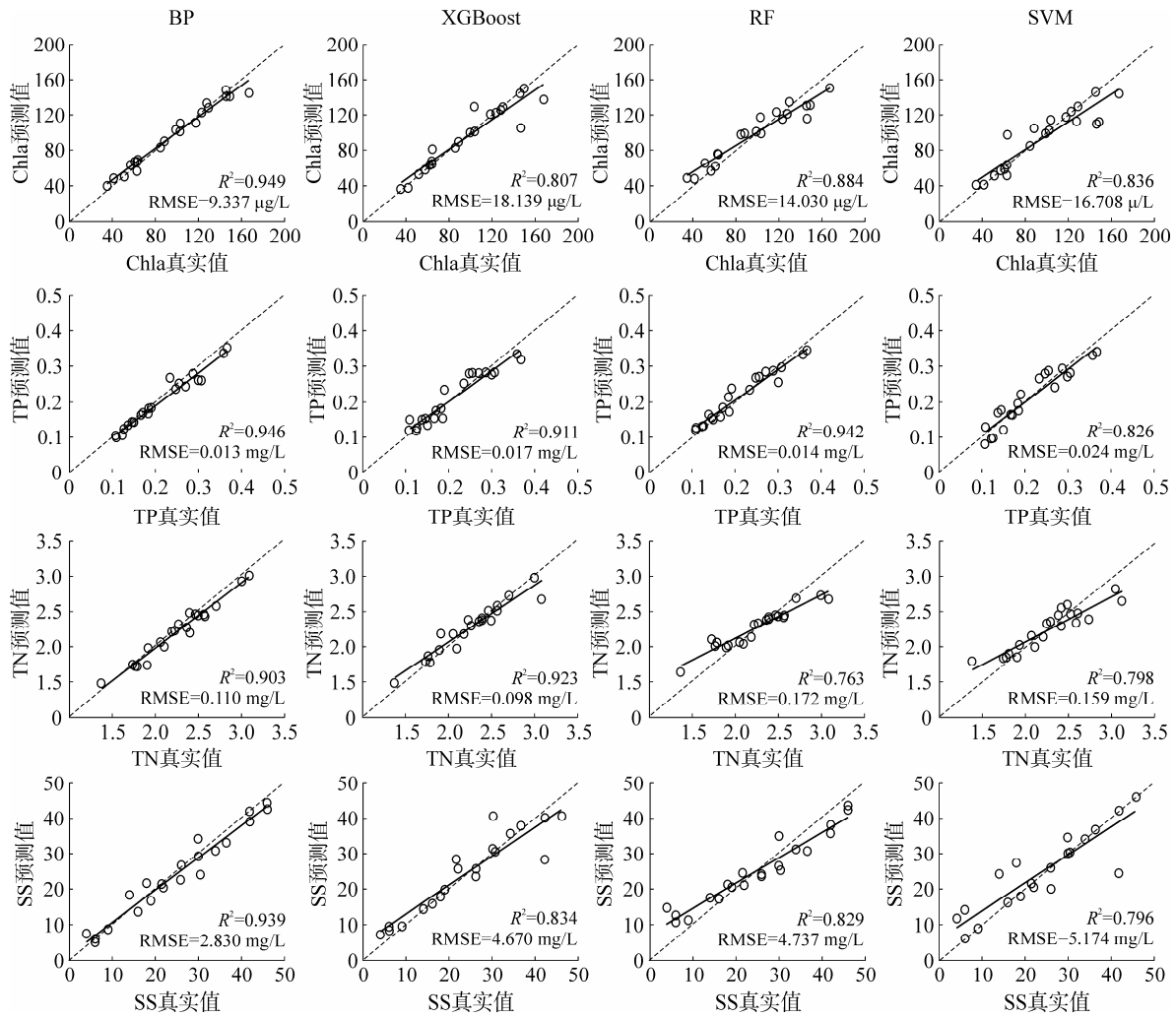


图 4 结合高光谱全波段和机器学习模型的水质参数反演结果

2.2.2 基于高光谱敏感波段的水质参数模型构建

基于高光谱敏感波段计算得到的 2D-SI 分别使用 BP、XGBoost、RF 和 SVM 进行水质参数反演建模。各机器学习模型和水质参数反演精度见表 2。对于 Chla 和 TP 参数反演建模，RF 取得了最高的反演精度， R^2 分别为 0.935 和 0.948，RMSE 分别为 10.551 $\mu\text{g/L}$ 和 0.013 mg/L 。相较于使用高光谱全波段的水质参数反演建模， R^2 分别提高了 0.051 和 0.006，RMSE 分别降低了 3.479 $\mu\text{g/L}$ 和 0.001 mg/L 。对于 TN 和 SS 参数反演建模，拟合精度最高的模型是 XGBoost， R^2 分别为 0.928 和 0.875，RMSE 分别为 0.095 mg/L 和 4.052 mg/L 。与使用高光谱全波段数据建模相比， R^2 分别提高 0.005 和 0.041，RMSE 分别降低 0.003 mg/L 和 0.618 mg/L 。

基于高光谱敏感波段计算得到的 2D-SI，RF 构建的 Chla、TP 水质参数反演模型比 BP 神经网络构建的模型取得的精度更高。对于水质参数 SS，使用 XGBoost 结合高光谱敏感波段的 2D-SI，获得

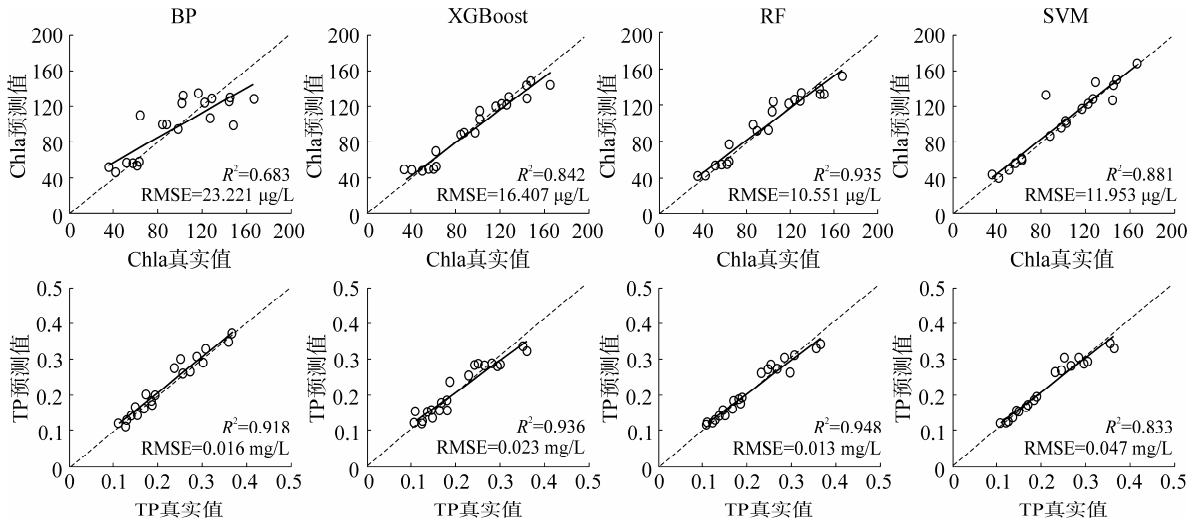
了比 BP 神经网络更高的精度。对于水质参数 TN，高光谱全波段和高光谱敏感波段反演最优的机器学习算法均为 XGBoost，且使用高光谱敏感波段数据反演的水质参数精度更高。

使用高光谱敏感波段数据进行建模，BP 神经网络的精度均有所下降，构建的 Chla、TP 和 SS 反演模型的 R^2 分别为 0.683、0.918、0.504，RMSE 分别为 23.221 $\mu\text{g/L}$ 、0.016 mg/L 、8.068 mg/L 。与使用高光谱全波段数据建模相比， R^2 分别下降了 0.266、0.028、0.435，RMSE 分别增长了 13.884 $\mu\text{g/L}$ 、0.003 mg/L 和 5.238 mg/L ，其中 Chla 和 SS 的反演模型性能降幅最大。因此，基于高光谱敏感波段数据，BP 神经网络不适用于对 Chla、TP 和 SS 进行反演建模。利用高光谱敏感波段数据结合各机器学习算法的水质参数反演结果如图 5 所示。

利用 XGBoost 分别基于高光谱全波段数据和高光谱敏感波段数据进行水质参数反演建模， R^2 均在 0.80 以上，XGBoost 构建的水质参数反演模

表 2 结合高光谱敏感波段数据和机器学习模型的水质参数反演模型精度对比

水质参数	指标	BP	XGBoost	RF	SVM
Chla	R^2	0.683	0.842	0.935	0.881
	RMSE/ $(\mu\text{g}\cdot\text{L}^{-1})$	23.221	16.407	10.551	11.953
TP	R^2	0.918	0.936	0.948	0.833
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	0.016	0.023	0.013	0.047
TN	R^2	0.878	0.928	0.911	0.875
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	0.124	0.095	0.106	0.125
SS	R^2	0.504	0.875	0.851	0.710
	RMSE/ $(\text{mg}\cdot\text{L}^{-1})$	8.068	4.052	4.429	6.168



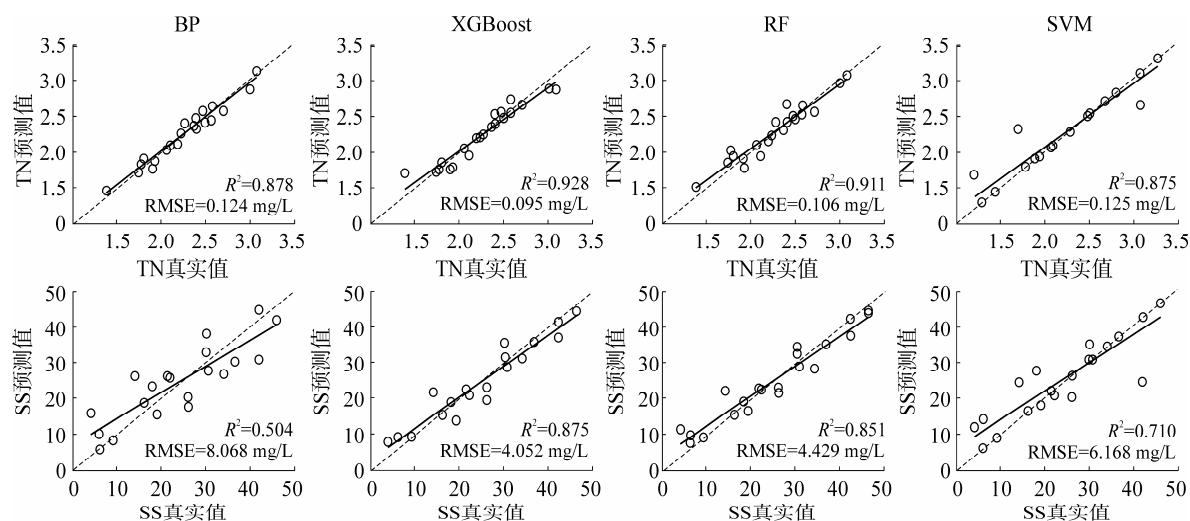


图5 结合高光谱敏感波段和机器学习模型的水质参数反演结果图

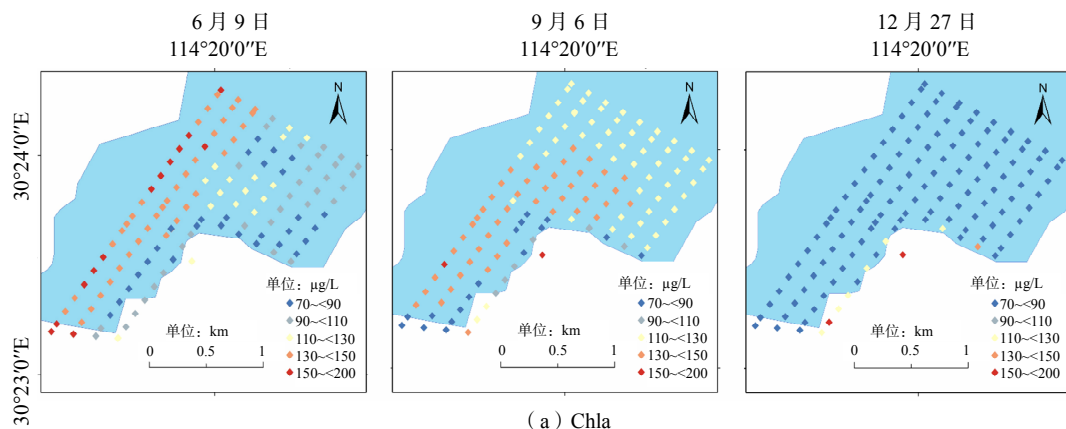
型的性能非常稳定。使用高光谱敏感波段数据建模后,模型性能都有小幅提升。RF和XGBoost水质参数反演建模的表现接近,RF在利用高光谱敏感波段数据建模时表现出了一定优势;但在利用高光谱全波段数据建模时,其稳定性稍低于XGBoost。BP神经网络适合于利用高光谱全波段数据进行水质参数反演建模,而使用高光谱敏感波段数据的表现较差。使用SVM构建的水质参数反演模型 R^2 均低于其他模型,表明SVM拟合遥感数据与水质参数之间的关联能力较差。综上可见,不同水质参数需要使用不同的机器学习算法进行建模。对于不同水质参数的反演,最优的机器学习算法具体选择如下:Chla适合使用BP对全波段高光谱数据进行建模($R^2=0.949$, $RMSE=9.337 \mu\text{g/L}$);TP适合使用RF对高光谱敏感波段数据进行建模($R^2=0.948$, $RMSE=0.013 \text{ mg/L}$);TN、SS适合使用XGBoost对高光谱敏感波段数据进行建模(TN: $R^2=0.928$,

$RMSE=0.095 \text{ mg/L}$; SS: $R^2=0.875$, $RMSE=4.052 \text{ mg/L}$)。

2.3 水质参数空间分布

结合研究区的高光谱全波段影像与各水质参数反演的最优机器学习模型,得到了研究区水质参数的空间分布图(图6)。从时间上来看,从6月9日—9月6日,Chla、TN的质量浓度呈现增大趋势,9月6日—12月27日,则呈现减少趋势,且减少幅度显著。可能是夏季温度逐渐升高,日照增强,水体稳定分层,藻类生长旺盛,致使Chla质量浓度上升。研究区的夏季为雨季,流域径流增强,陆源氮输入的可能性增加,导致TN质量浓度升高;而秋冬季水体温度降低,营养盐被稀释或沉降,Chla质量浓度下降,枯水期水体交换减少,营养盐可能被沉积或消耗,致使TN浓度下降。

TP浓度从6月9日—12月27日呈现下降趋势,可能是在夏季高温和强光照条件下,浮游植物



(a) Chla

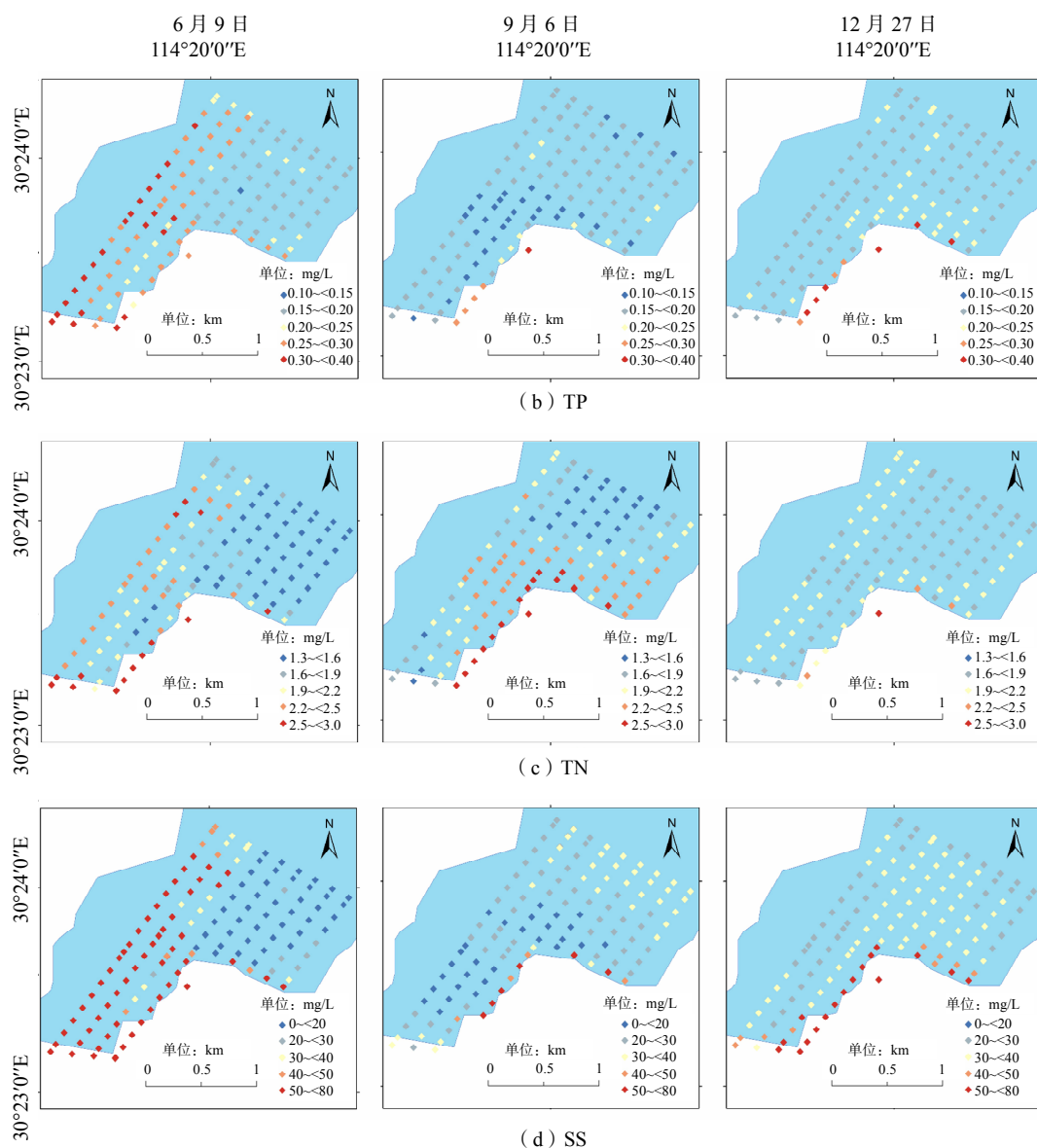


图6 研究区水质参数反演空间分布图

(如藻类)生长旺盛,会大量吸收水体中的磷元素,导致 TP 质量浓度降低。在秋季和冬季,随着水温降低,温跃层消失,造成表层和中层的水体混合,TP 质量浓度逐步降低。在雨季期间,流域地表径流的冲刷作用将大量泥沙和颗粒物输入湖泊,导致 SS 浓度达到峰值。但随着雨季结束,降水减少致使流域输沙量显著降低,SS 浓度随之呈下降趋势。

从空间上来看,东南地区湖岸的水体水质要明显差于其他湖区,可能是离岸较近,受人为干扰较大,造成该区域水质较差。其次,湖泊中心至研究区西南地区的水质明显差于其他湖区,可能是该部分区域向岸延伸,水体相对封闭,流动性差,造成

水中的营养物质不易扩散。

3 结论

本文基于无人机获取的汤逊湖典型区域高光谱影像,结合采样点水体样本得到的实测数据,采用机器学习方法构建了 Chla、TP、TN 和 SS 的反演模型。研究表明,Chla 敏感光谱指数为 RNDWI,对应敏感波段波长为 500~600 nm 和 700~800 nm; TP 敏感光谱指数为 RDWI,对应敏感波段波长为 700~900 nm; TN 敏感光谱指数为 RNDWI,对应敏感波段波长为 700~900 nm; SS 敏感光谱指数为 RRWI,对应敏感波段波长为 600~700 nm。Chla 的

敏感波段与植物色素的主要吸收区高度重叠, TP 和 TN 对近红外波段敏感表明颗粒物和有机物会增强近红外光的散射, SS 的光谱反射则主要集中在红光区域。在最优机器学习模型和高光谱数据的选择方面, Chla 适合使用 BP 神经网络结合高光谱全波段数据进行建模; TP 适合使用 RF 结合高光谱敏感波段数据进行建模; TN、SS 适合使用 XGBoost 结合高光谱敏感波段数据进行建模, 表明了 RF 和 XGBoost 在构建水质参数反演模型中的有效性。针对湖泊不同水质参数 (Chla、TP、TN、SS), 分别探索了其高光谱敏感波段和光谱指数, 并根据各水质参数的特点, 选择最适合的机器学习模型进行水质参数反演。研究结果表明, 本文构建的水质参数反演方法能够充分发挥不同模型的优势, 提高反演结果的准确性和可靠性, 为水质参数反演领域提供了新的思路和技术参考。

参考文献:

- [1] PEKEL J F, COTTAM A, GORELICK N, et al. High-resolution mapping of global surface water and its long-term changes[J]. *Nature*, 2016, 540(7633):418-422.
- [2] HOEKSTRA A Y, BUURMAN J, VAN GINKEL K C H. Urban water security: a review[J]. *Environmental research letters*, 2018, 13(5):053002.
- [3] ZENG F X, SONG C Q, CAO Z G, et al. Monitoring inland water via Sentinel satellite constellation: a review and perspective[J]. *ISPRS journal of photogrammetry and remote sensing*, 2023, 204:340-361.
- [4] PAGE B P, OLMANSON L G, MISHRA D R. A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems[J]. *Remote sensing of environment*, 2019, 231:12.
- [5] PAHLEVAN N, SARKAR S, FRANZ B A, et al. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: demonstrations and validations[J]. *Remote sensing of environment*, 2017, 201:47-56.
- [6] ZHU L D, CUI T W, RUNA A, et al. Robust remote sensing retrieval of key eutrophication indicators in coastal waters based on explainable machine learning[J]. *ISPRS journal of photogrammetry and remote sensing*, 2024, 211:262-280.
- [7] KUHN C, VALERIO A D, WARD N, et al. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity[J]. *Remote sensing of environment*, 2019, 224:104-118.
- [8] XIAO Y, CHEN J H, XU Y, et al. Monitoring of chlorophyll-a and suspended sediment concentrations in optically complex inland rivers using multisource remote sensing measurements[J]. *Ecological indicators*, 2023, 155:111041.
- [9] MAMUN M, KIM J J, ALAM M A, et al. Prediction of algal chlorophyll-a and water clarity in monsoon-region reservoir using machine learning approaches[J]. *Water*, 2020, 12(1):30.
- [10] MCCrackin M L, MULLER-KARULIS B, GUSTAFSSON B G, et al. A century of legacy phosphorus dynamics in a large drainage basin[J]. *Global biogeochemical cycles*, 2018, 32(7):1107-1122.
- [11] WANG X, WEN Z D, LIU G, et al. Remote estimates of total suspended matter in China's main estuaries using Landsat images and a weight random forest model[J]. *ISPRS journal of photogrammetry and remote sensing*, 2022, 183:94-110.
- [12] 刘曼, 付波霖, 何宏昌, 等. 基于多时相主被动遥感的漓江水面监测与水质参数反演(2016—2020 年)[J]. *湖泊科学*, 2021, 33(3):687-706.
- [13] GUO H W, TIAN S, HUANG J J, et al. Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive[J]. *ISPRS journal of photogrammetry and remote sensing*, 2022, 183:451-469.
- [14] ZENG C Q, RICHARDSON M, KING D J. The impacts of environmental variables on water reflectance measured using a lightweight unmanned aerial vehicle (UAV)-based spectrometer system[J]. *ISPRS journal of photogrammetry and remote sensing*, 2017, 130:217-230.
- [15] CHEN B T, MU X, CHEN P, et al. Machine learning-based inversion of water quality parameters in typical reach of the urban river by UAV multispectral data[J]. *Ecological indicators*, 2021, 133:108434.
- [16] 刘彦君. 基于无人机多光谱影像的小微水域水质要素反演[D]. 杭州: 浙江农林大学, 2018.
- [17] 赵琳. 基于多光谱遥感技术的白洋淀水质演变监测方法研究[J]. *现代信息科技*, 2023(3):106-109.
- [18] OLMANSON L G, BREZONIK P L, BAUER M E. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: the Mississippi River and its tributaries in Minnesota[J]. *Remote sensing of environment*, 2013, 130:254-265.
- [19] ABD-ELRAHMAN A, CROXTON M, PANDE-

- CHEETRI R, et al. In situ estimation of water quality parameters in freshwater aquaculture ponds using hyperspectral imaging system[J]. ISPRS journal of photogrammetry and remote sensing, 2011, 66(4):463-472.
- [20] MACIEL D A, BARBOSA C C F, NOVO E, et al. Water clarity in Brazilian water assessed using Sentinel-2 and machine learning methods[J]. ISPRS journal of photogrammetry and remote sensing, 2021, 182:134-152.
- [21] GUO Y H, XIAO Y, HAO F H, et al. Comparison of different machine learning algorithms for predicting maize grain yield using UAV-based hyperspectral images [J]. International journal of applied earth observation and geoinformation, 2023, 124:103528.
- [22] LI J, XIAO Z Q, SUN R, et al. A method to estimate leaf area index from VIIRS surface reflectance using deep transfer learning[J]. ISPRS journal of photogrammetry and remote sensing, 2023, 202:512-527.
- [23] HAFEEZ S, WONG M S, HO H C, et al. Comparison of machine learning algorithms for retrieval of water quality indicators in case-ii waters: a case study of Hong Kong[J]. Remote sensing, 2019, 11(6):617.
- [24] NEGRI R G, DUTRA L V, SANT'ANNA S J S. An innovative support vector machine based method for contextual image classification[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 87:241-248.
- [25] XIAO Y, LI X R, CHEN J H, et al. A novel approach for tracking the diurnal dynamics of chlorophyll-a in lakes using UAV hyperspectral images[J]. Journal of hydrology, 2025, 660:133403.

Water quality parameter inversion in Tangxun Lake using UAV hyperspectral imagery and machine learning algorithms

GUO Yahui^{1,2}, ZHOU Wei^{1,2}, LI Songze^{1,2}, MA Ruichen^{1,2}, GAO Jingdan^{1,2}, ZHANG Xuan³, HAO Fanghua^{1,2}

(1.Hubei Province Key Laboratory for Geographical Process Analysis and Simulation, Central China Normal University, Wuhan 430079, China; 2.College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China; 3.College of Water Sciences, Beijing Normal University, Beijing 100875, China)

Abstract: Timely and accurately monitoring of urban lake water quality is crucial for protecting urban ecosystems, safeguarding public health, and promoting sustainable development. This article takes the water quality parameter inversion of a typical lake area in Tangxun Lake, Wuhan as the research object, and conducts research on water quality parameter inversion based on unmanned aerial vehicle (UAV) hyperspectral images and machine learning algorithms. Error backpropagation (BP) neural network, Random Forest (RF), eXtreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) are used to construct the inversion models for Chlorophyll-a (Chla), Total Phosphorus (TP), Total Nitrogen (TN), and Suspended Sediment (SS), respectively. The main conclusions are as follows: (1) The optimal models for Chla, TP, TN, and SS are BP ($R^2=0.949$, RMSE=9.337 $\mu\text{g/L}$), RF ($R^2=0.948$, RMSE=0.013 mg/L), XGBoost ($R^2=0.928$, RMSE=0.095 mg/L), and XGBoost ($R^2=0.875$, RMSE=4.052 mg/L), respectively; (2) Comparative analysis reveals that BP and RF performed best for Chla and TP inversion, while XGBoost demonstrated superior performance for TN and SS inversion. The research results of this article provide technical support for the dynamic monitoring and management of urban lake water quality, which help to promote the refined management and protection of urban water environment.

Keywords: unmanned aerial vehicle remote sensing; hyperspectral imagery; water quality parameter inversion; machine learning algorithms

(责任编辑 章践立)