

基于特征优化的随机森林模型探究河岸带景观对入湖河流总氮浓度的影响*

李江¹, 王杰^{1,2**}, 崔玉环³, 张耀波⁴, 晏实江¹

(1: 安徽大学资源与环境工程学院, 合肥 230601)

(2: 安徽大学, 湿地生态保护与修复安徽省重点实验室, 合肥 230601)

(3: 安徽农业大学资源与环境学院, 合肥 230036)

(4: 安徽省测绘局, 合肥 230031)

摘要: 河岸带在拦截地表污染物进入河流方面起着重要作用, 探究河岸带景观特征对河流总氮(TN)浓度的影响机制对河流水质调控尤为重要。以受面源污染控制的巢湖入湖河流为研究对象, 根据实测水质数据与同期 Sentinel-2 MSI 影像构建机器学习模型反演河流 TN 浓度, 引入递归特征消除算法优化景观特征指标, 利用随机森林回归模型探究不同宽度河岸带景观对河流 TN 浓度的影响, 确定影响河流 TN 浓度的最有效河岸带宽度与关键景观指标。研究表明: (1) 适用于巢湖入湖河流的 TN 浓度反演模型为梯度提升回归树, 其反演精度 R^2 、均方误差和平均绝对百分误差分别达到 0.93、0.35 mg/L 和 28.86%; (2) 与传统的冗余分析等方法相比, 本文将递归特征消除算法与随机森林回归模型相结合能更为有效捕捉河岸带景观与河流水质间的复杂非线性关系, 其拟合优度 R^2 超过了 0.87; (3) 在干、湿季影响巢湖入湖河流 TN 浓度的最有效河岸带宽度分别为 1500、1000 m, 关键景观指标为农田破碎度、城镇面积比例、景观破碎度以及植被覆盖度。建议在上述有效河岸带宽度范围减少农田破碎度、城镇面积比例、景观破碎度并提高植被覆盖度, 以降低入湖河流及湖体 TN 浓度。本研究可为探究地表景观对河流水质的影响机制与受面源污染控制的河流水污染防治提供有效方法与科学依据。

关键词: 河岸带; 随机森林模型; 特征优化; 景观指标; 总氮; 巢湖

Effects of riparian zone landscape on riverine total nitrogen concentrations using a feature-optimized random forest model*

Li Jiang¹, Wang Jie^{1,2**}, Cui Yuhuan³, Zhang Yaobo⁴ & Yan Shijiang¹

(1: College of Resources and Environmental Engineering, Anhui University, Hefei 230601, P.R. China)

(2: Anhui Province Key Laboratory of Wetland Ecosystem Protection and Restoration, Anhui University, Hefei 230601, P.R. China)

(3: College of Resources and Environment, Anhui Agricultural University, Hefei 230036, P.R. China)

(4: Anhui Provincial Bureau of Surveying and Mapping, Hefei 230031, P.R. China)

Abstract: Riparian zone plays an important role in intercepting surface pollutants from entering rivers, so it is particularly important to explore the influence of riparian landscape on river total nitrogen(TN) concentration for the regulation of river water quality. However, it is difficult to quantitatively analysis the complex relationship between riparian landscape and river water quality to identify the key landscape metrics and optimal riparian strip. Taking the inlet rivers of Lake Chaohu controlled by non-point source pollution as the study area, a remote sensing inversion model was firstly constructed to retrieval TN concentration in rivers using the machine learning regression algorithms according to the measured water quality data and the synchronous Sentinel-2 MSI images,

* 2024-07-18 收稿; 2024-11-04 收修改稿。

国家自然科学基金项目(32171573)、安徽省自然科学基金项目(2308085MD114, 2308085Y29)和安徽省重点研究与开发项目(2022107020027)联合资助。

** 通信作者; E-mail: wangjie@ahu.edu.cn。

then the recursive feature elimination algorithm was introduced to optimize the landscape indices, and a new random forest regression model was lastly constructed to explore the influence of different width riparian landscape on river TN concentration, to determine the most effective riparian zone width and key landscape indices affecting river TN concentration. Results showed that (1) The retrieval model suitable to TN concentration in the inlet rivers of Lake Chaohu was gradient boosted regression model, and its inversion accuracy of R^2 , mean squared error and mean absolute percentage error reached 0.93, 0.35 mg/L and 28.86%, respectively. (2) Compared with the traditional methods such as redundancy analysis (RDA), combining the recursive feature elimination with random forest regression algorithms was more effective method to capture the complex nonlinear relationships between landscape and water quality, with the goodness of fit $R^2 > 0.87$. (3) The most effective widths of riparian zone for influencing river TN concentrations in the dry and wet seasons were 1500 m and 1000 m, respectively, and the key landscape metrics contained the farmland fragmentation, proportion of urban and town, landscape fragmentation and vegetation coverage. It is suggested to reduce the farmland fragmentation, urban construction proportion, landscape fragmentation and improve the vegetation coverage in the effective width of riparian zone, so as to reduce the TN concentration in the river and the lake. Our study can provide an effective method and scientific basis for investigating the influence of surface landscape on river water quality and the prevention of river water pollution controlled by non-point source pollution in the agricultural basin.

Keywords: Riparian zone; random forest model; feature optimization; landscape metrics; total nitrogen; Lake Chaohu

水体富营养化是一个日益严重的全球性问题,也是目前我国内陆水体所面临的主要问题,会对水生态系统造成一系列负面影响,如生物多样性减少、水质恶化、浮游植物大量繁殖^[1-3]。限制河流中的营养盐输入是遏制水体富营养化的有效方法^[4-5],河岸带在拦截地表营养盐进入河流方面起着重要作用。大量研究表明,流域土地利用、人类活动等因素会对受面源污染控制的河流营养盐的产生、迁移和转化产生重大影响^[5-6]。探究河岸带景观格局、植被盖度等景观指标对河流氮磷浓度的影响,确定对河流水质影响最有效的河岸带宽度与关键景观指标,已成为通过景观调控河流非点源污染的有效手段^[7-9]。

已有研究表明,流域景观格局与河流氮磷输出具有强相关性^[10-11]。然而,河流营养盐输出还受到流域的植被、气候、地形等因素的综合影响^[7,12]。因此,景观特征与河流氮磷浓度的关系表现为复杂非线性^[13]。目前研究大多采用冗余分析法(RDA)探究景观特征对河流氮磷浓度的影响作用,通过描述两者线性关系引起的变异在总变异中的比例,定量描述各景观指标对河流氮磷变异的贡献程度^[14-17],并利用多元线性回归寻找景观特征与水质关系的突变点,可为水质的景观调控提供依据^[17-19]。冗余分析法所需水质样点较少,是探究景观特征与河流水质关系的主流方法。然而,受限于冗余分析方法对自变量数据质量的要求,现有研究只能选用少量的景观指标用作输入变量^[7,20-21]。随着人工智能技术的发展,机器学习回归算法可以有效解决诸多变量之间的复杂非线性问题^[19,22],尤其是随机森林模型开发的递归特征消除算法(RFE),可以从大量输入中筛选出最优自变量子集,并通过机器学习方法预测河流水质,在流域景观特征与河流水质的关系分析中存在着优势^[13,18,23]。然而,利用机器学习回归分析河流景观与水质的关系往往受限于河流水质数据缺乏,难以满足大样本训练学习的需求。

现有研究在河流上布设若干采样点,选择样点上游河段探究河岸带景观对河流水质的影响^[16,24]。由于河流实地采样点大多分布于河流汇水口或入湖口,难以定量描述河流内部的水质变化。目前,基于水体自身的光学特性,通过构建水质遥感反演模型建立内陆水体水质参数与遥感反射率之间的关系实现区域水质监测的手段逐渐成熟,遥感技术在水体悬浮物、藻类色素、可溶性有机物等光学活性参数的定量监测方面得到广泛应用。氮、磷营养盐属于非水体光学活性参数,它虽与水体悬浮物、叶绿素等光学活性参数高度相关,但自身无明显光谱特征,因此建模难度较高且反演方法尚未统一。机器学习回归方法可以通过大量样本训练学习水体遥感反射率与各水质参数之间的复杂非线性关系,在水体营养盐浓度遥感反演方面具有很大的技术潜力。近年来,研究者尝试以 Landsat 系列、Sentinel-2 等卫星影像为遥感数据源,利用机器学习回归算法在天津市小型湖泊、北京密云水库以及淮河局部河段开展水体营养盐浓度遥感反演相关的研究^[25-27],取得了理想的成果,但在区域尺度河流水质监测中的应用不足。因而,开展水体营养盐浓度遥感反演可为区域尺度上河流水质的宏观监测提供技术支撑。若将遥感反演的河流水质数据与景观指标数据相结合可以有效获取大量的训练样本,有望满足机器学习分析景观与水质关系的大样本数量需求。

巢湖位于长江下游流域,是我国五大淡水湖之一。近年来,由于人类对该湖泊资源的过度利用和营养盐释放,湖泊富营养化日益严重^[5],其中入湖河流的 TN 负荷输入约占外源输入总量的 76.9%^[28]。基于此,本文以巢湖入湖河流为研究对象,利用河流实测水质数据与同期 Sentinel-2 MSI 影像构建了 TN 浓度遥感反演模型,生成巢湖入湖河流 TN 浓度数据集。而后提取河流不同河段的 TN 浓度与对应河岸带的景观指标,将递归特征消除算法与随机森林回归模型相结合拟合两者之间的关系,探究影响河流 TN 浓度变化的最有效河岸带宽度与关键景观指标,以期受面源污染控制的河流河岸带景观调控与水污染防治提供科学参考。

1 材料与方法

1.1 研究区域

巢湖位于安徽省中部合肥市东南侧,湖面约 750 km²,主要入湖河流包括杭埠河、十五里河、白石天河、派河、南淝河、烔炀河、柘皋河和兆河(图 1)。湖水主要来自于入湖河流与湖面降雨,集水面积约 9130 km²^[29]。巢湖流域年均气温为(16±4)℃,年降雨量在 500~1800 mm 之间。依据我国国家气候中心发布的 1 km 分辨率逐月降水栅格数据集,每年 5—10 月巢湖流域累积降水量约为 1050 mm,划为湿季;每年 11 月—次年 4 月累积降水量约为 350 mm,划为干季。

自 1959 年巢湖大坝建成以来,巢湖水文情势发生巨大变化,且随着当地经济快速发展,工业、农业和生活污水大量排放,导致湖泊水质持续恶化,富营养化问题严重。目前在地方政府的不断治理下,巢湖水质有所改善,但仍为我国污染较严重的淡水湖之一^[30]。

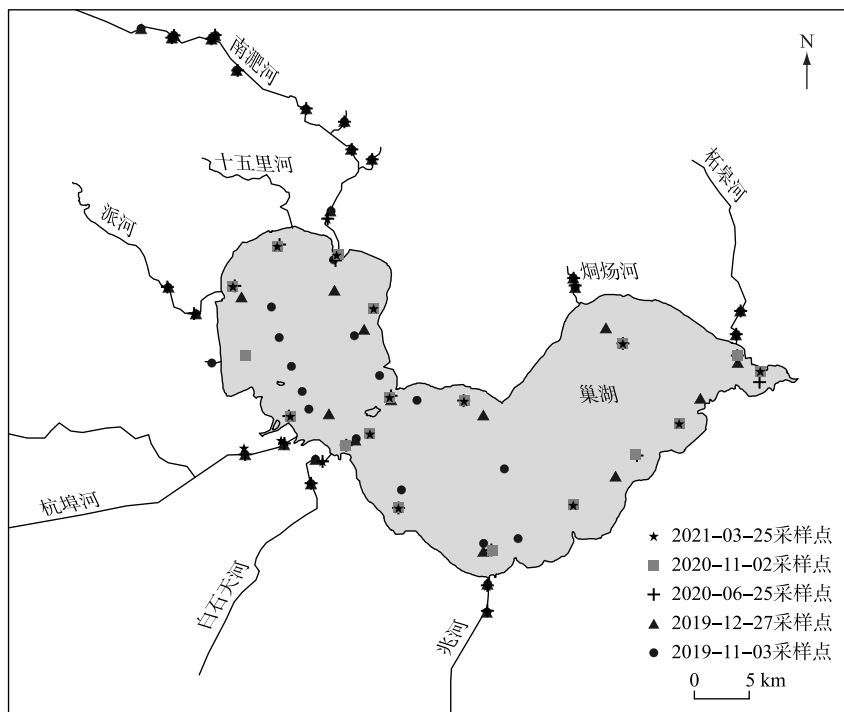


图 1 巢湖主要入湖河流与采样点分布

Fig.1 Distribution of main rivers entered Lake Chaohu and sampling sites

1.2 研究数据

考虑到地表水质的空间与季节性差异,本研究在巢湖主要入湖河流和湖体布设了 34 个采样点,共开展 5 次实地采样,采样日期分别为 2019 年 11 月 3 日、2019 年 12 月 27 日、2020 年 6 月 25 日、2020 年 11 月 2 日

和 2021 年 3 月 25 日,总计采集了 156 个水样(图 1)。而后将水样密封冷藏送往实验室,采用碱性过硫酸钾紫外分光光度法(HJ 636—2012)测定水体总氮(TN)浓度。同时,使用 Trimble GPS 接收机(GEOXH 2008)记录每个采样点的地理位置。

本研究所用 Sentinel-2 A/B 卫星影像来自欧空局网站(<https://dataspace.copernicus.eu/>),与水样采集日期保持同步。ACOLITE(atmospheric correction for OLI lite)是由比利时皇家自然科学研究所为海洋、沿海和内陆水域开发的大气校正算法,该算法可用于 Landsat-8 和 Sentinel-2 影像,已在全球多个地区得到应用^[31]。本研究用 ACOLITE 算法对原始遥感影像进行辐射定标和大气校正后,重采样至 10 m 分辨率,得到水体遥感反射率。考虑到 Sentinel-2 A、B 星在中心波长和光谱响应函数的差异,本文选用 ENVI 软件光谱库中自带的水体反射光谱曲线,将其重采样到 1 nm,分别与 A、B 星光谱响应函数卷积计算得到波段等效遥感反射率,并对相同波段的反射率数值进行散点回归分析,经拟合系数校正,实现了两种传感器遥感反射率间的转换。

本研究所用土地利用数据集为欧空局网站(<https://esa-worldcover.org/>)提供的 2020 年 World Cover 数据集,空间分辨率为 10 m。巢湖流域草地极少,可忽略不计,且不存在冰雪、荒漠、红树林、苔藓地衣等类型。考虑到不同地表类型对流域面源污染的影响程度,本文分别将耕地归为农用地、建筑归为城镇用地、灌木归为林地、水体归为湿地,并在 ArcGIS10.8 软件下将 World Cover 数据集重分类为 4 种景观类型:农用地、林地、湿地、城镇用地。

河岸带植被覆盖率(FVC)利用干、湿季成像的 Sentinel-2 MSI 影像反演得到。首先计算归一化植被指数(NDVI),进而利用像元二分法反演 FVC^[32]。

1.3 研究方法

本文的研究技术思路具体为:首先根据 Sentinel-2 MSI 影像构建水质遥感反演模型获取河流 TN 浓度数据,采用土地利用数据集获得不同宽度河岸带的景观指标,采用递归特征消除(RFE)算法从景观指标中筛选出最优自变量子集。而后利用随机森林回归算法分析河岸带景观指标与对应河段 TN 浓度的关系,探究影响入湖河流 TN 浓度的最有效河岸带宽度与关键性景观指标,并提出面向水污染防治的景观调控措施(附图 1)。

1.3.1 河湖遥感反射率提取 水体指数通常用于遥感影像中各类水体的提取,其中,改进的归一化差异水体指数(MNDWI)能有效减少建筑物等地物阴影对水体识别的干扰。本文采用 MNDWI 对研究区内河流和湖体进行提取(式(1))。考虑到近河岸水体遥感反射率会受到水陆混合像元的影响,本文选择以 0.1 作为阈值,并手动剔除细小水体,最终提取了杭埠河、十五里河、白石天河、派河、南淝河、炯炆河、柘皋河和兆河这 8 条入湖河流和湖体的遥感反射率。由于 Sentinel-2 MSI 影像空间分辨率仅为 10 m,本文所提取河流宽度不超过 3 个像元,即最小宽度为 30 m。

$$MNDWI = \frac{\rho(\text{Green}) - \rho(\text{SWIR})}{\rho(\text{Green}) + \rho(\text{SWIR})} \quad (1)$$

式中, $\rho(\text{Green})$ 、 $\rho(\text{SWIR})$ 分别表示绿光和短波红外波段的遥感反射率,即 Sentinel-2 MSI 影像的第 3 和第 11 波段。

1.3.2 构建河流 TN 浓度反演模型 首先利用 Pearson 相关系数法分析 Sentinel-2 MSI 影像单波段或波段组合与实测 TN 浓度的相关性,筛选出与 TN 浓度高度相关的特征波段(相关系数 r 大于 0.5),共计 12 个(B3、B3+B2、B4+B3、B2、B4+B2、B5+B3、B8/B2、B3×B2、B5+B2、B4、B4×B3、B8/B3)。本文所用特征波段主要集中在蓝、绿、红波段,与已有水体 TN 浓度反演研究存在一定不同^[25,27,33],这可能是由不同研究区水体物质组成的差异引起的。

为避免反演模型过拟合,本文采用四折交叉验证法开展反演模型的构建与精度评价。即将 156 个实测数据集随机排序并平均划分为 4 份,轮流将其中 3 份作为训练集,1 份作为测试集,以 12 个高相关的特征波段作为自变量,实测 TN 浓度作为因变量,分别采用 K 近邻(KNN)、随机森林(RF)、梯度提升树(GBT)、极端随机树(ET)、支持向量机(SVR)和反向传输神经网络(BPNN)等机器学习回归算法构建多种 TN 浓度反演模型。最后以确定系数(R^2)、均方误差(MSE)和平均绝对百分误差(MAPE)作为评价指标,将多次评价结果取平均值,综合训练和测试精度最高的模型作为河流最佳 TN 浓度反演模型。

1.3.3 河段与河岸带单元划分 考虑到河流样点 TN 浓度受到上游河岸带景观的影响^[16,20], 本文以杭埠河、十五里河、白石天河、派河、南淝河、炯扬河、柘皋河和兆河这 8 条入湖河流为研究对象, 从所提取河流的最上游开始顺着水流方向, 初始选择 1 km 作为一个河段, 往下游依次累加 1 km 作为一个新的河段, 直至靠近入湖口(图 2), 共将巢湖入湖河流划分成 180 个河段。而后利用 ArcGIS 10.8 软件对这 180 个河段单元进行缓冲区分析, 间隔宽度为 200~300 m, 缓冲区尺度包括 100、300、500、700、1000、1200、1500 m, 共得到 1260 个河岸带单元。

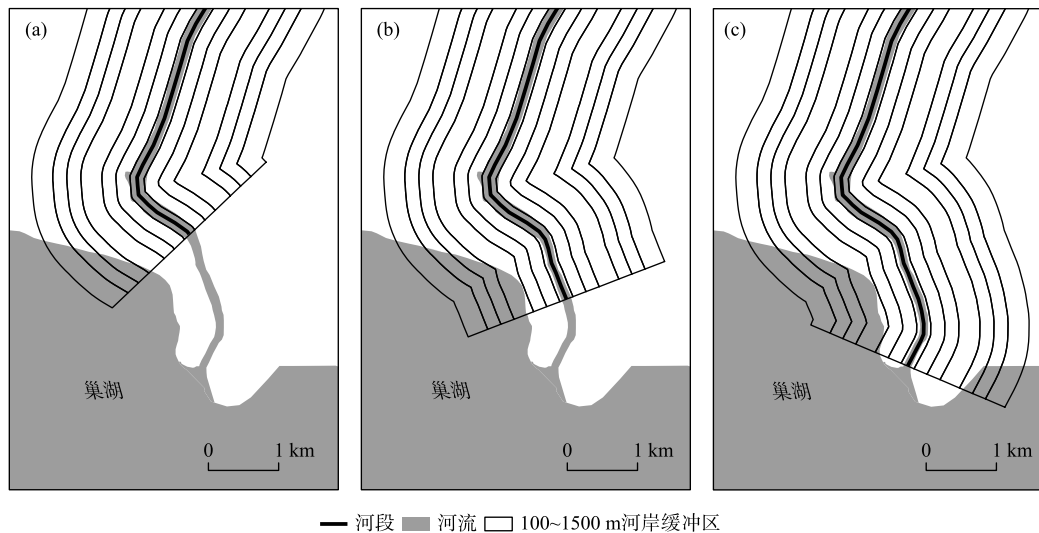


图 2 河段与河岸带单元划分示意: 第 N 个河段(a), 第 $N+1$ 个河段(b), 第 $N+2$ 个河段(c)

Fig.2 River sections and riparian zones division strategies in sections N (a), $N+1$ (b) and $N+2$ (c)

1.3.4 景观指标优选 从反映河岸带景观结构、形状特征、景观异质性及聚合性等方面, 本文参考已有研究^[14,16,18,34-35] 选用了多个影响河流水质的景观格局指标(附表 I)。考虑到地表植被生长对流域面源污染的影响^[36], 在景观格局指标之外还选择了植被覆盖度(FVC)指标。

本文利用 ArcGIS 10.8 软件提取不同河段的 TN 浓度以及对应河岸带单元的景观类型、FVC 等指标后, 利用 Fragstats 4.2 软件计算不同宽度河岸带的景观格局指数。考虑到不同景观指标之间的相关性, 本文采用随机森林模型开发的 RFE 算法进行特征优化, 从附表 I 中众多景观指标中筛选出最优景观指标子集。

1.3.5 随机森林回归与递归特征消除算法 随机森林回归(RF)算法是一种包含多个决策树作为分类器的集成学习算法, 通常用于具有高维特征和大数据集样本的建模, 对高维数据处理能力强且能够输出各特征的重要程度; 递归特征消除(RFE)算法是一种典型的特征选择算法, 其工作原理是通过从训练数据集中对给定的算法模型进行拟合, 按重要性对特征进行排序, 丢弃最不重要的特征, 并重新拟合模型。这个过程不断重复, 直到预设数量的特征被保留下来^[23]。

本文将以 RFE 算法筛选出来的最优景观指标子集作为自变量, 对应河段的 TN 浓度作为因变量, 随机选取 3/4 训练数据和 1/4 测试数据, 利用随机森林回归算法拟合两者间的关系, 并采用四折交叉验证保证最终拟合结果的稳定性, 输出最终拟合精度(拟合优度(R^2))、均方根误差(RMSE)与各自变量的重要性。

2 研究结果

2.1 环巢湖河流 TN 浓度遥感反演

根据 2019—2021 年环巢湖水体实测数据, 结合同期的 Sentinel-2 MSI 影像, 分别利用多种机器学习回归算法(包括 KNN、RF、GBT、ETR、SVR、BPNN)构建水体 TN 浓度反演模型。经过模型精度评价, 最终确定适用于环巢湖河流的最佳 TN 浓度反演模型为 GBT 模型, 对应超参数设置为 Random_state=0, Max_depth=3,

Learning_rate=0.072,其预测精度 R^2 、MSE 和 MAPE 分别达到 0.93、0.35 mg/L 和 28.86% (图 3a)。同时,利用 2020 年 6 月 25 日、2020 年 11 月 2 日实测 TN 浓度数据对同期的影像反演结果进行精度评价,其反演精度 R^2 、MSE 和 MAPE 分别达到 0.77、1.29 mg/L 和 39.75%,0.82、1.05 mg/L 和 35.74% (图 3b),反映了本文 TN 浓度遥感反演结果的可信性。

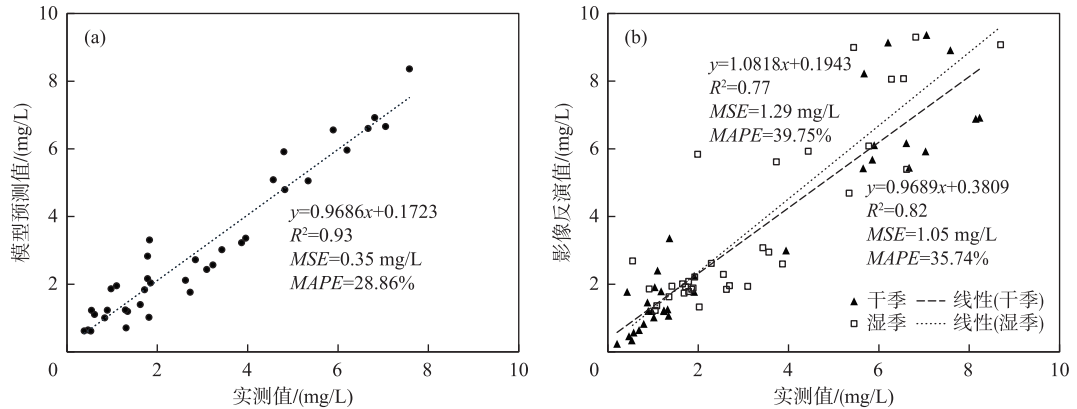


图 3 基于 GBT 模型的环巢湖河流 TN 浓度反演精度评价:(a) 模型预测精度,(b) 影像反演精度

Fig.3 Accuracy evaluation of TN concentration retrieval in rivers around Lake Chaohu based on the GBT model:
(a) Model prediction accuracy, (b) Image retrieval accuracy

而后将基于 GBT 的河流 TN 浓度反演模型应用于 2020 年不同时相 Sentinel-2 MSI 遥感影像,分别对反演结果取均值得到干、湿季节的环巢湖河流 TN 浓度分布图(图 4),分析主要入湖河流 TN 浓度的变化特征。结果表明,入湖河流 TN 浓度的空间分布表现为位于巢湖西北侧的南淝河、十五里河和派河大于其他入湖河流,这被认为是导致巢湖西部比东部富营养化严重的主要原因^[5,30],且从河流上游到入湖口呈增大的趋势。从季节上看,巢湖西北侧的南淝河、十五里河和派河的 TN 浓度季节差异不明显,其他入湖河流在干季均高于湿季。

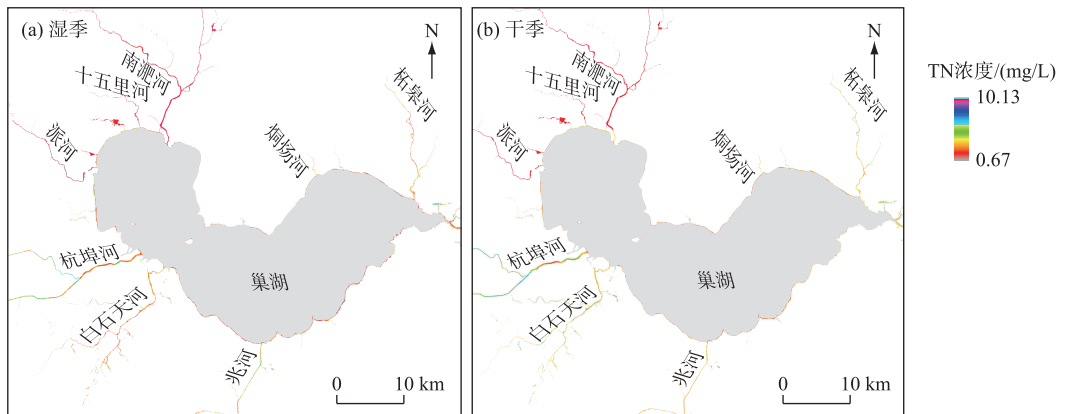


图 4 干、湿季环巢湖河流 TN 浓度空间分布

Fig.4 Spatial distribution of TN concentration in the rivers around Lake Chaohu in dry and wet seasons

2.2 影响河流 TN 浓度的最有效河岸带宽度与关键景观指标

为探究河岸带景观对河流 TN 浓度的影响,本文以不同宽度河岸带下最优景观子集作为自变量,对应河段单元的 TN 浓度作为因变量,在干、湿季分别利用随机森林回归算法拟合两者的关系,对应超参数设置为

$n_estimators=300$, $random_state=0$, $max_depth=3$, 并分析其拟合精度(图 5)。

从图 5 可看出,在湿季,河岸带景观与河流 TN 浓度的拟合精度随河岸带宽度变化明显,从 100 m 到 300 m 拟合精度下降到最低,在 500 m 处上升,至 700 m 处再次下降,到 1000 m 处明显上升,之后趋于稳定。其中,在 1000 m 宽度河岸带两者的拟合精度最高, R^2 为 0.92,因此, $RMSE$ 为 0.68 mg/L,因此,影响该季节河流 TN 浓度的最有效河岸带宽度应为 1000 m。在干季,河岸带景观与 TN 浓度的拟合精度在 100 m 处最低,从 300 m 到 700 m 再次下降,700 m 之后逐渐升高。其中,1500 m 宽度河岸带两者的拟合精度最高,对应 R^2 为 0.88, $RMSE$ 为 0.73 mg/L,因此,影响该季节河流 TN 浓度的最有效河岸带宽度应为 1500 m。

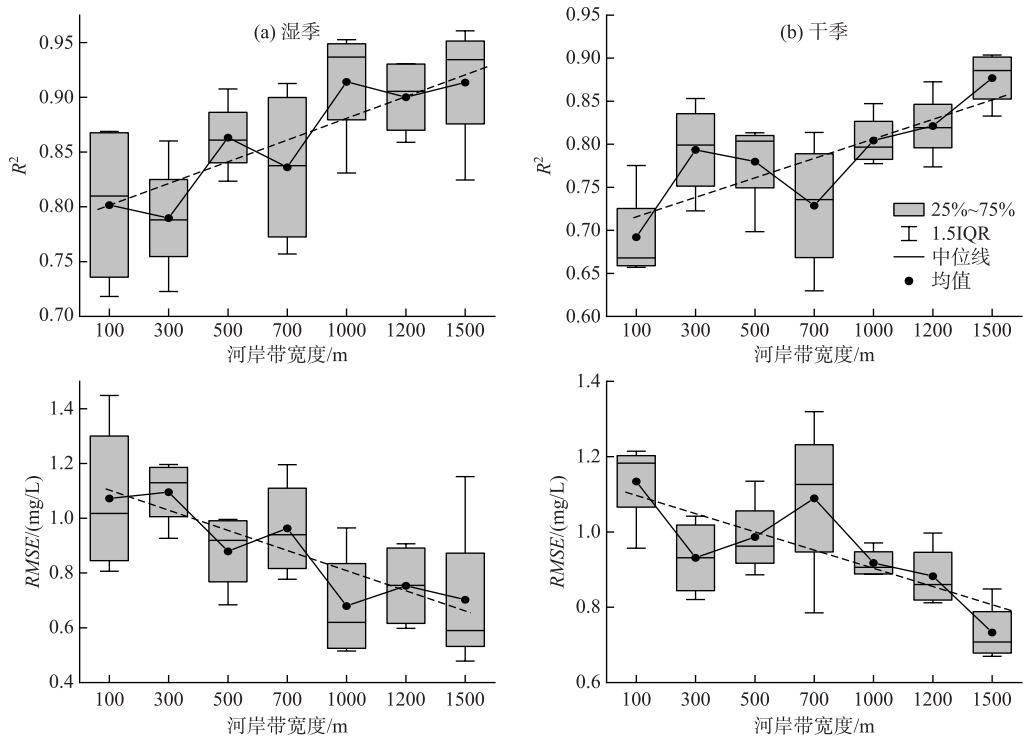


图 5 干、湿季不同宽度河岸带景观与 TN 浓度的拟合精度(R^2 和 $RMSE$)

Fig.5 Fitting accuracy (R^2 and $RMSE$) of landscape and TN concentration in riparian zones of different widths in dry and wet seasons

而后利用随机森林算法中的特征重要性评估模块计算在最有效宽度河岸带下各景观指标对河流水质的重要性系数,进而确定影响河流 TN 浓度的关键景观指标。结果表明,在湿季,1000 m 宽度河岸带的农田破碎度(A-PD)、城镇面积比例(UPLAND)以及景观破碎度(L-PD)的重要性系数超过 0.2,为该季节影响河流 TN 浓度的关键景观指标(图 6a)。干季时,1500 m 河岸带的 A-PD、UPLAND 和植被覆盖度(FVC)的重要性系数超过 0.2,为该季节影响河流 TN 浓度的关键景观指标(图 6b)。

2.3 随机森林回归算法在本研究中的适用性

为衡量所用基于特征优化的随机森林回归算法的适用性,本文利用主流的 RDA 方法分析河岸带景观与水质的关系,将最优景观指标子集作为环境变量,对应河段 TN 浓度作为解释变量,分析河岸带景观对河流 TN 浓度的总解释度随河岸带宽度的变化(图 7)。结果表明,当河岸带宽度在 1000 m 之内时,干、湿季河岸带景观对 TN 浓度的总解释度均呈增加的趋势;超过 1000 m 时,总解释度均超过了 0.7,且趋于稳定。因而,将 1000 m 宽度河岸带确定为影响河流 TN 浓度的最有效河岸带宽度。上述结论与采用随机森林回归算法所得的结论相一致。

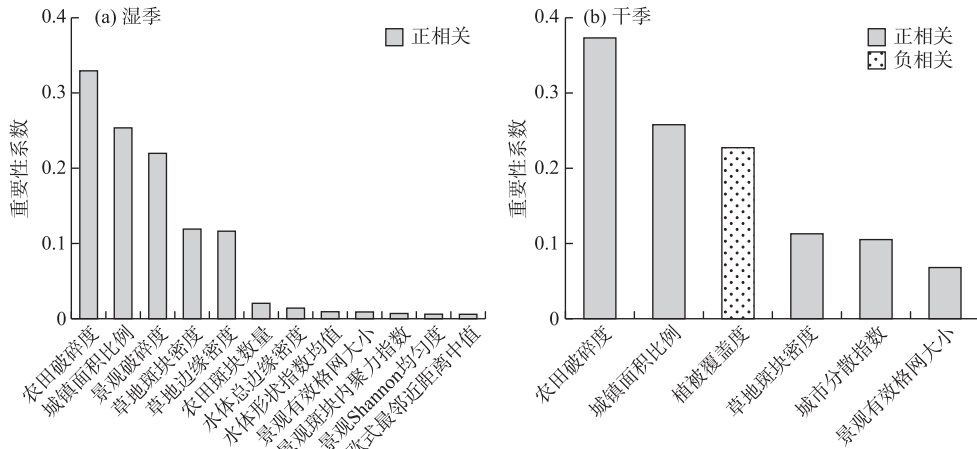


图6 最有效河岸带宽度下景观指标对河流 TN 浓度的重要性

Fig.6 Importance score of landscape metrics influencing TN concentration at the most effective riparian zone width

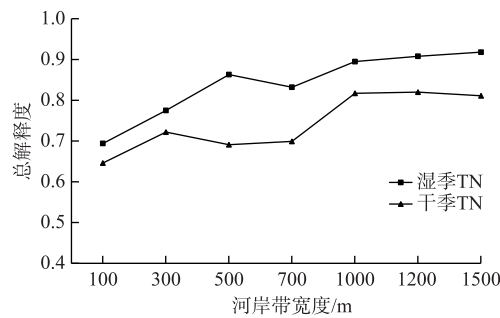


图7 不同宽度河岸带景观对河流 TN 浓度的总解释度

Fig.7 Total explanation of landscape metrics on river TN concentration at different riparian zone widths

在 RDA 得到的最有效河岸带宽度下,本文利用多元回归法拟合河岸带景观与 TN 浓度的关系,以拟合优度 R^2 作为精度评价指标。结果表明,本文所用基于特征优化的随机森林回归方法在干、湿季对景观与水质关系的拟合优度 R^2 (0.877、0.914) 均高于 RDA 后多元回归 (0.816、0.853),反映了该方法在分析景观与河流水质关系的优势。

3 讨论

3.1 机器学习回归在分析景观与水质关系的优势

河流营养盐浓度与地表景观密切相关,还会受地形、土壤、气候等地理环境因素的影响,两者之间关系具有复杂非线性^[13,37-38]。除 RDA 方法外,相关研究尝试通过多元线性、非线性函数等统计回归方法拟合景观与水质的关系^[38-39],然而受限于上述方法对于输入自变量数量的限制,难以同步分析多个景观指标对河流水质的协同影响。

本文通过与常用 RDA 方法的比较,证实了机器学习回归方法在干季、湿季均能有效提高对河岸带景观与入湖河流水质之间关系的拟合能力,为进一步量化地表景观格局对河流水质的影响、开展未来不同发展情景下的河流水质变化预测提供了有效手段。本文所提出基于特征优化的随机森林回归模型能有效处理具有高维特征的景观指标集^[40],在定量分析多景观指标与水质的关系方面具有一定的优势,同时避免了统计回归方法的稳定性易受样本数量的限制。此外,本研究还表明机器学习回归方法能够通过有效拟合非光

学活性水质参数与水体遥感反射率的复杂关系,适用于区域尺度的水体营养盐浓度遥感反演研究,其反演结果可作为实测水质数据的有效补充,为深入探究景观与水质的关系提供了大数量的样本。

3.2 河岸带景观指标对巢湖入湖河流 TN 浓度的影响

在景观指标上,本研究表明河流 TN 浓度与 A-PD 和 U-PLAND 呈正相关,与 FVC 呈负相关,这与已有在巢湖、洱海等农业流域的研究结论一致^[41-43]。在湿季,1000 m 河岸带的 A-PD 和 U-PLAND 是影响河流 TN 浓度变化的关键景观指标,这是由于该季节大气降水较多,高破碎度的农田斑块对化学肥料的保留和吸收能力较弱,导致农田中氮素更易随地表径流进入水体^[43-44];城镇面积比例与 TN 浓度呈正相关,反映了河流 TN 浓度明显受人类活动所产生废水和生活污水的影响^[45]。在干季,1500 m 河岸带的 A-PD 和 U-PLAND 也是影响 TN 浓度变化的关键景观指标。这是因为该季节河流受到降雨带来水文过程的影响很少^[46],冬小麦种植导致农业生产中产生大量的氮素,一部分汇入地表径流,而一部分则渗入地下,A-PD 通过影响地表径流等面源污染进入河流氮素的保留效率,进而对河流水质产生影响。U-PLAND 则反映了人类活动产生的生活污水与废水对河流 TN 浓度的影响。另外,FVC 对 TN 浓度变化的重要性表明了干季河岸带植被冠层对地表径流中污染物的吸附与拦截作用不可忽视。

此外,本研究还发现在湿季河岸带景观对巢湖入湖河流 TN 浓度的影响比干季更为显著,与 Wu 等在浙江省甬江流域的研究结论相一致^[47],其原因在于湿季大气降雨较多,大规模的地表径流快速流动加剧了景观对农田氮素迁移和交换作用的影响,使得氮素更易被地表径流带入水体中^[8]。影响河流 TN 浓度的最有效河岸带宽度在湿季(1000 m)小于干季(1500 m),这主要由于地表径流是流域污染物迁移的驱动力,景观斑块到河流的距离直接影响到面源污染的强度^[10,42]。

3.3 不足与展望

TN 浓度作为水体非光学活性参数,可能与某些光学活性参数(如叶绿素 α 浓度)高度相关^[26-27]。本文在采用机器学习模型反演时未能充分考虑所选特征波段与水体光谱特性的复杂关系,导致反演精度受限。在今后研究中,可通过探究 TN 浓度敏感波段与水体光谱性质的内在物理联系,并考虑不同机器学习算法模型对反演模型所选择波段组合的影响,提高河流 TN 浓度遥感模型的反演精度与普适性。

此外,流域面源污染强度通常与地表景观特征相关。然而,巢湖入湖河流 TN 浓度往往受到农业活动、大气沉降、城镇用水、地下水补给等多种因素的协同影响^[27,45,48]。尤其对于派河、南淝河与十五里河等城市河流,在干季大气降雨较少时,城市工业废水、居民生活用水等点源排放对河流 TN 浓度具有显著的影响。目前本文仅将城镇面积比例作为衡量人类活动强度的指标,难以量化污水厂尾水等点源污染对河流 TN 浓度的影响。考虑到机器学习自身的预测特性,可以通过纳入点源排污模型的方式,量化点源污染排放量作为输入自变量,进一步探究外源污染对河流 TN 浓度的影响。

4 结论

本研究根据实测水质数据与同期 Sentinel-2 MSI 影像构建机器学习模型反演巢湖入湖河流 TN 浓度,利用特征优化的随机森林回归模型探究入湖河流 TN 浓度与不同宽度河岸带景观之间的定量关系,确定影响河流 TN 浓度的最有效河岸带宽度与关键景观指标,得到以下结论:

1) 本文构建基于 GBT 的河流 TN 浓度遥感反演模型,反演精度 R^2 、MSE 和 MAPE 分别达到 0.93、0.35 mg/L 和 28.86%,为定量分析河岸带景观与河流水质的关系提供大量的样本。

2) 在干、湿季节,本文所用基于特征优化的随机森林模型在分析河岸带景观与 TN 浓度关系的拟合精度均高于 RDA 等统计分析方法,其拟合优度 R^2 在 0.87~0.92 之间,在分析地表景观与河流水质的关系上反映出更好的性能。

3) 湿季影响巢湖入湖河流 TN 浓度的最有效河岸带宽度为 1000 m,关键景观指标为农田破碎度、城镇面积比例以及景观破碎度;在干季影响河流 TN 浓度的最有效河岸带宽度为 1500 m,关键景观指标为农田破碎度、城镇面积比例以及植被覆盖度。建议在上述有效河岸带宽度范围内降低农田破碎度、城镇面积比例、景观破碎度并提高植被覆盖度,以降低入湖河流及湖体营养盐浓度。

5 附录

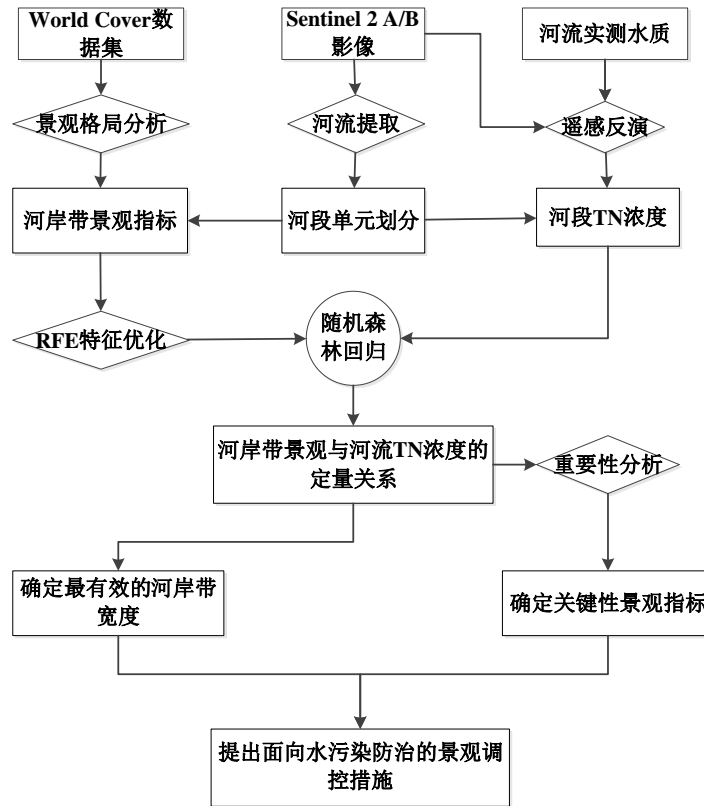
附图 I 和附表 I 见电子版(DOI:10.18307/2025.0423)。

6 参考文献

- [1] Qu X, Chen YS, Liu H *et al.* A holistic assessment of water quality condition and spatiotemporal patterns in impounded lakes along the eastern route of China's South-to-North water diversion project. *Water Research*, 2020, **185**: 116275. DOI: 10.1016/j.watres.2020.116275.
- [2] Schindler DW, Carpenter SR, Chapra SC *et al.* Reducing phosphorus to curb lake eutrophication is a success. *Environmental Science & Technology*, 2016, **50**(17): 8923-8929. DOI: 10.1021/acs.est.6b02204.
- [3] Sinha E, Michalak AM, Balaji V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, 2017, **357**(6349): 405-408. DOI: 10.1126/science.aan2409.
- [4] Kong FL, Xi M, Xu LH *et al.* Review of studies on the limitation of nutrients in the eutrophic water. *Journal of Earth Environment*, 2016, **7**(2): 121-129. DOI: 10.7515/JEE201602002. [孔范龙, 郗敏, 徐丽华等. 富营养化水体的营养盐限制性研究综述. 地球环境学报, 2016, **7**(2): 121-129.]
- [5] Yang X, Cui HB, Liu XS *et al.* Water pollution characteristics and analysis of Chaohu Lake basin by using different assessment methods. *Environmental Science and Pollution Research*, 2020, **27**(15): 18168-18181. DOI: 10.1007/s11356-020-08189-2.
- [6] Zhang ZM, Du JL, Chen DC *et al.* Effects of land use and landscape pattern characteristics on seasonal surface water quality in a typical reticulated river network area—A case study of Liyang City, Jiangsu Province. *J Lake Sci*, 2022, **34**(5): 1524-1540. DOI: 10.18307/2022.0509. [张志敏, 杜景龙, 陈德超等. 典型网状河网区域土地利用和景观格局对地表季节水质的影响——以江苏省溧阳市为例. 湖泊科学, 2022, **34**(5): 1524-1540.]
- [7] Liu HY, Meng C, Wang Y *et al.* From landscape perspective to determine joint effect of land use, soil, and topography on seasonal stream water quality in subtropical agricultural catchments. *Science of the Total Environment*, 2021, **783**: 147047. DOI: 10.1016/j.scitotenv.2021.147047.
- [8] de Mello K, Valente RA, Randhir TO *et al.* Effects of land use and land cover on water quality of low-order streams in Southeastern Brazil: Watershed versus riparian zone. *CATENA*, 2018, **167**: 130-138. DOI: 10.1016/j.catena.2018.04.027.
- [9] Zhang Q, Liu R, Zhang J *et al.* Effects of land use on river water quality at multiple spatial and temporal scales in the Three Gorges Reservoir area under extreme weather conditions. *J Lake Sci*, 2024, **36**(4): 1096-1114. DOI: 10.18307/2024.0423. [张琼, 刘睿, 张静等. 极端天气下三峡库区土地利用对河流水质的多时空尺度影响. 湖泊科学, 2024, **36**(4): 1096-1114.]
- [10] Lv LT, Gao XQ, Liu Q *et al.* Influence of landscape pattern on nitrogen and phosphorus output in the Dongjiang River Basin. *Acta Ecologica Sinica*, 2021, **41**(5): 1758-1765. [吕乐婷, 高晓琴, 刘琦等. 东江流域景观格局对氮、磷输出的影响. 生态学报, 2021, **41**(5): 1758-1765.]
- [11] Jia ZL, Chang X, Duan TT *et al.* Water quality responses to rainfall and surrounding land uses in urban lakes. *Journal of Environmental Management*, 2021, **298**: 113514. DOI: 10.1016/j.jenvman.2021.113514.
- [12] Kang GL, Qiu Y, Wang QX *et al.* Exploration of the critical factors influencing the water quality in two contrasting climatic regions. *Environmental Science and Pollution Research*, 2020, **27**(11): 12601-12612. DOI: 10.1007/s11356-020-07786-5.
- [13] Zhang ZY, Huang JL, Duan SW *et al.* Use of interpretable machine learning to identify the factors influencing the nonlinear linkage between land use and river water quality in the Chesapeake Bay watershed. *Ecological Indicators*, 2022, **140**: 108977. DOI: 10.1016/j.ecolind.2022.108977.
- [14] Chen JB, Lu J. Effects of land use, topography and socio-economic factors on river water quality in a mountainous watershed with intensive agricultural production in east China. *PLoS One*, 2014, **9**(8): e102714. DOI: 10.1371/journal.pone.0102714.
- [15] Chen XY, Pi YY, Zhang RQ *et al.* Influence of landscape factors on water quality of Ganjiang River. *Environmental Science & Technology*, 2024, **47**(S1): 207-213. [陈晓艺, 皮玉宇, 张睿麒等. 景观因子对赣江水体水质的影响分析. 环境科学与技术, 2024, **47**(S1): 207-213.]
- [16] Li K, Chi GQ, Wang L *et al.* Identifying the critical riparian buffer zone with the strongest linkage between landscape characteristics and surface water quality. *Ecological Indicators*, 2018, **93**: 741-752. DOI: 10.1016/j.ecolind.2018.05.030.
- [17] Zhang J, Li SY, Dong RZ *et al.* Influences of land use metrics at multi-spatial scales on seasonal water quality: A case study of river systems in the Three Gorges Reservoir Area, China. *Journal of Cleaner Production*, 2019, **206**: 76-85. DOI: 10.1016/j.jclepro.2018.09.179.
- [18] Wang RZ, Kim JH, Li MH. Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. *Science of the Total Environment*, 2021, **761**: 144057. DOI: 10.1016/j.scitotenv.2020.144057.
- [19] Xiong JF, Lin C, Cao ZG *et al.* Development of remote sensing algorithm for total phosphorus concentration in eutrophic lakes: Conventional or machine learning? *Water Research*, 2022, **215**: 118213. DOI: 10.1016/j.watres.2022.118213.

- [20] Xu QY, Wang P, Shu W *et al.* Impacts of landscape structure on water quality under different spatial scales in the Yuan River Basin. *Acta Scientiae Circumstantiae*, 2020, **40**(12): 4325-4337. [徐启渝, 王鹏, 舒旺等. 不同空间划分方式下袁河流域景观结构对水质的影响. 环境科学学报, 2020, **40**(12): 4325-4337.]
- [21] Fan YS, Yu WQ, Zhang J *et al.* Response of water quality to landscape pattern change in the water source area of upper reaches of Lake Taihu: A case study in the upper reaches of Dongtiaoxi River. *J Lake Sci*, 2021, **33**(5): 1478-1489. DOI: 10.18307/2021.0516. [范雅双, 于婉晴, 张婧等. 太湖上游水源区河流水质对景观格局变化的响应关系——以东苕溪上游为例. 湖泊科学, 2021, **33**(5): 1478-1489.]
- [22] Yuan QQ, Shen HF, Li TW *et al.* Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 2020, **241**: 111716. DOI: 10.1016/j.rse.2020.111716.
- [23] Liu ZR, Shurin GV, Bian L *et al.* A carbon nanotube sensor array for the label-free discrimination of live and dead cells with machine learning. *Analytical Chemistry*, 2022, **94**(8): 3565-3573. DOI: 10.1021/acs.analchem.1c04661.
- [24] Xu MZ, Xu GC, Qiao HL *et al.* Influence of different spatial scale landscape patterns on water quality in a small watershed at the southern foot of the Qinling Mountains. *Acta Scientiae Circumstantiae*, 2023, **43**(10): 396-406. DOI: 10.13671/j.hjkxxb.2023.0147. [徐明珠, 徐国策, 乔海亮等. 秦岭南麓小流域不同空间尺度景观格局对水质的影响分析. 环境科学学报, 2023, **43**(10): 396-406.]
- [25] Guo HW, Huang JJ, Chen BW *et al.* A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. *International Journal of Remote Sensing*, 2021, **42**(5): 1841-1866. DOI: 10.1080/01431161.2020.1846222.
- [26] Jiang Q, Xu L, Sun S *et al.* Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms—A case study in the Miyun Reservoir, China. *Ecological Indicators*, 2021, **124**: 107356.
- [27] Huangfu K, Li J, Zhang XJ *et al.* Remote estimation of water quality parameters of medium- and small-sized inland rivers using sentinel-2 imagery. *Water*, 2020, **12**(11): 3124. DOI: 10.3390/w12113124.
- [28] Wang SH, Jiang X, Jin XC. Classification and pollution characteristic analysis for inflow rivers of Chaohu Lake. *Environmental Science*, 2011, **32**(10): 2834-2839. DOI: 10.13227/j.hjlx.2011.10.024. [王书航, 姜霞, 金相灿. 巢湖入湖河流分类及污染特征分析. 环境科学, 2011, **32**(10): 2834-2839.]
- [29] Xi SS, Zhou CC, Liu GJ *et al.* Spatial and temporal distributions of nitrogen and phosphate in the Chaohu Lake. *Environmental Science*, 2016, **37**(2): 542-547. DOI: 10.13227/j.hjlx.2016.02.019. [奚姗姗, 周春财, 刘桂建等. 巢湖水体氮磷营养盐时空分布特征. 环境科学, 2016, **37**(2): 542-547.]
- [30] Zhang M, Shi XL, Yang Z *et al.* The variation of water quality from 2012 to 2018 in Lake Chaohu and the mitigating strategy on cyanobacterial blooms. *J Lake Sci*, 2020, **32**(1): 11-20. DOI: 10.18307/2020.0102. [张民, 史小丽, 阳振等. 2012—2018年巢湖水质变化趋势分析和蓝藻防控建议. 湖泊科学, 2020, **32**(1): 11-20.]
- [31] Caballero I, Stumpf RP. Atmospheric correction for satellite-derived bathymetry in the Caribbean waters: From a single image to multi-temporal approaches using Sentinel-2 A/B. *Optics Express*, 2020, **28**(8): 11742-11766. DOI: 10.1364/OE.390316.
- [32] Wu QY, Gao F, Li ZX *et al.* The contrast of Sentinel-2A and Landsat 8 data in remote sensing estimation of vegetation coverage. *Bulletin of Surveying and Mapping*, 2021, (S1): 104-108, 113. [吴青云, 高飞, 李振轩等. Sentinel-2A 与 Landsat 8 数据在植被覆盖度遥感估算中的比较. 测绘通报, 2021, (S1): 104-108, 113.]
- [33] Zhao L, Lu XY, Tan H *et al.* Water quality monitoring technology based on GF-1 satellite and XGBoost model. *Remote Sensing Information*, 2021, **36**(2): 96-103. [赵力, 卢修元, 谭海等. 利用高分一号卫星与 XGBoost 模型的水体总氮和总磷监测技术. 遥感信息, 2021, **36**(2): 96-103.]
- [34] Zhu AP, Yuan SY, Wen SS *et al.* Effects of landscape pattern on water quality at multi-spatial scales in the Liuxi River. *Acta Ecologica Sinica*, 2023, **43**(4): 1485-1495. [朱爱萍, 原升艳, 温姗姗等. 不同空间尺度的景观格局对流溪河水质的影响. 生态学报, 2023, **43**(4): 1485-1495.]
- [35] Wang J, Liu AA, Zhang JW *et al.* Effect of integrated landscape characteristics around Chaohu Lake on river water quality based on watershed units. *Environmental Science*, 2024, **45**(6): 3214-3224. [王杰, 刘安安, 张俊武等. 基于流域单元的环巢湖综合景观对河流水质的影响. 环境科学, 2024, **45**(6): 3214-3224.]
- [36] Shi ZQ, Wang M, Tan J *et al.* Effect of vegetation buffer zone on reducing agricultural non-point pollution in Wuliangsu Lake basin. *Journal of Soil and Water Conservation*, 2022, **36**(3): 51-56. [史中奇, 王猛, 谭军等. 植被缓冲带对乌梁素海区域农业面源污染的削减效果. 水土保持学报, 2022, **36**(3): 51-56.]
- [37] Zhou P, Huang JL, Pontius RG *et al.* New insight into the correlations between land use and water quality in a coastal watershed of China: Does point source pollution weaken it? *Science of the Total Environment*, 2016, **543**(A): 591-600.
- [38] Liu JF, Xu JJ, Zhang X *et al.* Nonlinearity and threshold effects of landscape pattern on water quality in a rapidly urbanized headwater watershed in China. *Ecological Indicators*, 2021, **124**: 107389. DOI: 10.1016/j.ecolind.2021.107389.
- [39] Zhong XC, Xu QL, Yi JH *et al.* Study on the threshold relationship between landscape pattern and water quality considering spatial scale effect—A case study of Dianchi Lake Basin in China. *Environmental Science and Pollution Research*, 2022, **29**(29): 44103-44118. DOI:

- 10.1007/s11356-022-18970-0.
- [40] Chen KY, Chen HX, Zhou CL *et al.* Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research*, 2020, **171**: 115454. DOI: 10.1016/j.watres.2019.115454.
- [41] Huang J, Zhan JY, Yan HM *et al.* Evaluation of the impacts of land use on water quality: A case study in the Chaohu Lake basin. *The Scientific World Journal*, 2013, **2013**(1): 329187. DOI: 10.1155/2013/329187.
- [42] Xu QY, Yan TZ, Wang CY *et al.* Managing landscape patterns at the riparian zone and sub-basin scale is equally important for water quality protection. *Water Research*, 2023, **229**: 119280. DOI: 10.1016/j.watres.2022.119280.
- [43] Wang YC, Du JJ, Zhang Y *et al.* Scale effects of landscape pattern on impacts of river water quality: A meta-analysis. *Environmental Science*, 2024, **45**(5): 2631-2639. DOI: 10.13227/j.hjlx.202306180. [王玉仓, 杜晶晶, 张禹等. 景观格局对河流水质影响的尺度效应 Meta 分析. *环境科学*, 2024, **45**(5): 2631-2639.]
- [44] Liu RM, Wang JW, Shi JH *et al.* Runoff characteristics and nutrient loss mechanism from plain farmland under simulated rainfall conditions. *Science of the Total Environment*, 2014, **468**: 1069-1077. DOI: 10.1016/j.scitotenv.2013.09.035.
- [45] Hu YF, Fan ZY, Chen ZT *et al.* Correlation analysis between landscape pattern and water quality in the Lianjiang River watershed in Shantou City. *Environmental Monitoring in China*, 2021, **37**(3): 126-133. [胡艳芳, 范中亚, 陈昭婷等. 汕头市练江流域景观格局与水质的关联分析. *中国环境监测*, 2021, **37**(3): 126-133.]
- [46] Gay ET, Martin KL, Caldwell PV *et al.* Riparian buffers increase future baseflow and reduce peakflows in a developing watershed. *Science of the Total Environment*, 2023, **862**: 160834. DOI: 10.1016/j.scitotenv.2022.160834.
- [47] Wu JH, Lu J. Spatial scale effects of landscape metrics on stream water quality and their seasonal changes. *Water Research*, 2021, **191**: 116811. DOI: 10.1016/j.watres.2021.116811.
- [48] Wu HJ, Yang T, Liu X *et al.* Towards an integrated nutrient management in crop species to improve nitrogen and phosphorus use efficiencies of Chaohu Watershed. *Journal of Cleaner Production*, 2020, **272**: 122765. DOI: 10.1016/j.jclepro.2020.122765.



附图 I 研究技术路线图

Attached Fig.I Research technology route

附表 I 本文选用的景观指标

Attached Tab.I Landscape metrics used in this study

一级类别	二级类别	景观指标
类型级别指标 (80) (包括农 业用地, 林地、城 镇用地和湿地)	面积指标 (28)	斑块面积占比 (PLAND), 斑块类型总面积 (CA), 斑块面积中位数 (AREA_MD), 展转半径中位数 (GYRATE_MD), 最大斑块面积占比 (LPI), 斑块数 (NP), 斑块密度 (PD)
	边缘指标 (8)	总边缘长度 (TE), 边缘密度 (ED)
	形状指标 (24)	边缘面积比中位数 (PARA_MD), 形状指数中位数 (SHAPE_MD), 形状指数平均值 (SHAPE_MN), 分维数中位数 (FRAC_MD), 相关外接圆中位数 (CIRCLE_MD), 景观形状指数 (LSI)
景观级别指标 (33)	聚集性指标 (20)	景观分离度 (DIVISION), 分散指数 (SPLIT), 散布与并列指数 (IJI), 斑块内聚力指数 (COHESION), 蔓延度指数中位数 (CONTIG_MD)
	面积指标 (6)	总面积 (CA), 最大斑块面积占比 (LPI), 斑块面积中位数 (AREA_MD), 展转半径中位数 (GYRATE_MD), 斑块数 (NP), 斑块密度 (PD)

边缘指标 (2)	总边缘长度 (TE), 边缘密度 (ED)
形状指标 (7)	边缘面积分维 (PAFRAC), 边缘面积比中位数 (PARA_MD), 形状指数中位数 (SHAPE_MD), 形状指数平均值 (SHAPE_MN), 分维数中位数 (FRAC_MD), 相关外接圆中位数 (CIRCLE_MD), 蔓延度指数中位数 (CONTIG_MD)
聚集性指标 (10)	景观分离度 (DIVISION), 分散指数 (SPLIT), 有效颗粒尺寸 (MESH), 散布与并列指数 (IJI), 景观形状指数 (LSI), 斑块内聚力指数 (COHESION), 蔓延度指数 (CONTAG), 相似毗邻百分比 (PLADJ), 聚集度指数 (AI), 欧几里得最临近距离中位数 (ENN_MD)
多样性指标 (8)	斑块丰富度 (PR), 斑块丰富度密度 (PRD), Shannon 多样性指数 (SHDI), Simpson 多样性指标 (SIDI), 修正的 Simpson 多样性指标 (MSIDI), Shannon 均匀度指标 (SHEI), Simpson 均匀度指标 (SIEI), 修正的 Simpson 均匀度指标 (MSIEI)
植被指标 (1)	植被覆盖指标 (1) 植被覆盖度 (FVC)
