DOI:10.11918/202509053

中国流域富营养化时空评价与智慧驱动因素解析

王沁茹[1,2],邱 瑞[3],寇 旭[1,2],陈 阳[1,2],郭洪光[1,2]

(1.四川大学 建筑与环境学院,成都 610065;2.深地工程智能建造与健康运维国家重点实验室(四川大学),成都 610065;3.四川大学 商学院,成都 610065)

摘 要:富营养化是目前地表水面临的最具挑战性的环境问题之一,现有研究多聚焦于区域尺度分析。为揭示大范围流域的富营养化格局及其复杂驱动机制,对中国范围内流域的富营养化状况进行了评估与等级划分,并结合机器学习和SHapley加性解释(SHAP)的可解释人工智能方法探究影响富营养状况变化的潜在驱动因素。结果显示:富营养现象在大多数流域中普遍存在,中国东部、东北部和中部地区的富营养化问题较西部地区严重,且呈现出显著的时空聚集特征;总磷和总氮是关键的水质因子,贡献度分别为20.6%和75.4%,二者与流域富营养化存在显著的空间正相关关系;轻量级梯度提升机(LightGBM)模型在所有模型中表现最优,准确率达91%。SHAP分析表明,相比人为因素,自然驱动因素在解释流域整体富营养状况变化中起着更为重要的作用,其中,地形条件控制营养物质的积累和流失,降水则起到调节作用。然而,随着富营养化程度的加剧,人口密度、国内生产总值和污水处理能力等人为因素的影响逐渐凸显。经济发展若能伴随基础设施的提高,能够在一定程度上缓解富营养化。鉴于内部营养负荷和外部流域特征对富营养化的协同作用,建议实施分区分类精准管控以有效应对这一环境挑战。本研究结合人工智能方法深化了对流域尺度富营养化驱动机制的理论理解,也为水环境治理提供了定量化、可解释的科学证据。

关键词:富营养;水质;流域;机器学习

中图分类号:X824;TP181 **文献标志码**:A **文章编号**:0367-6234(2026)06-0056-12

Spatial-temporal assessment of eutrophication in China's river basins and intelligent identification of driving factors

WANG Qinru[1,2], QIU Rui[3], KOU Xu[1,2], CHEN Yang[1,2], GUO Hongguang[1,2]

(1. College of Architecture & Environment, Sichuan University, Chengdu 610065, China; 2. State Key Laboratory of Intelligent Construction and Healthy Operation and Maintenance of Deep Underground Engineering (Sichuan University), Chengdu 610065, China; 3. Business School, Sichuan University, Chengdu 610065, China)

Abstract: Eutrophication is one of the most challenging environmental problems facing surface water at present, and existing studies mostly focus on regional-scale analysis. To reveal the eutrophication patterns and their complex driving mechanisms in large-scale river basins, this paper evaluated and classified the eutrophication status of river basins across China and explored the potential driving factors affecting the changes in eutrophication status by combining machine learning and the explainable artificial intelligence method of SHapley Additive exPlanations (SHAP). The results show that eutrophication is widespread in most river basins, and the eutrophication problem in eastern, northeastern, and central China is more severe than that in western regions, exhibiting significant spatial-temporal clustering characteristics; total phosphorus and total nitrogen are key water quality factors, with contributions of 20.6% and 75.4%, respectively, and both have a significant positive spatial correlation with river basin eutrophication; the Light Gradient Boosting Machine (LightGBM) model performs best among all models, with an accuracy of 91%. SHAP analysis reveals that compared with anthropogenic factors, natural driving factors play a more important role in explaining the overall changes in the eutrophication status of river basins, among which topographic conditions control the accumulation and loss of nutrients, while precipitation plays a regulatory role. However, with the aggravation of eutrophication, the impact of anthropogenic factors such as population density, gross domestic product, and wastewater treatment capacity gradually becomes prominent. If economic development is accompanied by the improvement of infrastructure, eutrophication can be mitigated to a certain

收稿日期:2025-09-12;**录用日期**:2025-10-17;**网络首发日期**:2025-12-08
网络首发地址:https://link.cnki.net/urlid/23.1235.t.20251205.2020.002
基金项目:国家重点研发计划(2023YFC3210100);国家自然科学基金(52470107)
作者简介:王沁茹(2001—),女,硕士研究生;邱 瑞(1990—),男,副研究员,博士生导师;郭洪光(1986—),男,教授,博士生导师
通信作者:郭洪光,hgguo@scu.edu.cn;邱 瑞,qiuruicd@scu.edu.cn

extent. Given the synergistic effect of internal nutrient loads and external river basin characteristics on eutrophication, it is recommended to implement precise management and control based on spatial zoning and classification to effectively address this environmental challenge. This study deepens the theoretical understanding of the driving mechanism of river basin-scale eutrophication by combining artificial intelligence methods and provides quantitative and interpretable scientific evidence for water environment governance.

Keywords: eutrophication; water quality; river basin; machine learning

富营养化是水生系统对过量养分输入的生物反应，这一过程会导致生物群落结构发生重大改变，损害生态系统功能，并严重破坏水体的生态平衡，是目前地表水面临的最具挑战性的环境问题之一[1-2]，对水质、人体健康以及经济社会等多方面造成威胁[3]。水体富营养化评价是水环境管理的重要基础，然而，目前尚未形成标准化的方法，现有研究多基于现场收集的少数监测点的多个指标数据[4-5]或某项相关指标（大多为藻类、叶绿素 a）的空间连续数据来评估[6-7]。并且，影响水体富营养化的关键因素在不同区域之间实际存在明显差异，除水体内部的变化外[8]，还包括由降水驱动的水文状况[9]、景观特征以及人类社会活动等外部影响，如土地利用[10]、污水处理[11]、化肥农药的使用[12]。水体内外部因素的同时作用，显著增加了富营养变化的不确定性，深入揭示这些因素的复杂作用，对于水生生态系统进行准确有效的监测和管理具有重要意义。

传统的统计或经验模型在捕捉因素与富营养化之间复杂的交互作用与非线性效应方面存在局限，而基于树的机器学习模型能够破译许多变量之间复杂的非线性关系[13]。此外，借助 SHapley 加性解释（SHAP）等可解释性人工智能技术，不仅能够量化目标变量与关键影响因子的正负作用及其关系模式，还能发现隐藏在数据之中的规律，为识别和理解水体富营养化的驱动因素提供了新的思路。

本研究旨在全国尺度系统评估中国流域的富营养化状况，揭示其时空变化规律并识别主要驱动因素。研究采取广泛应用于湖泊、水库和河流富营养化分析的对数型幂函数普适指数法（eutrophication index，I_E）进行富营养化评估，分析其变化规律并划分等级[14-15]。通过机器学习工具构建流域富营养化程度与反映流域背景下人类活动强度和自然特征之间关系的模型，并结合可解释的 SHAP 分析及特征依赖图，能够有效掌握流域中富营养化的变化趋势，进一步推进探究大范围水体富营养水平及潜在驱动因素，帮助识别富营养风险发生的关键时刻，从而制定考虑驱动因素变化规律的控制政策及优化修复措施。

1　材料与方法

1.1　研究区域及数据来源

研究基于中华人民共和国生态环境部发布的综合地表水水质数据（https://waterpub.cnemc.cn:10001/）。该数据集包含了分布在全国七大流域、浙闽片河流、西北诸河、西南诸河、重点湖泊、水库和入海口段的 3 646 个监测站的监测结果。数据集涵盖了 GB 3838—2002《地表水环境质量标准》中列出的 23 项水质指标（不包括粪大肠菌群）。选取了 2021—2023 年共 129 600 个数据样本进行分析。重点关注的指标为总磷（TP）、总氮（TN）、溶解氧（DO）、高锰酸盐指数（COD_{Mn}）、5 d 生化需氧量（BOD_5）和氨氮（NH_3-N），单位均为 mg/L。这些指标具有良好的可获取性、连续性和时空覆盖度，并且与水体富营养化密切相关。其中，TP 反映外源磷负荷，DO 指示水体的自净能力与耗氧过程，COD_{Mn} 和 BOD_5 可反映水体化学需氧量和生物可降解有机污染水平。TN 作为氮素负荷的综合指标，涵盖硝酸盐、铵盐、有机氮等多种形态。NH_3-N 作为补充指标，因其对水生生物具有较高的直接生物可利用性且具毒性[16-17]，能更敏感地反映近期污染输入及生态影响[18-19]。硝酸盐是河流中氮的主要存在形态，但由于全国尺度公开监测数据有限，为保障宏观尺度分析中数据的一致性与可比性，未将其纳入本研究的分析框架。

根据研究目标，从自然和人为两方面选取了合适的流域特征指标，其具体构成与数据来源如下。自然因素方面，降雨量（mm）、气温（℃）、归一化植被指数（NDVI，无量纲）数据来源于国家青藏高原科学数据中心，高程（m）数据从航天飞机雷达地形测量数据库中获得（http://srtm.csi.cgiar.org/srtmdata/），坡度（°）由高程数据得出。人为因素方面，人口密度（人/km^2）数据来自 LandScan 平台（https://landscan.ornl.gov）。耕地利用数据来自中国土地覆盖数据集（CLCD），仅取耕地面积（km^2）进

行分析。农药化肥使用量(t)数据从各地级市乡村振兴数据中获得,国内生产总值(GDP,亿元)和污水处理能力(10^4 m^3/d)数据分别源自《中国县域统计年鉴》和《中国城市建设统计年鉴》。

1.2 数据处理

为了评估不同环境、社会因素对流域富营养等级变化的影响,首先,建立了一个以流域特征为输入、以富营养等级为输出的综合数据库。耕地面积、农药化肥使用量、GDP 及污水处理能力在已有数据的基础上,通过地理信息系统(GIS)中的克里金插值法生成栅格数据。根据监测点的经纬度坐标,获得每个断面的气象条件、社会景观和人类活动数据,为模型构建提供解释变量。缺失值占比均小于 3%,使用 K 近邻(KNN)插补方法处理。为改进机器学习模型的训练过程以实现快速收敛,使用 Scikit-Learn 中的 StandardScaler 对输入特征进行标准化处理以获得相似的数据尺度。未公布的监测数据采取随机森林的方法填补断面缺失值,取每个断面的年均值用于富营养评估。

采用广泛应用于中国湖泊、水库和河流水体富营养化评价的对数型幂函数普适指数 I_E 划分富营养等级[14-15],I_E 的计算采用式(1),等级划分如表 1 所示。

$$I_E = \sum_{q=1}^{p} W_q \times I_{Eq} = 10.77 \times \sum_{q=1}^{p} W_q \times (\ln X_q)^{1.1826} \tag{1}$$

式中:W_q 为富营养特征 q 的标准化加权值,本研究中各特征权重相等;I_{E_q} 为特征 q 的富营养评价指数;X_q 为特征的规范值。特征包含 TP、TN、DO、COD_{Mn}、BOD_5 和 NH_3-N。

1.3 贡献度

为量化水质特征对富营养化评价的相对重要性,提出一种基于特征主导统计的贡献度计算方法。其核心思想是每个特征的贡献度可以用该特征的富营养评价指数(I_E)在所有样本中作为最大值出现的概率来表示。如式(2)~(4)所示:

$$I_{Ei,\max} = \max_{q \in Q}(I_{Eiq}) \tag{2}$$

$$q_{\max}(i) = \arg I_{Ei,\max} \tag{3}$$

$$P_q = \frac{1}{N}\sum_{i}^{N} I(q_{\max}(i) = q) \tag{4}$$

式中:$I_{Ei,\max}$ 为样本 i 对应的最大指标值;I_{Eiq} 为样本 i 在第 q 个指标上的取值;$q_{\max}(i)$ 为在样本 i 中取得最大值的指标;P_q 为指标 q 的贡献度;$I(q_{\max}(i) = q)$ 为指示函数,当且仅当样本 i 的最大值对应指标 q 时,$I(q_{\max}(i) = q)$ 为 1,否则为 0。

表 1 不同等级富营养化指数的范围

Tab. 1 Range of different grades of eutrophication index

富营养等级	Ⅰ贫营养	Ⅱ中营养	Ⅲ富营养	Ⅳ重富营养	Ⅴ极富营养
I_E	≤20	(20,39.42]	(39.42,61.29]	(61.29,76.28]	(76.28,99.77]

1.4 机器学习模型

为了确保机器学习建模过程的严谨性,并提高结果的可靠性和有效性,采用了 4 种常用的机器学习分类模型,即极端梯度提升(XGBoost)、自适应提升(AdaBoost)、随机森林(RF)和轻量级梯度提升机(LightGBM)。XGBoost 通过梯度提升决策树迭代优化残差[20],并结合正则化和并行计算防止过拟合,从而提高分类准确率。AdaBoost 迭代更新错误分类样本的权重,并通过加权投票聚合弱分类器,逐步强化分类边界[21]。RF 通过集成学习的思想整合多棵树,并通过使用装袋算法和特征选择实现快速训练和分类[22]。LightGBM 利用直方图分箱加速特征处理并采用深度受限的叶优先树生长策略[23],从而提高计算效率。

在本研究中,数据集被分为训练子集和测试子集。首先,通过训练子集进行 5 倍交叉验证(CV)训练模型,并采用网格搜索方法对模型进行超参数优化,以获得相对最优的模型配置。然后,使用多分类常用评价指标准确率(Accuracy,$A_{accuracy}$)、F_1(F1_score)和 AUC(roc_auc_score,C_{AU})3 个指标进行评估:

$$A_{accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{5}$$

$$P_{precision} = \frac{T_P}{T_P + F_P} \tag{6}$$

$$R_{recall} = \frac{T_P}{T_P + F_N} \tag{7}$$

$$F_1 = 2 \cdot \frac{P_{\text{precision}} \cdot R_{\text{recall}}}{P_{\text{precision}} + R_{\text{recall}}} \tag{8}$$

$$R_{\text{TP}} = R_{\text{recall}} \tag{9}$$

$$R_{\text{FP}} = \frac{F_{\text{P}}}{F_{\text{P}} + T_{\text{N}}} \tag{10}$$

$$C_{\text{AU}} = \int_0^1 R_{\text{TP}}(R_{\text{FP}}^{-1}(t))\,\mathrm{d}t \tag{11}$$

式中：T_{P} 代表正类被预测为实际为正的样本，T_{N} 代表负类被预测为负的样本，F_{P} 代表负类被预测为正类的样本，F_{N} 代表正类被预测为负类的样本，R_{TP}表示真阳性率，R_{FP}表示假阳性率。多分类问题视为多个二分类问题，计算各类别评价指标并考虑不同类别重要性，通过加权平均得到最终的评价结果：

$$S = \frac{1}{n}\sum_{k=1}^{n} w_k \cdot m_k \tag{12}$$

式中：n 为类别数，w_k 表示第 k 个类别的权重，m_k 为第 k 个类别的评价指标得分。

F_1是精确率和召回率的调和均值，能更加全面地评估模型，AUC 值可直观地反映分类器的预测能力。模型训练后，使用保留的测试子集进行测试，评估模型性能。

1.5　数据增强

类别不平衡是水质模型中的一个普遍问题，即其中一些类别与其他类别相比数量明显不足。大多数现有的相关模型没有明确解决这个问题，会使模型存在偏差，在少数类别上表现不佳[24]。合成少数类过采样技术（SMOTE）可以通过对少数类别的样本随机抽样解决类别不平衡问题，从而增加其在数据集中的代表性[25]。自适应合成采样方法（ADASYN）考虑数据集的动态变化和不平衡程度，以密度分布为标准调整合成样本的生成概率，避免将少数实例的副本添加到实际数据集时发生过拟合[26]，亦不会丢失有用信息。二者均是处理类别不平衡问题的常用方法。本研究以最少类别（重富营养）为基准，随机抽取其他类别原始样本形成测试集，应用于模型结果的验证过程。其余数据样本通过应用 SMOTE 或 ADASYN 方法，平衡所有类的计数形成训练集。新的训练样本数据集作为输入数据输入到机器学习算法中，在全部为实例的测试集上评估模型的性能。

1.6　SHAP 分析

与结构简单的单棵决策树相比，复杂集成决策树模型的预测过程难以直观地被人理解。SHAP 模型在 2017 年由 Lundberg 和 Lee 提出[27]，是一种基于博弈论和局部解释的分析方法，其核心思想是将每个特征的重要性分解为不同特征值的 SHAP 值的加权和，用来衡量各特征对模型输出的贡献度，可从全局和局部两个层面对模型进行解释[28]。SHAP 值通过 SHAP 图以可视化的方式展示，以显示每个特征对预测结果的正向或负向影响，并解释模型预测的原因和特征的重要性。对于所有样本而言，SHAP 值计算公式如式（13）、（14）所示：

$$\varphi(x_{i,j}) = c_{i,j}w_1 + c_{i,j}w_2 + \cdots + c_{i,j}w_g \tag{13}$$

$$\varphi_i = \varphi_{\text{mean}} + \varphi(x_{i,1}) + \varphi(x_{i,2}) + \cdots + \varphi(x_{i,g}) \tag{14}$$

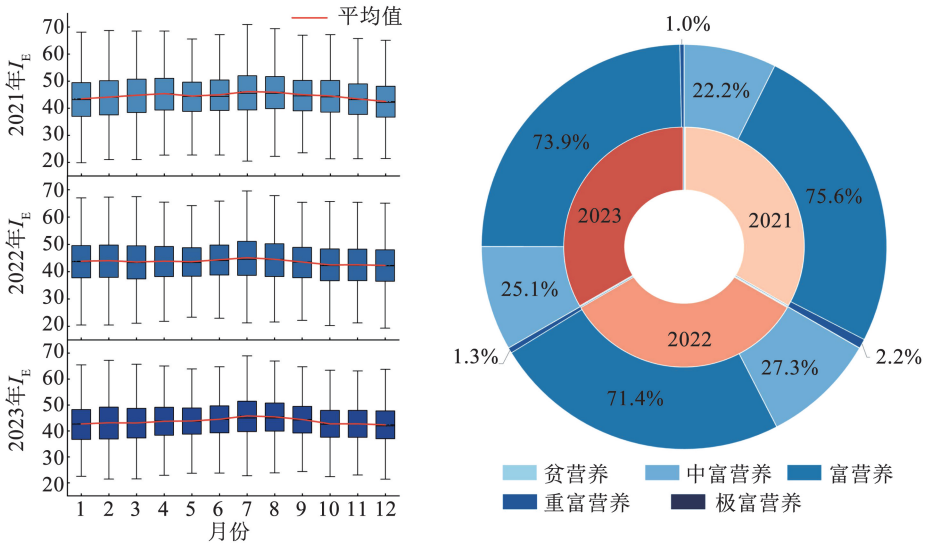
式中：x_i 为第 i 个样本；$x_{i,j}$为第 i 个样本的第 j 个特征；$c_{i,j}$为特征的边际贡献；w_j 为边的权重；$\varphi(x_{i,j})$为样本 $x_{i,j}$的 SHAP 值，当 $\varphi(x_{i,j}) > 0$ 时，说明该特征对预测结果具有正向作用，反之，该特征使预测值降低。模型对样本的预测结果为 φ_i，所有样本目标变量的均值为 φ_{mean}。

本研究在模型训练过程中未人为设定影响因子的权重，而是将因子作为输入变量，由机器学习模型自动学习变量与富营养化状况之间的关系。模型训练完成后，通过 SHAP 可解释人工智能方法量化各因子对预测结果的贡献度，以识别影响流域富营养化的重要因素。SHAP 分析使用 TreeExplainer 进行，数据挖掘和建模的所有元素均使用 Jupyter Notebook 编程环境在 Python 3.10 中进行。

2　结果与讨论

2.1　中国流域富营养化分析及其空间特征

如图 1（a）所示，研究期间各年的 I_{E}值分布较为稳定，中位数和平均值呈现逐年缓慢下降的趋势，说明流域总体营养状况有所改善。I_{E}值的月间波动幅度较小，箱体范围较窄，大部分值集中在 38 ~ 50，极端值分布范围较广，表明部分地区可能存在严重富营养化现象。各月 I_{E}平均值均超过 39.42，处于富营养化状态，其中，7—8 月夏季的 I_{E}值往往较高，一方面是由于流域内部水流速度减缓和停留时间增加[29]，另一方面与气温、降水、农业活动、人类活动等外部负荷相关[30]。进一步将 I_{E}值按富营养化等级标准划分（表 1），并绘制等级占比图，如图 1（b）所示。结果表明，2021—2023 年几乎没有监测点位处于贫营养和极富营养状态，总体状况以富营养化为主。



(a) 2021—2023年I_E月变化　　(b) 2021—2023年各富营养化等级占比

图 1　2021—2023 年中国流域富营养化状况评估

Fig. 1　Assessment of eutrophication in China′s river basins in 2021—2023

此外,中国流域富营养化在研究时间内呈现显著的地域差异性。如表 2 所示,淮河和海河流域的I_E值整体较高,富营养化严重,但随着时间推移,高值区域有减小的趋势。相比之下,西南诸河、北方内陆及新疆诸河流域营养状况总体较好。冷热点分析揭示了富营养化的空间集聚特征。由图 2 可以看出,各地之间相互关系较强。2021—2023 年,冷点和热点空间分布范围变化较小。热点呈现出明显的东部集聚趋势,几乎覆盖了整个东北和华北地区,部分华东和华中地区也被纳入。冷点主要集中在西北和西南地区。

表 2　2021—2023 年中国一级流域年均 I_E 值

Tab. 2　Annual average I_E values in China′s primary river basins in 2021—2023

一级流域	I_E		
	2021 年	2022 年	2023 年
东北诸河流域	47.39	46.40	46.79
海河流域	48.95	47.28	46.74
黄河流域	45.93	45.01	44.39
淮河流域	51.01	49.65	49.42
长江流域	40.78	39.75	40.01
东南沿海诸河流域	44.47	43.48	43.94
珠江流域	41.77	40.76	41.24
西南诸河流域	36.89	36.36	36.79
北方内陆及新疆诸河流域	38.28	37.77	37.37

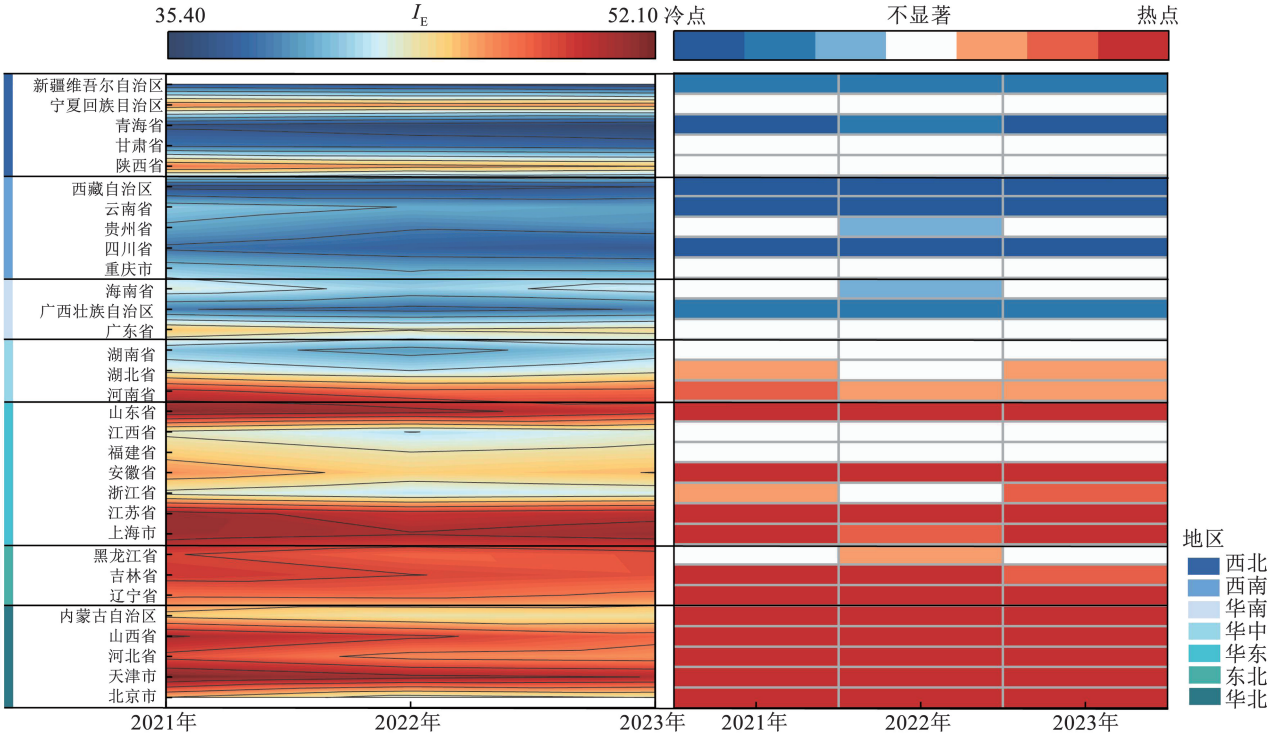


图 2　2021—2023 年富营养空间分布及冷热点分析

Fig. 2　Spatial distribution and hotspot-coldspot analysis of eutrophication in 2021—2023

2.2　富营养化的关键水质因子

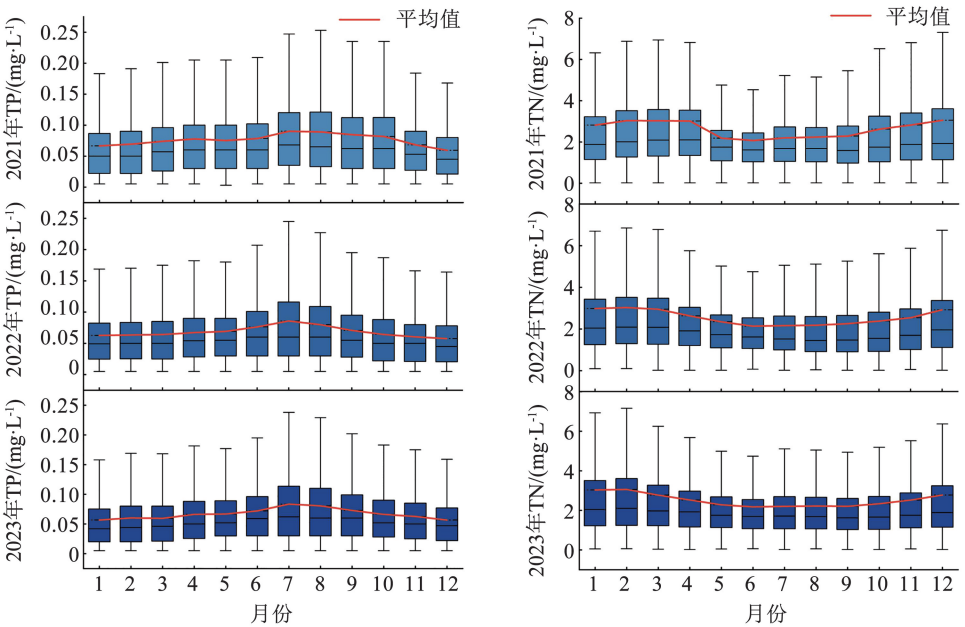
本研究进一步量化了各水质参数贡献度，结果如表 3 所示，TP（20.6%）和 TN（75.4%）是导致流域富营养化的主要因子。2021—2023 年，TP 质量浓度整体呈现先增加后下降的趋势（图 3（a）），6—8 月夏季质量浓度显著升高，达到全年最高值，而冬季（12 月—次年 2 月）质量浓度较低。各月平均值均达到 0.05 mg/L，处于易引发水体富营养化的较高风险水平[31]，大多流域集中在 0.025 ~ 0.1 mg/L。空间上，如图 4 所示，TP 质量浓度较高的区域主要集中在淮河流域，北方内陆及新疆诸河流域质量浓度普遍较低。北方平原的热点区域始终存在，且范围有逐渐扩大的趋势，然而热点的显著性有所下降，说明虽面临较高的污染压力，局部水质改善或污染空间分布的均匀化可能正在发生。冷点区域主要出现在华南地区、西北山区和部分西南地区，TP 的输入相对有限。

TN 质量浓度在一年中整体呈现先降后升的变化（图 3（b）），表现出较明显的季节特点。冬季 TN 质量浓度相对偏高，春季逐渐降低，夏季维持平稳状态，秋季开始上升并在冬季再次达到峰值。与 TP 相比，TN 的高质量浓度区在黄淮海平原表现出更为集中的核心分布，且热点范围更广，需同时关注平原的重点治理与边缘区的预警；冷点区域增加了部分西南山区（图 4）。此外，TP、TN 与 I_E 具有显著的空间正相关关系，且在时序上保持稳定，3 年平均莫兰指数（Moran′s I）分别为 0.51（$P < 0.01$）和 0.46（$P < 0.01$）。如表 4 所示，新疆维吾尔自治区、青海省、四川省和云南省是低-低集群的核心，江苏省和上海市的水体富营养化主要受 TP 驱动，内蒙古自治区、河北省和河南省则更受 TN 的影响，这在一定程度上揭示了中国东部沿海城市群与华北农业区在富营养化机制上的差异。值得注意的是，浙江省在研究期内始终呈现出“低 TP-高 I_E”的特殊格局，说明该省的其他营养物质或复合污染源在驱动富营养化过程中发挥了更重要的作用。

表 3　各水质因子对富营养化的贡献度

Tab. 3　Contribution of water quality factors to eutrophication

水质因子	贡献度/%		
	2021 年	2022 年	2023 年
TP	22.3	21.0	18.5
TN	73.6	75.0	77.5
COD_{Mn}	1.9	1.7	1.5
BOD_5	0.1	0	0.1
NH_3-N	0.2	0.1	0
DO	1.9	2.2	2.4



(a) 2021—2023年TP月变化　　(b) 2021—2023年TN月变化

图 3　2021—2023 年 TP 和 TN 的月变化趋势

Fig. 3　Monthly variation trends of TP and TN in 2021—2023

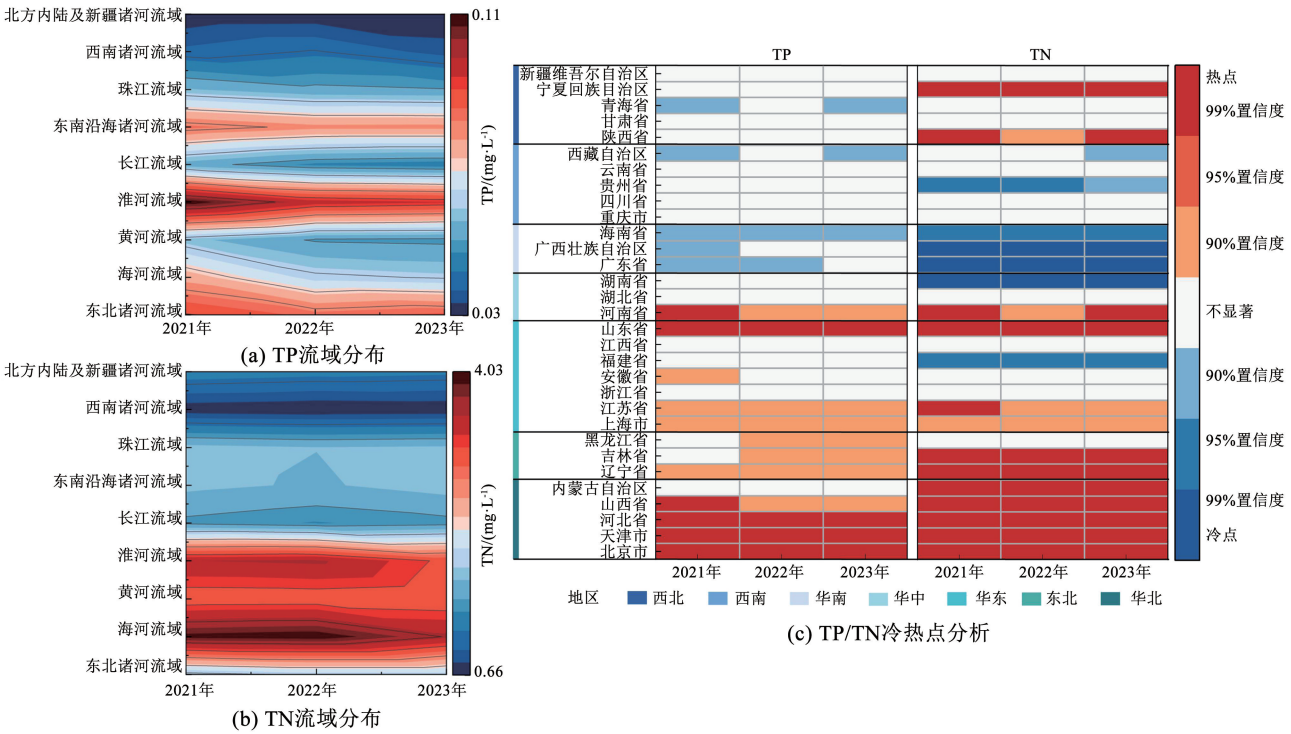


图 4 2021—2023 年 TP 和 TN 的多尺度空间格局

Fig. 4 Multi-scale spatial patterns of TP and TN in 2021—2023

表 4 2021—2023 年 TP 和 TN 同 I_E 的二元空间自相关分析

Tab. 4 Bivariate spatial autocorrelation analysis of TP and TN with I_E in 2021—2023

	年份	Moran's I	Z	P	高-高聚集	低-高聚集	低-低聚集	高-低聚集
TP-I_E	2021	0.523	5.053	0.001	a,b	c	d,e	—
	2022	0.465	4.595	0.001	a	c	d,e,f	—
	2023	0.529	5.073	0.001	a	c	d,e,f	—
TN-I_E	2021	0.498	5.139	0.001	h	—	d,g,i	—
	2022	0.448	4.731	0.001	h	—	d,g,i	—
	2023	0.446	4.737	0.001	h	—	d,g	—

注:P 值通过 999 次置换检验计算得出;"—"表示无该类聚集区。聚集区域代码:a:江苏省、上海市;b:山东省;c:浙江省;d:新疆维吾尔自治区、青海省、四川省、云南省;e:甘肃省、西藏自治区;f:重庆市;g:贵州省;h:内蒙古自治区、河北省、河南省;i:江西省。

先前研究表明,有效减少 TP、TN 对缓解流域富营养化至关重要,水体中藻华主要吸收过量的氮和磷进行繁殖[32],氮磷共限导致富营养化的概率很高[33]。然而,本研究发现不同区域受氮或磷限制的程度存在差异,应在氮主导区优先削减氮污染,在磷主导区强化磷治理,在低值区注重生态保护与预警。同时,还需关注季节差异,在污染高峰期(夏季 TP 质量浓度高、冬季 TN 质量浓度高)开展针对性的污染监测与排放管控。

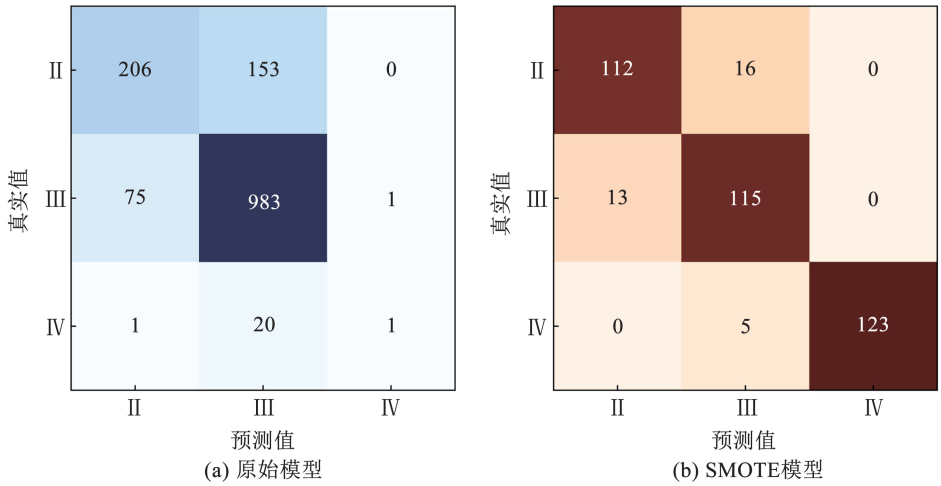
2.3 模型性能与比较

为探究潜在预测变量与富营养化等级之间的关系,构建了 4 种机器学习模型,结果如表 5 所示。综合各评价指标以及通过数据增强实现的改进效果,最终选择基于 SMOTE 数据增强的 LightGBM 分类模型。由该模型可知,流域特征的原始模型在交叉验证中很好地拟合了训练数据集,各评价指标均为 1。当应用于测试数据集时,其准确率为 0.826,F_1 为 0.814,AUC 值优于其他两个指标。然而,基于混淆矩阵的结果(图 5(a)),数据集中 II 级和 IV 级样本出现次数有限,导致这些类别在测试集中分布不均。这种不平衡使得模型的预测出现偏差,主要倾向于 III 级样本。而数据增强后,样本较少的等级也被关注(图 5(b)),测试集各指标提升幅度超过8.5%(表 5),模型效果显著改善,为后续评估和分析提供了更可靠的支持。

表 5　基于原始数据、SMOTE 与 ADASYN 采样方法的机器学习模型性能对比

Tab. 5　Performance comparison of machine learning models based on original data, SMOTE, and ADASYN sampling methods

模型		训练集			测试集		
		Accuracy	F1_score	roc_auc_score	Accuracy	F1_score	roc_auc_score
原始模型	XGBoost	1	1	1	0.826	0.813	0.869
	RF	1	1	1	0.829	0.818	0.880
	LightGBM	1	1	1	0.826	0.814	0.866
	AdaBoost	1	1	1	0.819	0.804	0.853
SMOTE 模型	XGBoost	1	1	1	0.898	0.899	0.971
	RF	1	1	1	0.901	0.902	0.976
	LightGBM	1	1	1	0.911	0.912	0.975
	AdaBoost	1	1	1	0.909	0.910	0.977
ADASYN 模型	XGBoost	1	1	1	0.888	0.888	0.975
	RF	1	1	1	0.893	0.894	0.974
	LightGBM	1	1	1	0.906	0.906	0.978
	AdaBoost	1	1	1	0.898	0.899	0.976



(a) 原始模型　　(b) SMOTE模型

图 5　LightGBM 原始模型及 SMOTE 模型在测试集上的混淆矩阵结果

Fig. 5　Confusion matrix results of LightGBM original model and SMOTE model on test set

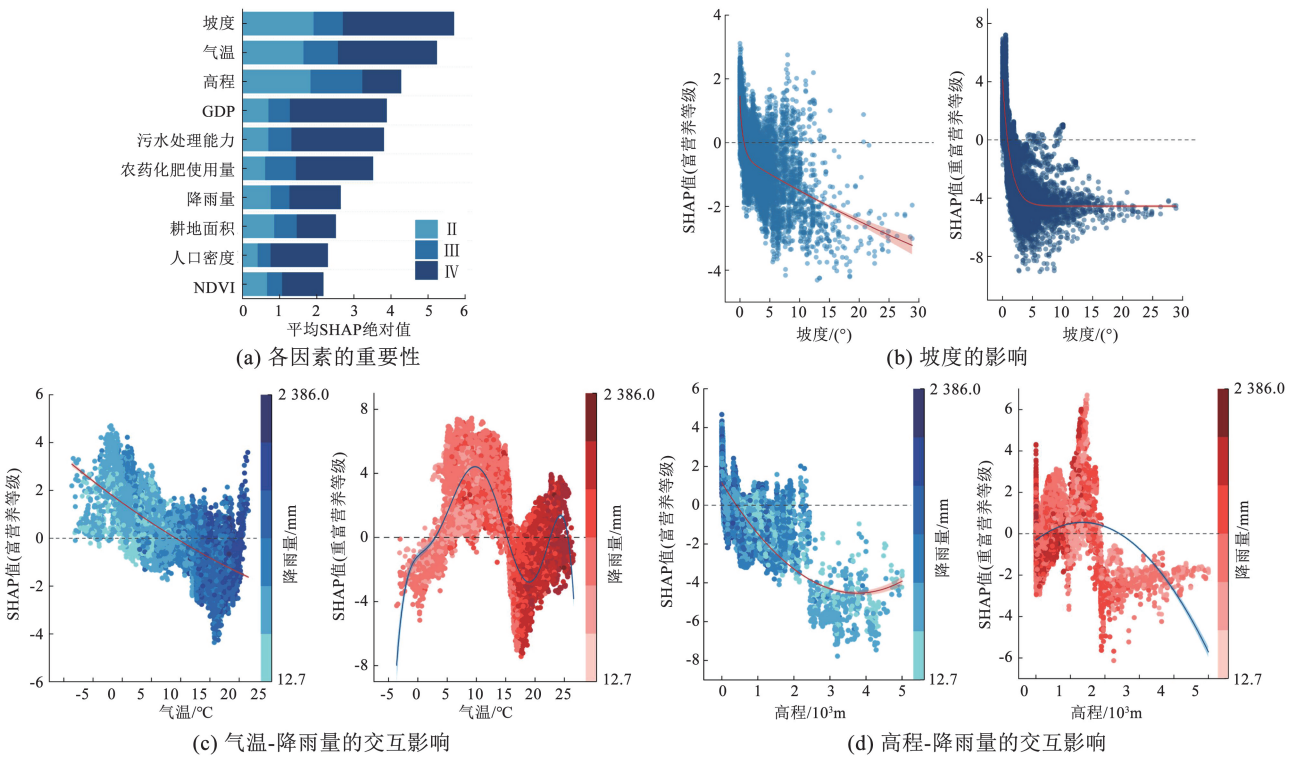
2.4　自然因素对富营养化的影响

如图 6(a)所示，在解释流域整体富营养状态变化中，自然驱动因素比人为因素更为重要，坡度、气温和高程是 3 个最重要的因素。由图 6(b)可以看出，坡度与富营养化呈负相关关系，在各营养状态中负驱动力大于正驱动力，高坡度更利于流域维持低等级的营养状态。坡度分布范围为 0° ~ 30°，较高的坡度可有效产生地表径流[34]，将养分输送到平坦地区。地势平坦地区更加适宜人类居住，通常城市化程度更高，养分输出更高(表 6)[35]。同时，水流缓慢和沉积作用增强易导致营养物质在流域中累积，有机物分解速率加快[36]，从而促进富营养化的过程。

降水往往和其他因素交互影响流域富营养化。如图 6(c)所示，气温集中在 0 ~ 25 ℃，对于富营养流域，随温度升高促进作用减弱，更利于流域朝中营养状态发展。以往研究表明，高温可能加剧藻类的生长和降解[37]，同时提升养分的再循环率，使其在流域中滞留[38]，但大量的降水能够一定程度抑制气温的影响。对于重富营养流域，气温呈现出先上升后下降再上升的波动趋势，温度处于 5 ~ 15 ℃时，极大程度地促进重富营养化，此后温度升高，降水量的增加使得促进作用减弱，但调节能力有限。这可能与降雨是否形成足够的径流有关，因为径流过程不仅决定了养分的迁移与冲刷[39]，也影响水体的稀释能力[40]，从而进一步调节气候因子对富营养化的作用。

此外，降雨与高程对富营养化也有一定的关联度。流域处于较低的地理位置，会受到强烈的农业活动、城市发展和人口聚集的影响[41]，显著增加非点源养分负荷。降水一方面提高水循环效率和自净能力[42]，稀释营养物质浓度，另一方面，降水径流中

含有的氮、磷等营养物质进入流域中[43]。如图 6(d)所示,低地(小于 2 000 m)高降水对大部分流域富营养化具有抑制作用,且流域中 TP、TN 质量浓度与高程、降水均呈负相关关系(表 6),大量降水对养分的稀释作用似乎更大。部分流域监测断面出现了高降水促进流域朝富营养方向发展的情况,可能与极端降水的发生有关[30,35],径流过程伴随的高强度输入可能抵消稀释效应[44]。高地流域受到的污染较少[45],降水量影响较小,不易造成富营养化,此时自然因素对水质的保护作用更为突出。



(a) 各因素的重要性　(b) 坡度的影响

(c) 气温-降雨量的交互影响　(d) 高程-降雨量的交互影响

图 6　自然因素对流域富营养化的影响

Fig. 6　Influence of natural factors on river basin eutrophication

表 6　TP 和 TN 与流域特征的标准化线性回归结果

Tab. 6　Standardized linear regression results of TP and TN with river basin characteristics

因素	TP			TN		
	系数	标准误差	*P*	系数	标准误差	*P*
气温	−0.012	0.019	0.511	0.009	0.018	0.635
降雨量	−0.022	0.018	0.208	−0.334	0.017	0
高程	−0.099	0.016	0	−0.135	0.016	0
坡度	−0.141	0.013	0	−0.076	0.013	0
NDVI	0.001	0.012	0.910	0.098	0.012	0
耕地面积	0.087	0.013	0	−0.097	0.013	0
农药化肥	−0.037	0.014	0.004	0.108	0.013	0
GDP	0.040	0.012	0.001	0.046	0.012	0
人口密度	0.029	0.012	0.016	0.037	0.012	0.001
污水处理能力	0.069	0.013	0	−0.003	0.013	0.796

2.5　人为因素对富营养化的影响

图 6(a)和图 7(a)表明,随着富营养化等级的提高,人为驱动因素的作用逐渐增强,且在重富营养阶段,人为因素的平均作用强于自然因素,其中,重要性排序变化最明显的是人口密度。人口密度可反映人类分布和社会活动情况,如图 7(b)所示,随着人口密度的增加,对重富营养的影响加大,当接近 2 000 人/km^{2}时,促进作用减弱直至趋于平稳状态,这意味着在一些城市化水平较高的地区,尽管高人口密度可能导致更高的排放污染,但同时也会有更

严格的污染控制措施，能够缓解富营养化程度的进一步加剧。

农药、化肥施用量在中营养和富营养时变化规律不明显，在重富营养状态下，随着施用量的增加，反而会抑制重富营养的发生（图 7（c）），可能的原因在于农药化肥的持续施用超过作物对营养物质的吸收，使其不断在土壤中积累，随径流输送到流域中[46]，从而促进藻类和植物的快速生长，加速流域重富营养化。当水体中的养分浓度过高时，藻类爆发性增长，水体浑浊，导致光照穿透深度下降，光合作用减少，使藻类繁殖受到抑制，逐渐进入饱和状态，此时流域更多地维持在富营养状态，而不是进一步朝重富营养化发展，也可能得益于精准施肥等现代农业管理措施的应用或其他社会因素制约了农药化肥施用产生的水体污染（图 7（d）），但此时流域水质处于较差的水平。耕地面积对流域富营养化的影响整体呈正相关，并且随耕地面积的增加，其影响显著增强（图 7（e））。耕地面积扩大会改变土壤结构，增加化肥农药总施用量，可能有更多的营养物质通过径流输送至水体[47]，加剧富营养化。

GDP 是影响最大的人为因素，其作用与城市基础设施建设和农业活动密切相关。如图 7（f）所示，随 GDP 水平的提高，对流域富营养和重富营养化的促进作用增强。当 GDP 在 2 000 亿元左右时，其正向作用减弱或趋于平衡。对于经济水平较低的地区，污水处理能力通常有限，经济活动的增加会显著加大污染物的排放，对富营养化的促进作用较大。当 GDP 达到较高水平时，这些地区往往有较强的农业和工业基础，为了提高农作物的产量，可能会大量使用化肥和农药，且频繁的工业和生活污水排放也会导致水体中营养物质浓度升高[48]。然而经济发展也伴随着基础设施尤其是污水处理设施的完善，能够有效减缓污染物排放对水体的负面影响[49]，对富营养化的促进作用减弱甚至逆转。

目前，多数城市的经济处于中等水平，对流域富营养化的治理和管理应遵循“控制—‘减轻 + 修复’”的原则，制定并严格执行氮、磷等营养物质排放标准。应持续提高污水处理技术水平，实现最低的富营养化潜力。还应采用更有效的施肥模式从源头上控制农药化肥的过度使用。对于已处于富营养或重富营养的流域，则要采用物理、化学、生物等多种方法修复流域内部生态。

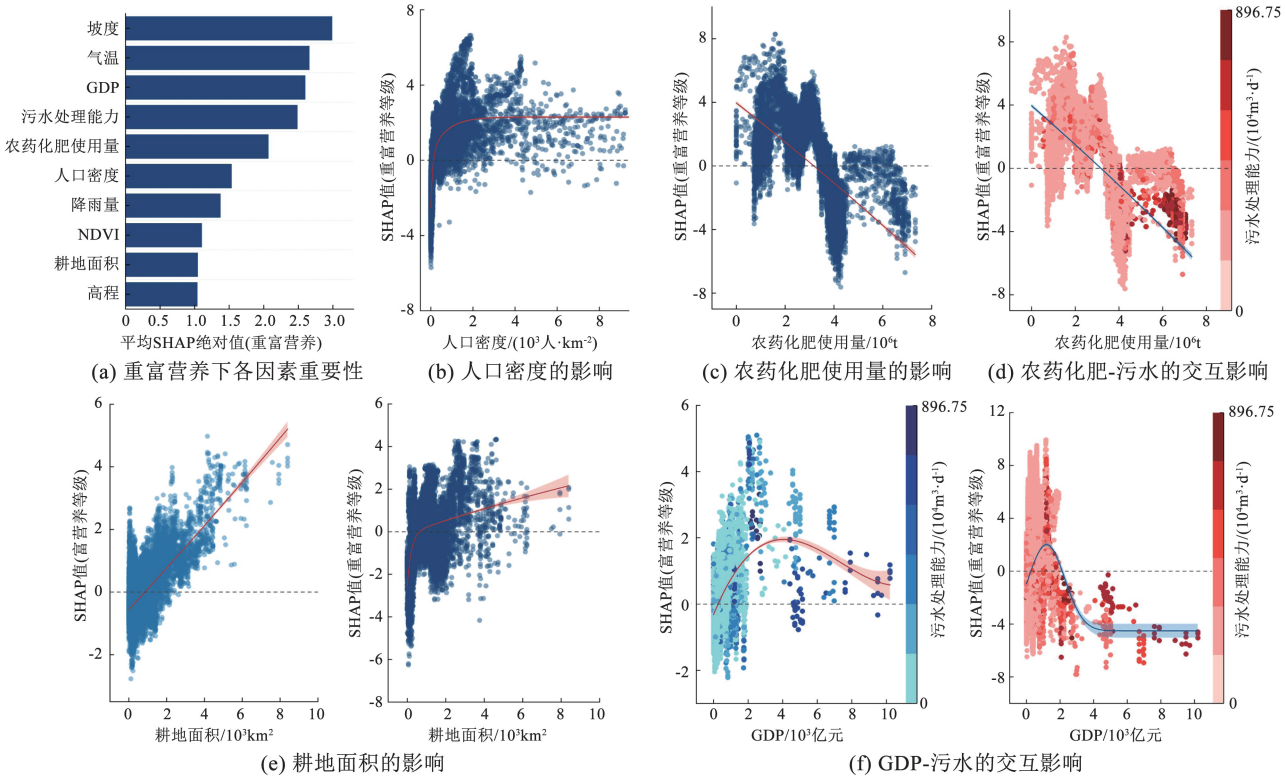


图 7　人为因素对流域富营养化的影响

Fig. 7　Influence of anthropogenic factors on river basin eutrophication

3　结　语

本研究在大空间尺度上评估了中国典型流域富营养化现状，并利用数据增强的方法有效处理了水质领域中类别不平衡问题，帮助机器学习模型更准确地识别影响流域富营养化的关键因素。研究结果

表明,大多数流域都出现了富营养化的现象,且表现出显著的空间异质性。总磷和总氮是富营养化的关键水质因子,与富营养化呈显著的正空间相关性。自然因素在阐明流域富营养化总体变化方面发挥了更大的作用,其中,坡度、温度和高程是3个最关键的因素。然而,随着富营养化等级的提高,人为因素的影响逐渐增强,尤其是人口密度、国内生产总值和污水处理能力。这表明流域富营养化是自然过程与人类活动共同作用的结果,并且随着富营养化程度加剧,人为因素的重要性逐渐凸显。本研究强调改善流域富营养状况应基于空间异质性的特点,并结合内部营养负荷和外部流域特征的变化规律制定针对性的控制政策和修复措施。这一思路不仅为中国流域富营养化的科学认知和治理实践提供了坚实基础,也为开展大尺度水环境评估和跨学科综合管理提供了可借鉴的技术路径与理论参考。

参考文献

[1] YANG Likun, PENG Sen, ZHAO Xinhua, et al. Development of a two-dimensional eutrophication model in an urban lake (China) and the application of uncertainty analysis[J]. Ecological Modelling, 2017, 345: 63. DOI:10.1016/j.ecolmodel.2016.11.014

[2] GENG Mengdie, ZHANG Weizhen, HU Ting, et al. Eutrophication causes microbial community homogenization via modulating generalist species[J]. Water Research, 2022, 210: 118003. DOI:10.1016/j.watres.2021.118003

[3] LE MOAL M, GASCUEL-ODOUX C, MÉNESGUEN A, et al. Eutrophication: a new wine in an old bottle? [J]. Science of the Total Environment, 2019, 651: 1. DOI: 10.1016/j.scitotenv.2018.09.139

[4] MA Tengfei, SHI Lin, LI Zhimin, et al. Eutrophication trend analysis and forewarning model construction of water source reservoirs: Gaozhou Reservoir, China[J]. Ecohydrology, 2022, 15(2): e2371. DOI:10.1002/eco.2371

[5] LIN J L, KARANGAN A, HUANG Yingmin, et al. Eutrophication factor analysis using Carlson trophic state index (CTSI) towards non-algal impact reservoirs in Taiwan[J]. Sustainable Environment Research, 2022, 32(1): 25. DOI:10.1186/s42834-022-00134-x

[6] GUAN Qi, FENG Lian, HOU Xuejiao, et al. Eutrophication changes in fifty large lakes on the Yangtze Plain of China derived from MERIS and OLCI observations[J]. Remote Sensing of Environment, 2020, 246: 111890. DOI: 10.1016/j.rse.2020.111890

[7] YANG Xujie, JIANG Yan, DENG Xuwei, et al. Temporal and spatial variations of chlorophyll a concentration and eutrophication assessment (1987—2018) of Donghu Lake in Wuhan using Landsat images[J]. Water, 2020, 12(8): 2192. DOI: 10.3390/w12082192

[8] WANG Junjie, BOUWMAN A F, VILMIN L, et al. Global inland-water nitrogen cycling has accelerated in the Anthropocene[J]. Nature Water, 2024, 2(8): 729. DOI:10.1038/s44221-024-00282-x

[9] WANG Ziteng, SUN Fuhong, SANG Yiwen, et al. Drivers analysis and future scenario-based predictions of nutrient loads in key lakes and reservoirs of the Yangtze River Catchment[J]. Journal of Environmental Management, 2025, 374: 124078. DOI:10.1016/j.jenvman.2025.124078

[10] YANG Ying, ZHANG Yujian, ZHANG Jixian, et al. Impacts of watershed nutrient loads on eutrophication risks under multiple socio-economic development scenarios in the Pearl River Estuary, China[J]. Journal of Cleaner Production, 2025, 496: 145133. DOI:10.1016/j.jclepro.2025.145133

[11] PREISNER M, NEVEROVA-DZIOPAK E, KOWALEWSKI Z. Analysis of eutrophication potential of municipal wastewater[J]. Water Science and Technology, 2020, 81(9): 1994. DOI:10.2166/wst.2020.254

[12] WIEDMANN T. Eutrophication's neglected drivers[J]. Nature Sustainability, 2018, 1(6): 273. DOI:10.1038/s41893-018-0084-2

[13] ZHU Junjie, YANG Meiqi, REN Z J. Machine learning in environmental research: common pitfalls and best practices[J]. Environmental Science & Technology, 2023, 57(46): 17671. DOI:10.1021/acs.est.3c00026

[14] JIN Xin, ZHANG Wenqiang, LI Simin. Complex responses of suspended particulate matter in eutrophic river and its indicative function in river recovery process[J]. Ecological Indicators, 2020, 115: 106397. DOI:10.1016/j.ecolind.2020.106397

[15] 李祚泳, 汪嘉杨, 郭淳. 富营养化评价的对数型幂函数普适指数公式[J]. 环境科学学报, 2010, 30(3): 664
LI Zuoyong, WANG Jiayang, GUO Chun. A universal exponential formula with logarithmic power function for eutrophication assessment[J]. Acta Scientiae Circumstantiae, 2010, 30(3): 664. DOI:10.13671/j.hjkxxb.2010.03.025

[16] CARLETTI M, BARBERA E, FILIPPINI F, et al. Effect of ammonium/nitrate ratio on microalgae continuous cultures: species-specificity of nutrient uptake and modelling perspectives[J]. Journal of Water Process Engineering, 2024, 58: 104762. DOI: 10.1016/j.jwpe.2023.104762

[17] WANG Xiaonan, LI Ji, CHEN Jin, et al. Water quality criteria of total ammonia nitrogen (TAN) and un-ionized ammonia (NH3-N) and their ecological risk in the Liao River, China[J]. Chemosphere, 2020, 243: 125328. DOI: 10.1016/j.chemosphere.2019.125328

[18] ZHANG Yang, LIU Liang, ZHANG Shenghong, et al. Monitoring and warning for ammonia nitrogen pollution of urban river based on neural network algorithms[J]. Analytical Sciences, 2024, 40(10): 1867. DOI:10.1007/s44211-024-00622-7

[19] EDWARDS T M, PUGLIS H J, KENT D B, et al. Ammonia and aquatic ecosystems: a review of global sources, biogeochemical cycling, and effects on fish[J]. Science of the Total Environment, 2024, 907: 167911. DOI:10.1016/j.scitotenv.2023.167911

[20] 沈晓峰, 孙进, 顾华, 等. 基于XGBoost算法的变压器油中溶解气体分析[J]. 电力大数据, 2022, 25(9): 20
SHEN Xiaofeng, SUN Jin, GU Hua, et al. Transformer dissolved gas analysis based on XGBoost algorithm[J]. Power Systems and Big Data, 2022, 25(9): 20. DOI: 10.19317/j.cnki.1008-083x.2022.09.001

[21] DING Yi, ZHU Hongyang, CHEN Ruyun, et al. An efficient AdaBoost algorithm with the multiple thresholds classification[J]. Applied Sciences, 2022, 12(12): 5872. DOI: 10.3390/app12125872

[22] MISHINA Y, MURATA R, YAMAUCHI Y, et al. Boosted random forest[J]. IEICE Transactions on Information and Systems, 2015, E98.D(9): 1630. DOI:10.1587/transinf.2014OPP0004

[23] 曹贞洋, 龚敏, 吴昊骏, 等. 基于GA-LightGBM算法的TBM掘进参数与岩体等级关系[J]. 哈尔滨工业大学学报, 2025,

57(7)：22
CAO Zhenyang, GONG Min, WU Haojun, et al. Relationship between TBM tunneling parameters and rock mass classification based on GA-LightGBM algorithm[J]. Journal of Harbin Institute of Technology, 2025, 57(7): 22

[24] KARTHICK K, KRISHNAN S, MANIKANDAN R. Water quality prediction: a data-driven approach exploiting advanced machine learning algorithms with data augmentation[J]. Journal of Water and Climate Change, 2024, 15(2): 431. DOI:10.2166/wcc.2023.403

[25] 辛永，刘燕秋，黄文思，等. 基于多模型融合的客户投诉风险预测方法[J]. 电力大数据，2018，21(11)：31
XIN Yong, LIU Yanqiu, HUANG Wensi, et al. Customer complaint risk prediction method based on multi-model fusion[J]. Power Systems and Big Data, 2018, 21(11): 31. DOI:10.19317/j.cnki.1008-083x.2018.11.006

[26] GOSAIN A, SARDANA S. Handling class imbalance problem using oversampling techniques: a review[C]//International Conference on Advances in Computing, Communications and Informatics (ICACCI). New York: IEEE, 2017: 79

[27] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associate, 2017: 4768

[28] 徐达梁，徐杭镔，靳心瑶，等. 基于机器学习的纳滤膜预测筛选模型构建与评估[J]. 哈尔滨工业大学学报，2024，56(6)：8
XU Daliang, XU Hangbin, JIN Xinyao, et al. Construction and evaluation of predictive screening model for nanofiltration membranes based on machine learning[J]. Journal of Harbin Institute of Technology, 2024, 56(6): 8

[29] CHARLTON M B, BOWES M J, HUTCHINS M G, et al. Mapping eutrophication risk from climate change: future phosphorus concentrations in English rivers[J]. Science of the Total Environment, 2018, 613: 1510. DOI:10.1016/j.scitotenv.2017.07.218

[30] TILAHUN A B, DÜRR H H, SCHWEDEN K, et al. Perspectives on total phosphorus response in rivers: examining the influence of rainfall extremes and post-dry rainfall[J]. Science of the Total Environment, 2024, 940: 173677. DOI:10.1016/j.scitotenv.2024.173677

[31] MOORE P A. Nutrient transformations in sediments as influenced by oxygen supply[J]. Journal of Environmental Quality, 1992, 21: 387. DOI:10.2134/jeq1992.00472425002100030014x

[32] ZHENG Xiaolei, LIU Gaoxiang, YANG Wei, et al. Dominant contribution of a lake′s internal pollution to eutrophication during rapid urbanization[J]. Bulletin of Environmental Contamination and Toxicology, 2021, 107: 904. DOI:10.1007/s00128-021-03236-1

[33] ZHOU Jian, HAN Xiaoxia, BROOKES J D, et al. High probability of nitrogen and phosphorus co-limitation occurring in eutrophic lakes[J]. Environmental Pollution, 2022, 292: 118276. DOI:10.1016/j.envpol.2021.118276

[34] WANG Feier, WANG Yixu, ZHANG Kai, et al. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation[J]. Environmental Research, 2021, 202: 111660. DOI:10.1016/j.envres.2021.111660

[35] HUANG Jiacong, ZHANG Yinjun, BING Haijian, et al. Characterizing the river water quality in China: recent progress and on-going challenges[J]. Water Research, 2021, 201: 117309. DOI:10.1016/j.watres.2021.117309

[36] WURTSBAUGH W A, PAERL H W, DODDS W K. Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum[J]. Wiley Interdisciplinary Reviews-Water, 2019, 6(5): e1373. DOI:10.1002/wat2.1373

[37] ARHEIMER B, ANDRÉASSON J, FOGELBERG S, et al. Climate change impact on water quality: model results from southern Sweden[J]. AMBIO: A Journal of the Human Environment, 2005, 34(7): 559. DOI:10.1579/0044-7447-34.7.559

[38] XIA Xinghui, ZHANG Sibo, LI Siling, et al. The cycle of nitrogen in river systems: sources, transformation, and flux[J]. Environmental Science-Processes & Impacts, 2018, 20: 863. DOI:10.1039/C8EM00042E

[39] SMITH J S, WINSTON R J, TIRPAK R A, et al. The seasonality of nutrients and sediment in residential stormwater runoff: implications for nutrient-sensitive waters[J]. Journal of Environmental Management, 2020, 276: 111248. DOI:10.1016/j.jenvman.2020.111248

[40] WANG Ai, YANG Dawen, TANG Lihua. Spatiotemporal variation in nitrogen loads and their impacts on river water quality in the upper Yangtze River basin[J]. Journal of Hydrology, 2020, 590: 125487. DOI:10.1016/j.jhydrol.2020.125487

[41] LYCHE SOLHEIM A, GLOBEVNIK L, AUSTNES K, et al. A new broad typology for rivers and lakes in Europe: development and application for large-scale environmental assessments[J]. Science of the Total Environment, 2019, 697: 134043. DOI:10.1016/j.scitotenv.2019.134043

[42] SHANG Wei, JIN Shuanggen, HE Yang, et al. Spatial-temporal variations of total nitrogen and phosphorus in Poyang, Dongting and Taihu Lakes from Landsat-8 data[J]. Water, 2021, 13(12): 1704. DOI:10.3390/w13121704

[43] OCKENDEN M C, DEASY C E, BENSKIN C M H, et al. Changing climate and nutrient transfers: evidence from high temporal resolution concentration-flow dynamics in headwater catchments[J]. Science of the Total Environment, 2016, 548: 325. DOI:10.1016/j.scitotenv.2015.12.086

[44] HU Jingyi, OUYANG Wei, HOU Congyu, et al. Effects of urban stormwater pollution on watershed diffuse loads under extreme precipitation conditions[J]. Journal of Hydrology, 2025, 654: 132802. DOI:10.1016/j.jhydrol.2025.132802

[45] XIA Xinghui, WU Qiong, MOU Xinli, et al. Potential impacts of climate change on the water quality of different water bodies[J]. Journal of Environmental Informatics, 2015, 25(2): 85. DOI:10.3808/jei.201400263

[46] MACCOUX M J, DOVE A, BACKUS S M, et al. Total and soluble reactive phosphorus loadings to Lake Erie[J]. Journal of Great Lakes Research, 2016, 42(6): 1151. DOI:10.1016/j.jglr.2016.08.005

[47] CUI Gannan, BAI Xinyu, WANG Pengfei, et al. Mechanism of response of watershed water quality to agriculture land-use changes in a typical fuel ethanol raw material planting area:a case study on Guangxi Province, China[J]. International Journal of Environmental Research and Public Health, 2022, 19(11): 6499. DOI:10.3390/ijerph19116499

[48] WEI Huaibin, QIU Haojie, LIU Jing, et al. Evaluation and source identification of water pollution[J]. Ecotoxicology and Environmental Safety, 2025, 289: 117499. DOI:10.1016/j.ecoenv.2024.117499

[49] BAK M P, KROEZE C, JANSSEN A B G, et al. Modelling future coastal water pollution: impacts of point sources, socio-economic developments & multiple pollutants[J]. Environmental Research Letters, 2024, 19(6): 064028. DOI:10.1088/1748-9326/ad489c

（编辑　刘　彤）