

降雨径流驱动下水体总氮关联特征变化与监测策略优化

郑凯丰¹, 崔国韬^{1,2}, 韩玉国³, 杜新忠⁴

(1. 中山大学 地理科学与规划学院, 广东 广州 510275; 2. 中山大学 粤北岩溶区碳水耦合野外科学观测研究站, 广东 广州 510275; 3. 北京林业大学 水土保持学院, 北京 10008; 4. 中国农业科学院农业资源与农业区划研究所 农业农村部面源污染控制重点实验室/北京昌平土壤质量国家野外科学观测研究站/北方干旱半干旱耕地高效利用全国重点实验室, 北京 100081)

摘要:【目的】为应对白洋淀流域频发的农业面源污染, 亟需揭示降雨径流下总氮(TN)的迁移机理并优化监测策略。【方法】以流域农业区(北河店站)为研究对象, 引入城市区(码头站)作为对比参照, 耦合随机森林(RF)模型与SHAP可解释性方法, 利用流量及常规易测水质指标, 解析TN的非线性关联特征, 评估不同监测频率对预测精度的影响。【结果】(1)在全年尺度下, RF模型在两站点均表现出高预测性, 纳什系数(NSE)分别为0.981和0.972。北河店站TN与电导率和水温呈协同关联, 码头站TN则与水温密切相关。(2)在降雨径流事件中, 农业区呈径流主导模式, 高流量事件后TN浓度存在48—72小时滞后正响应, 而短期流量变化则兼具冲刷与稀释效应; 城市区TN浓度关联特征发生转变, 水温在事件中不再主导, 而与电导率强正相关, 高流量主要为稀释作用。(3)降雨径流事件中监测频率对模型性能影响大, 日尺度模型预测能力显著下降(相较小时尺度NSE降幅超过42%), 码头站日尺度模型无法有效预测(NSE<0); 而在非降雨期两个尺度模拟精度相近(<1%)。【结论】明确了不同土地利用下TN的差异化关联特征, 揭示了高频监测在关键污染事件中的重要作用, 为构建流域面源污染防控及实时监测与预警提供科学依据。

关键词: 白洋淀流域; 总氮; 随机森林; SHAP; 降雨径流

中图分类号: TV697

文献标识码: A

人工氮肥的生产与使用剧烈改变了自然系统的氮循环过程, 中国使用了约全球三分之一的氮肥^[1]。氮肥的过量使用可造成水生生态系统中藻类过度生长, 引发富营养化和酸化, 导致水体水华、大范围缺氧等严重的水环境、水生态问题, 危害水生生物和人类健康^[2-6]。白洋淀作为华北平原最大的淡水湖泊具有重要生态价值与功能, 但是流域内农业面源与城镇生活点源交织, 其水体总氮(Total Nitrogen, TN)长期作为首要超标因子, 成为制约流域水质达标的关键瓶颈。然而, 该区域降雨径流过程短促且强度大, 传统低频监测难以捕捉动态变化的总氮浓度, 导致对流域内TN迁移变化规律认识存在盲区。因此, 解析白洋淀TN的高频动态特征

及其与常规监测指标的关联, 对流域水体总氮的精准监测与精细化管控具有重要意义。

地表面源污染中水体总氮的实时监测存在采样困难、成本高等问题^[7], 研究常采用污染物估算及模型进行预测^[4-6]。例如在流域尺度的研究常采用模型模拟方法, 利用易于获取的数据对污染物进行模拟预测。例如, Zhang L等^[8]基于HadCM3与SWAT模型模拟了石头口门水库流域面源污染负荷, 发现 $\text{NH}_4^+\text{-N}$ 负荷显著下降而总磷负荷呈不显著增加趋势; 侯玥等^[9]利用SWAT模型预测了气候变化情景下淮河上游流域未来地表水的氮素排放特征, 发现面源污染压力将进一步增加。基于物理机制和水体形态的模型有预测与机理解释价值, 但

收稿日期/Received: 2025-12-3

修回日期/Revised: 2026-02-04

基金项目: 国家重点研发计划资助项目(2024YFD1700800); 国家自然科学基金资助项目(42301012); 中山大学高校基本科研业务(24qnp020)。

第一作者: 郑凯丰(2002—), 男, 硕士研究生, 主要从事水文模型研究。E-mail: zhengkf@mail2.sysu.edu.cn。

通信作者: 崔国韬(1989—), 男, 副教授, 博士, 主要从事水文水资源研究。E-mail: cuigt@mail.sysu.edu.cn。

易受物理参数、数据缺失制约。而机器学习方法融合多元数据,构建非线性关系,能降低对专用参数的依赖,提升精度、拓展应用,是水质数据稀缺情况下的有效预测模型。

然而,机器学习模型存在“黑箱”特性,限制了其在机理解析中的应用。通过结合可解释性方法对非线性映射关系进行透视,可在保障模型预测性能的同时,有效弥补其机理不明的短板,提升模拟结果的可信度。例如,He B 等^[10]利用人工神经网络模型预测地表水体 TN,且通过敏感性分析解释了各个输入变量与 TN 之间的关系;肖焱等^[11]使用机器学习模型对地下水水质进行预测,讨论了不同滞后期下气象数据对预测性能的影响;Ha N-T 等^[12]使用随机森林模型(Random Forest, RF)对地表水体氮磷进行预测,并通过特征重要性分析揭示了关键水质变量如总悬浮固体、浊度等对负荷变化的影响。因此,机器学习与可解释性方法的结合,可为水体总氮的精准预测及其过程机理解析提供技术支撑。

建设用地占比和农业强度是影响水体总氮浓度变化的关键预测因子^[13]。然而,当前研究多关注静态土地利用对水质的长期影响,对“降雨-径流”这一高动态过程中不同景观格局下 TN 特征研究相对较少。在径流事件的影响下,农业面源的冲刷机制与城市的显著差异会影响常规水质指标与 TN 的关联模式,而固定频率监测可能无法精细地捕捉复杂污染过程。若忽略这种时空异质性,将难以构建具有普适性和高保真度的水质预警体系。鉴于此,本研究以白洋淀流域两个水质监测站点为研究对象,耦合 RF 模型与可解释性分析方法,使用流量与易监测水质指标(水温、pH、溶解氧、电导率、浊度、高锰酸盐指数)对 TN 浓度进行模拟。本研究有助于深化对农业面源污染和城市降雨汇流背景下 TN 差异化关联特征的认识,为建立 TN 实时预

测预警系统提供支撑,对优化水质监测方案和面源污染控制有重要实践价值。

1 研究方法 with 数据

1.1 研究区域

白洋淀流域属海河流域大清河子流域,西北部为山区,其他地区多为平原,呈西北高东南低地势,如图 1 所示。流域属温带大陆性气候,四季分明,雨热同期,其中东南部平原为主要的农业种植区,典型的农作物种植模式为小麦-玉米轮作,是氮施用、流失的主要地区。由地表径流进入水体的农田氮素是白洋淀流域农业面源污染的主要原因^[14]。选取代表性站点北河店、码头用于比较流域不同土地利用类型下的非点源污染特征,如表 1 所示。其中北河店上游集水区 60%为农田,契合东南部农业区属性,码头集水区 40%为城市建设用地,对应城市化区域,可分别代表农业面源污染区与城市汇流区。

表 1 北河店和码头站点集水区的土地利用类型占比

Table 1 Land use proportion of catchment area for the Beihedian and Matou stations.

| 土地利用类型 | 土地利用类型占比(%) | |
|--------|-------------|-------|
| | 北河店站 | 码头站 |
| 落叶阔叶林 | 1.11 | 11.05 |
| 混交林 | 0.03 | 0.03 |
| 密闭灌木林 | 0.86 | 3.75 |
| 木质稀树草原 | 0.03 | 2.09 |
| 稀树草原 | 8.97 | 16.73 |
| 草地 | 17.32 | 15.89 |
| 永久湿地 | 0.39 | 0.00 |
| 农田 | 62.28 | 10.16 |
| 城市和建成区 | 9.01 | 40.27 |
| 水体 | 0.00 | 0.03 |

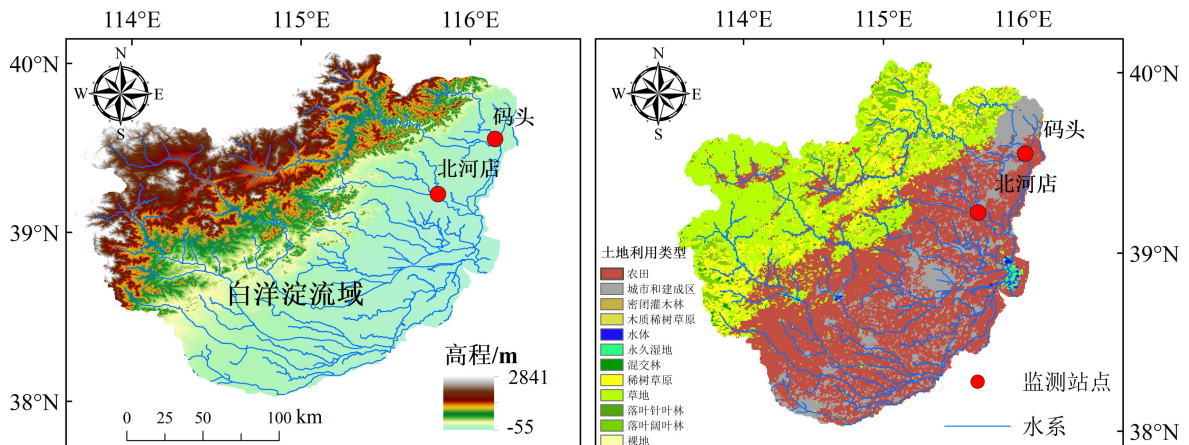


图 1 研究区示意图

Fig. 1 Location of the study region

1.2 数据来源与处理

本研究收集 2024 年 1 月 1 日至 2024 年 12 月 31 日国家地表水水质数据发布系统 (<https://www.cnemc.cn/sss/>) 的(北河店、码头)自动监测断面水质监测数据(如表 2 所示)与 GEOGLOWS 全球水文模型系统 (<https://geoglows.ecmwf.int/>) 1 小时分辨率流量数据。为适用于白洋淀流域,有效表征流域径流特征,根据流域内历史水文站径流观测数据,基于 MFDC-QM (Mass-conservative Frequency-Distribution Calibration with Quantile Mapping) 对 GEOGLOWS 流量进行校准^[15]。

表 2 水质水量数据及来源

Table 2 Water quality and runoff data and sources.

| 特征 | 单位 | 来源 |
|--------|-------------------|---|
| 水温 | °C | |
| pH | - | |
| 溶解氧 | mg/L | |
| 电导率 | μS/cm | 国家地表水水质数据发布系统 (https://www.cnemc.cn/sss/) |
| 高锰酸盐指数 | mg/L | |
| 浊度 | NTU | |
| 总氮 | mg/L | |
| 流量 | m ³ /s | 经白洋淀水文观测校准后的 GEOGLOWS 一小时数据 |

1.3 研究方法

选用随机森林模型并结合可解释性方法,以流量、水温、pH 等易观测变量作为模型输入特征,对总氮进行模拟预测,验证其预测精度并解析输入特征与 TN 的关联性,揭示总氮在不同时期(全年、降雨径流)下的关键关联特征。构建随机森林模型,划分训练、测试样本并训练测试模型,验证其精度并分析重要影响因子。以 95 分位数为界限识别并记录降雨径流事件,提取事件样本训练模型,并结合 SHAP (SHapley Additive exPlanations)、偏依赖分析 (Partial Dependence Plot, PDP)、个体条件期望 (Individual Conditional Expectation, ICE) 方法解释输入特征对总氮的影响。对比基于小时级数据及日尺度数据的模型预报精度,分析不同事件下(降雨径流时期、非降雨径流时期)的精度差异,探究自动监测的水质指标反演 TN 变化过程的可行性及监

测方案策略。

1.3.1 随机森林模型

RF 模型是一种基于决策树集成的监督学习算法。该方法通过 bootstrap 抽样生成多个训练数据子集,并在每个节点分裂时随机选择特征子空间,构建多棵去相关化的回归决策树。在预测阶段输出所有树预测值的算术平均。随机森林模型能够灵活处理离散、独立的样本,在预测连续性差、获取困难的水质指标方面优势明显^[16]。RF 模型构建主要包括:

1) 模型训练与超参数优化:根据水质指标有值时间点筛选流量数据并统一为模型样本集。使用五折交叉验证训练测试以优化模型超参数,每一折训练样本与测试样本比例为 4:1,优化目标为五折交叉验证的评价指标。使用 Optuna 超参数优化框架^[17],确定决策树数量、树深度等超参数;

2) 模型精度验证与特征重要性:基于优化后的超参数,在样本集中随机选取 20% 的样本作为测试集,实验模型预测性能,并基于平均不纯度减少 (Mean Decrease Impurity, MDI) 方法,通过统计节点分裂时不纯度下降数值来计算特征重要性。

1.3.2 可解释性分析方法

为全面解译流量、各类水质指标对 TN 预测的影响,建立了一个多层次的可解释性分析框架。该框架综合运用了 SHAP 值法、偏依赖图和个体条件期望三种方法,从“全局-局部”和“平均-个体”两个维度系统地揭示模型的内部决策机制,揭示输入特征在模型中的重要性,明晰特征的正向或负向影响以及在不同条件下 TN 可能的预测期望值。

SHAP 值法通过考察不同特征组合对模型输出的边际贡献来衡量每个特征的重要性^[18],用于评估特征的全局重要性和局部贡献。SHAP 值的计算基于 Shapley 值,显示集合中特征对目标变量的影响,其计算公式为:

$$\text{Shapley}_i = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \quad (1)$$

式中: N 为所有特征的集合; S 为不包含特征 j 的集合; $f(S \cup \{j\})$ 为使用子集 S 与特征 j 的并集作为模型输入时的性能; $f(S)$ 为仅使用子集 S 作为模型输入时的性能; $| \cdot |$ 为集合中的特征数量。

为明晰特定特征对 TN 预测的平均边际效应,PDP 通过固定其他特征并改变目标特征的取值,

计算模型预测的平均变化趋势,直观地表达预测模型依赖于特定变量的平均依赖关系^[19]。然而,PDP的平均趋势可能掩盖个体样本间的异质性。为此,进一步引入 ICE,为每个样本绘制一条独立曲线,展示各样本的预测值随目标特征变化特征。其中 PDP 与 ICE 计算公式分别为:

$$\widehat{f}_{x_s}(x_s) = E_{x_c}[\widehat{f}_{x_s}(x_s, X_c)] = \int \widehat{f}(x_s, x_c) dp(x_c) \quad (2)$$

$$\widehat{f}_{x_s}^{(k)}(x_s) = \widehat{f}(x_s, x_c^{(k)}) \quad (3)$$

式中: $\widehat{f}()$ 为训练完成的 RF 模型; x_s 为待分析的目标特征; x_c 为模型中除目标特征外的其他输入特征; X_c 为其他特征的随机变量; $E_{x_c}[]$ 是其他特征的期望; $p(x_c)$ 为概率分布; $x_c^{(k)}$ 为第 k 个样本的其他特征取值。

1.3.3 模型评价指标

为定量地评价模型的预测性能,选取了纳什效率系数 (Nash-Sutcliffe Efficiency, NSE) 与均方根误差 (Root Mean Square Error, RMSE) 作为评价指标。NSE 是水文模型应用广泛的指标,用于评估模型模拟值与观测值序列的拟合优度。其取值范围为 $(-\infty, 1]$, NSE 越接近 1,表明模型模拟结果与实测值越吻合,模型性能越好。RMSE 用于衡量模型预测值与观测值之间的偏差,即模型的预测精度。RMSE 越小,表明模型误差越小,预测精度越高。

NSE 与 RMSE 计算公式分别为:

$$NSE = 1 - \frac{\sum_{i=1}^n (\text{obs}_i - \text{pre}_i)^2}{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{obs}_i - \text{pre}_i)^2}{n}} \quad (5)$$

式中:obs 和 pre 分别为观测值和预报值;下标 i 为第 i 个时间步长; $\overline{\text{obs}}$ 和 $\overline{\text{pre}}$ 分别为观测值和预报值的平均值; n 表示样本总数。

2 研究结果

2.1 全年尺度的预测模型特征重要性与精度评估

在全年尺度上,采用 RF 模型评估北河店站点和码头站点整体的特征重要性和模型预测精度,结果如图 2、3 所示。从特征重要性排序来看,两个监测站点呈现出明显差异。如图 2 所示,北河店站点中,电导率的特征重要性最高,其次为水温,而码头站点则以水温的重要性最高,溶解氧次之。此外,流量在两个站点的重要性排序均较低。时间特征(年积日、月、日等)则在两个站点均显示出一定的重要性,其中年积日重要性最高。

如图 3 所示,两个站点的随机森林模型均表现出良好的预测能力。北河店站点的 NSE 为 0.981, RMSE 为 0.230 mg/L; 码头站点的 NSE 为 0.972, RMSE 为 0.223 mg/L。实测值与预测值散点图(图 2b, d)显示,两个站点的预测值均密布在 1:1 红线附近,表明模型能够准确预测 TN。

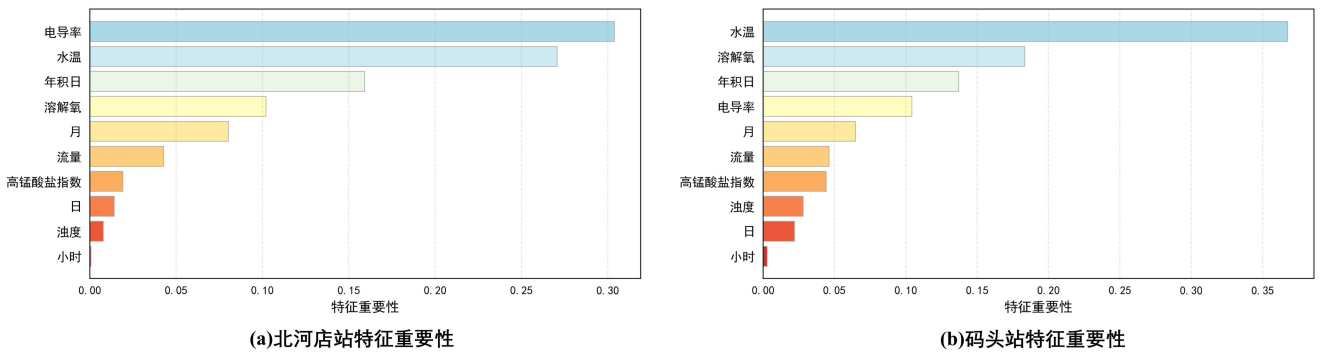


图 2 基于随机森林模型的整体特征重要性
Fig. 2 Overall feature importance based on RF model.

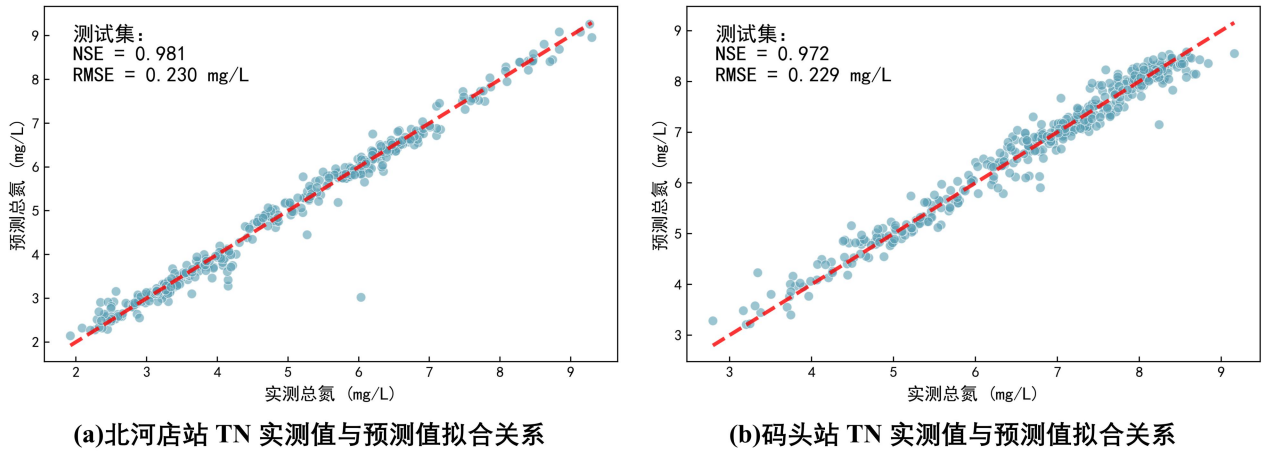


图 3 总氮预测结果

Fig. 3 TN prediction result.

2.2 降雨径流事件下总氮与水文指标关联性

基于流量 95 分位数阈值识别各站点降雨径流事件进行识别和提取。北河店和码头站点的流量均呈现出明显的季节性波动特征，降雨径流事件集中发生在汛期(7-9月)，其中北河店站点识别出 3 场事件，码头站点有 4 场事件。基于识别的降雨径流事件，对模型特征重要性以及可解释性进行分析。如图 4、5 结果所示，反映了两个站点不同降雨径流事件中各输入特征对总氮预测的重要性分布。为捕捉流量对总氮浓度

影响的时间滞后效应，加入滞后流量作为输入特征。其中“流量滞后”指当前 TN 样本对应时刻前若干小时的流量。该特征设置旨在反映径流过程中流量变化对水质影响可能存在的时间延迟特性。北河店站点呈现出明显的滞后流量主导特征。在三个径流事件中，60—72h 流量滞后特征始终具有较高重要性，而其他水质指标和近期流量的重要性则相对较低。而在码头站点(b)，水质相关指标特征的重要性更为突出。

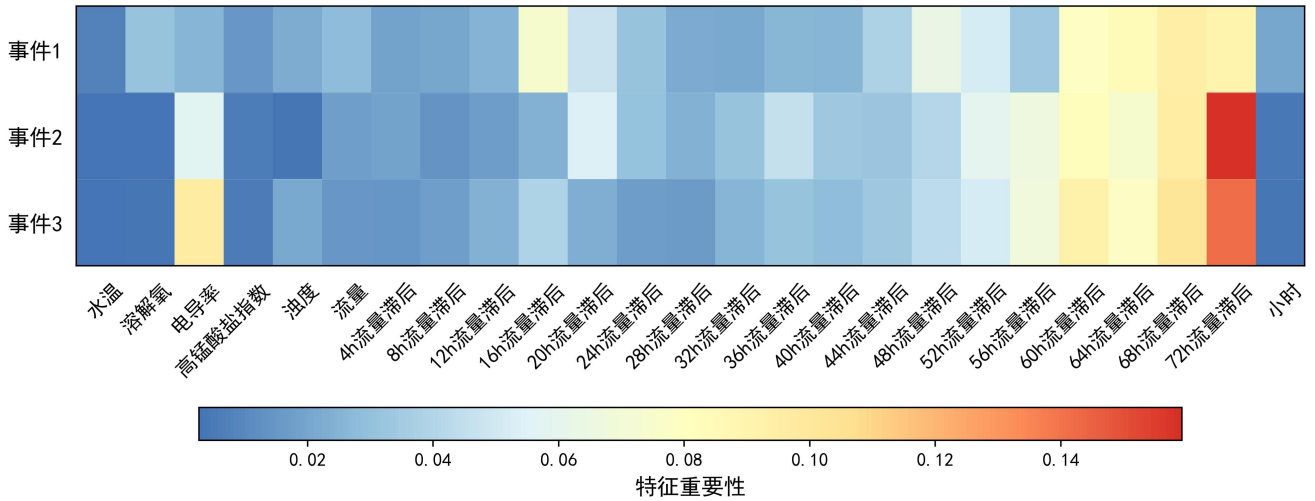


图 4 降雨径流事件中北河店站输入特征的重要性

Fig. 4 Importance of input features during rainfall-runoff events.

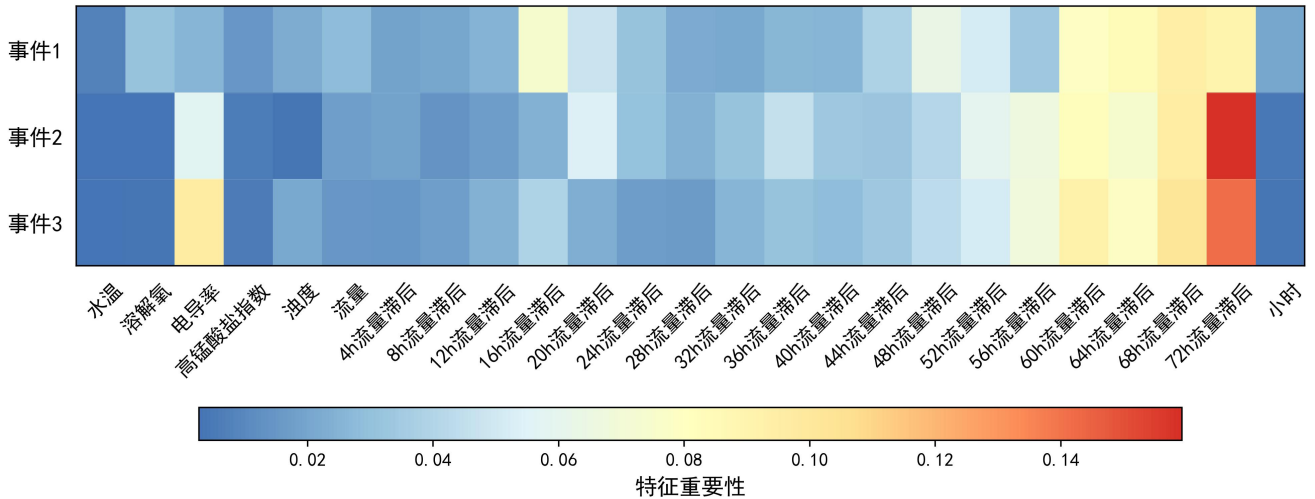


图 5 降雨径流事件中码头站输入特征的重要性

Fig. 5 Importance of input features during rainfall-runoff events.

选取两个站点径流峰值最高的典型事件(北河店的事件 3 和码头的 事件 4), 使用 SHAP 方法计算输入特征对 TN 预测的贡献分析, 其结果如图 6 所示。对于北河店站点, 较长期流量滞后(如 48—72h 流量滞后)下, 洪峰高流量对 TN 浓度表现出显著正向贡献, 事件初期、末期的相对低流量对 TN 的贡献则兼具正向与负向贡献, 表明高流量对 TN 起促进作用, 低流量同时存在促进与抑制效应。相比之下, 较短期流量滞后(如 4—16h 流量滞后)下, 洪峰高流量与相对低流量对 TN 的影响较为复杂, 兼具正向促进与负向抑制作用。水质指标对 TN 的影响不显著, 其中电导率与 TN 为正相关, 即高水平的电导率表现为促进 TN 上升。此外, SHAP 分析图仅展示了降雨径流事件中排名前 15 的重要特征, 北河店站点除电导率外, 其他水质指标因贡献度较低未在图中呈现, 表明该站点 TN 变化与水文过程变化关联性更强。

码头站点的 SHAP 解释结果与北河店站点中径流滞后特征主导的模式不同, 该站点以水质指标为关键主导因素。其中电导率、水温的特征贡献最高, 电

导率升高对 TN 有显著促进作用, 水温同样表现出促进效应。流量在码头站点的作用方向与北河店站点一致, 但贡献程度低。此外, 浊度、溶解氧等水质指标也有一定贡献。

使用 PDP 和 ICE 分析两个典型事件中 TN 随重要性最高的 72h 流量滞后变化下的条件期望值, 结果如图 7 所示。对于北河店站点, 贡献最高的 72h 流量滞后表现出对总氮浓度的促进作用。从 PDP 曲线可以看出, 当滞后流量超过约 $55\text{m}^3/\text{s}$ 后, TN 期望值开始呈现非线性急剧上升, 表明径流冲刷效应存在一个启动阈值。ICE 结果也显示出个体样本的差异, 其中低值 TN 样本对流量滞后变化呈现出更明显的上升, 而高值 TN 样本相对平缓。码头站点中, PDP 显示贡献最高的 72h 流量滞后在流量初步上升时, TN 出现升高。ICE 结果表明, 在流量上升至约 $380\text{m}^3/\text{s}$ 时, 高浓度的 TN 期望值呈现轻微下降的趋势, 这可能反映了极高流量下稀释作用开始对高浓度 TN 产生一定影响。

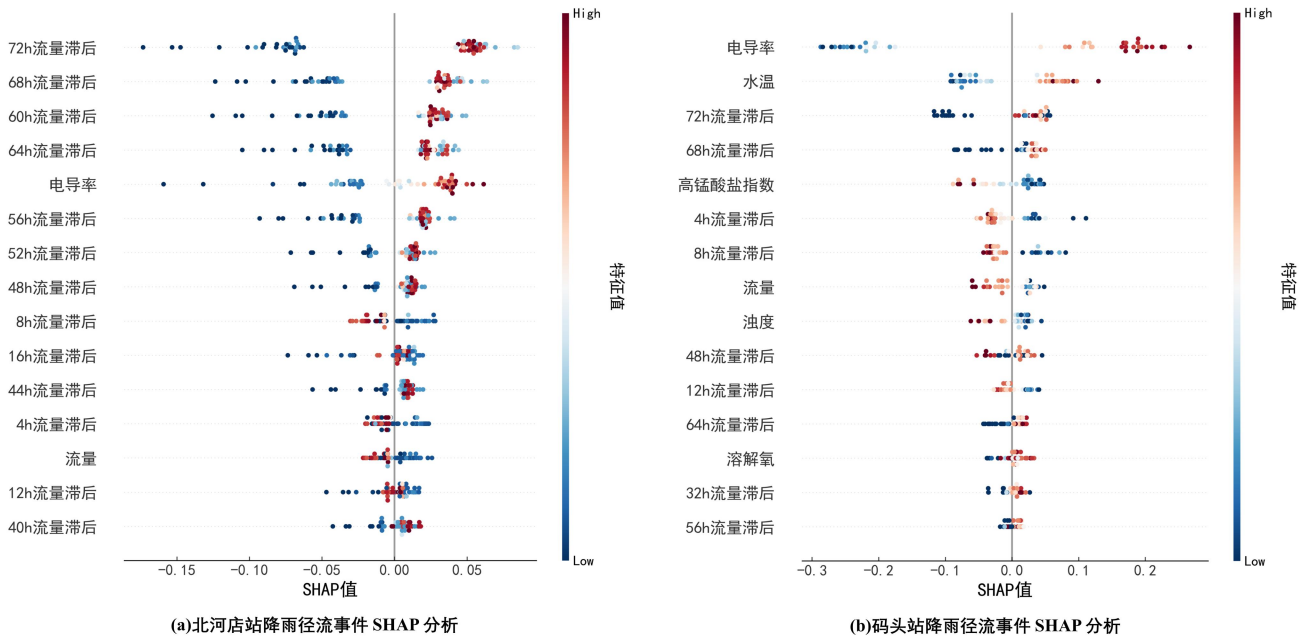


图 6 最高峰值径流事件 SHAP 解释图

Fig. 6 SHAP values for the highest peak runoff event.

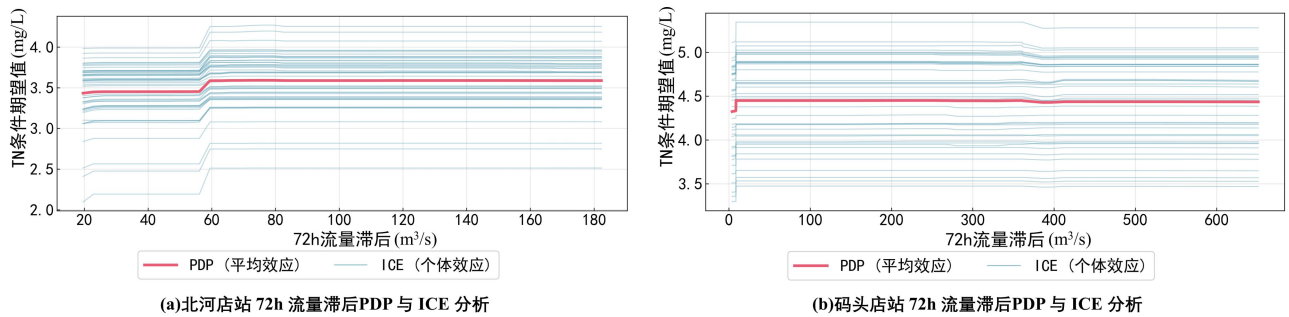


图 7 最高峰值径流事件 PDP 与 ICE 分析

Fig. 7 PDP and ICE results for the highest peak runoff event.

2.3 同时间频率下总氮预测性能差异

使用小时尺度和日尺度样本训练 RF 模型, 解析监测数据的时间分辨率对模型性能的影响, 研究结果如图 8 所示。结果表明, 发生降雨径流事件时预测精度对时间分辨率具有高度敏感性。在降雨径流时期, 两个站点均表现出小时尺度与日尺度模型精度的显著差异。对于北河店站点, 小时尺度模型的 NSE 达到 0.929, 而日尺度模型降至 0.536, 精度差异为 0.393, 表明日尺度模型无法精准刻画径流过程中的峰值波动。码头站点的差异更为显著, 其中小时尺度模型 NSE 为 0.702, 日尺度模型降至 -0.222, 精度差异高达

0.924, 其中 NSE 小于 0 表明该模型在降雨径流期的预测偏差已超过实测值的方差, 即日尺度模型完全无法捕捉码头站降雨径流过程中 TN 的脉冲变化, 其预测结果已不可信。在非降雨径流时期, 两个站点的小时尺度与日尺度模型精度接近, 均保持较高的预测精度。北河店站点两种时间尺度下的 NSE 分别为 0.987 和 0.979, 仅相差 0.008。码头站点的 NSE 分别为 0.946 和 0.948, 仅相差 0.002。这表明在平稳的基流期, 日尺度监测足以满足精准预测的需求, 而在降雨径流关键期, 高频监测则是保障数据质量与模型可靠性的必要条件。

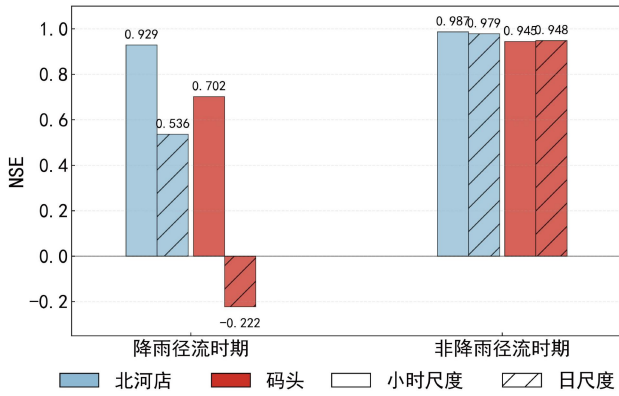


图 8 不同时间频率与降雨径流事件发生时模型预测性能差异

Fig. 8 Model-performance differences under different temporal resolutions and rainfall-runoff conditions.

3 分析讨论

3.1 氮全年变化归因差异

白洋淀两个监测站点在整体特征重要性上表现出显著差异,北河店站点电导率为最关键的影响因子,水温次之。码头站点则是水温作为主要表征指标。Han Y 等^[20]的研究表明,在气候变化背景下,气温通过影响水温间接导致白洋淀流域 TN 发生显著变化。气温通过直接调控水温,改变水体溶解氧溶解度及微生物酶活性,会直接影响着硝化与反硝化过程,从而影响水体 TN 浓度的变化^[21]。该温度环境因子的强控制作用,在本研究两站点的水温高特征重要性得以一定体现。北河店站位于农业区下游,农业活动中化肥施用不仅增加氮素输入,而且提高水体离子浓度。电导率作为水体离子浓度的指标,对该区域 TN 变化具有高度的表征能力,使其成为关键影响因子。码头站位于城市建设区下游,受点源污染影响,输入相对持续且稳定。因此,码头站 TN 浓度变化更多依赖于上游输入与水体自身的生物化学转化,而水温作为调控微生物活性和生化反应过程的关键特征,在码头站中表现出更高的重要性。此外,时间特征(年积日与月)在两个站点也显示出较高重要性,而华北地区降雨与农业活动的季节性耦合是驱动 TN 年内季节性变动的核心机制^[22]。

模型精度评估显示,两个站点 RF 模型均表现出优异的预测性能,表明通过 RF 模型能够有效利用常规水质指标和流量准确预测 TN 浓度。TN 传

统采样和实验室检测耗时长、成本高,阻碍高频连续监测的广泛应用。通过构建 RF 机器学习模型,可基于在线监测的常规参数实时估算 TN,降低监测成本并获取高时间分辨率的 TN 序列。例如, Wu 等^[23]采用机器学习算法基于多参数传感器实现农业区 TN 实时预测。

3.2 降雨径流事件中总氮与水文指标关联性

白洋淀流域两个水质监测站点上游河道附近的土地利用类型存在显著的差异。北河店站作为白洋淀流域典型农业区的下游监测站,其水体水质在降雨径流事件期显然更受农业面源污染的影响。码头站上游河道位于城市建设区,其水体水质更受降雨径流在城市建设区中的冲刷汇流影响。

农业区控制下的北河店站表现出明显的农业面源污染特征,流量与流量滞后基本成为了 TN 变化的主导因素,径流的冲刷作用将污染物带入水体导致水质恶化^[24]。污染物随径流进入水体的过程,体现了流域养分运移的“快响应”路径,并表现为随水文过程演变的冲刷、稀释的动态变化。短期的流量滞后(如 4—16h)对 TN 的影响较为复杂,包括促进或抑制作用。长期流量滞后(如 48—72h)下的洪峰高流量则表现出促进作用,反映地表冲刷基本完成且流量回落导致稀释作用减弱。相比之下,事件初期、末期的相对低流量滞后体现了冲刷、流量稀释的混合作用。在降雨径流初期,径流对水质的影响存在冲刷、污染物迁移与稀释共同作用的复杂关系。流量变化后期,随着地表冲刷的基本完成,径流与污染物迁移表现出显著的正向关系,即以冲刷后的污染物向地表水体的迁入为主导,这与地表产流的前中期以地表污染物优先迁移和稀释为主导的观点相一致^[25]。关于水体氮素对降雨的响应研究中发现,降雨的冲刷和稀释作用共同影响河流氮素的浓度,且该影响的滞后性常表现为氮素浓度的峰值出现在降雨后 1—5 天发生^[26]。这种数天的滞后是白洋淀流域农业面源污染“快响应路径”的定量表征,反映了降雨事件驱动下地表径流和壤中流迅速冲刷、淋溶农田表层移动性较强的氮素(如硝态氮),并通过快速通道在短时间内汇入河流。此外,有研究发现流域中 TN 输移与泥沙负荷存在高度的相关性^[27],即随着流量增大,含氮泥沙大量输移至下游河段,流量变化后期表现出对 TN 的显著促进作用。

与北河店站不同,城市区控制下的码头站的 TN 未呈现流量主导的模式,水质指标成为关键影

响因子,特别是电导率的重要性高于流量滞后特征。城市降雨冲刷效应分析结果表明,TN与降雨历时及前期降雨量存在较强关联且变动性大,表明城市TN变化同样存在滞后性^[28]。本研究结果进一步显示,河流中TN在不同事件中表现出不同的滞后特点,且其随流量变化的增减程度存在差异。城市化一般通过加速地表径流,促使地表污染物汇入水体。对于码头站,TN浓度变化并非主要由径流量驱动,而与电导率等水质指标密切相关,反映出氮污染机制与农业区的显著不同。这揭示了城市氮污染的主要来源可能并非地表径流的冲刷,而可能是持续性的、高离子强度和高氮浓度的点源或准点源(如污水处理厂尾水、城市排水)输入。在此情境下,电导率成为了污染源输入的示踪剂,流量升高伴随的稀释作用比农业面源污染区的更为明显。

3.3 监测频率对降雨与非降雨时期总氮预测影响

两个监测站点在降雨径流时期与非降雨径流时期表现出的时间分辨率敏感性差异明显,模型的日尺度和小时尺度的预测精度差异十分显著。结果与已有研究相符^[29],即在基流条件下,河流主要由浅层地下水补给,此时污染物浓度变化缓慢。但在降雨事件中,低频监测方案无法捕捉到降雨和河流短期变化中水文与生物主要驱动因子所引发的水化学响应,导致了模型预测精度的显著下降。结果表明关键径流事件的捕捉对于农业面源监测尤为重要,在白洋淀流域少数几场降雨径流贡献了大部分年度污染负荷。基于此,本研究提出白洋淀流域监测策略应从传统的“定时定频采样”向“水文响应触发的智能化变频监测”转变。在非降雨径流时期,鉴于模型精度对时间分辨率不敏感,采用日尺度或更低频次监测,主要用于跟踪水质长期基线变化,以最大限度降低运维成本与数据冗余。在降雨径流事件期间切换至小时尺度高频监测,不仅能够有效捕捉关键污染输出过程,还可在保证监测精度的前提下优化采样成本和数据存储分析量^[30]。采用该基于水文事件触发的变频监测策略,能够确保捕捉污染的关键变化过程,解决低频漏测问题及降低高频监测成本,可实现监测精度与经济效益的双重优化。

4 结论

1)RF模型在北河店(NSE=0.981)和码头(NSE=0.972)均表现出较好的预测性能,可基于常规水质指标和流量精准预测TN动态变化。全年尺

度上,北河店站TN与电导率和水温高度相关,码头站则与水温密切相关。两个站点均表现出明显的季节性特征。

2)在降雨径流时期,农业区控制的北河店站在降雨径流事件中表现为流量主导,反映了白洋淀流域农业区氮运移的“快响应”路径。且流域农业面源污染的汇入存在一定滞后性,滞后约48—72h的高流量对TN有显著促进作用。而短期流量滞后(如4—16h)对TN的影响复杂,即在流量变化前期,污染物冲刷与径流稀释作用存在动态变化。

3)在降雨径流时期,城市区控制的码头站在降雨径流事件中与水质指标有强关联的变化模式。其他水质指标对TN浓度变化有明显的指征作用,特别是电导率作为城市污水等高浓度污染源输入的有效示踪指标,与TN浓度呈现强正相关。高流量在该站点主要表现稀释作用。

4)时间分辨率对模型性能的影响在不同时期存在显著差异。降雨径流时期,小时尺度与日尺度模型精度差异显著,两站点NSE分别相差0.393、0.924;非降雨径流时期,模型精度接近,NSE差异小于0.01。建议在白洋淀流域非降雨径流期采用低频监测、降雨径流期采用高频监测的变频采样方案,可有效捕捉关键污染过程同时优化采样成本。为北方平原湖泊流域的水生态安全与高质量发展提供支撑。

参考文献

- [1] Yu C, Huang X, Chen H, et al. Managing nitrogen to restore water quality in China[J]. *Nature*, 2019, 567(7749): 516-520.
- [2] De Vries W. Impacts of nitrogen emissions on ecosystems and human health: A mini review[J/OL]. *Current Opinion in Environmental Science & Health*, 2021, 21: 100249.
- [3] Medina M, Kaplan D, Milbrandt E C, et al. Nitrogen-enriched discharges from a highly managed watershed intensify red tide (*Karenia brevis*) blooms in southwest Florida[J]. *Science of The Total Environment*, 2022, 827: 154149.
- [4] 门宝辉,申耀铎,李宣瑾,等.基于ReNuMa模型的武烈河氮磷负荷模拟及来源解析[J]. *华北水利水电大学学报(自然科学版)*, 2024, 45(5): 69-77, 99. [Men B H, Shen Y D, Li X J, et al. Simulation and Source Analysis of Nitrogen and Phosphorus Loading of Wulie River Based on ReNuMa Model[J]. *Journal of North China University of Water Resources and Electric Power(Natural Science*

- Edition), 2024, 45(5):69-77, 99.]
- [5] 史冲, 王浩宇, 王慧亮, 等. 基于 SWAT 的鲇鱼山水库流域氮磷面源污染时空变化研究[J]. 华北水利水电大学学报(自然科学版), 2023, 44(2):39-48. [Shi C, Wang H Y, Wang H L, et al. Study on Spatio-temporal Variation of Nitrogen and Phosphorus Non-point Source Pollution in Nianyushan Reservoir Watershed Based on SWAT[J]. Journal of North China University of Water Resources and Electric Power(Natural Science Edition), 2023, 44(2):39-48.]
- [6] 陈华君, 付景保, 陈玉辉, 等. 南水北调中线水源区农业面源污染负荷及驱动因素[J]. 华北水利水电大学学报(自然科学版), 2025, 46(6):153-160. [Chen H J, Fu J B, Chen Y H, et al. Agricultural Non-Point Source Pollution Load and Driving Factors in Water Source Area of Middle Route of South-to-North Water Diversion Project[J]. Journal of North China University of Water Resources and Electric Power(Natural Science Edition), 2025, 46(6):153-160.]
- [7] 王萌, 杨生光, 耿润哲. 农业面源污染防治的监测问题分析[J]. 中国环境监测, 2022, 38(2):61-66. [Wang M, Yang S G, Geng R Z. Analysis on Monitoring and Practices of Agricultural Non-point Source Pollution Control[J]. Environmental Monitoring in China, 2022, 38(2):61-66.]
- [8] Zhang L, Lu W, An Y, et al. Response of non-point source pollutant loads to climate change in the Shitoukoumen reservoir catchment[J]. Environmental Monitoring and Assessment, 2012, 184(1):581-594.
- [9] 侯玥, 徐成东, 刘伟, 等. 气候变化情景下淮河上游流域氮排放预测研究[J]. 地球信息科学学报, 2022, 24(8):1558-1574. [Hou Y, Xu C D, Liu W, et al. Prediction of Nitrogen Emission in the Upper Reaches of the Huai River Basin Under Climate Change Scenarios[J]. Journal of Geo-information Science, 2022, 24(8):1558-1574.]
- [10] He B, Oki T, Sun F, et al. Estimating monthly total nitrogen concentration in streams by using artificial neural network[J]. Journal of Environmental Management, 2011, 92(1):172-177.
- [11] 肖焱, 郭亚会, 李明蔚, 等. 基于机器学习的地下水水质预测研究[J]. 北京师范大学学报(自然科学版), 2022, 58(2):261-268. [Xiao Y, Guo Y H, Li M W, et al. Machine learning to predict groundwater quality[J]. Journal of Beijing Normal University (Natural Science Edition), 2022, 58(2):261-268.]
- [12] Ha N-T, Nguyen H Q, Truong N C Q, et al. Estimation of nitrogen and phosphorus concentrations from water quality surrogates using machine learning in the Tri An Reservoir, Vietnam[J]. Environmental Monitoring and Assessment, 2020, 192(12):789.
- [13] 朱加应, 彭双云, 林之强, 等. 土地利用类型、强度和景观格局对河流水质的多尺度影响[J]. 湖泊科学, 2025, 38(1):154-170. [Zhu J Y, Peng S Y, Lin Z Q, et al. Multi-scale impacts of land use type, intensity, and landscape pattern on river water quality[J]. Journal of Lake Sciences, 2025, 38(1):154-170.]
- [14] 杨紫薇, 尹高飞, 赵海璇, 等. 白洋淀流域种植结构及氮盈余的时空变化特征[J]. 农业工程学报, 2023, 39(6):111-119. [Yang Z W, Yin G F, Zhao H X, et al. Spatiotemporal variations in planting structure and nitrogen surplus in the Baiyangdian basin[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(6):111-119.]
- [15] Sanchez Lozano J L, Rojas Lesmes D J, Romero Bustamante E G, et al. Historical simulation performance evaluation and monthly flow duration curve quantile-mapping (MFDC-QM) of the GEOGLOWS ECMWF streamflow hydrologic model[J]. Environmental Modelling & Software, 2025, 183:106235.
- [16] Virro H, Kmoch A, Vainu M, et al. Random forest-based modeling of stream nutrients at national level in a data-scarce region[J]. Science of The Total Environment, 2022, 840:156613.
- [17] Akiba T, Sano S, Yanase T, et al. Optuna: A Next-generation Hyperparameter Optimization Framework[PP/OL]. arXiv(2019-07-25) [2025-10-25].
- [18] 徐泽华, 刘柏音, 李斌, 等. 基于可解释集成机器学习的巢湖流域溶解氧驱动因素[J]. 环境科学, 2025, 46(9):5641-5649. [Xu Z H, Liu B Y, Li B, et al. Investigation of Dissolved Oxygen Drivers in the Chaohu Basin Using Explainable Integrated Machine Learning[J]. Environmental Science, 2025, 46(9):5641-5649.]
- [19] Han J W, Kim T, Lee S, et al. Machine learning and explainable AI for chlorophyll-a prediction in Namhan River Watershed, South Korea[J]. Ecological Indicators, 2024, 166:112361.

- [20] Han Y, Bu H. The impact of climate change on the water quality of Baiyangdian Lake (China) in the past 30 years (1991–2020)[J]. *Science of The Total Environment*, 2023, 870: 161957.
- [21] Jiang J, Wang Z, Zhang Z, et al. Extreme drought-heatwave exacerbates water quality deterioration in China[J]. *Ecological Indicators*, 2025, 170: 113008.
- [22] 宁全坤, 王晓燕, 王俊, 等. 华北半干旱半湿润流域总氮变化驱动机制研究 [J/OL]. *环境科学与技术*, 2026-01-06. [Ning C K, Wang X Y, Wang J, et al. Driving Mechanism of Total Nitrogen Variations in Semi-arid and Semi-humid Watersheds of North China[J/OL]. *Environmental Science & Technology*, 2026. [2026-01-28].]
- [23] Zhuang Y, Wen W, Ruan S, et al. Real-time measurement of total nitrogen for agricultural runoff based on multiparameter sensors and intelligent algorithms[J]. *Water Research*, 2022, 210: 117992.
- [24] Wilson C O. Land use/land cover water quality nexus: quantifying anthropogenic influences on surface water quality[J]. *Environmental Monitoring and Assessment*, 2015, 187 (7) : 424.
- [25] 王安侯, 谢志宜, 陈多宏, 等. 广州市典型小流域降雨时期农业面源污染特征分析 [J]. *生态环境学报*, 2025, 34 (10) : 1633-1643. [Wang A H, Xie Z Y, Chen D H, et al. Characterizing Agricultural Non-point Source Pollution in Representative Small Watershed of Guangzhou under Rainfall Conditions[J]. *Ecology and Environmental Sciences*, 2025, 34(10): 1633-1643.]
- [26] 周波, 李晓光, 童思陈, 等. 辽河流域氮素时空分布及其对土地利用和降雨的响应 [J]. *环境科学*, 2024, 45 (4) : 2373-2384. [Zhou B, Li X G, Tong S C, et al. Spatial and Temporal Distribution of Nitrogen in the Liaohe River Basin and Its Responses to Land Use and Rainfall[J]. *Environmental Science*, 2024, 45(4): 2373-2384.]
- [27] 郑子帅, 徐童, 高宇昂, 等. 基于 SWAT 与机器学习 LSTM-MLP 混合模型的白洋淀流域氮通量研究[J]. *环境科学学报*, 2025, 45(12) : 219-232. [Zheng Z S, Xu T, Gao Y A, et al. Nitrogen flux in Baiyangdian basin based on SWAT and LSTM-MLP hybrid model[J]. *Acta Scientiae Circumstantiae*, 2025, 45(12): 219-232.]
- [28] Jeung M, Baek S, Beom J, et al. Evaluation of random forest and regression tree methods for estimation of mass first flush ratio in urban catchments[J]. *Journal of Hydrology*, 2019, 575: 1099-1110.
- [29] Neal C, Reynolds B, Rowland P, et al. High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge[J]. *Science of The Total Environment*, 2012, 434: 3-12.
- [30] 王超, 申田田, 陈邵柠, 等. 基于降雨径流事件的小流域农业面源自动监测采样策略: 微综述 [J]. *湖泊科学*, 2025, 37 (2) : 354-367. [Wang C, Shen T T, Chen S N, et al. Research progress on automatic monitoring and sampling strategies for agricultural nonpoint source in small watersheds based on rainfall runoff events: A mini-review[J]. *Journal of Lake Sciences*, 2025, 37(2): 354-367.]

Characteristics Variation and Monitoring Optimization of Total Nitrogen Driven by Rainfall-Runoff

ZHENG Kaifeng¹, CUI Guotao^{1,2}, HAN Yuguo³, DU Xinzhong⁴

(1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; 2. Carbon-Water Research Station in Karst Regions of Northern Guangdong, Sun Yat-sen University, Guangzhou 510275, China; 3. College of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China; 4. Key Laboratory of Non-point Source Pollution Control, Ministry of Agriculture and Rural Affairs/Changping Soil Quality National Observation and Research Station/State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: [Objective] To address the frequent agricultural non-point source pollution in the Baiyangdian Basin,

there is an urgent need to elucidate the migration mechanisms of Total Nitrogen (TN) under rainfall-runoff conditions and to optimize monitoring strategies. **【Methods】** This study focuses on a typical agricultural watershed (Beihedian Station), with an urban watershed (Matou Station) serving as a comparative reference, and employs a Random Forest (RF) model coupled with the interpretable SHapley Additive exPlanations (SHAP) method. Using flow rate and readily observable conventional water quality indicators as inputs, we quantitatively analyzed TN response mechanisms and evaluated the impact of different monitoring frequencies on prediction accuracy. **【Results】** (1) The RF model demonstrated good predictive performance at both sites (with Nash-Sutcliffe Efficiency, NSE = 0.981 and 0.972, respectively). On an annual scale, TN at the Beihedian station showed a synergistic association with electrical conductivity and water temperature, while water temperature was dominant at the Matou station, with both exhibiting distinct seasonal patterns. (2) During rainfall-runoff events, the agricultural area exhibited a runoff-dominated pattern, with a significant positive TN response lagging 48–72 hours after high-flow events. Short-term flow variations produced a mixed effect of scouring and dilution. The urban area displayed a significant shift in association characteristics where water temperature was no longer dominant, instead, TN concentration was positively correlated with electrical conductivity, while high flow primarily induced a dilution effect. (3) Monitoring frequency had a substantial impact on model performance. During rainfall-runoff periods, the accuracy of the daily-scale model decreased significantly compared to the hourly-scale model (NSE reduction > 42%), while the daily-scale model at the Matou station showed limited prediction performance (NSE < 0), whereas accuracy was similar during non-runoff periods (<1% difference). **【Conclusion】** The study elucidated the differentiated response mechanisms of TN under varying land-use types, highlighted the crucial role of high-frequency monitoring in capturing and predicting pollutant dynamics during key events, and provided a scientific foundation for advancing non-point source pollution control, real-time monitoring, and early warning systems.

Keywords: Baiyangdian Basin; Total Nitrogen; Random Forest; SHAP; Rainfall-Runoff