汪 昊,高凯璐,尹高飞,等.海河流域河流氮素变化趋势及关键因子识别 [J]. 中国环境科学, 2025,45(7):3968-3977.
Wang H, Gao K L, Yin G F, et al. Identification of river nitrogen trends and key factors in the Haihe River [J]. China Environmental Science, 2025,45(7):3968-3977.

海河流域河流氮素变化趋势及关键因子识别

汪 昊[1],高凯璐[1],尹高飞[1],孙学莹[2],刘一航[1],张旭琳[1],王 凤[3],翟梦瑜[4],王佳宝[1],李昭慧[1],文宏达[1],李文超[1*]
(1.河北农业大学资源与环境科学学院,省部共建华北作物改良与调控国家重点实验室,河北省农田生态环境重点实验室,河北保定 071000；2.天津大学电气自动化与信息工程学院,天津 300072；3.北京师范大学地理科学学部, 地表过程与资源生态国家重点实验室,北京 100875；4.北京工业大学循环经济研究院,北京 100124)

摘要：基于 2021~2024 年海河流域 69 个水质监测站的监测数据,通过耦合随机森林模型(RF)和长短时记忆神经网络(LSTM)构建了 TN 浓度预测的 RF-LSTM 数据驱动模型,并探究了 TN 浓度的时空变化规律及主要影响因素.结果表明:2021 年 TN 浓度较高,2022 年 TN 浓度较低,年均值变化在 4.97~5.22mg/L;5 个子流域 TN 浓度年内呈 U 型的季节变化特征,TN 浓度最高值出现在 1 月和 2 月,最低值出现在 8 月;水温和溶解氧浓度是影响海河流域水体 TN 浓度的关键因子,水温与 TN 浓度呈极显著负相关,溶解氧浓度与 TN 浓度呈极显著正相关;RF-LSTM 数据驱动模型的预测效果(R^2:0.830~0.939)优于传统单一 LSTM(R^2:0.818~0.931)模型,且 RF-LSTM 数据驱动模型具有更强的泛化性能.

关键词：海河流域；时空变化规律；影响因子；机器学习
中图分类号：X522 **文献标识码**：A **文章编号**：1000-6923(2025)07-3968-10

Identification of river nitrogen trends and key factors in the Haihe River. WANG Hao[1], GAO Kai-lu[1], YIN Gao-fei[1], SUN Xue-ying[2], LIU Yi-hang[1], ZHANG Xu-lin[1], WANG Feng[3], ZHAI Meng-yu[4], WANG Jia-bao[1], LI Zhao-hui[1], WEN Hong-da[1], LI Wen-chao[1*] (1.Key Laboratory of Ecological Environment of Farmland in Hebei, State Key Laboratory of North China Crop Improvement and Regulation, College of Resources and Environmental Sciences, Hebei Agricultural University, Baoding 071000, China；2.School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China；3.State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China；4.Institute of Circular Economy, Beijing University of Technology, Beijing 100124, China). *China Environmental Science*, 2025,45(7)：3968~3977

Abstract：This study constructed a data-driven RF-LSTM model to predict Total Nitrogen (TN) concentrations in the Haihe River, based on monitoring data from 69 water quality stations collected between 2021 and 2024. By coupling the Random Forest (RF) algorithm with a Long Short-Term Memory (LSTM) neural network, the model also explored the spatiotemporal variation and key influencing factors of TN. Results showed that TN concentrations were relatively high in 2021 and lower in 2022, with annual averages ranging from 4.97 to 5.22mg/L. Across the five watersheds, TN levels displayed a U-shaped seasonal pattern, peaking in January and February and reaching the lowest levels in August. Temperature and dissolved oxygen were identified as the main influencing factors. TN concentration showed a highly significant negative correlation with temperature and a highly significant positive correlation with dissolved oxygen. The RF-LSTM model outperformed the traditional standalone LSTM model in prediction accuracy (R^2:0.830~0.939, R^2:0.818~0.931) and demonstrated stronger generalization performance.

Key words：Haihe River；temporal and spatial patterns of change；influence factors；machine learning

近年来,对水体总氮(TN)的研究已成为国内学者关注的重点.传统研究方法主要依赖卫星遥感技术[1]、模型分析[2]和统计分析[3].通过获取水体及其周边环境因子(如土地利用、气象、水文、土壤等)的监测数据,构建与 TN 浓度相关的因果关系模型.这类方法在水质空间分布可视化、趋势分析和因子识别方面发挥了积极作用,为水资源管理提供了有力支撑.然而,这些传统方法普遍依赖于静态的监测数据,难以捕捉 TN 浓度的动态变化特征.此外,传统方法应对复杂非线性关系、多源异构数据融合以及高维环境变量处理等方面存在一定局限.无法满足现代水环境治理对实时性和精细化管理的需求.

收稿日期：2024-12-02
基金项目：云南省重大科技专项(202402AE090024);国家自然科学基金资助项目(42477441);河北省引进留学人员资助项目(C20230337);河北省中央引导地方科技发展资金资助项目(246Z3607G)
* 责任作者, 副教授, dachao279@126.com

随着人工智能技术的不断发展,机器学习(ML)方法因其在自动化特征提取、高维非线性建模、时间序列预测及多模态数据融合等方面的优势,逐步成为水质研究领域的重要工具[4].当前,越来越多的研究将机器学习模型应用于污染物来源解析[8]、水质动态监测[9]和未来水质浓度预测[10]等任务中.

研究表明,在众多河流中,海河流域的氮污染问题严重,水体 TN 浓度在近几十年呈现出明显的升高趋势,在未来可能进一步加剧[11].对此,本研究利用机器学习模型强大的数据分析能力,通过内联式耦合随机森林(RF)模型和长短时记忆神经网络(LSTM)模型开发了 RF-LSTM 数据驱动模型.通过解析海河流域水体 TN 浓度的时空变化特征和关键影响因子,为海河流域的水环境治理提供参考.

1　研究方法

1.1　研究区域概况

海河流域地处华北平原北部(35°~43°N,112°~120°E),流域面积达 $3.19\times10^5 km^2$,是中国北方重要的水系和生态区域,为周边提供了丰富的水资源[13].根据水文情况可将海河流域分为:永定河水系、潮白河水系、子牙河水系、漳卫南运河水系以及大清河水系 5 个子流域(图 1).海河流域属于温带大陆性气候,夏季高温多雨水,冬季寒冷干燥,多年平均降水约 539mm,80%以上的降水集中在 5~10 月,年均气温 1.5~14℃[14].该流域主要种植小麦、玉米和大豆,是国家重要的粮食生产基地[15].目前,海河流域居住着超过 1 亿人,主要集中在北京、天津和河北省的核心城市及周边地区.

近年来,城市化和工业化进程加快,海河流域的水资源短缺和水质污染问题日益严重,其中 TN 污染尤为突出[16].水资源供需矛盾突出,地下水过度开采导致水位下降;同时,工业废水、生活污水及农业面源污染也成为该流域水质恶化的主要因素[17].

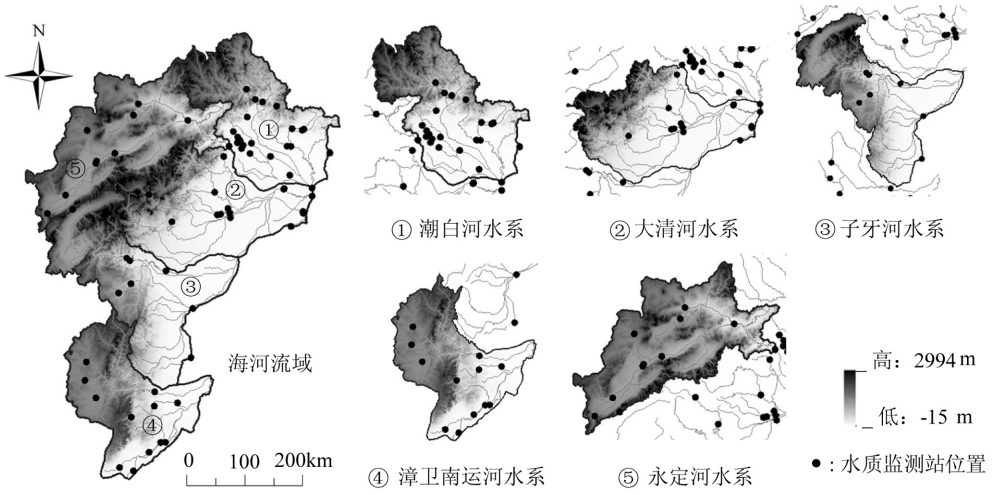


图 1　研究区域概况及水质监测站分布情况

Fig.1　Overview of the study area and distribution of water quality monitoring stations

审图号:GS(2016)2923 号

1.2　数据来源

数据来源于中国环境监测总站(https://www.cnemc.cn/)提供的海河流域 69 个监测断面、8 个水质指标的监测数据,覆盖了从 2021 年 1 月~2024 年 4 月的监测周期,共计 1216d.原始数据的总量为 1091766 条,采样频次为每日 3 次.然而,由于监测设备的维护,部分监测时间点存在数据缺失.为了保证数据的连续性,本文采用每日监测数据的平均值作为日监测数据.经过这一处理,最终得到的数据总条数为 671232 条.

1.3　数据处理

数据预处理是确保数据准确、去除噪音干扰的关键环节[18].数据预处理的步骤基于 Python 编程语言实现,使用 Jupyter Notebook 编码器进行可视化展示.首先,检查原始数据集中是否有相同时间点的重复数据,将重复数据去除.使用四分位距(IQR)方法去除异常值,IQR 用于衡量数据集的分散程度,它表示第三四分位数(Q3)与第一四分位数(Q1)之间的差值.

通过计算每一列的第一四分位数(Q1)和第三四分位数(Q3),然后根据 IQR 计算异常值的上下界,并过滤掉不在此范围内的值.

多元统计方法广泛应用于流域水环境研究.利用拟合线性曲线来比较五大子流域在 2021 年~2023 年水体 TN 浓度演变趋势.采用反距离加权(IDW)插值法,利用断面监测站数据对水体 TN 浓度进行空间特征分析.IDW 方法将有限的监测数据扩展至整个研究区域,生成连续的水质空间分布图,从而更全面地揭示流域内水质的空间分布特征.为了探究影响水体 TN 的关键因子,首先采用 RF 模型,通过构建多个决策树并集成它们的预测结果,识别各个影响因子对 TN 的贡献占比.其次,结合 Pearson 相关性分析,进一步量化这些影响因子与 TN 之间的关联强度,从而更深入地理解各个影响因子在预测模型中的作用.利用 LSTM 构建水体 TN 预测模型,以 2021 年 1 月~2023 年 12 月水质数据为训练样本,训练模型并保存参数.选取 2024 年 1 月~2024 年 4 月数据为测试样本,输入到训练好的模型中,输出预测结果.为全面评估 RF−LSTM 数据驱动模型的优越性,以传统 TN 预测方法(单一 LSTM 模型)为对比实验,两者均使用相同的数据集进行训练和测试.

1.4 水体总氮预测模型搭建

1.4.1 随机森林模型 RF 是一种用于分类和回归任务的集成学习方法[19].通过训练多个决策树并结合其预测结果,RF 能够提升模型的准确性和稳健性[20].RF 算法从训练数据中随机抽取子集构建多个独立的决策树.尽管单棵决策树可能会产生过拟合问题,但通过平均多个树的预测结果,RF 有效减少了过拟合风险,提升了整体模型的稳健性[21].因此,RF 常用于识别水质的主要污染因素[22].在 RF 模型中,初始训练数据集被划分为两个部分:训练数据和测试数据.由于自举采样的独立性和随机性,RF 模型无需额外的交叉验证[24].此外,RF 还可以用于因变量的分类,并评估每个变量的相对重要性.

1.4.2 长短时记忆神经网络 LSTM 是一种递归神经网络(RNN)的变体,专为解决传统 RNN 在长序列处理中的梯度消失问题而设计.LSTM 通过引入遗忘门、输入门和输出门来控制信息流,从而更好地记住长期依赖关系[25].凭借这些门控机制,LSTM 在处理长距离依赖方面表现优异,广泛应用于时间序列预测、自然语言处理和语音识别等领域.

LSTM 主要分为 3 个计算过程:

(I)遗忘门

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

遗忘门的主要作用是决定保留多少先前单元状态的信息.它通过控制信息的流动来帮助网络选择性地遗忘不再需要的信息,从而优化记忆和学习过程.通过读取 h_{t-1} 和,x_t,经由 sigmoid,输入一个在 0~1 之间数值给每个在记忆单元 C_{t-1} 中的数字,1 表示完全保留,0 表示完全舍弃.

(II)输入门

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

输入门的目的是确定什么样的信息内存放在记忆单元中.这里包含两个部分.sigmoid 层同样输出[0,1]的数值,决定候选状态 $\tilde{C}_t$ 有多少信息需要储存,tanh 层会创建候选状态 $\tilde{C}_t$.随后更新旧的细胞状态,将 C_{t-1} 更新为 C_t.

(III)输出门

$$O_t = \sigma(W_O[h_{t-1}, x_t] + b_O) \tag{5}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{6}$$

输出门的目的是将内部状态的信息传递给外部状态.sigmoid 层决定了哪些单元状态的信息被输出.然后,通过 tanh 层处理单元状态,并将其与输出门的结果相乘,从而确保外部状态只接收到输出门所确定的部分信息.

1.4.3 水体总氮预测模型搭建 为了更好探究水体 TN 浓度的变化特征,简化数据收集过程、减低水质预测成本.基于迭代式特征筛选与模型反馈机制,开发了 RF-LSTM 数据驱动模型(图 2).模型开发流程包括以下两个步骤:第一,采用 RF 进行初步特征筛选.使用 Python 编程语言中的 Scikit-learn 库构建 RF 模型,将多个水质指标数据集输入 RF,通过计算各指标的特征重要性来筛选特征.初始筛选标准设定为重要性占比大于 10%的特征.随后,将筛选出的特征输入到 LSTM 模型中,并通过模型性能反馈对特征重要性阈值进行调整,迭代优化特征集.第二,使用改进后的特征集构建 LSTM 模型.采用滑动时间窗口技术生成输入和输出样本,按照 7:3 比例划分训练集和测试集.将训练集输入 LSTM 模型进行训练,

根据反馈动态调整模型参数,不断优化 LSTM 模型的预测精度,直到达到满意的性能指标.这一迭代流程通过 RF 与 LSTM 之间的反馈机制,实现了特征筛选与模型训练的深度协同,使模型更具准确性和稳定性.在模型效果评估阶段,本研究选择了决定系数(R^2)和均方根误差(RMSE)这两个常用指标来全面评价模型的性能[28].

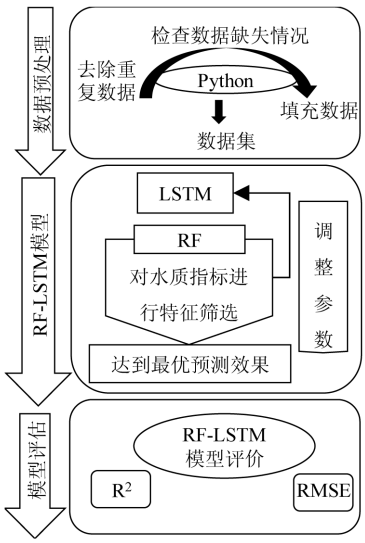


图 2　研究框架

Fig.2　Research framework

2　结果与分析

2.1　海河流域主要河流 TN 浓度时空变化特征

2.1.1　时间变化特征　从年际变化趋势来看, 2021 年 TN 浓度为 5.22mg/L,2022 年降至 4.97mg/L (下降 5%),在 2023 年增加到 5.07mg/L(增加 2%).5 个子流域在 3a 内 TN 变化情况各不相同,子牙河水系 TN 浓度持续升高,永定河水系 TN 浓度呈现先下降后上升的趋势,漳卫南运河水系、潮白河水系和大清河水系 TN 浓度持续降低.具体而言,子牙河水系由 6.54mg/L 增长至 8.56mg/L,总体增长了 31%;永定河水系先由 4.98mg/L 降至 4.67mg/L 后升至 4.7mg/L,总体下降了 6%.漳卫南运河水系由 5.58mg/L 降至 4.18mg/L,总体降低了 25%;潮白河水系由 4.73mg/L 降至 4.09mg/L,总体降低了 14%;大清河水系由 4.27mg/L 降至 3.79mg/L,总体降低了 11%.

从年内变化趋势来看,5 个子流域的 TN 浓度在 3a 内呈现相似的季节性变化特征,整体变化趋势呈现 U 型.一年中最高值出现在 1 月和 2 月,最低值出现在 8 月.具体而言,潮白河水系、漳卫南运河水系和永定河水系从 2~8 月持续下降,随后逐渐上升.而大清河水系和子牙河水系从 1~8 月开始下降,随后持续上升.通过计算 5 个子流域 TN 浓度月均值来比较其变化幅度:子牙河水系从 10.68mg/L 下降至 4.01mg/L(下降 62%)>大清河水系从 5.90mg/L 下降至 2.37mg/L(下降 60%)>永定河水系从 6.63mg/L 下降至 3.01mg/L(下降 55%)>漳卫南运河水系从 5.97mg/L 下降至 3.33mg/L(下降 44%)>潮白河水系从 5.68mg/L 下降至 3.43mg/L(下降 40%).

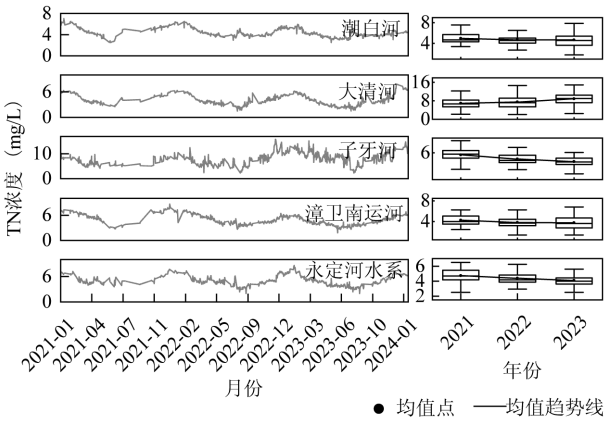


图 3　2021 年~2024 年每个流域的 TN 浓度变化情况

Fig.3　Changes in TN concentration in each watershed from 2021 to 2024

2.1.2　空间变化特征　2021~2023 年海河流域 TN 浓度年均值分布呈现出显著的空间异质性,如图 5.基于自然间断点法,根据 TN 浓度划分为 4 个浓度梯度(0~2mg/L、2~4mg/L、4~6mg/L、>6mg/L).如图 4(a)所示,大清河水系、子牙河水系、漳卫南运河水系和永定河水系的水体 TN 浓度超过 6mg/L 的区域占比最大(43%~83%),而潮白河水系水体 TN 浓度在 2~4mg/L 的区域占比最大(48%).5 个子流域中,水体 TN 浓度在小于 2mg/L 的区域占比最小(0~2%).对于漳卫南运河水系水体 TN 浓度大于 6mg/L 的区域占比远超其他流域,为 25%.

如图 4(b)所示,其中,0~2mg/L 和>6mg/L 的两个浓度梯度中,漳卫南运河水系的区域占比最高(分别是 33%和 34%).2~4mg/L 浓度梯度中,永定河水系的区域占比最高(36%),子牙河水系的区域占比最低(6%).4~6mg/L 浓度梯度中,子牙河水系的区域占比最高(29%),漳卫南运河水系和永定河水系的区域占比最低(均为 11%).

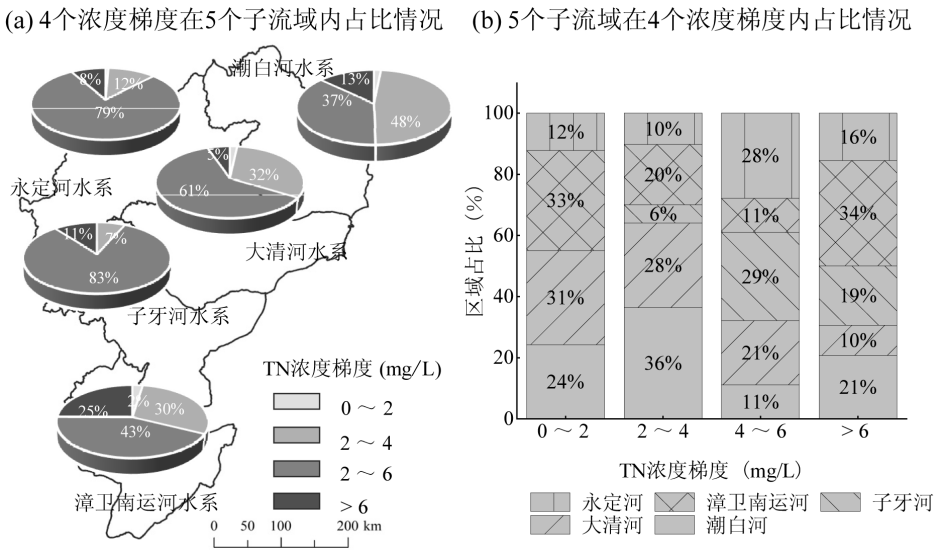


图 4 海河流域 TN 浓度空间分布

Fig.4 Spatial distribution of TN concentrations in the Haihe River

2.2 海河流域主要河流 TN 浓度变化的影响因子

基于 RF 算法的特征筛选结果如图 5 所示.对于 TN 而言,除子牙河水系和漳卫南运河水系外,水温在五个子流域中对于水体 TN 贡献度占比最高(贡献占比 21%~76%).子牙河水系的电导率贡献度占比最高(贡献占比 41%),漳卫南运河水系的高锰酸盐指数贡献度占比最高(贡献占比 36%).pH 值在 5 个子流域中对于水体 TN 贡献度最低(贡献占比 2%~6%).

基于 Pearson 相关性分析结果如图 5 所示,其中使用圆圈边缘颜色来表示环境因子与 TN 浓度之间的相关性(正相关和负相关).*P* 的负对数(−log10)用于表示相关性的显著性水平,通过圆圈大小进行可视化.分别计算环境因子平均值并进行归一化处理,结果通过色带来表示.在五个子流域内,水温、高锰酸盐指数和浊度与水体 TN 浓度之间呈极显著的负相关,而溶解氧浓度与水体 TN 浓度之间则呈极显著的正相关.水温与水体 TN 浓度的相关性最为显著,而氨氮与水体 TN 浓度的相关性最弱.

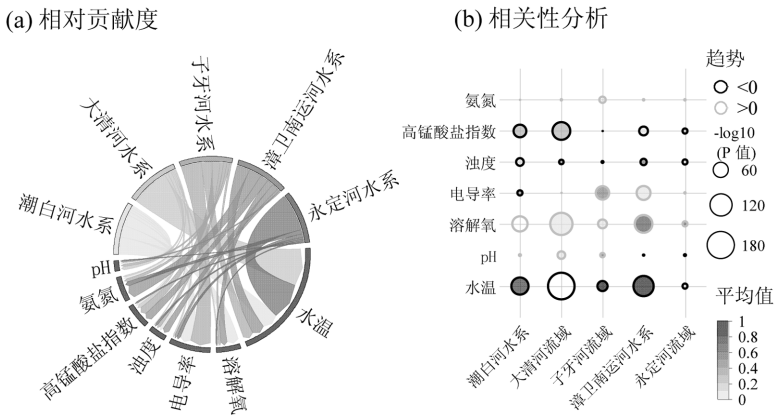


图 5 5 个子流域中各水质指标对 TN 相对贡献度以及相关性分析

Fig.5 Relative contribution of each water quality indicator to TN in the five watersheds and correlation analysis

2.3 基于机器学习的海河流域河流 TN 浓度预测模型构建

根据 2021~2023 年水质数据构建 RF−LSTM 数据驱动模型.RF 特征筛选结果发现 5 个子流域的输入特征各不相同,具体如下:在潮白河流域,选取水温、溶解氧、电导率和浊度;在永定河水系,选择水温、氨氮和电导率;在大清河水系,选择水温、高锰酸盐指数和 pH 值;在漳卫南运河水系,选择高锰酸

盐指数、氨氮、溶解氧、浊度、水温和电导率;在子牙河水系,选择电导率、水温、高锰酸盐指数、浊度和氨氮作为输入特征.

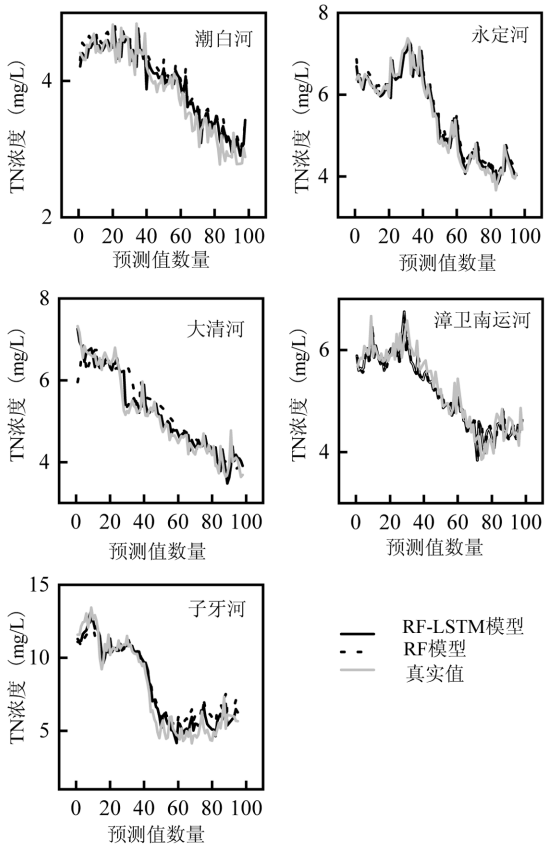


图 6　两种模型的预测值与真实值比较

Fig.6　Comparison of predicted values and true values of two models

表 1　单一 LSTM 模型与 RF−LSTM 数据驱动模型的预测效果评价

Table 1　Evaluation of the prediction effect of single LSTM model and RF−LSTM data−driven model

流域名称	单一 LSTM 模型		RF−LSTM 数据驱动模型	
	R^2	RMSE	R^2	RMSE
潮白河水系	0.848	0.312	0.933	0.262
永定河水系	0.931	0.254	0.936	0.245
大清河水系	0.818	0.595	0.830	0.461
漳卫南运河水系	0.840	0.426	0.850	0.418
子牙河水系	0.858	0.337	0.939	0.318

如图 6 和表 1 所示,在 5 个子流域中,RF−LSTM 数据驱动模型的 R^2 都比单一的 LSTM 的 R^2 要高,提升的区间在 0.005(永定河水系)~0.081(子牙河水系).同时,RF−LSTM 数据驱动模型的 RMSE 都比单一的 LSTM 的 RMSE 要低,降低的区间在 0.009(永定河水系)~0.134(大清河水系).从两个模型预测效果上看,RF−LSTM 数据驱动模型比单一的 LSTM 效果好.从 RF−LSTM 数据驱动模型在五个子流域的预测效果来看,R^2 在 0.830~0.939 之间,RMSE 在 0.245~0.461 之间.

表 2　模型预测性能与时间步长的关系

Table 2　Relationship between model prediction performance and time−steps

流域名称	时间步长(d)	R^2	RMSE
潮白河水系	1	0.833	0.262
	4	0.713	0.34
	8	0.748	0.322
	12	0.783	0.347
永定河水系	1	0.936	0.275
	4	0.889	0.362
	8	0.838	0.437
	12	0.739	0.555
大清河水系	1	0.670	0.682
	4	0.688	0.663
	8	0.690	0.661
	12	0.690	0.661
漳卫南运河水系	1	0.750	0.418
	4	0.725	1.330
	8	0.631	1.940
	12	0.469	2.330
子牙河水系	1	0.899	1.018
	4	0.782	1.493
	8	0.735	1.647
	12	0.605	2.012

实验中设置了 epochs=100,批次大小(batch_size)为 128.在训练过程中,随着 epochs 的增加,模型的训练损失和验证损失均逐渐降低.当 epochs 达到 10 时,训练损失趋于稳定.实验结果表明,设置 epochs 为 100 时,能够有效降低模型的损失值,实现最优的模型表现.如表 2 所示,在不同步长设置下,分别使用前 1,4,8,12d 的 TN 数据来预测未来 TN 的变化.随着预测步长的增加,模型的预测误差逐渐增大.当预测步长为 1 时,模型的预测效果最佳(其中大清河水系的最佳预测步长为 8).

3　讨论

3.1　河流 TN 时空变化的驱动机制分析

本研究发现,5 个子流域 TN 浓度在 2021 年较高,而在 2022 年较低.在年内呈现 U 型变化趋势,1 月和 2 月 TN 浓度最高,8 月最低.该发现与熊伟光等[30]对 2018~2022 年海河流域 TN 浓度分析结果相似,产生

这种变化可能与降水量波动密切相关.2021 年和 2023 年是特丰水年,其中 2021 年平均降水量为 838.5mm,远高于 2023 年的 608.5mm 和 2022 年的 554.4mm.降水量的增加直接导致了地表径流量的增加,而地表径流是污染物进入水体的主要途径之一.尤其在强降水和极端天气条件下,降水带来的径流迅速将氮素污染物从农田、城市等区域带入河流中,从而显著提高水体中的 TN 浓度[31].在年内变化中,海河流域受到枯水期降雨量减少的影响,流量和径流量也相应减少.根据相关研究,海河流域的水体污染主要来源于点源污染[32],而点源污染的污染物排放量不受降雨时期的影响.浓缩效应使枯水期水体的 TN 浓度通常维持在较高水平(大于 5.68mg/L).相比之下,丰水期由于水体的流量和径流量增加,水体的稀释作用使得水体 TN 浓度下降,尤其在 8 月时降至最低值(小于 4.01mg/L).

海河流域水体 TN 浓度呈现明显的空间异质性.TN 浓度大于 6mg/L 的区域主要集中在潮白河水系、永定河水系、子牙河水系和大清河水系.这一空间分布特征与 Zha 等[33]对海河流域的氮磷排放特征的研究结果相类似,氮磷的高排放区域与农业和工业分布区域高度重合.对于本研究而言,漳卫南运河水系南部焦作、新乡,子牙河湿地和永定河西北部的张家口地区均是农业和各类工业聚集地,且农业和工业污水排放严重超标[34].而对于 TN 浓度较低的潮白河水系和大清河水系,前者是因为其上游密云水库作为重要的水源涵养区,生态环境较好,没有农业和工业的排放影响[37].后者则是得益于该流域对农业面源污染和点源污染显著的治理成效,尤其白洋淀流域通过源头控制,对上游生活、农业、畜禽养殖业排污治理.对排污企业进行整改.对沿途的生态进行修复提高植被覆盖度、涵养水源、防治水土流失等[38].总体上,海河流域 TN 浓度分布的主要原因是农业源和人为因素共同作用的结果.

3.2 河流 TN 浓度变化的关键影响因子辨识与作用解析

水温和溶解氧溶度是影响海河流域 TN 浓度的重要影响因子[39].较高的水温促进浮游植物和藻类的生长.藻类通过光合作用可以吸收水中的氮源,从而降低水体中的 TN 浓度[40].溶解氧一方面对硝化作用起促进作用.硝化作用是氮循环中的重要过程,在硝化细菌和亚硝化细菌的作用下,将氨氮逐步氧化为硝酸盐氮.这一过程分为两个阶段:首先,亚硝化细菌将氨氮氧化为亚硝酸盐;接着,硝化细菌将亚硝酸盐进一步氧化为硝酸盐.在这一过程中,较高溶解氧浓度(大于 8mg/L)可以显著提高硝化细菌的代谢效率,加速氨氮的转化,从而增加了 TN 浓度[41].另一方面溶解氧作为水体中最重要的氧化剂之一,可以促进有机氮的分解和转化.当氧气充足时,需氧微生物的代谢活性增强,能够快速分解有机氮化合物(如植物残留、动物排泄物等),将大分子物质转化为氨氮[43].这些释放的氨氮在溶解氧的作用下进入硝化过程.因此,高溶解氧浓度通过促进有机氮分解和硝化作用,是水体中 TN 浓度增加的重要原因,也会因硝酸盐的积累引发水体富营养化等生态问题.

3.3 模型评估与泛化性分析

RF 进行特征筛选后的模型相较于传统预测模型在预测精度上有明显的提升,特征筛选可以有效的提高模型的预测精度.这一结论与大多数研究结果相一致[44].其原因可以从特征筛选的理论基础和随机森林算法的特性两方面进行分析.首先,特征筛选的核心在于通过算法评估变量的重要性,剔除冗余或无关的特征,从而优化模型的学习能力.过多的无关特征会导致模型在训练时关注无意义的模式,进而增加模型的复杂性和过拟合风险.通过特征筛选剔除这些变量,模型能够更高效地聚焦于关键特征,从而提高对数据模式的捕捉能力[45].此外,减少无关特征的输入还可以减小模型参数的搜索空间,加速模型的收敛过程,提升计算效率.其次,随机森林作为一种基于集成学习的算法,在特征选择中具有天然优势.它通过引入随机性和多样性,在样本和特征子集上进行训练,不仅降低了单一模型可能出现的偏差,还能通过“多数投票”机制提升模型的稳健性和泛化能力.在特征筛选中,随机森林能够通过分析每个特征对分类或回归任务的贡献,综合评估变量的重要性分数,从而科学地选择出最相关的特征子集[46].

特征筛选还可以提高模型的解释性.高维数据往往隐藏着大量复杂的相关性和多重共线性问题,这可能导致模型在理解变量之间关系时出现偏差.通过特征筛选,可以剔除具有高度共线性的变量,减少变量之间的干扰,增强模型对变量间因果关系的

解释能力.

实验结果证明,RF-LSTM 数据驱动模型的性能与预测步长密切相关,RF-LSTM 数据驱动模型具有更强的泛化性[47].较短的预测步长(如 1d)可以为模型提供更加准确的历史数据,从而提高模型的预测精度.随着步长的增加,模型的误差逐渐增大,这可能是由于历史数据对远期预测的影响逐渐减弱,导致模型难以捕捉到未来更复杂的变化趋势[48].该实验为未来水质预测中步长选择提供了参考依据.总体而言,RF-LSTM 数据驱动模型在不同预测步长(1~12)下均能保持良好的预测效果,证明了其具有较强的泛化能力.

3.4　研究局限性

本文研究重点集中在水质指标对 TN 浓度变化的影响上,主要关注了水温、溶解氧等直接与水体氮循环相关的物理化学指标.这些指标在模型中表现出较高的相关性和预测能力,但这并不意味着它们是唯一的决定性因素.实际上,TN 浓度的变化不仅与水质参数密切相关,还受到外部环境因素的综合作用[49],例如流域内的农业活动、土地利用类型、气候变化、人口密度及排污行为等.因此,仅从水质指标角度出发,可能未能全面捕捉影响 TN 浓度变化的复杂机制.未来研究可以进一步整合更多的环境因素,构建更加全面的预测框架.农业活动,尤其是氮肥的过量施用,是河流氮污染的重要来源[50].这主要源于氮肥在土壤-水体系统中的迁移行为和作用特性.农作物对氮肥的吸收能力有限,过量施用的氮肥常以硝酸盐等形式滞留于土壤中,并在降雨或灌溉过程中渗入地下水,或通过地表径流进入河流.同时,土壤中氮肥经过硝化和反硝化作用生成的中间产物也会进一步流向水体,导致河流氮含量不断增加.此外,长期施用过量氮肥易引发区域土壤“氮饱和”现象,显著加速氮元素向水体的累积性迁移,从而加剧河流氮污染的程度.气候因素是影响河流 TN 浓度变化的另一个关键方面[51].本研究中虽然初步分析了水温对 TN 浓度的直接影响,但对气候因素(如降水量和气温波动等)的作用分析不足.降水量的增加通常会导致地表径流量增加,从而加剧氮流失并输入河流;而极端干旱条件则可能通过减少水体稀释效应,导致 TN 浓度显著升高.

本文仍需在未来研究中进一步整合多学科视角,结合更长时间尺度、更广空间范围以及更全面的环境数据,从而提升 TN 浓度预测模型的科学性和适用性,为流域水质管理提供更可靠的决策支持.

4　结论

4.1　5 个子流域 TN 浓度在 2021 年较高,在 2022 年较低.年内表现为U型波动.TN 浓度的最高值出现在 1 月和 2 月,而最低值出现在 8 月.

4.2　5 个子流域 TN 浓度的空间分布具有明显异质性,不同子流域的浓度分布规律有所不同.

4.3　水温和溶解氧浓度是影响海河流域 TN 浓度的关键因素,其中水温与 TN 浓度呈极显著负相关,而溶解氧浓度与 TN 浓度呈极显著正相关.

4.4　RF-LSTM 数据驱动模型的预测结果与实际观测值的变化趋势高度一致,且相较于传统 LSTM 模型,该模型在 TN 浓度预测的精准性和泛化性方面具有显著优势.

参考文献：

[1] Zhang Y, He X, Lian G, et al. Monitoring and spatial traceability of river water quality using Sentinel-2satellite images [J]. Science of the Total Environment, 2023,894:164862.

[2] Wang A, Yang D, Tang L. Spatiotemporal variation in nitrogen loads and their impacts on river water quality in the upper Yangtze River basin [J]. Journal of Hydrology, 2020,590(1):125487.

[3] 徐建锋,尹　炜,闫峰陵,等.农业源头流域景观异质性与溪流水质耦合关系 [J]. 中国环境科学, 2016,36(10):3193-3200.
Xu J F, Yin W, Yan F L, et al. The coupling relationship between landscape heterogeneity and stream water quality in an agricultural catchment [J]. China Environmental Science, 2016,36(10):3193-3200.

[4] Kaveh K, Kaveh H, Bui M D, et al. Long short-term memory for predicting daily suspended sediment concentration [J]. Engineering with Computers, 2021,37:2013-2027.

[5] Nourani V, Behfar N. Multi-station runoff-sediment modeling using seasonal LSTM models [J]. Journal of Hydrology, 2021,601:126672.

[6] Fang L, Shao D. Application of long short-term memory (LSTM) on the prediction of rainfall-runoff in karst area [J]. Frontiers in Physics, 2022,9:790687.

[7] 杜尚海,古成科,张文静.随机森林理论及其在水文地质领域的研究进展 [J]. 中国环境科学, 2022,42(9):4285-4295.
Du S H, Gu C K, Zhang W J. A review on the progresses in random forests theory and its applications in hydrogeology [J]. China Environmental Science, 2022,42(9):4285-4295.

[8] Xue J, Yuan C, Ji X, et al. Predictive modeling of nitrogen and phosphorus concentrations in rivers using a machine learning framework: A case study in an urban-rural transitional area in Wenzhou China [J]. Science of the Total Environment, 2024,910: 168521.

[9] Hasani S S, Arias M E, Nguyen H Q, et al. Leveraging explainable machine learning for enhanced management of lake water quality [J]. Journal of Environmental Management, 2024,370:122890.

[10] Zhi W, Baniecki H, Liu J, et al. Increasing phosphorus loss despite widespread concentration decline in US rivers [J]. Proceedings of the National Academy of Sciences, 2024,121(48):e2402028121.

[11] Zhang W, Jin X, Liu D, et al. Assessment of the sediment quality of freshwater ecosystems in eastern China based on spatial and temporal variation of nutrients [J]. Environmental Science and Pollution Research, 2017,24:19412-19421.

[12] 冯爱萍,吴传庆,王雪蕾,等.海河流域氮磷面源污染空间特征遥感解析 [J]. 中国环境科学, 2019,39(7):10.
Feng A P, Wu C Q, Wang X L, et al. Spatial character analysis on nitrogen and phosphorus diffuse pollution in Haihe River Basin by remote sensing [J]. China Environmental Science, 2019,39(7):10.

[13] Yao T, Liu S, Hu S, et al. Response of vegetation ecosystems to flash drought with solar-induced chlorophyll fluorescence over the Hai River Basin, China during 2001~2019 [J]. Journal of Environmental Management, 2022,313:114947.

[14] 成星露,孙永华,张王宽,等.气候干湿冷暖化复合类型的识别与量化方法 [J]. 地球信息科学学报, 2023,25(11):2204-2217.
Cheng X L, Sun Y H, Zhang W K, et al. Method to identify and quantify compound dry(wet)-cool(warm) climate trend [J]. Journal of Geo-information Science, 2023,25(11):2204-2217.

[15] 赵安周,李子洋,承达瑜.海河流域饱和水汽压差时空演变及影响因素分析 [J]. 生态学报, 2025,45(1):385-394.
Zhao A Z, Li Z Y, Cheng D Y. Temporal and spatial variation and influencing factors of saturated water vapor pressure difference in the Hai River Basin [J]. Acta Ecologica Sinica, 2025,45(1):385-394.

[16] 王庆明,赵 勇,王 浩,等.海河流域地表水资源衰减归因与规律 [J]. 中国科学:地球科学, 2024,54(5):1573-1587.
Wang Q M, Zhao Y, Wang H, et al. Surface water resource attenuation attribution and patterns in Hai River Basin [J]. Scientia Sinica(Terrae), 2024,54(5):1573-1587.

[17] Yua X, Jun Z. Water resource risk assessment based on non-point source pollution [J]. Water, 2021,13(14):1907.

[18] 李海林,周文浩,吴炳毅,等.基于云模型理论的数据预处理及校准分析方法研究 [J]. 系统工程理论与实践, 2025:1-19.
Li H L, Zhou W H, Wu B Y, et al. Research on data preprocessing and calibration method based on cloud model [J]. Systems Engineering-Theory and Practice, 2025:1-19.

[19] Zou T, Yang K, Pan M, et al. Analysis of the temporal and spatial evolution of turbidity in Tonle Sap Lake and its influencing factors [J]. Science of The Total Environment, 2024,943(000):16.

[20] 郝玉莹,赵 林,孙 同,等.基于RF-LSTM的地表水体水质预测 [J]. 水资源与水工程学报, 2021,32(6):41-48.
Hao Y Y, Zhao L, Sun T, et al. Surface water quality prediction based on RF-LSTM [J]. Journal of Water Resources and Water Engineering, 2021,32(6):41-48.

[21] Wan D. Distribution, sources and main controlling factors of nitrate in a typical intensive agricultural region, northwestern China: Vertical profile perspectives [J]. Environmental Research, 2023,237(1): 116911-116911.

[22] Podgorski J E, Labhasetwar P, Saha D, et al. Prediction modeling and mapping of groundwater fluoride contamination throughout India [J]. Environmental Science and Technology, 2018,52(17):9889-9898.

[23] Wei W, Nghiem A, Ma R, et al. Factors controlling iodine enrichment in a coastal plain aquifer in the North Jiangsu Yishusi Plain, China [J]. Journal of Contaminant Hydrology, 2021,243:103894.

[24] Tesoriero A J, Gronberg J A, Juckem P F, et al. Predicting redox - sensitive contaminant concentrations in groundwater using random forest classification [J]. Water Resources Research, 2017,53(8):7316-7331.

[25] Luo L, Zhang Y, Dong W, et al. Ensemble empirical mode decomposition and a long short-term memory neural network for surface water quality prediction of the Xiaofu River, China [J]. Water, 2023,15(8):1625.

[26] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural computation, 2019,31(7): 1235-1270.

[27] 林天宙,彭 杨,罗诗琦,等.基于 RF-LSTM 模型的长江上游汛期日含沙量预测 [J]. 中国农村水利水电, 2024,(10):32-39.
Lin T Z, Peng Y, Luo S Q, et al. Prediction of daily sediment concentrations in the upper reaches of Yangtze River during flood season based on RF-LSTM model [J]. China Rural Water and Hydropower, 2024,(10):32-39.

[28] Adamowski J, Chan H F. A wavelet neural network conjunction model for groundwater level forecasting [J/OL]. Journal of Hydrology, 2011,407(1-4):28-40.

[29] Dayev Z A. Application of artificial neural networks instead of the orifice plate discharge coefficient [J], Flow Measurement and Instrumentation, 2020,71:101674.

[30] 熊伟光,彭嘉玉,李晓光,等.2018~2022 年海河流域总氮浓度时空变化特征及驱动因素分析 [J/OL]. 环境科学, 2025,46(1):184-192.
Xiong W G, Peng J Y, Li X G, et al. Spatial-Temporal Variations and Driving Factors Analysis of Total Nitrogen Concentration in the Haihe River Basin from 2018 to 2022 [J/OL]. Environmental Science, 2025,46(1):184-192.

[31] Li S, Wang X, Qiao B, et al. First flush characteristics of rainfall runoff from a paddy field in the Taihu Lake watershed, China [J]. Environmental Science and Pollution Research, 2017,24(9):1-16.

[32] 韦英怀,胡敏鹏,陈丁江.我国主要河流水系硝态氮污染特征及定量源解析 [J]. 环境科学, 2024,45(02):755-767.
Wei H Y, Hu M P, Chen D J. Nitrate pollution characteristics and its quantitative source identification of major river systems in China. Environmental Science [J], 2024,45(2):755-767.

[33] Zhao Z M L. Agricultural nitrogen and phosphorus emissions to water and their mitigation options in the Haihe Basin, China [J]. Agricultural Water Management, 2019,212:262-272.

[34] 徐华山,徐宗学,唐芳芳,等.漳卫南运河流域水质时空变化特征及其污染源识别 [J]. 环境科学, 2012,33(2):359-369.
Xu H S, Xu Z X, Tang F F, et al. Spatiotemporal variation analysis and identification of water pollution sources in the Zhangweinan River Basin [J]. Environmental Science, 2012,33(2):359-369.

[35] 孙 雪,于 格,刘汝海,等.海河南系子牙河水系湿地生态系统健康评价研究 [J]. 中国海洋大学学报, 2019,49(11):120-132.

Sun X, Yu G, Liu R H, et al. Wetland ecosystem health assessment in Ziya River Basin, Haihe River [J]. Periodical of Ocean University of China, 2019,49(11):120−132.

[36] 张厌非.试谈永定河上游水土流失和水土保持 [J]. 中国水土保持, 1984,(2): 44−46.
Zhang Y F, Discussion on soil and water loss and conservation in the upper reaches of Yongding River [J]. Soil and Water Conservation in China, 1984,(2):44−46.

[37] 温燕华.密云水库上游潮白河流域典型小流域治理研究 [D]. 邯郸: 河北工程大学, 2020.
Wen Y H. Study on governance of typical small watersheds in Chaobai River Basin in the upper reaches of Miyun Reservoir [D]. Handan: Hebei University of Engineering, 2020.

[38] 李琳琳,王国清,秦　攀,等.白洋淀水环境状况与治理保护对策 [J]. 科技导报, 2019,37(21):14−25.
Li L L, Wang G Q, Qin P, et al. Water environmental quality of Baiyangdian Lake and its treatment and protection countermeasures [J]. Science and Technology Review, 2019,37(21):14−25.

[39] Huang J, Zhang Y, Bing H, et al. Characterizing the river water quality in China: Recent progress and on−going challenges [J]. Water Research, 2021,201(4):117309.

[40] Liu L, Yang J, Yu X, et al. Patterns in the composition of microbial communities from a subtropical river: Effects of environmental, spatial and temporal factors [J]. PLOS ONE, 2017,8(11):81232.

[41] Lei Y, Dong F, Liu X, et al. Short−term variations and correlations in water quality after dam removal in the Chishui river basin [J]. Journal of Environmental Management, 2023,327:116917.

[42] Wang Y T, Xu H, Shen Z, et al. Effects of dissolved oxygen on water quality and biofilms in the raw water distribution pipeline for a reserved water source [J]. Environmental Science: Water Research and Technology, 2022,8(10):2277−2288.

[43] Thach T T, Harada M, Oniki A, et al. Experimental study on the influence of dissolved organic matter in water and redox state of bottom sediment on water quality dynamics under anaerobic conditions in an organically polluted water body [J]. Paddy and Water Environment, 2017,15(4):889−906.

[44] 张朋超,赵有明,刘　翔,等.麻雀搜索算法优化极端梯度提升模型的岩石爆破块度预测 [J]. 科学技术与工程, 2024,24(24):10212−10219.
Zhang P C, Zhao Y M, Liu X, et al. Prediction for blasting fragmentation of rocks using extreme gradient boosting optimized bysparrow search algorithm [J]. Science Technology and Engineering, 2024,24(24):10212−10219.

[45] 麦嘉琪.基于机器学习的偶氮染料最大吸收波长规律研究 [D]. 上海:上海大学, 2023.
Mai J Q. Research on the Maximum Absorption Wavelength of Azo Dyes based on Machine Learning [D]. Shanghai: Shanghai University, 2023.

[46] 李振今,王志勇,刘晓彤,等.互花米草治理对黄河三角洲湿地利用及景观格局的影响——利用随机森林和改进的 CA−Markov 模型 [J]. 生态学报, 2024,44(18):8366−8382.
Li Z J, Wang Z Y, Liu X T, et al. Using random forest and improved CA−Markov model to evaluate the effect of Spartina alterniflora clearing on wetland use and landscape pattern in the Yellow River Delta wetland [J]. Acta Ecologica Sinica, 2024,44(18):8366−8382.

[47] 向　玲,陈锦鹏,付晓孟婷,等.基于 ViT 和 LSTM 的风速多步预测 [J]. 太阳能学报, 2024,45(9):525−533.
Xiang L, Chen J P, Fu X M T, et al. Multi−step wind speed forecasting based on VIT and LSTM [J]. Acta Energiae Solaris Sinica, 2024,45(9): 525−533.

[48] 汪　敏,荣腾飞,李　茜,等.基于可学习小波自注意力模型的海上风电功率超短期预测 [J]. 高电压技术, 2025,51(3):1422−1433.
Wang M, Rong T F, Li Q, et al. Ultra−short term prediction of offshore wind power based on learnable wavelet self−attention model [J]. High Voltage Engineering, 2025,51(3):1422−1433.

[49] Ling X, Lu G, Xue C. Environmental and anthropogenic factors affect bacterial community and nitrogen removal in the Yarlung Zangbo River [J]. Environmental science and pollution research international, 2022,29(56):84590−84599.

[50] Feng S, Wang M, Xu W, et al. The impact of emissions controls on atmospheric nitrogen inputs to Chinese river basins highlights the urgency of ammonia abatement [J]. 2024,10(37):2558.

[51] Ferrier R C, Whitehead P G, Sefton C, et al. Modelling impacts of land use change and climate change on nitrate−nitrogen in the River Don, North East Scotland [J]. Water Research, 1995,29(8):1950−1956.

[52] Ballard T C. Past and future changes in river temperatures and nitrogen export [M]. Stanford: Stanford University, 2021:1–100.

作者简介：汪　昊(1996−),男,辽宁锦州人,河北农业大学硕士研究生,主要从事农业面源污染研究.wangh04282014@163.com.