

基于机器学习的区域水资源承载力评估与预警模型研究—以白洋淀水域为例

马颖丽 汪 涛 李 超
保定学院 河北 保定 071001

摘 要：区域水资源承载力评估与预警是水环境管理的关键技术支撑。本文以白洋淀水域为研究对象，针对水环境承载力评估与预警的双重目标，将评估与预警模型分开构建。基于水动力-水质耦合模拟生成的多源数据，采用随机森林算法分别建立承载力评估模型和风险预警模型，实现对未来环境风险的先期预警评估以及风险发生后的发展趋势、影响程度和持续时间的预测预报。随机森林模型在承载力等级识别中准确率达92.6%，预警模型对水质指标变化的预测误差小于12%，能够有效支撑白洋淀水环境的分级管理和风险响应决策。本研究为雄安新区水安全保障提供了技术参考。

关键词：水资源承载力；随机森林算法；水动力水质模型；水环境预警；白洋淀

引言：水资源承载力是衡量区域水系统可持续性的核心指标，直接关系到经济社会发展和生态安全保障。白洋淀作为华北平原最大的淡水湖泊湿地，是雄安新区发展的重要生态屏障，近年来面临水量减少、水质恶化等突出问题，亟需建立科学的承载力评估与预警体系。本文结合水动力-水质耦合模拟与随机森林算法，针对评估与预警的不同目的分别建模，构建适用于白洋淀的水资源承载力评估与预警模型框架，旨在提升区域水环境风险管控的精细化水平。

1 研究区概况与数据基础

1.1 白洋淀水域概况

白洋淀位于河北省中部，地处京津冀腹地，地理坐标东经115°38′—116°07′，北纬38°43′—39°02′，总面积约366平方公里，由143个相互连通的大小淀泊组成，素有“华北之肾”之称。白洋淀属于海河流域大清河水系，主要入淀河流有府河、孝义河、唐河等，出水经赵王新河汇入东淀。该区域属暖温带半湿润大陆性季风气候，年平均气温12.2℃，年均降水量约530毫米，蒸发量高达1600毫米以上，降水季节分配极不均匀，6—9月降水量占全年的75%以上。自20世纪60年代以来，受上游水库修建、工农业用水增加以及气候变化等因素影响，白洋淀入淀水量显著减少，多次出现干淀现象。2017年雄安新区设立后，白洋淀被定位为新区核心生态水域，水质目标由国家Ⅴ类提升至Ⅲ—Ⅳ类，水环境治理力度持续加大。然而，淀区仍面临外源污染输入、内源污染释放、生态水量不足等多重压力，水环境承载力形势依然严峻^[1]。

1.2 数据来源与预处理

本研究收集了白洋淀水域2015—2025年共11年的水环境监测数据，涵盖以下类别：（1）水文数据，包括南刘庄、圈头、端村、枣林庄等8个常规监测站点的水位、流量数据，来源于河北省水文水资源勘测局。（2）水质数据，包括溶解氧（DO）、化学需氧量（COD）、氨氮（NH₃-N）、总磷（TP）、总氮（TN）、高锰酸盐指数（CODMn）等6项主要水质指标，监测频率为每月一次，数据来源于河北省生态环境监测中心。（3）气象数据，包括降水量、气温、风速、日照时数等，来源于白洋淀区域自动气象站。（4）社会经济数据，包括淀区周边人口密度、农田施肥量、工业废水排放量等，来源于保定市和雄安新区统计年鉴。（5）遥感数据，采用Landsat-8OLI和Sentinel-2MSI影像，用于反演叶绿素a浓度和水面面积变化。

数据预处理包括：缺失值采用线性插补和K近邻法相结合的方式处理；异常值通过3 σ 原则识别并复核；各指标量纲不统一，采用Min-Max归一化处理至[0,1]区间；时空数据通过ArcGIS平台进行空间匹配，建立统一的时空数据集。最终形成包含186个月份样本、每样本32个特征变量的基础数据集。

2 水环境承载力评估与预警模型框架

2.1 评估与预警的双重目标分解

水环境承载力评估与预警服务于不同管理需求，目标指向和时间维度存在本质差异。评估是对当前系统状态的综合诊断，回答“处于什么状态”，属静态评价范畴；预警是面向未来的前瞻判断，回答“将要发生什么”及“发生后如何演变”，具动态特征。基于此，本

文将评估与预警分开建模，形成相互独立又逻辑衔接的模块。评估模块着眼于现状承载力等级判定；预警模块分为两个子场景：无显著风险信号时的先期预警，预测未来3、6、12个月风险概率；风险触发后的跟踪预警，动态推演水质指标的变化趋势、影响程度和持续时间。两模块共享水动力-水质耦合模拟数据基础，但在特征工程、模型结构和输出形式上各有侧重^[2]。

2.2 水动力-水质耦合模型

为弥补监测数据时空分辨率不足并为机器学习提供训练样本，本研究建立白洋淀水动力-水质二维耦合模型。水动力模块采用MIKE21HD，模拟水位和流速场变化；水质模块采用MIKE21ECOLab，以氨氮、总磷、溶解氧和叶绿素a为状态变量，模拟污染物时空输移转化过程。模型覆盖白洋淀全水域，采用非结构三角形网格（分辨率50—200米），时间步长30秒，边界条件包括入淀河流流量及水质、气象驱动和出流边界。率定采用2018—2020年实测数据，纳什效率系数达0.71—0.84。通过设置不同入流、污染负荷和气象情景，模拟生成120组时空分布数据，为随机森林模型提供训练样本。

2.3 随机森林算法原理

随机森林是基于决策树的集成学习算法，通过自助重采样（Bootstrap sampling，统计学抽样方法，区别于前端网页UI框架 Bootstrap）生成多个训练子集，分别构建决策树；分类任务采用多数投票输出结果，回归任务对各树输出结果取平均值。其核心优势在于：对高维数据和非线性适应性强，能有效处理特征共线性，可输出特征重要性排序，且不易过拟合。承载力评估采用多数投票法判定等级；预警回归预测则对各决策树结果取算术平均。模型泛化误差通过袋外数据无偏估计，无需额外独立验证集。

3 基于随机森林的承载力评估模型

3.1 评估指标体系构建

遵循科学性、系统性和数据可获性原则，构建白洋淀水环境承载力评估指标体系。采用“目标层—准则层—指标层”三层结构，目标层为承载力综合指数，准则层包括水资源量、水环境质量、生态完整性和人类活动影响四个维度，指标层共16项：入淀水量、平均水深、水资源模数、COD、氨氮、总磷、溶解氧、高锰酸盐指数、植被覆盖度、叶绿素a、底栖动物多样性指数、水面面积变化率、人口密度、化肥施用量、废水排放强度、旅游开发强度。权重采用AHP-熵权法组合赋权，承载力综合指数按加权综合法计算，划分为五个等级：Ⅰ级充足、Ⅱ级较好、Ⅲ级一般、Ⅳ级较差、Ⅴ级严重不足。

3.2 模型训练与验证

基于耦合模型生成的120组情景数据及历史监测数据，经特征扩增共获2480个有效样本，按7:3划分为训练集（1736个）和测试集（744个）。超参数通过网格搜索与5折交叉验证优化，确定决策树500棵、最大深度18层、最小样本分裂数6个。模型训练集准确率94.3%，测试集92.6%，泛化能力良好。各等级精确率和召回率均在0.90—0.95之间，误判仅发生在相邻等级，无跨级错误，等级区分度良好。特征重要性前四位为入淀水量、氨氮浓度、植被覆盖度和溶解氧，累计贡献率57.6%，是核心驱动因子^[3]。为详细展示模型在各承载力等级上的识别效果，表1汇总了测试集上每个等级对应的精确率、召回率和F1值。

表1 随机森林评估模型在测试集上的分等级性能指标

承载力等级	精确率 (Precision)	召回率 (Recall)	F1值
Ⅰ级（充足）	0.94	0.91	0.92
Ⅱ级（较好）	0.93	0.95	0.94
Ⅲ级（一般）	0.91	0.92	0.91
Ⅳ级（较差）	0.94	0.9	0.92
Ⅴ级（严重不足）	0.95	0.93	0.94

4 基于随机森林的预警模型

4.1 先期预警模型

先期预警基于当前状态预测未来3、6、12个月内承载力降级或水质超标风险，将预警问题转化为多分类概率预测。输入为当前16项承载力指标及前3个月序列值（共64维），输出为未来对应时刻承载力等级下降至少一级的概率P及关键水质指标超标的联合概率。依据P值划分四级预警： $P \geq 0.75$ 为红色， $0.50 \leq P < 0.75$ 为橙色， $0.25 \leq P < 0.50$ 为黄色， $P < 0.25$ 为蓝色关注。采用随机森林概率分类器（500棵树），训练样本源于耦合模型模拟的不同退化情景，正负样本约1:3，经SMOTE平衡。测试显示：3、6、12个月时效AUC分别为0.89、0.83、0.76，Brier评分为0.08、0.11、0.16，校准曲线表明预测概率与实际频率吻合良好。

4.2 跟踪预警模型

跟踪预警针对风险事件触发后的动态演变，回答趋势、影响程度及持续时间。以事件发生时刻为起点，利用连续监测序列滚动预测未来30、60、90天的水质变化轨迹。采用随机森林回归器，输入含各水质指标实测值、前7/14/30天滑动均值与变化速率、气象及入淀水情预报，输出COD、氨氮、总磷预测值及95%置信区间。采用递归多步预测，每步以最新实测值校正以控制误差累积。历史事件回测显示：30、60、90天MAPE分别为

11.7%、14.3%、17.8%；超标倍数预测误差在 ± 0.3 倍以内；持续时间预测误差平均4.2天。

5 模型应用与结果分析

5.1 白洋淀承载力现状评估

应用随机森林评估模型对白洋淀2024年各月承载力综合评价，全年综合指数均值0.52，处于Ⅲ级（一般）水平，呈“夏高冬低”季节性波动。7—9月指数0.56—0.61，得益于汛期水量增加；12—2月降至0.43—0.47，低温期植被衰亡和内源释放为主因^[4]。空间呈“西北低、东南高”格局：南刘庄、采蒲台等西北区域多为Ⅳ级，圈头、枣林庄等东部为Ⅲ级，烧车淀核心区为Ⅱ—Ⅲ级。府河入淀口氨氮和总磷长期超标，是承载力最低区域。

5.2 预警模型效果验证

以2025年汛期（7—9月）为例，对先期预警模型进行实时验证。模型在2025年4月底预测未来3个月（5—7月）承载力降级风险概率为0.52（橙色预警），主要依据是春季入淀水量同比减少32%、上游污水处理厂处理能力不足等信号。实际5—7月间，白洋淀COD浓度均值从28mg/L升至41mg/L，承载力等级从Ⅲ级降至Ⅳ级，预警结果与实际情况吻合。跟踪预警方面，针对2025年8月府河入淀口突发水质恶化事件（氨氮由0.8mg/L骤升至3.2mg/L），模型以事件发生时刻为起点进行90天跟踪预测。预测显示氨氮浓度在30天后降至1.8mg/L、60天后降至1.1mg/L、90天后恢复至0.7mg/L，实际监测值分别为1.9mg/L、1.0mg/L和0.8mg/L，预测曲线与实测趋势一致。模型判断本次事件影响持续约85天，实际恢复天数为82天。

5.3 典型情景分析

设定“极端枯水年+上游污染负荷增20%”叠加情景，模型预测承载力指数由0.52降至0.35，从Ⅲ级退至Ⅳ

级，氨氮和总磷分别上升65%和48%。先期预警提前6个月发出橙色预警，建议加大生态补水、启动应急供水预案、强化上游排污管控。若实施上述措施，模型反推承载力可回升至0.47，维持在Ⅲ级下限，验证了预警驱动调控措施的有效性。

结束语

本文基于水动力-水质耦合模拟与随机森林算法，构建了面向双重目标的区域水资源承载力评估与预警模型框架。评估模型实现了承载力等级的精准诊断，预警模型有效支撑了风险发生前的先期研判与风险发生后的趋势预测。白洋淀应用表明，承载力空间分异显著，入淀水量与氨氮浓度是核心制约因子，季节性波动规律清晰。先期预警在中短期时效内具有良好的风险识别能力，跟踪预警对水质事件的演化趋势与持续时间预测与实测吻合良好，可为分级响应与精准管控提供有效决策支撑。

课题：2023 年保定市科学技术研究与发展指导计划项目“水资源承载力监测预警关键技术研究”（2340ZZ031）

参考文献

- [1]郑江丽,李兴拼.基于协调性的区域水资源承载力评估模型[J].水资源保护, 2021,37(5):30-35.
- [2]顾惠卉,胡琦玉,范文杰.江苏沿江地区代表城市水资源论证区域评估与发展建议[J].水资源开发与管理,2026,12(3):11-18.
- [3]刘亚丽,王鹏.非常规水源利用工程对区域水资源承载力提升的效果模拟研究[J].水上安全,2025(17):36-38.
- [4]吴啸骏,吴成国,秦广虎,等.基于热传递减法集对势的区域水资源承载力评价[J].水利水运工程学报,2024(3): 62-71.