能昌信,张慧敏,孙晓晨,等.机器学习模型精细表征填埋场渗漏风险及其不确定性 [J]. 中国环境科学, 2025,45(9):4986-4996.
Nai C X, Zhang H M, Sun X C, et al. Machine learning models for fine characterization of landfill leakage and uncertainty analysis [J]. China Environmental Sciences, 2025,45(9):4986-4996.

机器学习模型精细表征填埋场渗漏风险及其不确定性

能昌信[1,2],张慧敏[1,2],孙晓晨[1*],徐 亚[2,3*],刘景财[2],刘玉强[2] (1.山东工商学院信息与电子工程学院,山东 烟台 264005；2.中国环境科学研究院,北京 100012；3.甘肃省工业废弃物循环与调控重点实验室,甘肃 兰州 730000)

摘要：针对传统基于解析解与 Monte Carlo 耦合的填埋场地下水污染风险预测方法难以准确刻画渗漏不确定性的问题,提出了一种高性能仿真模拟与不确定性分析耦合的填埋场渗漏风险及其不确定性表征技术.为解决传统不确定性分析框架下,高仿真模拟数值模型计算负荷大,不确定性分析效率低的问题,引入粒子群优化算法(PSO)与极限梯度提升算法(XGBoost)模型,通过学习精细模拟模型计算获得的多组样本,建立替代模型来预测;同时,为了解决 PSO-XGBoost 对抽样数量的过度依赖,提出了改进的抽样方法.实验结果表明,与以 Landsim 为代表的解析解-monte carlo 耦合算法相比,本研究构建基于数值模拟模型样本的机器学习替代模型更好的捕捉了参数不确定性对结果的影响,预测的污染浓度分布区间平均为传统模型的 1.36 倍;基于 PSO-XGBoost 的替代模型及改进的抽样方法,相比于其他机器学习模型,预测精度(R^2=0.99)和稳定性更好,且模型训练阶段计算时间减少 70%,大幅提升了不确定性分析的效率.基于该模型的应用研究表明,第 10a 特征污染物 COD 超标的概率达到 56.43%,存在超标可能的距离为 54m,表明研究区存在一定污染风险.研究成果为填埋场渗漏污染风险及不确定性分析提供了高效、精准的科学方法,并为进一步开展相关研究提供了借鉴.

关键词：不确定性分析；替代模型；数值模拟；填埋场；风险评估
中图分类号：X523 **文献标识码**：A **文章编号**：1000-6923(2025)09-4986-11

Machine learning models for fine characterization of landfill leakage and uncertainty analysis. NAI Chang-xin[1,2], ZHANG Hui-min[1,2], SUN Xiao-chen[1*], XU Ya[2,3*], LIU Jing-cai[2], LIU Yu-qiang[2] (1.School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China；2.Chinese Research Academy of Environmental Sciences, Beijing 100012, China；3.Key Laboratory of Industrial Waste Recycling and Regulation, Gansu Province, Lanzhou 730000, China). *China Environmental Sciences*, 2025,45(9)：4986~4996

Abstract：This study addresses the limitations of traditional landfill groundwater pollution risk prediction methods by integrating high-performance simulation with uncertainty analysis to more accurately capture the effects of concealed leakage. The limitations of high computational load and low efficiency were addressed by combining a Particle Swarm Optimization (PSO) algorithm with an Extreme Gradient Boosting (XGBoost) model. By learning from multiple samples generated by detailed simulation models, a surrogate model was established to improve prediction efficiency. Additionally, an improved sampling method was proposed to reduce the reliance of PSO-XGBoost on sample size. The results indicate that the machine learning surrogate model developed in this study better captures the impact of parameter uncertainty compared to traditional analytical–Monte Carlo methods such as Landsim. The predicted pollution concentration range is 1.36 times greater than that of conventional models. The PSO-XGBoost surrogate model, combined with the improved sampling method, outperforms other machine learning models in prediction accuracy (R^2= 0.99) and stability, while reducing the computational time for model training by 70%, significantly improving the efficiency of uncertainty analysis. The application research based on this model shows that the probability of the characteristic pollutant COD exceeding the standard in the 10th year reaches 56.43%, and the possible distance of exceeding the standard is 54meters, indicating that there is a certain pollution risk in the study area. These findings provide an efficient and precise scientific method for analyzing landfill leakage risks and uncertainties and offer valuable references for further related research.

Key words：uncertainty analysis；surrogate model；numerical simulation；landfill；risk assessment

填埋作为固体废物处理的无害化手段,凭借其低成本和低技术门槛的特点,是当前固体废物(SW)治理的主要方式[1-2].在全球范围内,人口增长、城市化进程加快以及经济发展共同导致废物的产生量大幅增加[3].在此背景下,垃圾填埋场被广泛用于固体废物的处理,尤其是在中国、墨西哥等发展中国家,填埋在固体废物处理中的占比更大[4].然而,研究表

收稿日期：2025-02-21
基金项目：国家重点研发计划项目(2023YFC3708902)
* 责任作者, 讲师, 202414174@sdtbu.edu.cn; 研究员, xuya@craes.org.cn

明,大多数填埋场都存在渗滤液防渗系统和排水系统堵塞等问题,且垃圾填埋场产生的渗滤液中含有大量的有毒有害物质,如重金属[5]、有机物[6]等.如果处理不当,会对地下水造成污染,进而对垃圾填埋场附近的人类健康构成威胁[7].

鉴于上述潜在风险,亟待开展填埋场的风险评估和预警工作,深入探究渗滤液的产生、泄漏及其对环境质量和人类健康的潜在影响,指导渗漏及其污染风险的管控决策[8-9].其中,参数不确定性分析是地下水风险评估中的关键[10],也是本文研究的重点之一.目前,蒙特卡罗(Monte Clarlo, MC)取样方法因其广泛的适用性以及方法独立性,已成为最常用的不确定性分析手段[11].然而,其收敛速度慢,需要大量模拟随机生成的参数场,计算负荷巨大.为缓解这一挑战,Lahmar S 等[12]结合人工神经网络(Artificial Neural Network, ANN)和克里金替代模型,提出了自适应采样方法,在保证模型精度的同时显著降低了计算成本.陈骏骏等[13]采用 LHS 方法,通过实施分层随机抽样策略保证样本点分布的均匀性和代表性,显著提升了采样效率并减轻了计算负荷.尽管上述工作已经显著改善了传统不确定性分析方法的效率,然而在面对运行周期较长的高精度地下水模型时,所需的计算时间仍然过长,因此亟需开发更高效、快速的解决方法以应对这一挑战.

近年来,极限梯度提升算法(eXtreme Gradient Boosting, XGBoost)模型[14]在替代模型的构建中展现出了巨大潜力,被广泛应用于地下水不确定性分析与风险评估等领域[15].例如,董广齐等[16]采用 XGBoost 方法拟合涉及多维不确定变量的模拟模型,评估其在复杂情况下的性能,结果表明 XGBoost 能够有效逼近模拟模型的输入输出关系.此外,吉林大学卢文喜教授团队[17-19]提出了一种基于 XGBoos 算法的替代模型,该模型可用于解决参数不确定条件下的地下水污染源溯源问题,并通过一系列相关实验验证了模型的可靠性,为地下水污染修复及风险评估提供了科学依据.然而,现有研究主要集中于低维问题,高维不确定性分析仍面临“维数灾难”的挑战[20],而且现有方法难以提供有效的解决方案.

因此,本文旨在借助 XGBoost 算法,针对地下水模型中的高维不确定性分析问题,寻求更为高效的解决方案.本文提出了一种融合地下水高性能仿真模拟数值模型与 PSO−XGBoost 的多维输入输出浓度回归模型,通过粒子群优化算法(Particle Swarm Optimization, PSO)对 XGBoost 模型的超参数进行优化,显著提升了模型在高维数据处理中的精度与效率.研究进一步将该方法应用于复杂地下水模拟系统的高维不确定性分析问题预测中,建立了高维不确定性参数和化学需氧量(Chemical Oxygen Demand, COD)浓度分布之间的映射关系,进一步提高了复杂模型的应用能力.

1　模型及方法

1.1　过程模型构建与求解

过程模型是模拟渗滤液渗漏后在地下水中迁徙扩散的过程.其中渗漏过程采用 HELP 模型模拟[24],计算渗漏液通过漏洞进入包气带与地下水时的渗漏速率和污染组分浓度(渗滤液源强).迁移扩散过程则采用有限差分法求解地下水水流方程和地下水溶质运移方程建立地下水高性能仿真模拟数值模型,进而模拟污染物在地下水中的迁移扩散过程.

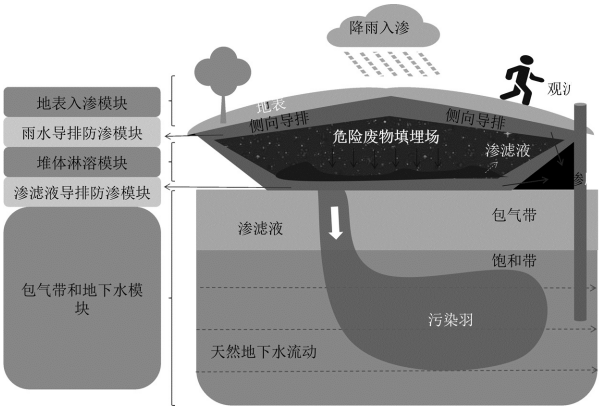


图 1　填埋场渗滤液渗漏风险的演化过程

Fig.1　Temporal evolution of leachate leakage risk in landfill systems

基于对地下水文地质条件的分析,并结合达西基础渗流方程,得到地下水流的数学模型,用于模拟研究区内地下水流动过程[25-26],如下式所示:

$$\begin{cases} \dfrac{\partial}{\partial x}\left(K_{xx}\dfrac{\partial H}{\partial x}\right)+\dfrac{\partial}{\partial y}\left(K_{yy}\dfrac{\partial H}{\partial y}\right)+\dfrac{\partial}{\partial z}\left(K_{z}\dfrac{\partial H}{\partial z}\right)+\omega=\mu_s\dfrac{\partial H}{\partial t} & (x,y,z)\in D,t>0 \\ H_0(x,y,z,t)_{t=0}=H_0(x,y,t) & (x,y,z)\in\Omega \\ H(x,y,z,t)|_{\tau_1}=H_1(x,y,t) & (x,y,z)\in\tau_1,t>0 \\ k_n\dfrac{\partial H}{\partial n}|_{\tau_2}=q(x,y,t) & (x,y,z)\in\tau_2,t>0 \end{cases} \tag{1}$$

式中:D 表示地下水渗流区域;K_{xx}、K_{yy} 和 K_{zz} 分别代

表在 x、y 和 z 方向上的渗透系数分量;H 代表水位标高;ω代表单位体积流量,即源汇项;μ_s 代表贮水率;$H_0(x,y,z,t)$代表初始时刻水头函数;t 代表时间;$H_0(x,y,t)$代表初始地下水水头; $H_1(x,y,t)$代表第一类边界已知地下水水头函数;τ_1 代表第一类水头边界条件; τ_2 代表第二类流量边界条件;n 指的是边界τ_2上的外法线方向;$q(x,y,t)$为已知流量函.

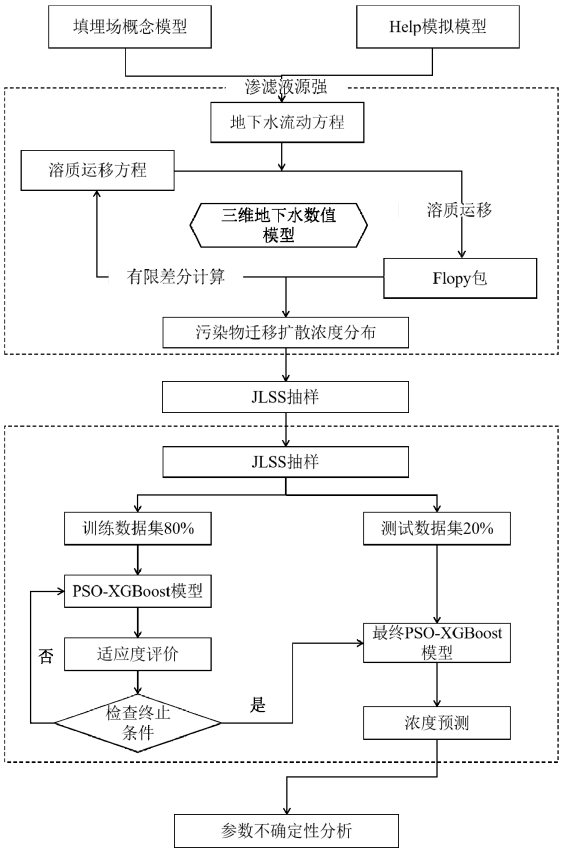


图 2 技术路线

Fig.2 Schematic of the research methodology

采用三维溶质运移方程描述生活垃圾填埋场渗滤液及其中所含污染物在地下水中的扩散与迁移过程,计算浓度时考虑了多个参数取值.这些参数取值的数据来源于现有资料及实地测量.地下水溶质运移的计算模型可以表示如下:

$$\begin{cases} \dfrac{\partial \theta C}{\partial t}=\dfrac{\partial}{\partial x_i}\left(D_{ij}\dfrac{\partial C}{\partial x_j}\right)-\dfrac{\partial(\theta C_{v_i})}{\partial x_i}+S & (i,j=1,2,3) \\ C(x,y,z,t)|_{t_0}=C_0(x,y,z) & \\ H(x,y,z,t)|_{\tau_1}=C_1(x,y,z,t) & (x,y,z)\in\tau_1 t_0 \\ D_{ij}\dfrac{\partial C}{\partial x_i}n_i|_{\tau_2}=-q_2(x,y,z,t) & \end{cases} \tag{2}$$

式中:C 表示溶质浓度,m/s;D_{ij} 为水动力弥散系数张量,m²/s;v_i 为平均孔隙流速,m/s;S 为源汇项;θ为含水率;C_0 为初始时刻的浓度分布;$C(x,y,z,t)$为空间位置的已知函数;第一类边界τ_1 上的浓度 $C_1(x,y,z,t)$是已知函数;第二类边界τ_2上的水动力弥散通量 $q_2(x,y,z,t)$是已知函数.

本研究利用 Python 中的 FloPy 模块构建并求解地下水流动和溶质运移模型,通过数值模拟预测典型污染物(如 COD 等)浓度的动态变化.

1.2 JLSS 抽样策略

在本文设计的测试算例中,考虑了抽水井并对源汇项特征进行了处理.在抽水实验初期,抽水井周边地下水位波动显著,而远离抽水井的边界区域水位变化较为平缓[27].因此,合理选取时空采样点对模型的训练至关重要.

为了获得更精确的水头预测数据,本文提出了一种能够灵活适应各种复杂情况的采样策略,即 LHS 采样和时空加密采样策略联合的联合拉丁时空采样(JLSS)策略,并基于地下水高性能仿真模拟模型完成采样.

1.2.1 空间采样策略 当源汇项中包含抽水井时,地下水水位的变化情况主要取决于与抽水井的距离.具体而言,与抽水井的距离越短,地下水水头的变化越剧烈.反之,当远离抽水井时,水头变化则呈现出逐渐平缓的趋势.鉴于这种空间变化特性,为了准确捕捉地下水位的动态变化,本文采用空间局部加密采样(LRS)策略,将空间区域划分为内部区域和边界区域.边界区域采用统一的 LHS 采样,抽水井附近则采用 LRS 策略,以精准捕捉水位的剧烈变化.图 3 展示了在多个抽水井共存的情况下源汇项采样分布布局.通过这一精细的采样设计,能够更有效地收集和分析地下水位的动态数据,为后续模型的训练和预测提供坚实基础.

1.2.2 时间采样策略 鉴于地下水问题与时间维度的密切相关性,合理的时间采样策略与空间采样同样重要.图 4 展示了两种不同的时间区域采样方式,图中每一个黑点表示一次数值模拟的采样时刻,用于提取污染物浓度值.

根据对区域地下水问题的先验知识[28],含水层的水头会随着时间的推移逐渐趋于一个相对稳定的状态.然而,当含水层中存在抽水井时,特别是在井

的周边区域及初始阶段,水头会发生显著变化.为了准确捕捉这些复杂变化,需要在空间上采用 LRS 策略的同时,在时间上执行类似策略,如图 4(b)所示.综上,空间采样与时间采样相结合,构建出了一个全面且精细的时空采样点集.该点集即涵盖了地下水系统的空间特征,还准确反映了水头随时间变化的趋势.该数据集将用于替代模型的训练,进一步提升地下水动态预测的准确性和可靠性.

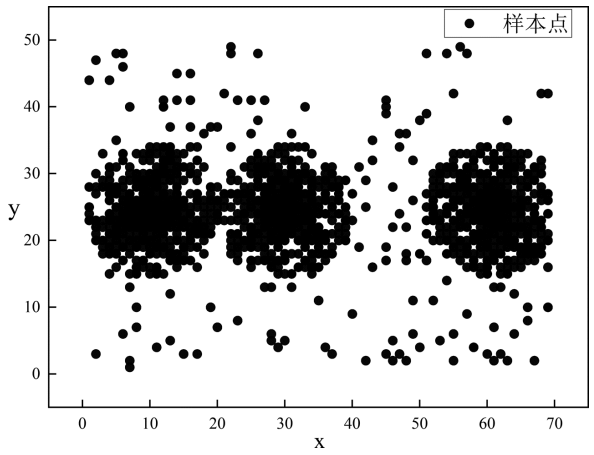


图 3　多井加密采样结合拉丁超立方采样的样本点示意

Fig.3　Schematic diagram of sample points from multi-well refinement combined with Latin hypercube sampling

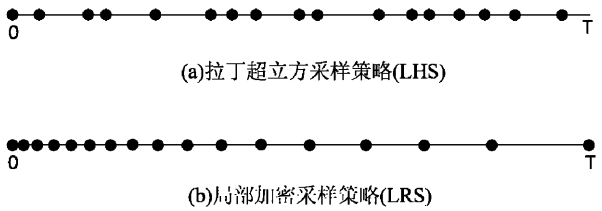


图 4　时间区域采样点示意

Fig.4　Schematic diagram of time domain sample point

1.3　基于 PSO-XGBoost 的替代模型构建方法

本文提出的 PSO-XGBoost 模型充分利用 XGBoost 算法在分类及回归任务中展现出的高效学习能力,同时结合了 PSO 算法的全局搜索寻优能力,旨在解决参数不确定性分析问题.通过采用 XGBoost 的线性增强器(XgbLinear),显著提升模型在处理稀疏数据和提升计算效率方面的能力,能够在保持精度和速度的前提下,用于大规模的参数不确定性分析[29].

PSO-XGBoost 模型的开发流程涵盖从数据处理到模型部署的各个步骤,详述如下:

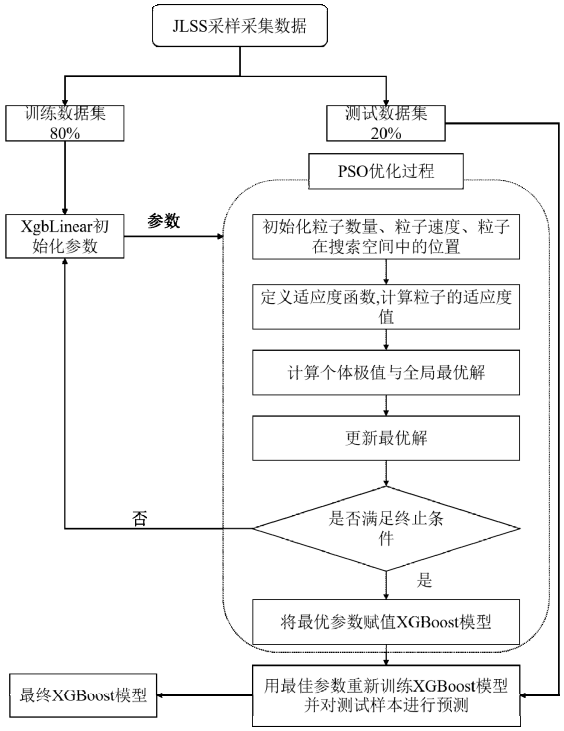


图 5　PSO-XGBoost 模型开发流程

Fig.5　Development process of PSO-XGBoost model

(1)数据分割　将采样得到的模拟数据按 8:2 比例划分为训练集和测试集,确保模型在测试阶段能客观地评估其泛化能力.

(2)特征选择和初始化　通过 XGBoost 算法自动选择重要特征,并对模型的关键超参数进行初始化设置.

(3)超参数优化　应用 PSO 算法对学习率、最大树深等超参数进行迭代优化,通过粒子在搜索空间中的位置更新,逐步逼近最优解.

(4)适应度评估　采用交叉验证等评估策略,在每轮迭代中评价模型性能并动态调整参数,确保模型在不同数据组合下的鲁棒性.

(5)终止判定　设置最大迭代次数和性能阈值,当满足收敛条件后,输出当前最优超参数组合.

在满足终止条件后,构建 PSO-XGBoost 模型.随后,将该模型用于预测和评估不同参数下的污染物浓度,并通过测试数据集验证其在浓度预测中的性能.

本研究选取了 9 个不确定性参数作为模型的输入变量,具体包括渗透系数、孔隙度、纵向弥散系数、水平方向横向弥散度、垂直方向横向弥散度、空间位置(x,y,z)及时间变量 t.模型的输出变量为监测井

的污染物的浓度值.

1.3.1 粒子群优化算法(PSO) 针对 XGBoost 模型超参数空间广泛且参数相互作用复杂的问题,引入了粒子群优化算法(PSO),其特点在于其能够在超参数空间中进行全局搜索,高效找到 XGBoost 模型的最优超参数集,显著提高模型性能和计算效率[30-31].作为一种元启发式算法,PSO 通过个体之间的协作与信息共享,有效避免局部最优,从而更快速地找到全局最优解[32].粒子群算法的最优搜索过程描述如下[33-34]:

(1)随机初始化粒子的数量、速度和位置.

(2)定义适应度函数,每个粒子根据其当前位置计算适应度值,寻找各自的个体最优解和全局最优解.

(3)比较并更新全局最优解.

(4)每个粒子的速度和位置按照以下公式进行更新:

$$V_{id} = \omega V_{id} + C_1 a(P_{id} - X_{id}) + C_1 b(P_{gd} - X_{id}) \quad (3)$$

$$X_{id} = X_{id} + V_{id} \quad (4)$$

式中:C_1 和 C_2 是学习因子,分别体现粒子的个体学习和社会学习;P_{id} 表示第 i 个粒子的局部最优的第 d 维,而 P_{gd} 表示全局最优的第 d 维;a 和 b 代表(0,1)之间的随机数;ω表示惯性权重系数,其线性缩减过程可描述为:

$$\omega = \frac{\omega_{\max} + (\mathrm{iter} - \mathrm{iter}_i) \times (\omega_{\max} - \omega_{\min})}{\mathrm{iter}} \quad (5)$$

式中:iter 表示最大迭代次数;iter_i表示当前迭代次数;$\omega_{\max}$ 和$\omega_{\min}$ 分别表示ω的最大值和最小值.

(5)持续迭代更新,直到满足终止条件,输出最优解.

PSO 在 XGBoost 中超参数优化中展现了高适配性,特别是在处理稀疏数据和具有线性关系的任务时效果显著.通过全局优化迭代次数、学习率、正则化参数等关键超参数[35],显著提升了模型预测性能,充分展现了其在微调 XGBoost 超参数方面的高度适应性和优势.

1.3.2 评价指标 模型验证的核心在于通过对其预测结果的误差进行精确估计,并运用一系列统计指标来全面评估其性能.随机从测试数据集中选取 200 组样本,利用均方误差(MSE)、均方根误差(RMSE)和决定系数(R^2)3 个性能指标来评价模型的性能[36-37].具体计算公式如下:

MSE(均方误差)表示数据的离散程度,取值越小则说明替代模型的性能更加稳定,计算公式为:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (6)$$

式中:n 为样本个数,y_i表示传统方法对数值模型的求解结果,$\overline{y}_i$表示替代模型的预测结果.

确定性系数(R^2)用于评估模型对数据的拟合程度,表示预测值与真实值之间的相关性.其计算公式为:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{\sum_{i=1}^{n}(\overline{y}_i - y_i)^2} \quad (7)$$

式中:$\overline{y}_i$ 表示传统方法对数值模型求解结果的平均值.

RMSE 能够更直观地反映预测值与真实值之间的差异,值越小,模型的预测效果越好,计算公式为:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)} \quad (8)$$

2 案例应用

本研究以北京市大兴区安定镇生活垃圾填埋场(以下简称“安定填埋场”)为研究对象,开展地下水污染迁移建模与风险评估工作.该填埋场始建于 1996 年,位于华北冲积平原核心区,总占地面积约 49hm^2,其中填埋库区面积约为 38hm^2,地势平坦,海拔约 22m.区域土壤类型主要为淤泥质土和黏土,具备一定的渗透性与储水性.

安定填埋场所处区域为典型的自由潜水含水层,地下水埋深浅,主要接受大气降水与地表径流补给.年均降水量约为 568mm,降雨主要集中在 6~8 月.长期运行过程中,由于防渗系统局部老化及地下水位波动等因素,存在渗滤液渗漏进入地下水系统的潜在风险[38].

根据北京市历年环境监测报告,填埋场周边地下水中化学需氧量(COD)浓度常年处于较高水平,部分观测井已多次超出《地下水质量标准》(GB/T 14848-2017)中的限值[39].COD 作为衡量有机污染负荷的关键指标,广泛应用于生活垃圾及危险废物填埋场污染评价中,具有检测简便、响应灵敏等优势,因此在本研究中被选定为主要污染

物模拟因子.

为系统模拟污染物在地下水系统中的迁移规律,本文基于实地调查与文献数据,构建了三维地下水流-质耦合高性能数值模型.模拟区域被概化为一个 240m×143m×8m 的三维长方体结构,水文地质参数参考现场实测与工程文献设定,具体如表 1 所示.模型网格划分为 70 行×49 列×4 层,共计 13,720 个网格单元.

表 1　研究区水文地质参数的详细描述

Table 1　Detailed description of hydrogeological parameters in the study area

参数	描述	取值	单位
K	水平水力传导系数	N(0.30,0.05)	m/d
α	纵向弥散度	N(10,1)	m
trpv	垂向弥散系数	N(3,0.5)	m
trpt	横向弥散系数	N(0.3,0.05)	m
ζ	孔隙度	N(0.3,0.05)	-
μ	给水度	N(0.2,0.09)	-

含水层被视为均质各向同性介质,区域水流由西向东,边界条件设置如下:西侧(Γ_1)和东侧(Γ_2)为定水头边界(分别为 48m 与 45m),南侧(Γ_3)和北侧(Γ_4)为不透水边界,东北与西南两侧为弱透水边界,以模拟填埋场边界阻隔特性.

为模拟污染源行为及控制抽水方案,模型中设置一口注水井及两口抽水井(井 1、井 2),布设位置依据实际工程案例进行优化设计.污染源设于模型中部,采集安定填埋场 6 个不同时间的渗滤液样品,进行 COD 浓度测试,平均浓度为 4060mg/L,因此取该值为渗漏源强浓度,泄漏速率为 0.58m^3/d,模拟时长为 20 年,划分为 5 个阶段(1a、4a、10a、14a、20a).抽水井抽水速率分别为 0.46m^3/d 和 0.48m^3/d,参考北京市南部地下水调控工程运行经验.污染源及观测井布置如图 6 所示.

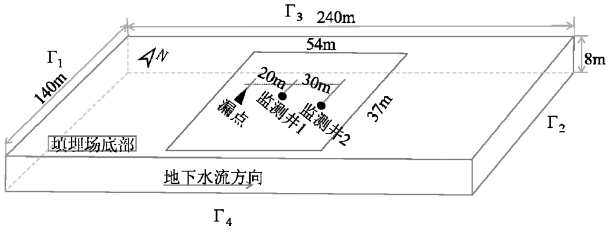


图 6　研究区示意

Fig.6　Schematic diagram of the study area

本研究以 COD 为主控因子,开展污染迁移路径与时空变化分析,为城市填埋场地下水污染防控与风险管理提供理论支撑.

3　结果与分析

3.1　模型验证

通过与 Landsim 软件分析结果的对比,验证了本研究提出的三维地下水仿真模型的准确性及其应用可行性.在案例场景中,分别使用本模型与 Landsim 的 Monte Carlo 模块对地下水污染进行的 Monte Carlo 模块对地下水污染进行参数不确定分析.以三维模型第一层第 10a 的污染物浓度预测结果为例,图 7 展示了两种模型在统计指标上的对比.结果表明,本研究模型与 Landsim 在不确定性分析结果上基本一致,能够有效开展参数不确定性分析.

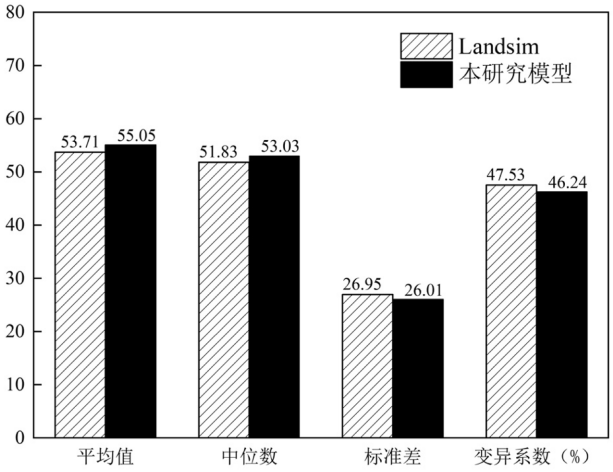


图 7　本研究模型与 Landsim 不确定性分析的比较

Fig.7　Comparison of the uncertainty analysis between the proposed model and Landsim

实验结果如表 2 所示,直接使用模拟模型进行不确定性分析需要进行 1000 次计算,总耗时约为 33.3h.而 PSO-XGBoost 替代模型仅需 300 次模拟(约 10h)以获取输入-输出样本,且替代模型的计算耗时约为 1min,节省了约 70%的计算时间,极大提高了不确定性分析的效率,对地下水的风险评估具有重要意义.

为进一步验证模型性能,本研究对比了 1 号和 2 号监测井第 10a 污染物浓度的预测结果,并绘制了浓度累积分布及其拟合曲线(见图 8).结果显示,与 Landsim 模型相比,本研究构建的三维模型具有更宽

的浓度分布范围,预测的污染浓度分布区间平均为传统模型的 1.36 倍,且拟合曲线更平缓.

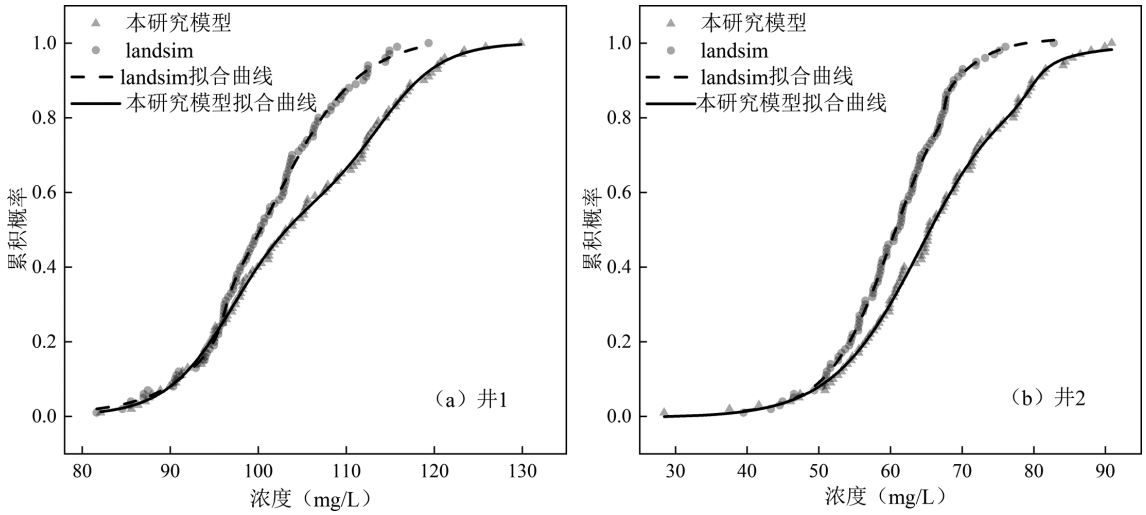


图 8 本研究模型与 Landsim 模型的浓度累积分布及拟合曲线

Fig.8 Cumulative concentration distributions and fitted curves of the proposed model and landsim Model

表 2 数值模型与 PSO−XGBoost 模型的计算时间对比

Table 2 Computational time comparison between the numerical model and PSO−XGBoost model

方法	模拟次数	抽样单次模拟时间(h)	总模拟时间(h)	模型计算时间(h)	节省时间(%)
模拟模型	1000 组	0.033	33.3	–	–
PSO−XGBoost	300 组	0.033	10	0.017	70%

通过与 Landsim 软件的多维度对比,本研究的三维模型在处理地下水污染的参数不确定性方面表现出更高的敏感性和准确性.此外,三维模型还考虑了空间层次的动态特性,增强了对污染物浓度随空间和时间变化的捕捉能力,为地下水污染预测与管理提供了更有力的支持.

3.2 模型精度分析

替代模型的构建是从数据集的收集开始.首先,基于安定镇生活垃圾填埋场的地下水高性能仿真模拟模型,采用 JLSS 抽样方法在不确定变量分布范围内抽取 1000 组样本,生成在不确定条件下的 1000 组污染物数据.随后,将数据集按 80%训练集和 20%测试集划分,以 9 个不确定性参数作为输入向量,COD 浓度作为输出向量,建立 PSO−XGBoost 替代模型,对不同参数下的 COD 浓度进行预测.利用训练集构建模型后,通过测试集评估其性能.图 9 展示了 PSO−XGBoost 模型与未优化 XGBoost 模型对 COD 浓度的预测结果及误差分布.从对比图中可以看出,相比于未优化模型,PSO−XGBoost 模型的误差更集中在零附近,呈正态分布,显示出更高的预测一致性和精度.

为进一步验证所提出 JLSS 抽样策略的有效性,本文在不同模型与样本量条件下进行了系统对比实验(见表 3).结果表明,在相同的样本量下,采用 JLSS 采样策略显著优于传统的 LHS 采样,能够提高模型的拟合精度与稳定性.

以样本量为 300 组为例,本文在 PSO−XGBoost 模型框架下,分别采用联合拉丁时空采样(JLSS)和传统拉丁超立方采样(LHS)进行对比.结果表明,采用 JLSS 策略时,模型的 R^2 达到 0.986,MSE 为 0.005,RMSE 为 0.071;而采用 LHS 策略时,R^2 仅为 0.901,MSE 为 0.034,RMSE 为 0.184.

与传统采样相比,JLSS 策略使得 MSE 降低了约 85.3%,RMSE 降低了约 61.4%,R^2 提升约 9.4%.表明所提采样策略在有限数据条件下仍具有良好的泛化能力.

进一步分析发现,即使在样本数量减少至原始 1/3 的情况下,PSO−XGBoost 模型的预测性能也优于未优化模型在 1000 组数据下的结果.这进一步证明,PSO 算法优化的模型在小样本条件下仍具有较高的精度与稳定性.

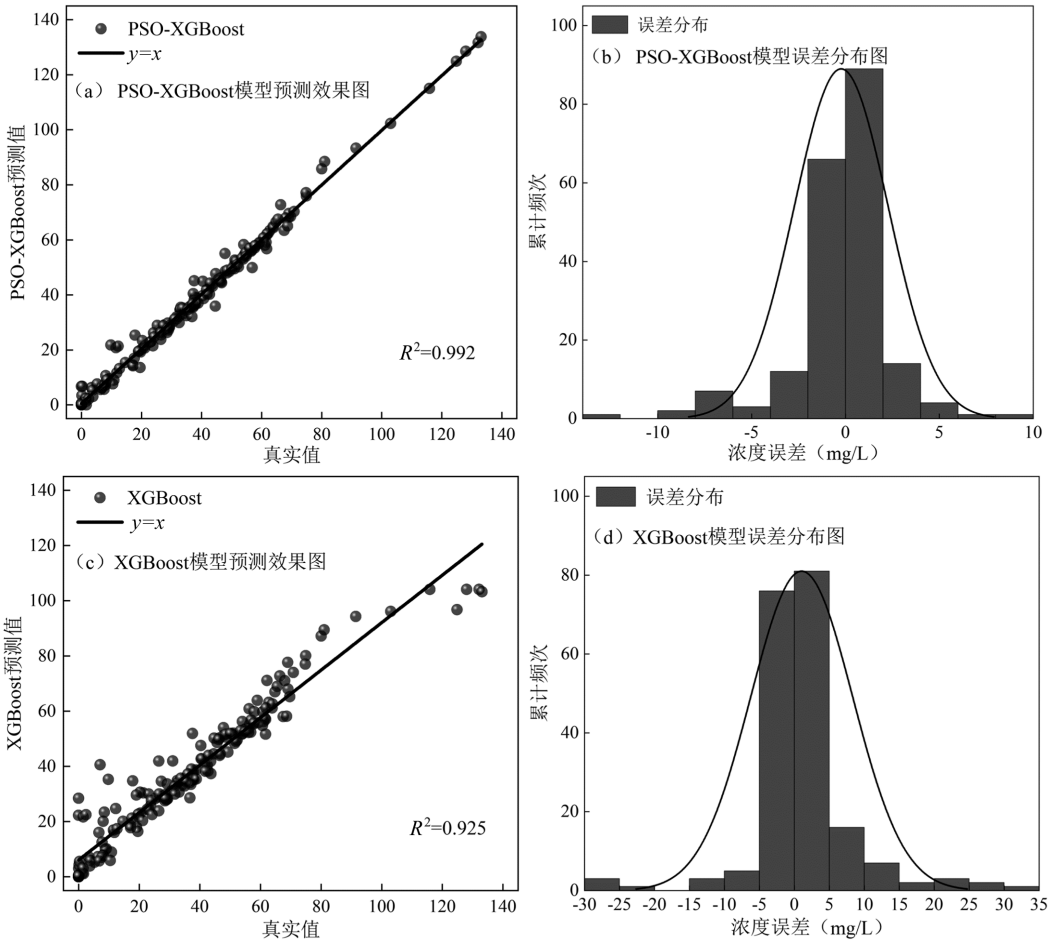


图 9　1000 组样本下 PSO−XGBoost 与 XGBoost 模型的预测效果及误差分布对比

Fig.9　Comparison of prediction performance and error distribution between PSO−XGBoost and XGBoost models with 1000 samples

表 3　PSO−XGBoost 模型与其他模型预测结果的评价指标

Table 3　Evaluation metrics of PSO−XGBoost model and other models' prediction results

方法	样本量	MSE	R^2	RMSE
XGBoost +JLSS	1000 组	0.013	0.925	0.114
PSO−XGBoost + JLSS	1000 组	0.002	0.992	0.044
XGBoost + JLSS	300 组	0.089	0.821	0.298
PSO−XGBoost +JLSS	300 组	0.005	0.986	0.071
PSO−XGBoost + LHS	300 组	0.034	0.901	0.184
XGBoost + LHS	300 组	0.070	0.764	0.265

3.3　模型对比实验

近年来,随着人工智能技术在环境领域的广泛应用,地下水污染物浓度预测方法逐渐从传统的数值模拟转向数据驱动模型[40-41].为验证本文所提出的 PSO−XGBoost 模型的预测效果,本研究选取了 3 种常见的数据驱动模型作为对比对象.

首先,选取两种单一结构模型——卷积神经网络(CNN)与长短期记忆网络(LSTM),它们在时间序列建模与特征提取方面被广泛用于环境污染预测任务;其次,进一步引入一种组合建模方法—变分模态分解与极端梯度提升模型(VMD−XGBoost),该模型通过信号分解与集成学习相结合,具备较强的特征表示能力和预测精度.

上述 3 种模型覆盖了从深度学习到集成学习的典型建模路径,能够从结构复杂度、建模策略和数据敏感性等多个维度,全面验证所提出 PSO−XGBoost 模型在污染物浓度预测任务中的适应性与优越性.

本实验在统一的数据样本、参数范围与评价指标(参见 1.3.2 节)的基础上,对上述四类模型进行了性能对比.图 10 展示了各模型在测试集上的预测值与真实值的拟合效果、误差分布以及 90%置信预测带.

结果显示,PSO−XGBoost 模型在精度和稳定性方面均优于对比模型,其预测结果分布最为集中,误差波动最小,大部分样本点落在 90%预测带内,说明

模型具备较强的泛化能力和对不确定性的表达能力.相比之下,LSTM 和 CNN 在中高浓度段出现较多偏离,VMD−XGBoost 预测结果整体波动更大,误差分布不均,表现出一定的不稳定性.

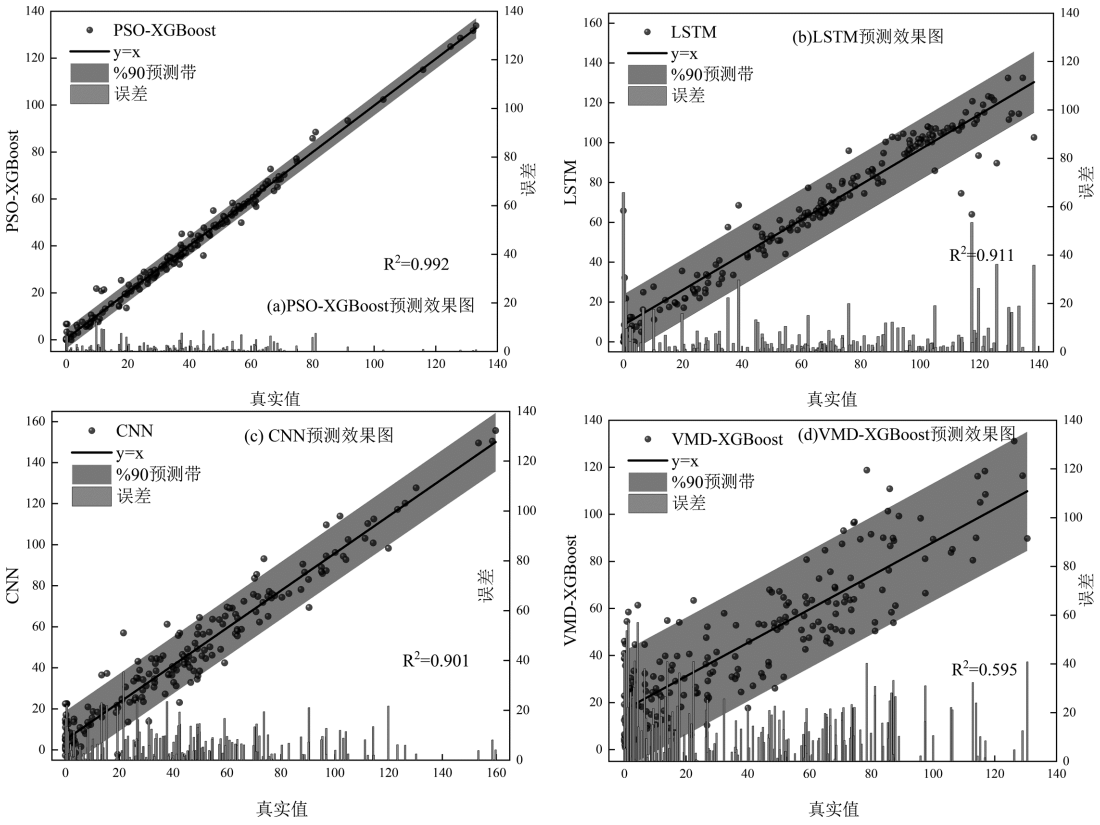


图 10 各模型预测效果

Fig.10 Prediction performance diagrams of different models

图中展示了 PSO−XGBoost、LSTM、CNN、VMD−XGBoost 四种模型的预测值与真实值对比(散点图),以及误差分布(柱状图),同时绘制了 90%预测带

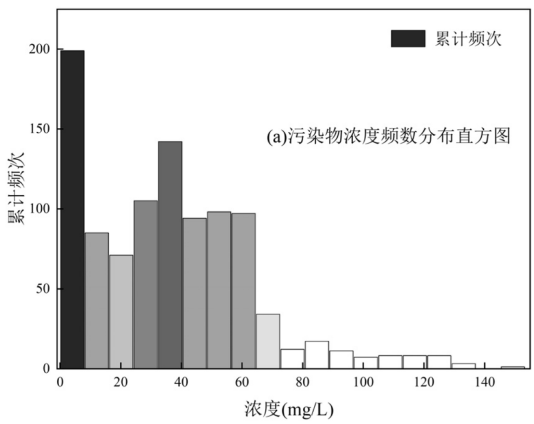
综上,PSO−XGBoost 模型在本研究中表现出最优的浓度预测能力,结合本文提出的时空采样策略,可为地下水污染物迁移的快速评估与风险预测提供有效的技术支撑.

3.4 地下水参数不确定性分析

通过精度分析验证了 PSO−XGBoost 模型在不同参数条件下进行 COD 浓度不确定性预测的有效性.以北京市安定镇典型生活垃圾填埋场第 10a 污染物扩散为例,采用 JLSS 方法生成 1000 组参数组合,输入训练好的 PSO−XGBoost 模型进行 预测,并对结果进行风险分析(见图 11).

对这 1000 组不确定参数下的浓度预测结果进行统计,并绘制了频数分布柱状图.结果显示,污染物浓度在 3.13~62.54mg/L 范围内的概率最大(73.9%).且由图(b)可以看出,污染物浓度变化范围较大,表明预测结果存在较大的不确定性.

根据国家有关规定,没有划定地下和地表水功能区的执行四类标准.依据《地表水环境质量标准》,四类标准的 COD 限值为 30mg/L.基于 1000 组浓度预测结果绘制的累积分布函数(图 11)显示,浓度超过 30mg/L 的概率为 56.43%,表明研究区污染风险较高.经统计,超标区域的最大范围为 54m,进一步表明该区域污染风险较高,该结果可为类似场地污染防控提供风险预判支持.



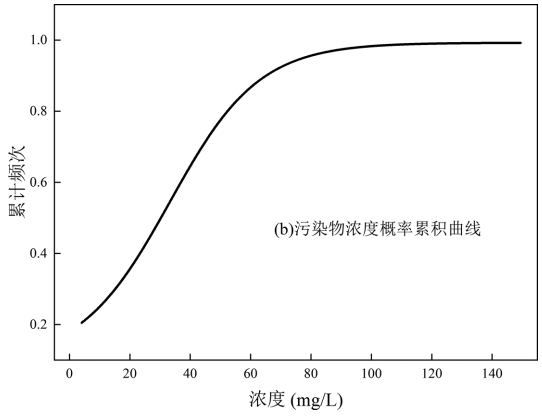


图 11　污染物浓度频数分布直方图及概率累积曲线

Fig.11　Histogram and cumulative probability curve of pollutant concentration

4　结论

4.1　与以 Landsim 为代表的解析解−Monte Carlo 耦合算法相比,本研究构建基于数值模拟模型样本的机器学习替代模型更好的捕捉了参数不确定性对结果的影响,预测的污染浓度分布区间平均为传统模型的 1.36 倍.

4.2　将 JLSS 抽样方法应用于 PSO−XGBoost 模型的构建,不仅显著提升了采样的覆盖率和代表性,还有效降低了模型训练阶段的计算成本,使用 JLSS 抽样构建 PSO−XGBoost 模型仅需 300 次数值模拟(10h),相较于传统方法节省了 70%的计算时间,大大提高了不确定性分析的效率,对地下水的风险评估具有重要意义.

4.3　基于 PSO−XGBoost 方法构建的替代模型展现出较高的预测精度(R^2=0.99),能够有效逼近模拟模型的输入输出关系,并显著降低了计算负担.最终,将该模型用于地下水污染风险进行了评估.结果表明,研究区内 COD 浓度超过 30mg/L 的概率为 56.43%,最大超标范围为 54m,表明研究区污染风险较高.

参考文献：

[1] Xu Y, Xue X, Dong L, et al. Long−term dynamics of leachate production, leakage from hazardous waste landfill sites and the impact on groundwater quality and human health [J]. Waste Management, 2018,82:156−166.

[2] Vinti G, Bauza V, Clasen T, et al. Health risks of solid waste management practices in rural Ghana:A semi−quantitative approach toward a solid waste safety plan [J]. Environmental Research, 2023, 216:114728.

[3] Aydin M E, Aydin S, Bahadir M, et al. Toxicity assessment of landfill leachates with a battery of bioassays [J]. Fresenius Environmental Bulletin, 2015,24(11):3584−3589.

[4] 张　帅,周弋铃,彭靖宇,等.生活垃圾填埋场渗滤液渗漏场地污染风险评价体系 [J]. 环境科学研究, 2024,37(07):1583−1591.DOI:10.13198/j.issn.1001−6929.2024.03.04.
Zhang S, Zhou G L, Peng J Y, et al. Contamination risk assessment system for leachate leakage from municipal solid waste landfill sites [J]. Research of Environmental Sciences, 2024,37(7):1583−1591. DOI:10.13198/j.issn.1001−6929.2024.03.04.

[5] Sun X C, Xu Y, Liu Y Q, et al. Evolution of geomembrane degradation and defects in a landfill:Impacts on long−term leachate leakage and groundwater quality [J]. Journal of Cleaner Production, 2019,224:335−345.

[6] Saheli P T, Rowe R K, Petersen E J, et al. Diffusion of multiwall carbon nanotubes through a high−density polyethylene geomembrane [J]. Geosynthetics International, 2017,24(2):184−197.

[7] 张鲁玉.基于深度学习的危废填埋渗漏风险预测 [D]. 山东:山东工商学院, 2023.DOI:10.27903/d.cnki.gsdsg.2023.000154.
Zhang L Y. Risk Prediction of Hazardous Waste Landfil Leakage Based on Deep Learning [D]. Shandong:Shandong Technology and Business University, 2023.DOI:10.27903/d.cnki.gsdsg.2023.0001:54.

[8] Sheng D, Wen X, Wu J, et al. Comprehensive Probabilistic Health Risk Assessment for Exposure to Arsenic and Cadmium in Groundwater [J]. Environmental Management, 2021,67(4):779−792.

[9] Ciriello V, Lauriola I, Bonvicini S, et al. Impact of Hydrogeological Uncertainty on Estimation of Environmental Risks Posed by Hydrocarbon Transportation Networks [J]. Water Resources Research, 2017,53(11):8686−8697.

[10] Demissie Y, Valocchi A, Cai X, et al. Parameter estimation for groundwater models under uncertain irrigation data [J]. Groundwater, 2015,53(4):614−625.

[11] 莫绍星.基于深度学习的地下水模拟高维不确定性分析和反演 [D]. 南京:南京大学, 2019.DOI:10.27235/d.cnki.gnjiu.2019.000108.
Mo S X. Towards efficient high−dimensional uncertainty quantification and inverse analysis in groundwater modeling using deep learning [D]. Nanjing: Nanjing University, 2019.DOI:10.27235/d.cnki.gnjiu.2019.000108.

[12] Lahmar S, Maalmi M, Idchabani R. Investigating adaptive sampling strategies for optimal building energy performance using artificial neural networks and kriging surrogate models [J]. Journal of Building Engineering, 2024,82:108341.

[13] 陈骏骏.基于深度学习的地下水反应性溶质运移模型参数反演研究 [D]. 吉林:吉林大学, 2023.DOI:10.27162/d.cnki.gjlin.2023.000764.
Chen J J. Deep learning based inverse modeling of groundwater reactive transport models [D]. Jilin: Jilin University, 2023.DOI:10.27162/d.cnki.gjlin.2023.000764.

[14] 罗苑萍,孙世泰,黄旭斌,等.基于机器学习的变电站地下水埋深智能预报预警模型 [J/OL]. 水利水电技术(中英文)[2024−10−30].
Luo Y P, Sun S T, Huang X B, et al. Intelligent forecasting and early warning models for substation drainage systems based on [J/OL]. Water Resources and Hydropower Engineering machine learning, [2024−10−30].

[15] Alsaif A,Abbas Y M. Interpretable constitutive compressive stress−strain model for rubberized aggregate concrete − Integrating comprehensive empirical database and efficient XGBoost ensemble learning [J]. Case Studies in Construction Materials, 2024,21:e03382.

[16] 董广齐,卢文喜,范　越,等.不确定性条件下地下水污染监测井网优化设计——基于 XGBoost 替代模型 [J]. 中国环境科学, 42(5):2144−2152.
Dong G Q, Lu W X, Fan Y, et al. Optimal design of groundwater pollution monitoring network under uncertainty [J]. China Environmental Science. 42(5):2144−2152.

[17] 罗建男,李雪利,王　鹤,等.基于局部−全局混合自适应替代模型的地下水污染源反演识别 [J]. 中国环境科学, 2023,43(7):3664−3671.
Luo J N, Li X L, Wang H, et al. Groundwater pollution sources inversion based on local−global hybrid adaptive surrogate model [J].

China Environmental Science. 2023,43(7):3664-3671.

[18] 葛渊博,卢文喜,白玉堃,等.基于 SSA-BP 与 SSA 的地下水污染源反演识别 [J]. 中国环境科学, 2022,42(11):5179-5187.
Ge Y B, Lu W X, Bai Y K, et al. Inversion and identification of groundwater pollution sources based on SSA-BP and SSA [J]. China Environmental Science. 2022,42(11):5179-5187.

[19] 贾顺卿,卢文喜,李久辉,等.基于 U-D 分解卡尔曼滤波地下水污染源溯源辨识 [J]. 中国环境科学, 2021,41(2):713-719.
Jia S Q, Lu W X, Li J H, et al. Inversion identification of groundwater contamination source based on U-D factorization Kalman filter [J]. China Environmental Science. 2021,41(2):713-719.

[20] Lin G,Tartakovsky A M.An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media [J]. Advances in Water Resources, 2009,32(5):712-722.

[21] Kavvas M, Chen Y, Strelkoff T. Performance assessment of a subsurface drip irrigation system using the HELP model [J]. Irrigation Science, 2002,21(3):129-137.

[22] Nai C, Tang M, Liu Y, et al. Potentially contamination and health risk to shallow groundwater caused by closed industrial solid waste landfills: Site reclamation evaluation strategies [J]. Journal of cleaner production, 2021,286:125402.

[23] Wu C M, Yeh T C J, Zhu J F, et al. Traditional analysis of comparing apples to oranges [J]. Water Resources Research, 2005,41(9)W09402.

[24] Berger K U. On the current state of the Hydrologic Evaluation of Landfill Performance (HELP) model [J]. Waste management, 2015,38: 201-209.

[25] 王新港,杨 昱,徐祥健,等.单井抽出-回渗同步循环地下水水力控制模型研究 [J]. 环境科学研究, 2023,36(01):180-187.DOI:10.13198/j.issn.1001-6929.2022.08.05.
Wang X G, Yang Y, Xu X J, et al. Hydraulic control model research on single well pumping-recharge synchronous cyclical groundwater [J]. Research of Environmental Sciences, 2023,36(1):180-187.DOI:10.13198/j.issn.1001-6929.2022.08.05.

[26] Pan Y, Zeng X, Xu H, et al. Use of stacked Gaussian processes regression method to improve prediction of groundwater solute transport model [J]. Journal of Hydrology, 2023,620:129530.

[27] 王 静.基于深度学习的地下水问题数值模拟方法 [D]. 武汉:武汉大学, 2021.
Wang J. Numerical simulation of ground water problem using a deep learning approach [D]. Wuhan:Wuhan University, 2021.

[28] 张雨佳,杨蕴,龚绪龙,等.油水混合层 CO_2 地质封存与利用数值模拟及参数全局敏感性分析 [J/OL]. 中国环境科学, 1-12[2025-07-25]. https://doi.org/10.19674/j.cnki.issn1000-6923.20250321.003.
Zhang Y J, Yang Y, Gong X L, et al. Numerical simulation and parameter global sensitivity analysis of CO_2 geological storage and utilization in oil-water mixed layers [J]. China Environmental Science. 1-12[2025-07-25].https://doi.org/10.19674/j.cnki.issn1000-6923.20250321.003.

[29] Dai X M,Liu L X,Cheng Z.Elucidating price variability drivers in highway electromechanical equipment using CV predictions with PSO-XGBoost[J].Alexandria Engineering Journal,2024,109:754-767.

[30] 邵 萌,潘正中,孙金伟,等.基于 EMD-PSO-BP 模型的短期潮流流速预测 [J]. 中国海洋大学学报(自然科学版), 2024,54(11):134-141.DOI:10.16441/j.cnki.hdxb.20220321.

[31] Shao M, Pan Z Z, Sun J W, et al. Short-term tidal current speed prediction based on EMD-PSO-BP model [J]. Periodical of Ocean University of China, 2024,54(11):134-141.

[32] 邓田丰,李梓铭,张少波,等.基于机器学习和时间序列算法融合的北京地区花粉浓度预测技术研究 [J/OL]. 中国环境科学,1-23[2025-07-25].https://doi.org/10.19674/j.cnki.issn1000-6923.20250624.001.
Deng T F, Li Z M, Zhang S B, et al. Research on pollen concentration prediction technology in Beijing based on the fusion of machine learning and time series algorithms. [J]. China Environmental Science. 1-23[2025-07-25].https://doi.org/10.19674/j.cnki.issn1000-6923.20250624.001.

[33] Zhang J, Cao X, Li C, et al. Experimental analysis of combustion characteristics of corn starch dust clouds under the action of unilateral obstacles and machine learning modeling based on PSO-XGBoost [J]. Advanced Powder Technology, 2024,35(11): 104641.

[34] Sheikhi S, Kostakos P. Safeguarding cyberspace: Enhancing malicious website detection with PSOoptimized XGBoost and firefly-based feature selection [J]. Computers & Security, 2024,142:103885.

[35] Bo Y, Guo X, Liu Q, et al. Prediction of tunnel deformation using PSO variant integrated with XGBoost and its TBM jamming application [J]. Tunnelling and Underground Space Technology, 2024,150:105842.

[36] Luo J, Ma X, Ji Y, et al. Review of machine learning-based surrogate models of groundwater contaminant modeling [J]. Environmental Research, 2023:117268.

[37] 王永财,万华伟,高吉喜,等.基于深度学习语义分割的数码相机图像草地植被盖度估算 [J/OL]. 环境科学研究[2024-10-30].DOI:10.13198/j.issn.1001-6929.2024.09.02.
Wang Y C, Wan H W, Gai J X, et al. Grassland vegetation coverage estimation in digital camera images based on deep learning semantic segmentation [J/OL]. Research of Environmental Sciences. [2024-10-30].DOI:10.13198/j.issn.1001-6929.2024.09.02.

[38] 徐亚宁,卢文喜,王梓博,等.考虑参数和边界条件不确定性的地下水污染随机模拟 [J]. 中国环境科学, 2022,42(7):3244-3253.
Xu Y N, Lu X W, Wang Z B, et al. Stochastic simulation of groundwater pollution considering uncertainty of parameters and boundary conditions [J]. China Environmental Science. 2022,42(7): 3244-3253.

[39] 刘志刚,郭婧婷,佘祺锐,等.北京安定生活垃圾填埋场扩容边坡稳定性分析 [C]//中国环境科学学会环境工程分会.中国环境科学学会 2021 年科学技术年会——环境工程技术创新与应用分会场论文集(二).北京:《工业建筑》杂志社有限公司, 2021:234-238,105.DOI:10.26914/c.cnkihy.2021.022070.
Liu Z G, Guo J T, Yu Q R, et al. Slope stability analysis of anding municipal solid waste landfill in Beijing [C]. Environmental Engineering Branch of the Chinese Society for Environmental Sciences. Proceedings of the sub - session on innovation and application of environmental engineering technology at the 2021 scientific and technological annual conference of the Chinese Society for Environmental Sciences (II). Beijing: Industrial Construction Magazine Co., Ltd., 2021:234-238,105.DOI:10.26914/c.cnkihy.2021.022070.

[40] 张永贺;刘亚淘;蓝智钢.北京市安定卫生填埋场渗沥液处理工艺 [J]. 环境卫生工程, 2013,21(6):61-64.
Zhang Y H, Liu Y T, Lan Z G, et al. Leachate treatment technology of anding waste sanitary landfill site in Beijing [J]. Environmental Sanitation Engineering. 2013,21(6):61-64.

[41] 张 弦.基于 LSTM-BP 的填埋场污染扩散监测方法研究 [D]. 烟台:山东工商学院, 2022.DOI:10.27903/d.cnki.gsdsg.2022.000091.
Zhang X. Research on monitoring method of landfill pollution diffusion based on LSTM-BP [D]. Yantai: Shandong Technology and Business University, 2022.DOI:10.27903/d.cnki.gsdsg.2022. 000091.

[42] 陈亚松,刘家雯,赵云鹏,等.基于机器学习的人工湿地出水水质预测与影响因素 [J]. 中国环境科学, 2025,45(06):3161-3170.DOI:10.19674/j.cnki.issn1000-6923.2025.0085.
Chen Y S, Liu J W, Zhao Y P, et al. Prediction of effluent water quality and analysis of influencing factors in constructed wetlands based on machine learning. [J]. China Environmental Science. 2025,45(6): 3161-3170.DOI:10.19674/j.cnki.issn1000-6923.2025.0085.

作者简介：能昌信(1965-),男,山东临沂人,教授,博士,主要从事环境监测技术研究.发表论文 100 余篇.naicx@126.com.