

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]. 水利水电技术(中英文), 2025, 56(8): 118-130. DOI: 10.13928/j.cnki.wrahe.2025.08.009

LUO Yuanping, SUN Shitai, HUANG Xubin, et al. Intelligent prediction and early warning model for substation groundwater depth based on machine learning[J]. Water Resources and Hydropower Engineering, 2025, 56(8): 118-130. DOI: 10.13928/j.cnki.wrahe.2025.08.009

# 基于机器学习的变电站地下水埋深 智能预报预警模型

罗苑萍<sup>1</sup>, 孙世泰<sup>1</sup>, 黄旭斌<sup>1</sup>, 郑泽举<sup>1</sup>, 蔡榕鸿<sup>1</sup>, 梁伟强<sup>2</sup>, 胡立堂<sup>3</sup>

(1. 广东电网有限责任公司揭阳供电局, 广东 揭阳 510000; 2. 广州市奔流电力科技有限公司, 广东 广州 510700; 3. 北京师范大学水科学研究院, 北京 100875)

**摘要:**【目的】设计一套准确且高效的地下水位动态预测模型对于变电站排水系统的智能监控与预警系统的有效应用以及确保变电站安全稳定运行至关重要。【方法】聚焦于220kV园区变电站的试点研究项目, 针对极致梯度提升树(XGBoost)、随机森林(RF)和长短期记忆模型(LSTM)三种机器学习模型进行了综合评估, 重点分析了三种模型在预测暴雨情景下地下水埋深动态方面的性能表现。模型的训练数据来源于经过校准和验证的地下水流数值模型, 并结合多种暴雨情景下的地下水埋深动态预测结果作为基准参考值。为了较好地评估这些模型的预测准确性和可靠性, 采用了Nash-Sutcliffe效率系数(NSE)、均方根误差(RMSE)、Pearson相关系数和平均绝对误差(AE)作为评价指标。【结果】研究结果显示, XGBoost、RF和LSTM三种模型均能在时间尺度上模拟出与基准结果相近的地下水埋深动态, NSE、RMSE和Pearson相关系数分别达到0.9998、0.0031m和0.9999, 但在空间上表现差异大, RF模型模拟的AE小于0.01m, XGBoost模型模拟的AE小于0.26m, LSTM模型给出的AE小于0.12m。使用20%网格点的模型数据进行机器学习训练输入, RF模型的性能依然表现最佳, 同时模型训练和预测的时间效率提升了5倍。【结论】基于机器学习模型构建的地下水埋深动态预测模型表现良好, 在排水系统的智能监控与预警系统中具有良好的应用前景。

**关键词:** 变电站; 地下水埋深预警系统; 机器学习; 智能监控; 数值模拟

DOI: 10.13928/j.cnki.wrahe.2025.08.009

开放科学(资源服务)标志码(OSID):

中图分类号: TU99

文献标志码: A

文章编号: 1000-0860(2025)08-0118-13



## Intelligent prediction and early warning model for substation groundwater depth based on machine learning

LUO Yuanping<sup>1</sup>, SUN Shitai<sup>1</sup>, HUANG Xubin<sup>1</sup>, ZHENG Zeju<sup>1</sup>, CAI Ronghong<sup>1</sup>, LIANG Weiqiang<sup>2</sup>, HU Litang<sup>3</sup>

(1. Jieyang Power Supply Bureau of Guangdong Power Grid Co., Ltd., Jieyang 510000, Guangdong, China; 2. Guangzhou Power Electrical Technology Co., Ltd., Guangzhou 510700, Guangdong, China; 3. College of Water Sciences, Beijing Normal University, Beijing 100875, China)

收稿日期: 2024-07-08; 修回日期: 2024-08-08; 录用日期: 2024-08-12; 网络出版日期: 2024-10-15

基金项目: 清华大学-宁夏银川水联网数字治水联合研究院专项统筹重点项目(SKL-IOW-2023TC2307); “十四五”国家重点研发项目(2023YFC3708903); 广东揭阳惠来园区输变电工程项目(0352004823220005)

作者简介: 罗苑萍(1976—), 女, 高级经济师, 学士, 主要研究方向为电力建设。E-mail: 17624052757@163.com

通信作者: 梁伟强(1999—), 男, 助理工程师, 学士, 主要研究方向为智能电网规划与运行控制。E-mail: 19587442470@163.com

©Editorial Department of Water Resources and Hydropower Engineering. This is an open access article under the CC BY-NC-ND license.

**Abstract:** [Objective] Designing an accurate and efficient groundwater depth dynamics prediction model is crucial for the effective application of intelligent monitoring and early warning systems for substation drainage systems and for ensuring the safe and stable operation of substations. [Methods] Focusing on the pilot study project of the 220kV substation in the industrial park, a comprehensive evaluation was conducted on three machine learning models: Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Long Short-Term Memory (LSTM). The performance of these models in predicting groundwater depth dynamics under heavy rainfall scenarios was analyzed in detail. The training data for the models were derived from a calibrated and validated groundwater flow numerical model, using prediction result of groundwater depth dynamics under various rainfall scenarios as benchmark reference values. To thoroughly assess the prediction accuracy and reliability of these models, the Nash-Sutcliffe Efficiency Coefficient (*NSE*), Root Mean Square Error (*RMSE*), *Pearson* Correlation Coefficient, and Mean Absolute Error (*AE*) were used as evaluation indicators. [Results] The research result showed that XGBoost, RF, and LSTM models could simulate groundwater depth dynamics consistent with the benchmark result over the time scale, with *NSE*, *RMSE*, and *Pearson* correlation coefficients reaching 0.999 8, 0.003 1 m, and 0.999 9, respectively. However, the spatial performance varied significantly. The *AE* simulated by the RF model was less than 0.01 m, the *AE* simulated by the XGBoost model was less than 0.26 m, and the *AE* given by the LSTM model was less than 0.12 m. When using model data from 20% of the grid points for machine learning training, the RF model still showed the best performance, and the time efficiency of model training and prediction improved by 5 times. [Conclusion] The groundwater depth dynamics prediction model based on machine learning models demonstrates excellent performance and shows promising application prospects in the intelligent monitoring and early warning systems for drainage systems.

**Keywords:** substation; groundwater depth early warning system; machine learning; intelligent monitoring; numerical simulation

## 0 引言

电力系统的稳定与安全对于区域电力供应的可靠性具有决定性作用, 而变电站作为该系统的关键组成部分, 其运行的稳定性至关重要。然而, 在变电站的日常运维过程中, 地下水排水问题往往未受到应有的重视, 尽管其潜在的负面影响不容忽视。地下水位的异常波动或排水系统的缺陷, 如排水不畅或堵塞, 均有可能引起变电站地基的不稳定性, 这种不稳定性可能进一步对整个设施的安全性能造成影响。因此, 对地下水排水系统的优化设计和管理, 对于确保变电站长期安全稳定运行具有重要的实践意义。变电站排水系统的智能监控与预警系统是确保变电站安全运行的关键研究领域。其核心目标在于预防因排水不畅导致的积水问题, 进而避免设备损坏、电力中断及其他潜在的安全隐患。传统的预警方法主要依赖于人工监测地下水埋深和基于经验的排水系统判断, 这种方法难以实现对变电站排水状态的准确、及时和有效诊断。随着在线监测技术、数值模拟和机器学习技术的飞速发展, 开发变电站排水系统的智能监控模块成为可能, 这将显著提升系统的准确性和智能化水平, 及时预警潜在风险, 确保变电站的安全运行<sup>[1-2]</sup>。

在真实的地下水系统中, 地下水流动的非线性特性、介质参数的空间异质性以及补给的延迟效应等因素, 使得基于物理过程的地下水位预测模型变得极为

复杂。机器学习方法能够有效处理这些影响因素与地下水位之间的高度非线性相互作用<sup>[3]</sup>。机器学习技术能够自动学习、分析数据规律, 并利用这些规律对未知数据进行预测<sup>[4]</sup>, 无须深入理解地下水系统的复杂物理机制, 即可充分利用现有数据, 并实现自我改进, 从而实现地下水位的准确预测。然而, 若缺乏可靠的监测数据, 机器学习模型将无法建立<sup>[5]</sup>, 因此仍需结合地下水数值模型以提供足够的数据量<sup>[6]</sup>。

近年来, 随着信息技术的迅猛发展, 一系列先进的技术如大数据分析、物联网、人工智能以及云计算等, 已经在土木工程和水利工程领域内实现了广泛的应用。这些技术通过提供实时数据收集、处理和分析的能力, 提高了工程项目的规划、设计、施工和维护过程的效率和精确度<sup>[7]</sup>。特别地, 以人工智能和深度学习技术为基础的数据驱动的机器学习模型应用广泛, 并被用来解决水文问题<sup>[8-10]</sup>。SERRADILLA等<sup>[11]</sup>详细回顾了各种最先进的深度学习架构, 并分析了它们如何与预测性维护阶段集成, 以满足工业公司从需求角度简化这一任务。VANCE等<sup>[12]</sup>从宏观层面说明了水文学、海洋学和大气科学中近期大数据应用的进展, 强调大数据提供了前所未有的机会, 以增强我们对地球复杂模式和相互作用的理。在与变电站排水密切相关的城市洪水预报方面, 张然等<sup>[13]</sup>综合了国内外在机器学习领域的研究成果, 深入探讨了机器学习模型的基本原理及其独特的技术特性, 进一

步评估了机器学习模型在洪水预测领域的应用优势, 并对其存在的局限性进行了总结。在使用物理机制模型来模拟城市洪涝灾害时, 由于缺乏足够的校准和验证数据且模型运行时间长, 存在一定的局限性<sup>[14]</sup>。而机器学习为快速准确地解决城市洪水预测问题提供了新的见解。CHANG 等<sup>[15]</sup>建立了 3 个人工神经网络模型, 基于一种先进的因子选择方法提取的输入, 根据当前蓄洪池水位和降雨信息, 对蓄洪池的水位演变进行预测, 以便有足够的时间提前预热抽水系统, 从而防止城市发生洪水。WU 等<sup>[16]</sup>通过比较不同指标组合方案对积水深度的预测精度, 确定了以降雨量、降雨持续时间、峰值降雨量、区位系数、降雨强度方差、峰值倍数、浓度偏度为组合的方案为最适合进行水深预测的指标方案, 利用深度学习算法建立了积水点积水过程的预测模型。ZHOU 等<sup>[17]</sup>从研究可行性、方法适用性和实际应用三个方面系统阐述了构建积水点积水过程预测预警方法, 并以郑州市 2019 年 8 月 1 日降雨事件为例, 构建了降雨预报数据驱动的城市洪涝过程精细化实时预测模型, 结果表明预警总体准确率在 80% 以上, 且随着预测周期的缩短, 预警结果准确率呈上升趋势, 能够满足城市防洪要求。以上研究成果都实现了基于降雨预报数据的实时城市洪水和水浸过程预测与预警。这些研究成果表明, 机器学习模型能够在实时暴雨情景下对变电站排水系统状态进行准确预测。然而, 在变电站地下水埋深智能监控与预警方面的研究仍相对较少<sup>[18]</sup>, 尚没有讨论数据数量对模型时空精度的影响。

本研究立足于实际需求, 拟以广东省某 220 kV 园区变电站项目区为研究区, 利用专业地下水数值模型模拟不同暴雨条件下地下水埋深变化, 再将多情景模拟结果作为训练数据, 以 Nash-Sutcliffe 效率系数 (Nash-Sutcliffe Efficiency Coefficient, *NSE*)、均方根误差 (Root Mean Square Error, *RMSE*) 以及 *Pearson* 相关系数作为评价指标, 比较三种典型的机器学习模型, 分别利用观测井数据、20% 网格结点和所有网格结点数据构建模型的模拟效果, 进一步探索构建的模型在预测区域地下水埋深空间分布的模拟效率, 旨在更有效地管理和控制变电站的排水风险, 确保变电站的安全稳定运行。

## 1 研究区和研究方法

### 1.1 研究区概况

研究区位于广东省揭阳市某地, 计划建设一座 220 kV 园区变电站及其配套设施。220 kV 园区变电

站主变压器规模为  $4 \times 180$  MVA, 220 kV 终期出线 10 回, 110 kV 终期出线 14 回, 10 kV 终期出线 36 回, 配套建设线路通信光缆以及通信设备。该区域位于冲积平原, 属于河口三角洲堆积地貌, 地面高程在  $-0.50 \sim 2.50$  m。场地气候属于亚热带季风气候类型, 多年平均气温为  $21.40$  °C; 多年平均降雨量为  $1723.00$  mm, 年最大降雨量达到  $2582.60$  mm, 年最小降雨量为  $1260.60$  mm; 最大日降雨量高达  $360.00$  mm; 多年平均水面蒸发量为  $1478.10$  mm。年内集中暴雨的主要时间段为 4—9 月, 平均每月发生约一次暴雨。前汛期 (4—6 月) 以锋面雨和低槽雨为主, 而后汛期 (7—9 月) 则以台风雨为主。4—9 月多年平均降雨量相对较高, 其中 4—5 月和 10—11 月通常为水位平稳期, 6—9 月则为水位较高的丰水期, 而 12 月—次年 3 月则为水位较低的枯水期。年最大洪水发生在上述几个月的概率高达 84.30%。

场区地下水类型为孔隙水和裂隙水, 地下水主要接受大气降雨补给, 并以蒸发和侧向径流为排泄项。根据已有的监测数据, 研究区的地下水埋深为  $1.20 \sim 1.55$  m。

### 1.2 变电站地下水埋深预警模型思路

220 kV 园区变电站地下排水预警模型主要分为两类: 一类是基于实时监测数据的预警模型, 另一类是基于数值模型的替代模型。当拥有大量反映时序变化的地下水埋深数据时, 可以采用机器学习技术构建模型, 以揭示暴雨强度和历时与地下水埋深之间的关系, 这种模型适用于预测预报。然而, 要确保模型的准确性, 必须依赖多次暴雨事件下的地下水埋深监测数据, 这些数据将作为训练集, 以提高模型的预测精度。本预警系统将在 220 kV 园区变电站进行试点建设应用。

地下水流数值模型能够综合考虑水文地质条件、历次暴雨的监测数据、排水管网设计以及地下水运动的物理机制, 从而精确预测研究区域内不同位置的地下水埋深动态变化。尽管这类模型功能强大, 但它们需要大量的输入信息, 运行时间较长, 并且需要对数据进行后处理, 以获取关键区域或关键点的地下水埋深动态信息。为了提高模型的应用效率, 许多研究采用地下水流数值模型的现状和多情景模拟结果作为机器学习模型的训练数据, 通过建立替代模型来提升模型的应用效率。

鉴于变电站尚未竣工, 缺乏实际的地下水埋深观测数据, 本研究选择采用通过数值模型在预设观测井位置模拟得到的地下水埋深数据, 采用多种暴雨情景

下的数值模型模拟获得不同网格不同时间的地下水埋深数据, 将这些数据作为基于监测数据的模型训练集。为了探究不同网格点密度对模型性能的影响, 构建的替代模型将考虑使用数值模型在全平面网格以及通过地理信息系统对全网格点进行 20% 抽稀处理后得到的网格点模拟数据作为训练数据。通过这种方式, 本研究旨在评估网格点密度对模型预测精度的影响, 并优化模型的训练过程。使用 4 个指标来评估模型精度, 包括 Nash-Sutcliffe 效率系数 ( $NSE$ )、均方根误差 ( $RMSE$ )、Pearson 相关系数和平均绝对误差 (Mean Absolute Error,  $AE$ )。 $NSE$  直接测量了预测时间序列和目标时间序列之间在幅度和相位方面的匹配质量,  $NSE$  越接近 1, 表示模型的模拟精度更高。计算公式为

$$NSE = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

式中,  $x_i$  和  $y_i$  分别为观测值和模拟值;  $i$  为数据的序列号;  $n$  为数据总数;  $\bar{x}$  和  $\bar{y}$  分别为  $x$  和  $y$  的均值。

### 1.3 机器学习模型

机器学习模型种类繁多, 其中表现优异且广泛应用的模型包括极致梯度提升树 (Extreme Gradient Boosting, XGBoost)<sup>[17]</sup>、随机森林 (Random Forest, RF)<sup>[19]</sup> 和长短期记忆模型 (Long Short-Term Memory, LSTM)<sup>[20]</sup>。本研究选取这 3 种模型算法来构建模型, 并通过比较结果来挑选出稳定可靠的数据驱动模型。3 种机器学习模型的参数设置基于模拟结果与基准值的对比, 经过试算和经验值的调整确定。具体而言, XGBoost 模型和 RF 模型的决策树数量设定为 100, 随机状态数均设为 42; LSTM 模型的 epochs 参数设定为 20, 桶大小为 32。

XGBoost 模型是梯度提升回归 (Gradient Boosting Regression, GBR) 模型的增强版, 其核心在于沿着损失函数的负梯度方向拟合生成 CART 树。XGBoost 模型被认为是用于监督学习的性能最好的算法之一<sup>[21]</sup>。XGBoost 模型以 CART 树为基本单元, 能够构建多棵决策树, 且这些树是相互依赖的, 来自先前树的最新残差或误差用作下一树的基础, 每个样本的期望值是通过将树叶的值相加来计算的, 学习误差低, 对参数变化具有较强的鲁棒性<sup>[22]</sup>。随机森林 (RF) 是一种基于决策树分类器的集成算法<sup>[23]</sup>, 主要用于分类与回归, 是一种流行的建模工具<sup>[24]</sup>。它结合了决策树模型的预测, 以产生更准确的预测结果<sup>[25]</sup>。其基本原

理是通过 bootstrap 重采样技术, 从原始训练数据集中随机抽取数据点, 生成多个新的数据集, 进而构建多个独立的弱决策树分类器。通过整合这些分类器, 随机森林回归模型得以形成, 最终预测结果取所有决策树的平均值<sup>[26]</sup>。随着决策树数量的增加, RF 模型不易过拟合, 能够有效处理高维数据, 且对异常值不敏感。通过平均多个决策树和对特征的随机扰动, 该算法显著提升了模型的泛化能力, 使其具有较强的抗干扰能力。长短期记忆模型 (LSTM) 由 Hochreiter 和 Schmidhuber 于 1997 年提出, 是一种用于解决传统递归神经网络长期依赖问题<sup>[27]</sup>的模型。LSTM 与传统循环神经网络具有相似的链式结构, 但其内部操作更为复杂, 能够有效学习长期和短期依赖关系。LSTM 在长序列中表现出色, 其独特之处在于通过门的控制保留远古特征。LSTM 结构主要包括输入门、遗忘门和输出门, 这些结构作为输入数据的过滤器, 用于调整单元格和隐藏状态<sup>[28]</sup>。

### 1.4 变电站地下排水的地下水流数值模型

研究区地下水位大体呈南高北低的态势, 地下水由西南向东北流动。为了合理模拟变电站与周边地下水的互动, 模型的模拟范围围绕核心场地向外扩展了 57.00 m (见图 1), 模拟区的面积约为 14 473.00 m<sup>2</sup>。设计的 220 kV 园区变电站核心区范围为 85.00 m×64.00 m。依据含水层分布及边界地下水流的特点, 模型西南侧接受地下水侧向径流补给, 东北侧向下游排泄。因此, 模型西南、东北侧分别确定为流入、流出边界, 其他外边界采用默认边界 (零流量边界)。根据初步勘测的结果, 地层呈现分层特性, 因此模型被简化为一个三维的多层含水层结构, 模型的顶部接受大气降雨的入渗补给, 底部则以未风化的花岗岩为阻水层。模型关注的地下水垂直深度约为 50.00 m。地下水的主要补给来源为降雨, 通过上层较为透水的填土层向下渗透。研究区域从西北部接收侧向流入的地下水, 而场地本身没有地下水开采活动。地下水的主要排泄方式是通过径流向东北方向流出。该研究区的地下水系统处于一个天然的平衡状态, 在多年的平均条件下, 地下水流呈稳定状态。因此, 排水效果检测模型被概化为一个非均质、各向异性的地下水三维稳定流动的数值模型。该模型采用商业软件 FEFLOW<sup>[29]</sup> 创建而成, 模型的参数识别基于未建变电站的监测数据, 并通过模拟不同降雨条件来检测排水能力。

排水系统的管网排水采用等效渗透系数法<sup>[30]</sup>进行模拟。在处理积水时考虑了地表积水的超渗产流和

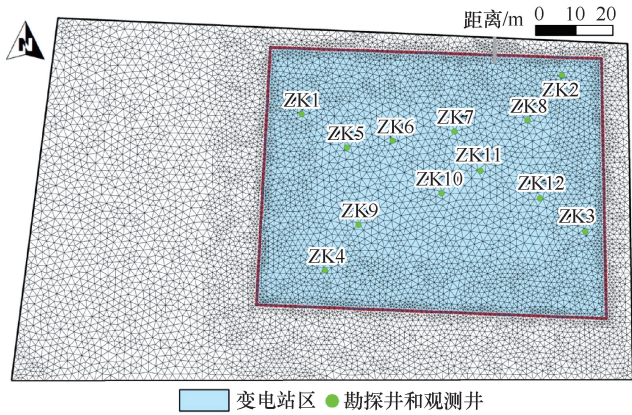


图1 研究区钻孔和排水管网分布

Fig. 1 Distribution of boreholes and drainage pipelines in study area

蓄满产流模式<sup>[31]</sup>。变电站修建之前,地下水埋深在1.80 m以内。变电站修建后,变电站建设范围地面标高抬高至4.80 m,同时变电站建设区由于地面硬化使得渗透系数和降雨入渗补给系数降低。在高降雨强度入渗补给情景下,区域地下水位会逐渐上升并超出地表,本研究在构建模型中对地下水位超高地表不作约束。

## 2 数值模型构建

### 2.1 地层结构和参数

利用区内12眼钻孔记录数据,构建了研究区三维层状地质结构模型。不同岩性呈现较均匀的层状分布,部分地段粉细砂层厚度较小。模型分为8个模拟层,与实际钻孔岩性分布一致,分别对应素填土、淤泥质土(夹砂和植物根系)、淤泥质土、粉质黏土、中粗砂、砂质黏性土、全风化花岗岩、强风化花岗岩(见图2)。垂向上不同含水层岩性差异导致含水层水文地质参数不同,因此含水层被概化为非均质各向同性介质。根据岩性经验渗透系数,进行垂向分区并赋初始参数。

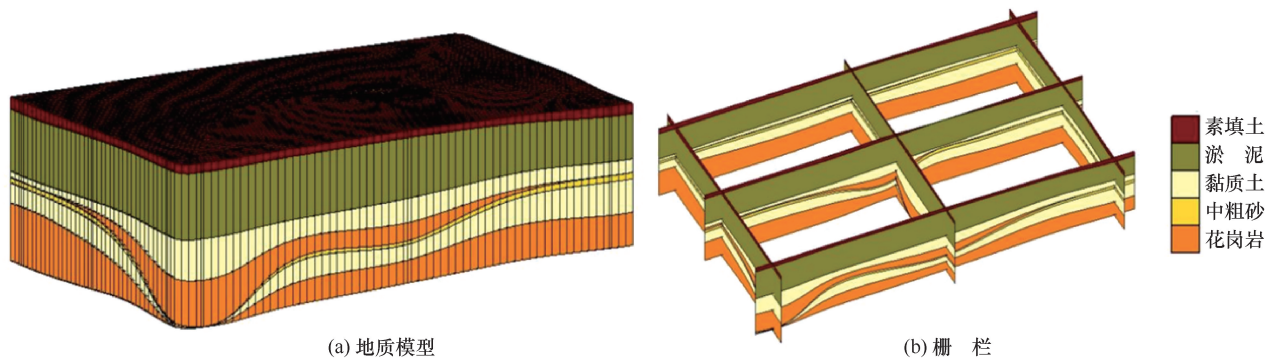


图2 变电站范围岩性空间分布

Fig. 2 Spatial distribution of lithology in substation area

### 2.2 排水管网和变电站建立前后的参数设置

变电站建成后,核心区域高程变为4.80 m,外围区域高程变为4.50 m,变电站周边范围的地面高程不变。因此,在建立变电站建成以后的地下水流模型时,将考虑这种高程变化。其次,变电站建成以后,核心区域由于路面硬化导致地下水下渗能力减弱,渗透系数和入渗补给系数相比于建站之前将大大减小,模拟区其他地区的渗透系数和入渗补给系数不变,渗透系数仍为初始值,入渗补给系数为初始设置的0.20。当年降雨量为1804.00 mm时,入渗补给地下水的强度为1.00 mm/d。

### 2.3 模型参数识别

平面上,将模拟区离散为7628个结点,15024个单元。根据水文地质概念模型,模拟变电站建立前的地下水稳定流场作为初始条件,调整各分层渗透系数取值。12眼勘探井有10眼井观测到了静止地下水埋深数据。10眼统测井模拟水位和实测水位之间的对比如图3所示,模拟水位与实测水位相近,模拟相对误差在3%~17%,与实测水位相比,模拟水位更加平缓,观测水位的突然升高(ZK9)或降低(ZK4)可能与测量误差或局部非均质介质有关。模拟流场基本反映了变电站建立前地下水流场变化特征。经识别后素填土、淤泥质土、粉质黏土、中粗砂、砂质黏性土、全风化花岗岩和强风化花岗岩的渗透系数分别为0.69 m/d、0.52 m/d、1.30 m/d、34.56 m/d、0.35 m/d、2.59 m/d、0.60 m/d。

### 2.4 不同暴雨强度下的模拟情景设置

为判断极端降雨事件下的变电站排水效果,按50 a一遇和100 a一遇的暴雨频率两种情景考虑。根据当时气象资料,考虑天、小时和10 min三种时间尺度。因此共设计3种情景,分别表示为S1、S2和S3情景。S1、S2和S3表示100 a一遇情景,降雨量分别为422.3 mm/d(持续7 d)、108.36 mm/h(持续

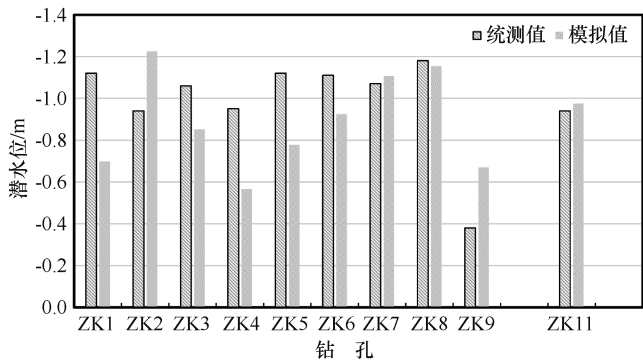


图3 典型观测井统测和模拟地下水水位对比

Fig. 3 Comparison between measured and simulated groundwater levels in typical monitoring wells

8 h)、4.052 min(持续 60 min), 同时考虑多年平均降雨情景, 每个情景模拟 90 d。模拟区不同暴雨情景剖分网络所有结点的地下水埋深数据将作为机器学习模型的数据源。由于该数据源数据量大, 近 458 万条数据, 模型训练和应用效率低, 因此考虑选择剖分网络中的 12 眼钻井点、20% 网格点和所有网格点的地下水埋深数据分别作为机器学习模型的样本训练数据。机器学习模型的样本训练数据的个数分别为 7 428、916 120 和 4 579 362 个, 而模型测试的数据均为 1 480 个。在本研究的机器学习模型应用中, 暴雨数据来自不同假定的暴雨情景, 后期可以根据布置的雨量站实时监测和气象预报进行确定。

### 3 结果分析

#### 3.1 机器学习模型精度对比分析

在本研究中, 采用 XGBoost、随机森林(RF)和长短期记忆网络(LSTM)三种机器学习模型对监测点的地下水埋深动态进行预测, 模型的模拟值与基准值的相关曲线如图 4 所示, 结果展示了模拟和基准地下水埋深在 4.60~6.00 m 之间的变化, 三个模型的 Nash-Sutcliffe 效率系数(NSE)均超过了 0.97, 显示出高度的预测准确性, 但平均绝对误差在 0.001 2~0.027 7 m。特别值得关注的是, RF 模型在模拟结果的准确性方面表现最好, 其 NSE 值高达 0.999 8, 平均绝对误差 AE 为 0.001 2 m, 几乎完美地捕捉了地下水埋深的真实动态。XGBoost 模型也展现了较高的精度, 而 LSTM 模型虽然精度略逊一筹, 但仍保持了良好的预测能力。为更清晰地展示出 12 眼钻井对应的 3 种机器学习模型精度, 对其精度进行了统计(见表 1)。从表中可以发现, 12 眼井的 XGBoost、RF 和 LSTM 模型的 NSE 变化范围分别为 0.888 9~0.995 9、

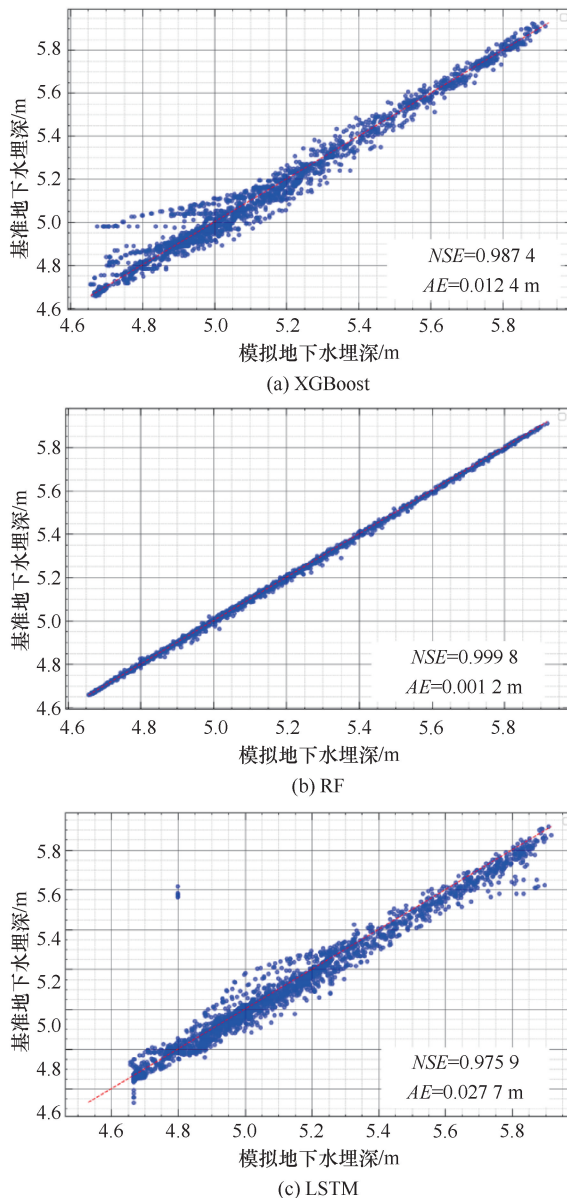


图4 基于观测井数据训练的 3 种机器学习模型模拟效果  
Fig. 4 Simulation performance of three machine learning models trained on monitoring well data

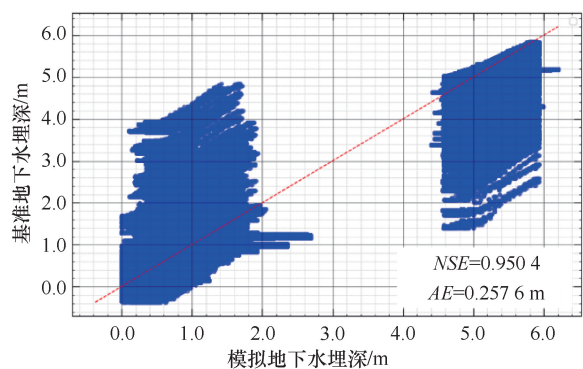
0.999 3~0.999 9 和 0.813 3~0.983 1, RMSE 变化范围分别为 0.011 3~0.061 5 m, 0.002 1~0.005 3, 0.018 7~0.101 2 m, Pearson 相关系数都在 0.91 以上。RF 模型在 12 眼井的表现效果都很好, 但个别井(例如 ZK2)在 XGBoost 和 LSTM 模型中表现较差, 影响了机器学习模型的效果。

进一步探讨了基于数值模型全网格(共 7 628 个结点)的机器学习模型性能, 特别是 XGBoost、随机森林(RF)和长短期记忆网络(LSTM)三种模型的表现, 模拟结果如图 5 所示。所有模型的地下水埋深模拟值均稳定分布在 0.00~6.50 m 的范围内。在这

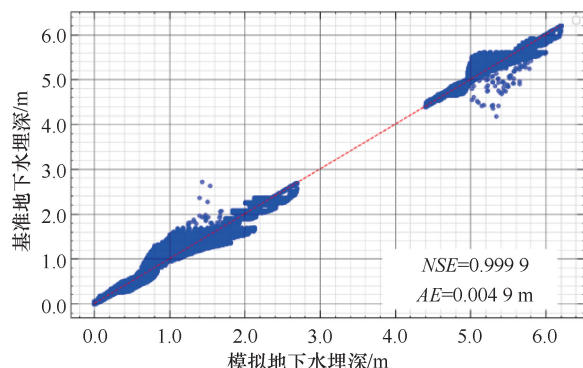
表 1 基于观测井数据训练的 3 种机器学习模型精度统计

Table 1 Accuracy statistics of three machine learning models trained on monitoring well data

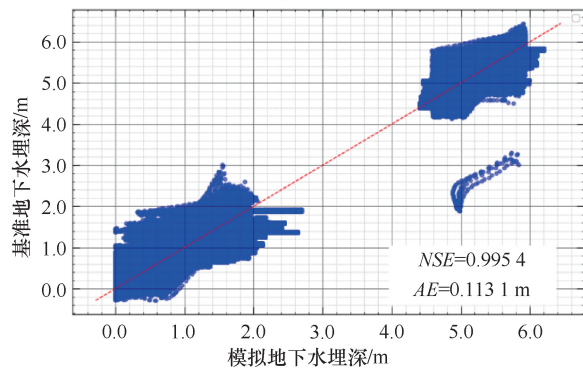
井名	Nash-Sutcliffe 效率系数 <i>NSE</i>			均方根误差 <i>RMSE</i> /m			<i>Pearson</i> 相关系数		
	XGBoost 模型	RF 模型	LSTM 模型	XGBoost 模型	RF 模型	LSTM 模型	XGBoost 模型	RF 模型	LSTM 模型
ZK1	0.987 4	0.999 5	0.813 3	0.026 3	0.005 2	0.101 2	0.993 8	0.999 8	0.911 9
ZK2	0.888 9	0.999 3	0.815 4	0.061 5	0.004 9	0.079 3	0.959 4	0.999 7	0.913 8
ZK3	0.975 6	0.999 4	0.948 9	0.032 5	0.005 3	0.047 0	0.987 8	0.999 7	0.975 7
ZK4	0.992 4	0.999 8	0.967 6	0.023 5	0.004 2	0.048 5	0.996 2	0.999 9	0.993 2
ZK5	0.994 2	0.999 7	0.976 8	0.014 3	0.003 5	0.028 6	0.997 1	0.999 8	0.994 3
ZK6	0.994 3	0.999 6	0.941 2	0.011 3	0.002 9	0.036 1	0.997 4	0.999 8	0.990 4
ZK7	0.961 3	0.999 6	0.952 2	0.023 5	0.002 3	0.026 1	0.981 4	0.999 8	0.984 3
ZK8	0.948 5	0.999 5	0.948 2	0.028 6	0.002 8	0.028 7	0.976 2	0.999 8	0.974 4
ZK9	0.995 9	0.999 9	0.983 1	0.013 8	0.002 4	0.027 9	0.998 0	0.999 9	0.997 4
ZK10	0.993 3	0.999 6	0.957 4	0.012 7	0.003 1	0.032 2	0.997 3	0.999 8	0.994 2
ZK11	0.979 0	0.999 8	0.980 8	0.019 6	0.002 1	0.018 7	0.991 0	0.999 9	0.991 6
ZK12	0.991 6	0.999 5	0.958 6	0.014 6	0.003 5	0.032 5	0.996 0	0.999 8	0.987 7



(a) XGBoost



(b) RF



(c) LSTM

图 5 基于数值模型全网格数据训练的 3 种机器学习模型模拟效果

Fig. 5 Simulation performance of three machine learning models trained on full-grid data from numerical models

一系列模型中, RF 模型再次证明了其卓越的性能, 其 Nash-Sutcliffe 效率系数 (*NSE*) 高达 0.999 9, 平均绝对误差为 0.004 9 m, 几乎达到了理论上的完美匹配。相比之下, LSTM 模型和 XGBoost 模型的 *NSE* 分别达到了 0.995 4 和 0.950 4, 平均绝对误差 *AE* 分别为 0.113 1 m 和 0.257 6 m, 虽然空间分布上的 *AE* 有较大出入, 但在时间上的 *NSE* 仍显示出模型具有较高的预测精度。为了评估模型在不同数据密度下的适应性, 将网格点数缩减至全网格的 20%, 重新进行了模拟, 结果如图 6 所示。尽管网格点数大幅减少, 但总体模拟结果仍与全网格模型保持一致, RF 模型依旧表现最佳。这一发现表明, 即使在数据较为稀疏的情况下, RF 模型仍能保持其预测性能。

### 3.2 典型井地下水埋深动态变化对比

选取了 ZK11 作为典型钻孔位的地下水监测井, 以深入分析不同情景下的地下水埋深动态。针对 S1、S2 和 S3 三种情景, 采用地下水流数值模型模拟了 90 d 内 ZK11 孔的地下水埋深动态变化, 结果如图 7 所示。从这些模拟结果可以看出, 无论是 XGBoost 模型、RF 模型还是 LSTM 模型, 它们所模拟的地下水埋深动态均与基准值保持了良好的趋势一致性。为了更精确地评估这些模型的预测精度, 统计了三种模型在 S1、S2 和 S3 情景下的 Nash-Sutcliffe 效率系数 (*NSE*)、均方根误差 (*RMSE*) 和 *Pearson* 相关系数, 如表 2 所列。由表 2 可以清晰地看到, RF 模型在所有三种情景下均表现出色, 其 *NSE* 和 *Pearson* 相关系数接近 1.00, 显示出极高的预测准确性, 而其 *RMSE* 更是控制在 0.005 m 以内。紧随其后的是 XGBoost 模型, 其 *NSE* 系数在 0.94~0.99 之间, *Pearson* 相关系数超过 0.98, *RMSE* 在 0.02~0.04 m 之间, 表明该模型在预测精度上同样表现优异。相比之下, LSTM 模型的表现略显逊色, 其 *NSE* 系数在 0.89~0.96 之间,

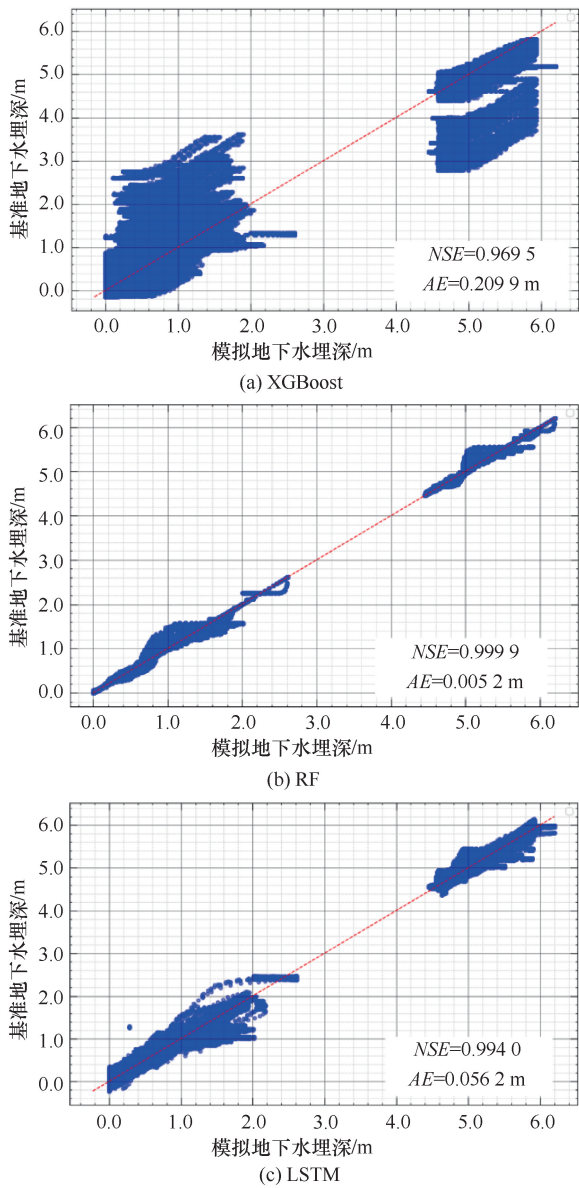


图6 基于数值模型20%网格数据训练的3种机器学习模型模拟效果

Fig. 6 Simulation performance of three machine learning models trained on 20% grid data from numerical models

Pearson 相关系数在 0.95 ~ 0.98 之间, RMSE 在 0.04~0.06 m 范围内。

### 3.3 地下水埋深空间分布的比较分析

在本研究中, 全网格和 20% 网格模型在 S1 暴雨情景结束时间(第 7 天)的基准地下水埋深以及 XGBoost、RF 和 LSTM 模型模拟误差的空间分布情况如图 8 和图 9 所示。分析这些图表, 我们可以观察到全网格和 20% 网格模型所显示的地下水埋深分布具有很好的相似性。具体来看, 在变电站核心区域, 地下水埋深较大, 超过 4.00 m; 而在变电站周边区域, 地下水埋深则在 0.25 ~ 1.25 m 之间; 向外围区

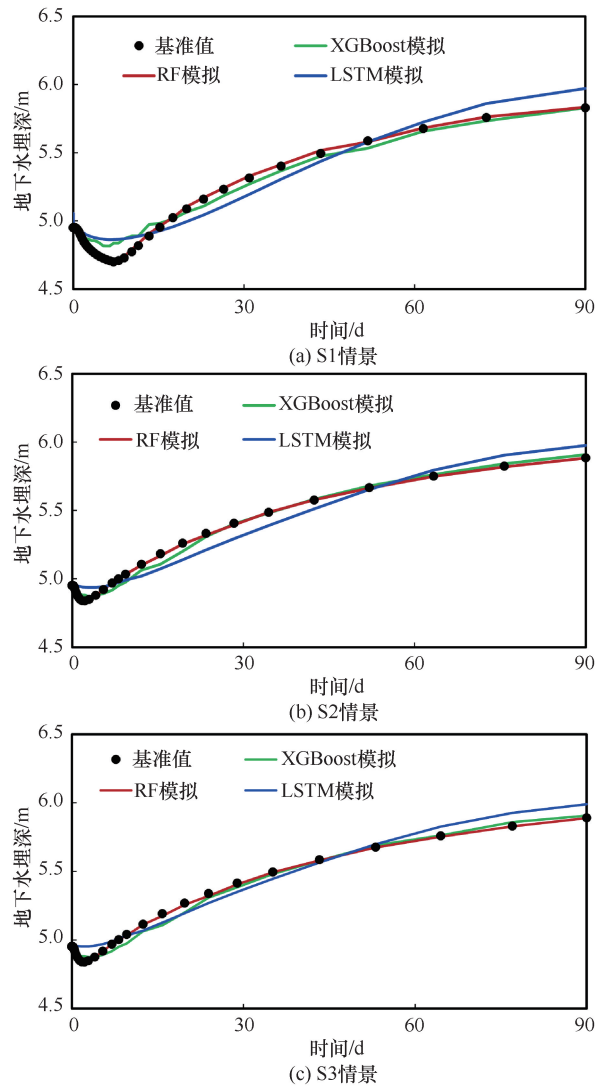


图7 不同情景下 ZK11 孔 3 种机器学习模型模拟地下水埋深动态变化对比曲线

Fig. 7 Comparison of variation curves of groundwater depth dynamics simulated by three machine learning models for well ZK11 under different scenarios

表2 3种机器学习模型在3种情景下的模型精度统计  
Table 2 Accuracy statistics of three machine learning models under three scenarios

情景	模型	Nash-Sutcliffe 效率系数 NSE	均方根误差 RMSE/m	Pearson 相关系数
S1	XGBoost	0.942 6	0.041 9	0.983 2
	RF	0.999 4	0.004 7	0.999 8
	LSTM	0.890 4	0.060 8	0.957 5
S2	XGBoost	0.991 1	0.019 3	0.995 6
	RF	0.999 8	0.003 1	0.999 9
	LSTM	0.953 4	0.042 8	0.979 1
S3	XGBoost	0.989 8	0.020 7	0.995 0
	RF	0.999 7	0.003 5	0.999 9
	LSTM	0.964 0	0.039 3	0.985 5

域, 地下水埋深则降至 0.25 m 以内。RF 模型在空间上的模拟精度表现出色, 其绝对误差控制在  $-0.05 \sim 0.05$  mm 范围内, 显示出极高的空间一致性和预测准确性。相比之下, XGBoost 模型在空间上的模拟误差表现出较大的不均匀性, 尤其是在变电站核心区周边, 模拟值偏低; 在 220kV 园区变电站核心区内, 模拟值则存在高估或低估的情况, 最大误差在 0.25 m 以内; 而在变电站核心区外围, 模拟值的误差较大, 超过 0.25 m。LSTM 模型则显示出更大的不均匀性, 其模拟结果基本倾向于高估地下水埋深, 最大误差超过 0.25 m。

为了更全面地评估这三种机器学习模型的计算效率, 统计了它们的训练和预测耗时情况(见表 3)。总体而言, XGBoost 模型因其高效的训练速度而表现突出, RF 模型紧随其后, 而 LSTM 模型由于其复杂的结构, 训练所需时间最长。具体数据显示, XGBoost 模型在构建全网格(458 万个数据量)和 20% 网格(92 万个数据量)时的耗时最少, 其模型训练耗时仅为 RF 模型的约 1/4, LSTM 模型的约 1/25; 同时,

其模型训练时间与另两种模型的比值也更小。LSTM 模型在 20% 网格的模型训练时长为 1 688 s(约 30 min), 模型预测时间为 71 s, 而对于全网格时, 模型训练和预测时间增加了约 5 倍。

#### 4 结果讨论

变电站排水系统的智能监控与预警系统是确保变电站安全稳定运行的关键研究内容。其中, 准确而且高效的地下水位动态预测是预警系统应用的核心。设计一套既可靠又快速的计算方法, 是一项具有挑战性的任务。考虑到传统物理机制模型的局限性, 研究基于数值模型的基准数据集评估和分析了三种机器学习模型的模拟效果。

上述结果表明随机森林模型在模拟精度和耗时方面均表现出色, 因此被选定为变电站排水风险实时预报的理想模块。为了实现这一目标, 首先利用现有的监测数据和数值模型, 生成了不同暴雨情景下地下水埋深的时空动态数据。在模型训练阶段, 我们不仅对模型进行了充分的训练, 还保存了这些训

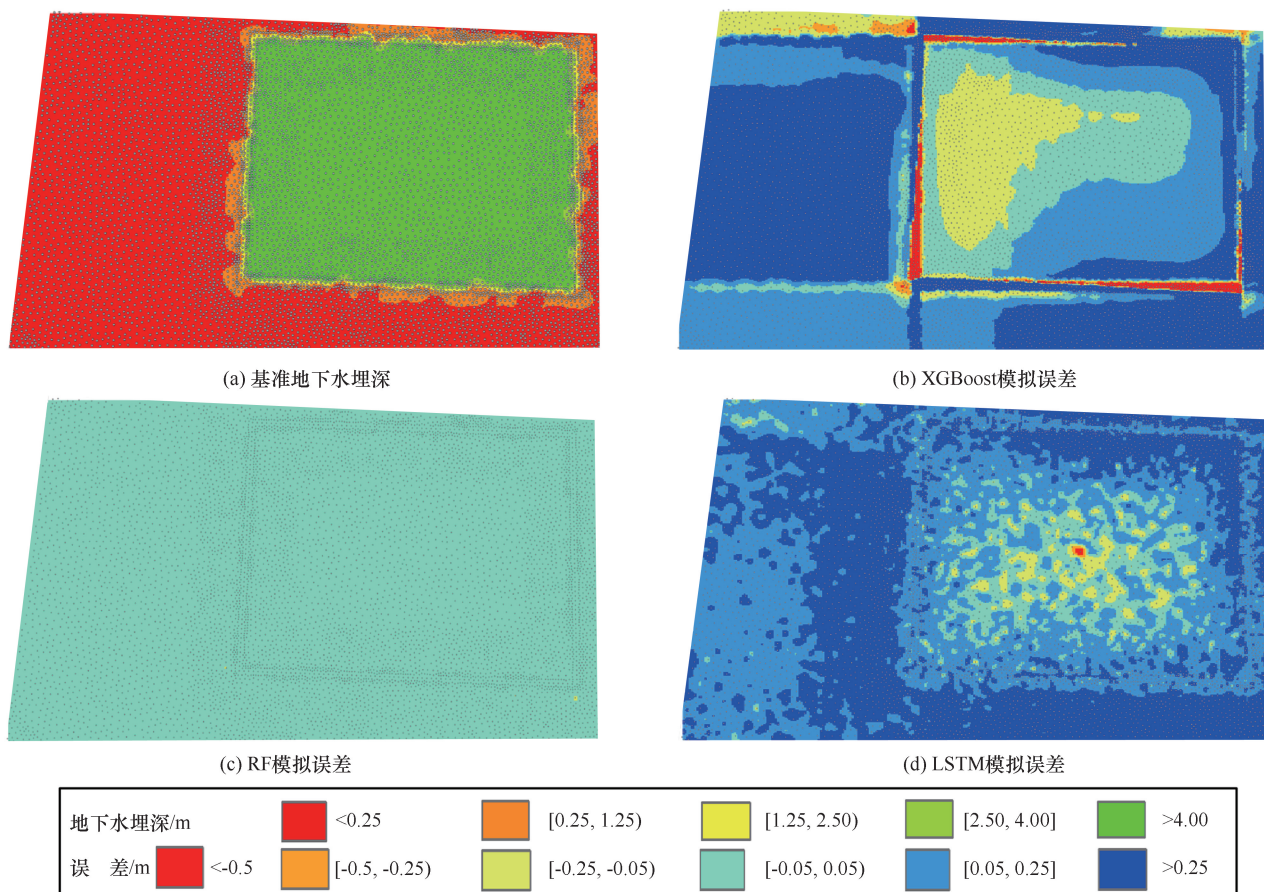


图 8 S1 情景 7 d 时基于数值模型全网格数据训练的 3 种机器学习模型模拟效果

Fig. 8 Simulation performance of three machine learning models trained on full-grid data from numerical models on day 7 under scenario S1

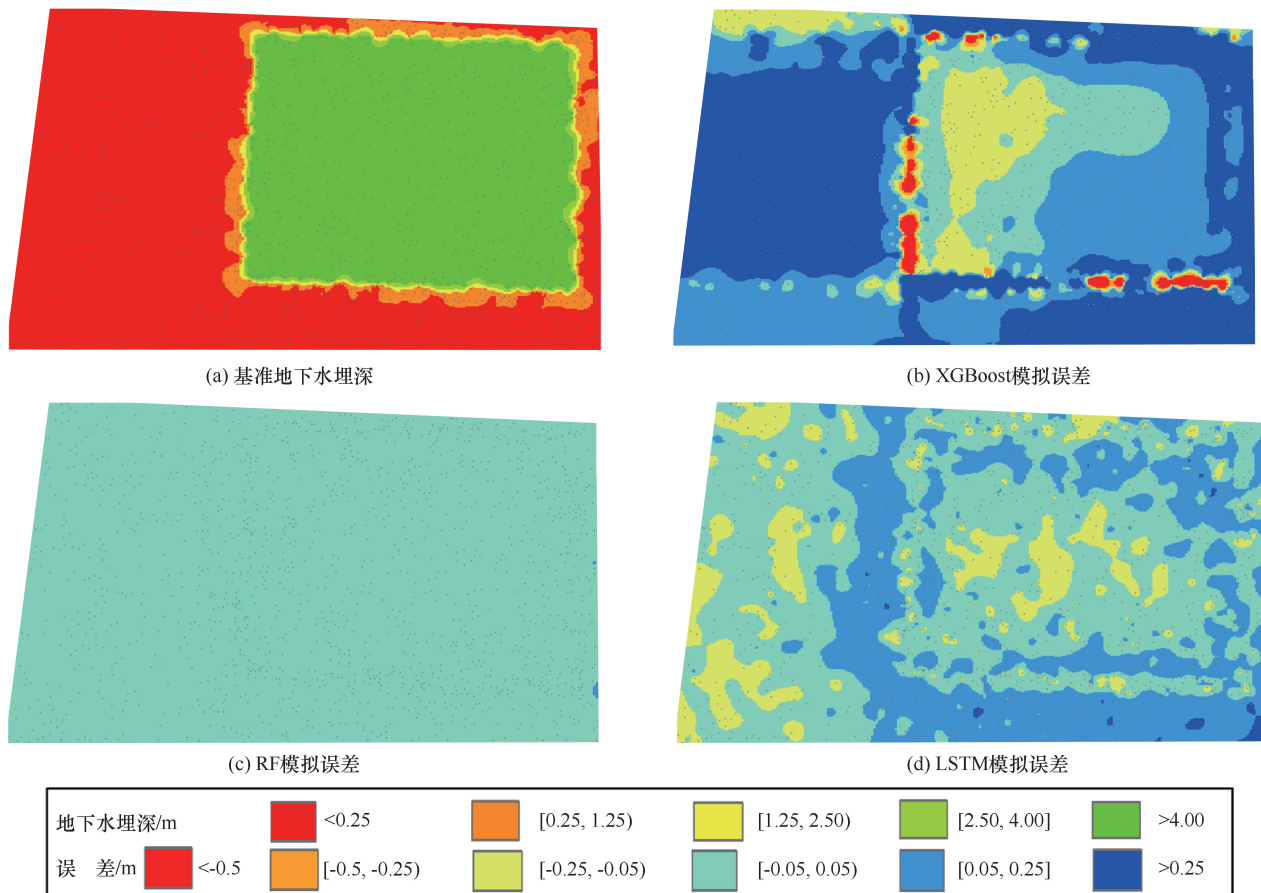


图9 S1 情景 7 d 时基于数值模型 20%网格数据训练的 3 种机器学习模型模拟效果

Fig. 9 Simulation performance of three machine learning models trained on 20% grid data from numerical models on day 7 under scenario S1

表 3 S1 情景不同网格的 3 种机器学习模型训练和预测耗时统计

Table 3 Statistics of training and prediction time of three machine learning models with different grids under scenario S1

数据量	模型步骤	XGBoost 模型	RF 模型	LSTM 模型
458 万条 (全网格)	模型训练	402	1731	8 653
	模型预测	5	153	360
92 万条 (20% 网格)	模型训练	63	250	1 688
	模型预测	1	23	71

练好的模型参数,以便在实际应用中快速调用。在实时预测时,只需加载这些预先训练好的模型参数,即可迅速进行预测,大大提高了预测的效率和响应速度。

为了进一步减少模型实时预测的时间,我们可以从已有的实践经验或数值模型中筛选出重点关注点,这样不仅减少了预测点的数量,也使得实时预测成为可能。为对比显示区域特征点数据量对机器学习模型的影响,选择了所有网格的 5% (370 个点) 对应的数

据作为训练数据,3 种机器学习模型对于 S1 情景在第 7 天的模拟结果如图 10 所示。相比图 9 显示的 20% (1 480 个点), 5% 网格数的 3 种机器学习模型的模拟误差明显较大,特别是变电站附近的模拟值普遍偏低,误差大于 0.50 m,因此选取的训练数据应该适中,采用空间数量少而且空间分布稀疏的网格点数据进行训练,难以反映区域地下水埋深的空间变化趋势。

通过空间分布适量的数据输入模型训练,可更有效地给出空间变化趋势,更将有效地应对变电站排水风险,确保预测的及时性和准确性。建立的这一模块将成为 220 kV 园区变电站排水风险监控平台的重要组成部分。通过整合在线监测数据和视频监控,我们可以实时监控变电站的排水情况,并逐步验证和优化模型,从而不断提高模型的实用性和预测精度。

## 5 结论

变电站排水系统的监控与预警是确保变电站安

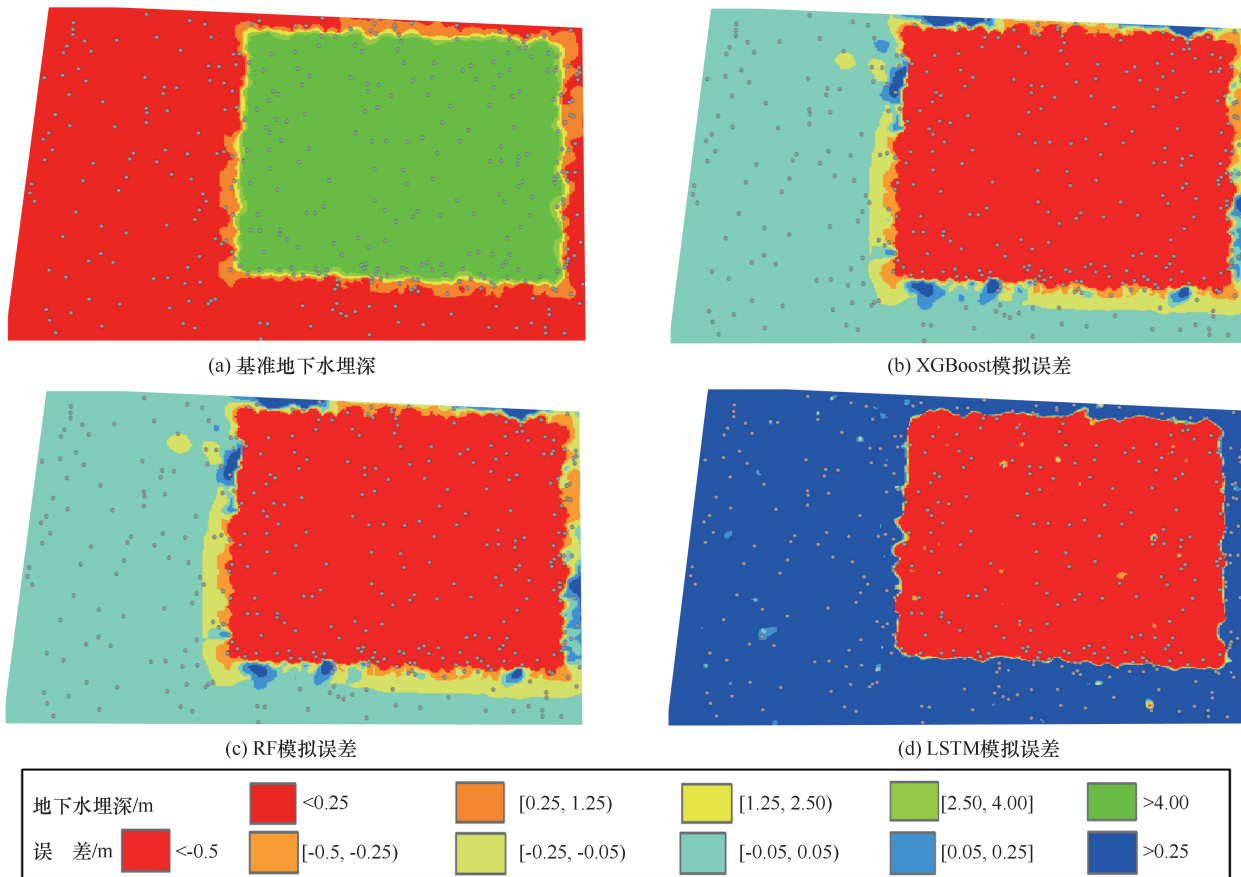


图 10 S1 情景 7 d 时基于数值模型 5% 网格数据训练的 3 种机器学习模型模拟效果

Fig. 10 Simulation performance of three machine learning models trained on 5% grid data from numerical models on day 7 under scenario S1

全运行的重点研究内容。本研究聚焦于机器学习模型在 220 kV 园区变电站地下水埋深预警系统中的试点应用, 通过深入分析, 得出了以下主要结论。

(1) 经过严格验证的地下水流数值模型可提供可靠的数据源, 因此采用了地下水流数值模型来模拟不同暴雨情景下的基准地下水埋深时空动态数据作为模型训练输入, 模型网格的大量时空数据对于模型的训练和验证至关重要。

(2) XGBoost、RF 和 LSTM 三种模型均能模拟出与基准结果相似的地下水埋深时空动态。然而, RF 模型在时间和空间的模拟精度上表现最为出色, 同时在计算耗时方面也相对较为优越, 表明 RF 模型在实时预警系统中具有较高的应用潜力。

(3) 为了更精确地捕捉 220kV 园区变电站研究范围内的地下水埋深时空动态, 采用了 20% 网格点的模型数据进行训练, 取得了良好的效果, 但 5% 网格点的模型数据构建模型的模拟精度不佳。根据已有的模型可以确定积水区(如地下水埋深小于 0.20 m)的具体分布, 同时减少网格点的数量可以

节省模型计算耗时。此外, 持续的在线监测点应优先设置在积水点范围内, 以便更有效地监测和预警潜在的排水风险。

本研究构建的机器学习模型部署在变电站试点的智能预报预警平台上, 在实际应用中可以结合变电站实时降雨预报进行地下水埋深动态预测, 再结合变电站确定的地下水埋深预警阈值, 实时确定重点监控区, 为变电站安全稳定运行提供技术服务。

参考文献 (References):

[1] 周刚, 徐伟, 陈刚, 等. 变电站智能排水系统的研制[J]. 华东电力, 2013, 41(7): 1555-1557.  
 ZHOU G, XU W, CHEN G, et al. Development of substation smart drainage system [J]. East China Electric Power, 2013, 41(7): 1555-1557.

[2] 胡杨. 变电站电缆层汛情监测及智能排水系统的研制[J]. 农村电气化, 2024(6): 65-68.  
 HU Yang. Development of flood monitoring and intelligent drainage

- system for substation cable layer[J]. *Rural Electrification*, 2024 (6): 65-68.
- [3] SAHOO S, RUSSO T A, ELLIOTT J, et al. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U. S[J]. *Water Resources Research*, 2017, 53(5): 3878-3895.
- [4] SUN A Y, SCANLON B R. How can Big Data and machine learning benefit environment and water management: A survey of methods, applications, and future directions [J]. *Environmental Research Letters*, 2019, 14(7): 073001.
- [5] KONAPALA G, KAO S C, PAINTER S L, et al. Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US[J]. *Environmental Research Letters*, 2020, 15(10): 104022.
- [6] WANG N Z, ZHANG D X, CHANG H B, et al. Deep learning of subsurface flow *via* theory-guided neural network [J]. *Journal of Hydrology*, 2020, 584: 124700.
- [7] 邓凡杰, 司乔瑞, 李方忠, 等. 山区和边远灾区应急供水系统提水装备发展现状[J]. *排灌机械工程学报*, 2023, 41(6): 614-624.
- DENG Fanjie, SI Qiaorui, LI Fangzhong, et al. Review on water carrying equipment for emergency water supply system in mountainous areas and remote disaster area[J]. *Journal of Drainage and Irrigation Machinery Engineering*, 2023, 41(6): 614-624.
- [8] 陈雪, 胡宏昌, 刘慧, 等. 基于机器学习算法的清涧河流域径流及含沙量预测研究[J]. *中国水利水电科学研究院学报(中英文)*, 2024, 22(5): 520-529.
- CHEN Xue, HU Hongchang, LIU Hui, et al. Prediction of runoff and sediment concentration in Qingjian River basin based on machine learning algorithm[J]. *Journal of China Institute of Water Resources and Hydropower Research*, 2024, 22(5): 520-529.
- [9] SHEN C P. A transdisciplinary review of deep learning research and its relevance for water resources scientists [J]. *Water Resources Research*, 2018, 54(11): 8558-8593.
- [10] BOO K B W, EL-SHAFIE A, OTHMAN F, et al. Groundwater level forecasting with machine learning models: A review [J]. *Water Research*, 2024, 252: 121249.
- [11] SERRADILLA O, ZUGASTI E, RODRIGUEZ J, et al. Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects [J]. *Applied Intelligence*, 2022, 52(10): 10934-10964.
- [12] VANCE T C, HUANG T, BUTLER K A. Big data in earth science: emerging practice and promise [J]. *Science*, 2024, 383(6688): eadh9607.
- [13] 张然, 柴志勇, 张婷, 等. 基于机器学习模型的洪水预报研究进展[J]. *水利水电技术(中英文)*, 2023, 54(11): 89-101.
- ZHANG R, CHAI Z Y, ZHANG T, et al. Research progress of flood forecasting based on machine learning models [J]. *Water Resources and Hydropower Engineering*, 2023, 54(11): 89-101.
- [14] HOU J M, WANG N, GUO K H, et al. Effects of the temporal resolution of storm data on numerical simulations of urban flood inundation [J]. *Journal of Hydrology*, 2020, 589: 125100.
- [15] CHANG F J, CHEN P N, LU Y R, et al. Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control [J]. *Journal of Hydrology*, 2014, 517: 836-846.
- [16] WU Z N, ZHOU Y H, WANG H L. Real-time prediction of the water accumulation process of urban stormy accumulation points based on deep learning [J]. *IEEE Access*, 2020, 8: 151938-151951.
- [17] ZHOU Y H, WU Z N, JIANG M M, et al. Real-time prediction and ponding process early warning method at urban flood points based on different deep learning methods [J]. *Journal of Flood Risk Management*, 2024, 17(1): e12964.
- [18] THESSLER A, PÉREZ-VELÁZQUEZ J, KETTELGERDES M, et al. Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry [J]. *Reliability Engineering & System Safety*, 2021, 215: 107864.
- [19] SUN J C, HU L T, LI D D, et al. Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management [J]. *Journal of Hydrology*, 2022, 608: 127630.
- [20] ZHANG J F, ZHU Y, ZHANG X P, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas [J]. *Journal of Hydrology*, 2018, 561: 918-929.
- [21] IBRAHEM AHMED OSMAN A, NAJAH AHMED A, CHOW M F, et al. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia [J]. *Ain Shams Engineering Journal*, 2021, 12(2): 1545-1556.
- [22] ALI S, RAN J J, KHORRAMI B, et al. Downscaled GRACE/GRACE-FO observations for spatial and temporal monitoring of groundwater storage variations at the local scale using machine learning [J]. *Groundwater for Sustainable Development*, 2024, 25: 101100.
- [23] BREIMAN L. Random Forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [24] KOCH J, BERGER H, HENRIKSEN H J, et al. Modelling of the shallow water table at high spatial resolution using random forests [J]. *Hydrology and Earth System Sciences*, 2019, 23(11): 4603-4619.
- [25] LIU Q, GUI D W, ZHANG L, et al. Simulation of regional groundwater levels in arid regions using interpretable machine learning models [J]. *Science of The Total Environment*, 2022, 831: 154902.
- [26] WANG X H, LIU T L, ZHENG X L, et al. Short-term prediction of groundwater level using improved random forest regression with a combination of random features [J]. *Applied Water Science*, 2018, 8(5): 125.

- [27] XIANG Z R, YAN J, DEMIR I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning[J]. *Water Resources Research*, 2020, 56(1): e2019WR025326.
- [28] CHU H B, BIAN J M, LANG Q, et al. Daily groundwater level prediction and uncertainty using LSTM coupled with PMI and bootstrap incorporating teleconnection patterns information [J]. *Sustainability*, 2022, 14(18): 11598.
- [29] DIERSCH H J G. In *FEFLOW: Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media* [R]. Berlin, Heidelberg: Springer, 2014.
- [30] 陈崇希, 唐仲华, 胡立堂. 地下水流数值模拟理论方法及模型设计[M]. 北京: 地质出版社, 2014.
- CHEN C X, TANG Z H, HU L T. *Theoretical Methods and Model Design for Groundwater Flow Numerical Simulation* [M]. Beijing: Geological Publishing House, 2014.
- [31] SUN K N, HU L T, LIU X M. The influences of sponge city construction on spring discharge in Jinan city of China[J]. *Hydrology Research*, 2020, 51(5): 959-975.

(责任编辑 王璐)