

DOI:10.3876/j.issn.1000-1980.2025.05.015

# 基于 PCA-AVOA-LightGBM 的混凝土坝应力预测模型

常留红<sup>1,2</sup>, 朱 勇<sup>1,2</sup>, 曾子彬<sup>1,2</sup>, 尹光景<sup>3</sup>, 高宏宇<sup>1,2</sup>, 邬传峰<sup>1,2</sup>

(1. 长沙理工大学水利与环境工程学院, 湖南 长沙 410114; 2. 长沙理工大学水沙科学与水灾害防治湖南省重点实验室, 湖南 长沙 410114; 3. 中国水利水电第八工程局有限公司科研设计院, 湖南 长沙 410004)

**摘要:** 基于主成分分析 (PCA) 方法、非洲秃鹫优化算法 (AVOA) 和轻量级梯度提升学习机 (LightGBM) 模型构建了 PCA-AVOA-LightGBM 混凝土坝应力预测模型, 模型采用 PCA 方法挖掘降维应力预测的主要影响因素, 引入 AVOA 优化 LightGBM 模型超参数。依托某混凝土坝应力监测数据, 将 PCA 方法应用于向量回归机、随机森林、极端梯度提升、LightGBM 等模型中, 并与 PCA-AVOA-LightGBM 模型进行了对比分析。结果表明, PCA 方法有效降低了各模型影响因素间多重共线性, PCA-AVOA-LightGBM 模型相较于其他模型在预测精度和效率中表现出更优异的性能, 可在类似混凝土坝的应力监测中推广应用。

**关键词:** 混凝土坝; 超参数; 应力预测; 主成分分析方法; 非洲秃鹫优化算法; 极端梯度提升

**中图分类号:** TV698.1 **文献标志码:** A **文章编号:** 1000-1980(2025)05-0127-09

## Stress prediction model of concrete dam based on PCA-AVOA-LightGBM

CHANG Liuhong<sup>1,2</sup>, ZHU Yong<sup>1,2</sup>, ZENG Zibin<sup>1,2</sup>, YIN Guangjing<sup>3</sup>, GAO Hongyu<sup>1,2</sup>, WU Chuanfeng<sup>1,2</sup>

(1. School of Hydraulic and Environmental Engineering, Changsha University of Science & Technology, Changsha 410114, China;

2. Key Laboratory of Water & Sediment Science and Water Hazard Prevention of Hunan Province, Changsha University of Science & Technology, Changsha 410114, China;

3. Research and Design Institute of Sinohydro Engineering Bureau 8 Co., Ltd., Changsha 410004, China)

**Abstract:** Based on the principal component analysis (PCA) method, the African vulture optimization algorithm (AVOA), and lightweight gradient learning machine (LightGBM) model, a stress prediction model of concrete dam based on PCA-AVOA-LightGBM was constructed. PCA method was used to mine the main influencing factors of dimension reduction-based stress prediction, and AVOA was introduced to optimize the hyperparameters of the LightGBM model. Based on the stress monitoring data of a concrete dam, the PCA method was applied to vector regression machine (SVR), random forest (RF), extreme gradient boosting (XGboost), LightGBM models, and a comparative analysis was conducted with the PCA-AVOA-LightGBM model. The results show that the PCA method effectively reduces the multicollinearity among the influencing factors of each model. The PCA-AVOA-LightGBM model shows better performance in accuracy and efficiency than other prediction models and can be applied in stress monitoring of similar concrete dams.

**Key words:** concrete dam; hyperparameter; stress prediction; principal component analysis; African vulture optimization algorithm; extreme gradient boosting

我国已建混凝土坝的数量与高度均居世界首位, 未来仍将建设大量混凝土坝<sup>[1]</sup>。混凝土坝工作环境复杂, 在服役过程中需要承受多种动、静循环荷载, 随着服役时间的推移, 易出现不同程度的隐患乃至险情<sup>[2]</sup>, 大坝安全监测已成为掌握大坝工作状态、保障大坝运行安全的主要措施。其中, 应力是大坝安全监测中主要的监测量之一, 一旦应力超出材料或基岩的强度极限, 将直接导致大坝失稳或破坏。因此, 构建大坝应力预测模型, 结合预测值与监测数据判断大坝运行是否异常, 对保障大坝安全具有重要意义<sup>[3]</sup>。

大坝的应力预测模型主要包括回归模型、确定性模型、混合模型、灰色模型及机器学习模型等<sup>[4-5]</sup>。随着

**作者简介:** 常留红 (1979—), 女, 副教授, 博士, 主要从事水工结构与水生态修复研究。E-mail: claire886@163.com

**引用本文:** 常留红, 朱勇, 曾子彬, 等. 基于 PCA-AVOA-LightGBM 的混凝土坝应力预测模型[J]. 河海大学学报(自然科学版), 2025, 53(5): 127-135.

CHANG Liuhong, ZHU Yong, ZENG Zibin, et al. Stress prediction model of concrete dam based on PCA-AVOA-LightGBM[J]. Journal of Hohai University(Natural Sciences), 2025, 53(5): 127-135.

机器学习技术的发展,基于随机森林(random forest, RF)<sup>[6]</sup>、支持向量机(support vector machine, SVM)<sup>[7-8]</sup>、极端梯度提升(extreme gradient boosting, XGBoost)<sup>[9]</sup>的集成机器学习模型,凭借其在预测精度与效率上的优势,已被应用于大坝应力预测领域。其中, XGBoost 模型在可扩展性与准确性上均优于 RF 和 SVM 模型<sup>[10-12]</sup>,但 XGBoost 模型的迭代次数与特征维度相关,每次迭代都会搜索整个数据集的数据,存在内存消耗较大、计算耗时较长的问题。轻量级梯度提升学习机(light gradient boosting machine, LightGBM)算法在保留 XGBoost 模型优势的基础上,进一步提升了训练速度与预测准确度,更适用于处理大坝监测数据等大规模数据集。众多研究表明,在滑坡预测<sup>[13-14]</sup>、大坝变形预测<sup>[15]</sup>等场景中, LightGBM 模型相较于其他机器学习模型,在训练速度、预测准确性及可行性方面均有显著提升。尽管基于 LightGBM 的预测模型具备一定的优越性,但仍存在两方面问题:①输入的影响因子过多,因子间的多重共线性会降低模型的预测精度与泛化程度;②超参数数量多于其他模型,传统寻优方法难以找到最优参数。

降低影响因子复杂度的方法主要有特征提取<sup>[16-17]</sup>与降维<sup>[18-19]</sup>等。主成分分析(principal component analysis, PCA)方法可在不改变数据结构的前提下对高维数据进行降维,既能有效消除原始数据的相关性与冗余信息,又能提升模型预测的准确度与效率<sup>[20]</sup>。刘娣等<sup>[20]</sup>采用 PCA 方法筛选与土壤湿度相关性显著的气象因子,分别基于 SVM、BP 与 RF 等 3 种算法构建多层土壤温度反演模型,结果显示 RF 算法的反演效果最优。王仁超等<sup>[21]</sup>引入 PCA 方法对大坝变形影响因子进行主成分提取,优化了输入变量;同时采用布谷鸟搜索(cuckoo search, CS)算法优化核极限学习机(kernel extreme learning machine, KELM)的核参数与正则化系数,应用结果表明, PCA-CS-KELM 模型的预测精度与泛化能力均优于传统神经网络与逐步回归模型。可见, PCA 方法可有效减少输入变量间的信息重叠,降低影响因子间多重共线性对预测结果的干扰。

为解决 LightGBM 模型超参数寻优难的问题,相关学者采用粒子群优化(particle swarm optimization, PSO)<sup>[22]</sup>算法、灰狼优化(grey wolf optimizer, GWO)<sup>[23]</sup>算法、鲸鱼优化算法(whale optimization algorithm, WOA)<sup>[24]</sup>等启发式算法,对 LightGBM 模型的超参数进行优化。非洲秃鹫优化算法(African vulture optimization algorithm, AVOA)与其他启发式算法相比,在算法开发过程中融入了随机性,在保障开发效率的同时,具有更强的全局搜索能力。高永涛等<sup>[25]</sup>提出利用 AVOA-XGBoost 模型预测岩爆烈度,并将其应用于三山岛金矿工程,结果表明, AVOA 可高效确定 XGBoost 算法的超参数。牛传乐等<sup>[26]</sup>提出基于 AVOA 优化卷积神经网络(convolutional neural network, CNN)的重叠峰解析方法,解析结果显示, AVOA-CNN 可成功分解重叠峰且准确率较高,证实 AVOA 的全局寻优特性可有效减少人为调整超参数带来的误差。大量研究证明, AVOA 具有强大的寻优能力<sup>[27]</sup>,为 LightGBM 等超参数复杂的模型寻找最优参数提供了可行思路。

针对混凝土坝应力预测模型中影响因子间存在多重共线性、超参数寻优困难等问题,本文应用 PCA 方法进行主成分分析,以减少各因素间的多重共线性效应;同时,基于 AVOA 强大的寻优能力优化 LightGBM 模型超参数,构建耦合 PCA 方法、AVOA 和 LightGBM 模型的混凝土坝应力预测模型——PCA-AVOA-LightGBM 模型,以有效提升模型预测精度和效率。

1 模型构建

1.1 模型理论

1.1.1 LightGBM 模型

LightGBM 是具有开源、高效的分布式梯度提升树算法,通过限制梯度提升框架 LightGBM 的参数,有效避免了决策树易过拟合的问题<sup>[28]</sup>。LightGBM 模型以提升树(boosting tree)作为基本学习器,采用梯度提升(gradient boosting)损失函数的负梯度作为回归问题提升树残差的近似值。LightGBM 模型相对于一般集成机器学习模型而言,具备较快的训练速度、较小的内存需求和较高的预测精度,因此,在大坝各效应量预测等通常需要处理大规模数据时, LightGBM 模型处理效果更好<sup>[29]</sup>。

LightGBM 模型通过 GOSS(gradient-based one-side sampling)算法、EFB(explosive feature bundling)算法及直方图算法,有效解决了高维大数据场景下兼顾预测精度与运算效率的核心难题。这 3 种算法从不同层面协同提升模型的学习效率与训练速度,其中, GOSS 算法可以大幅减少训练样本的数目; EFB 算法与直方图算法则对特征抽取进行缩减,在降低特征数量规模的同时进一步提高学习效率。此外, LightGBM 模型中的 leaf-wise 生长策略还能显著增强模型的预测准确度。

1.1.2 PCA 方法

将影响因子输入预测模型时,若影响因子过多,将使泛化程度降低<sup>[16]</sup>。在大坝预测模型中,PCA 方法能够在减少信息损耗的情况下,对具有相同效应量且存在相关性的影响因子进行压缩与重构。最终采用 PCA 方法获得少数几个表达式,以及相互独立的主成分,从而有效减少或消除影响因子间的相关性。

在保持原有信息不变的前提下,PCA 方法对数据降维精简,降低了模型数据之间多重共线性的影响,获得了预测模型的输入变量。

1.1.3 AVOA

AVOA 由 Abdollahzadeh 等<sup>[30]</sup>于 2021 年提出,其灵感源于非洲秃鹫的独特觅食行为,是一种元启发式搜索算法,且在相关试验中取得了良好的寻优效果。AVOA 的基本思路是:用目标函数表征秃鹫的饥饿程度——适应度越低,代表秃鹫的饥饿程度越高;算法通过持续迭代优化,最终以逼近最优目标函数值的方式完成寻优过程。LightGBM 模型存在超参数复杂的问题,而引入 AVOA,利用秃鹫的群体搜索策略和个体学习能力寻找最优解,具有全局搜索能力强、快速收敛等优点,为寻找 LightGBM 模型最优参数提供了路径。

1.2 模型构建步骤

基于 LightGBM 模型、PCA 方法和 AVOA,深度挖掘影响因子间的主成分降维输入变量,再寻找 LightGBM 模型最优参数,进而构建 PCA-AVOA-LightGBM 模型。具体步骤如下:

**步骤 1** 预测模型各参数输入。对原始应力时间序列数据进行预处理,采用拉依达准则检测监测数据中的异常值并剔除突变点,以减少数据问题对预测结果的干扰,得到初始的数据集。通过分析各效应量的影响因素,选取相关影响因子,依据统计模型公式计算各影响分量,随后将所有影响因子整合至同一数据集并进行划分。将划分后的数据集导入 SPSS 软件进行主成分分析,将高维影响因子转化为低维主成分,以此作为预测模型的输入变量,实测值则作为输出变量。

**步骤 2** 模型训练。采用 Python 深度学习工具箱搭建 LightGBM 模型,初始化模型参数。通过 AVOA 生成初始种群并计算其适应度,划分第一、二类非洲秃鹫种群后选取种群领导者,对各超参数进行迭代寻优并更新秃鹫位置,直至满足预设种群数量条件。最终基于 AVOA-LightGBM 模型的最优参数,得到模型超参数的最优解,从而获得最优 AVOA-LightGBM 模型。

PCA-AVOA-LightGBM 模型构建流程如图 1 所示。

1.3 模型评价

构建 RF、SVR、XGBoost、LightGBM 模型并进行参数调整,将经 PCA 方法降维后的影响因子输入各模型,同时对比未采用 PCA 方法时各模型的预测结果,以验证 PCA 方法在降低影响因子间多重共线性方面的有效性。为准确评价 PCA-AVOA-LightGBM 模型的预测精准度和计算效率等性能,引入平均绝对误差 (MAE)、平均绝对百分误差 (MAPE)、均方根误差 (RMSE) 及运算时间等指标进行综合评价,比较该组合模型与其他对比模型的各项性能指标。

2 实例验证

2.1 工程概况及数据选取

某水电站为大(2)型水库,主要功能以发电为主。工程区域属于热带季风气候,年平均气温 27℃ 左右,全年气温变化较小。水电站大坝为混凝土碾压重力坝,从左至右依次为:左岸非溢流坝段(1~4 号),长 81.75 m;进水口坝段(5 号~7 号),长 66 m;底孔坝段(8 号),长 22.75 m;1 号溢流表孔坝段(9 号),长 21 m;隔墙坝段(10 号),长 12 m;2 号~4 号溢流表孔坝段(11 号~12 号),长 57 m;右岸非溢流坝段(13 号~

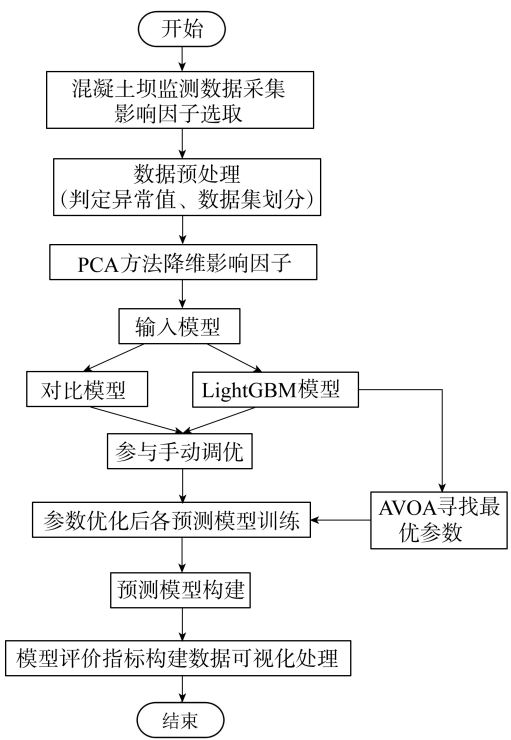


图 1 PCA-AVOA-LightGBM 模型构建流程  
Fig.1 Construction of PCA-AVOA-LightGBM model

15号),共15个坝段,长80 m。

选取11号溢流表孔坝段为研究对象,该坝段已布设倒垂系统、多点位移计、基岩变位计、测缝计、钢筋计、压应力计等监测设备,选取压应力计监测点B11-C-01(图2)数据进行分析。根据2015—2023年实测数据,测点原始数据较为完整,大坝于2015年11月8日关闸蓄水,运行期大坝上游最高水位为441.33 m(2018年10月26日),最低水位为390.85 m(2015年11月9日),水位变幅为23.34 m。B11-C-01监测点应力变幅较小,数据稳定无异常。11号溢流表孔坝段上下游水位和压应力时间序列曲线见图3。

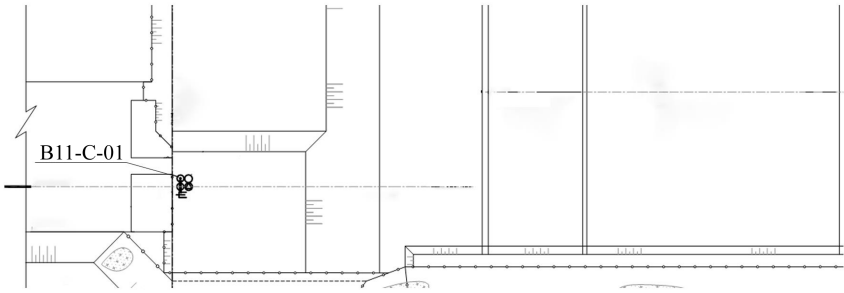


图2 11号溢流表孔坝段监测点布置

Fig. 2 Monitoring point layout on No. 11 overflow crest dam section

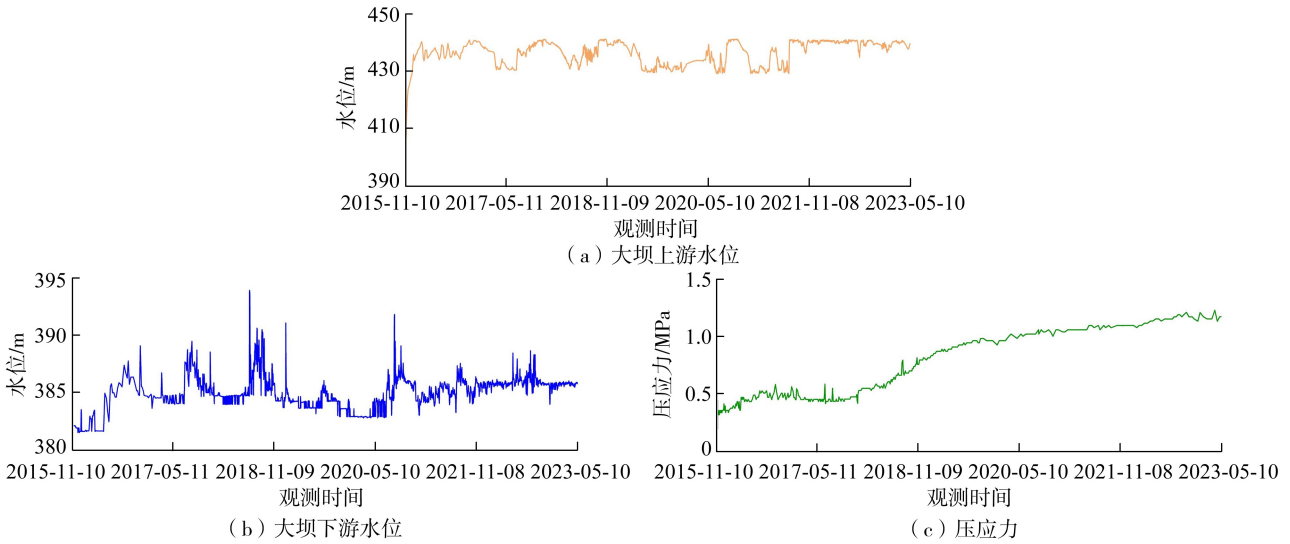


图3 水位与压应力时间序列过程曲线

Fig. 3 Time series curves of water level and compressive stress

2.2 数据预处理

以2015年11月10日至2020年11月4日的应力数据作为训练集,2020年11月7日至2023年9月10日的应力监测数据作为测试集,监测数据间隔时间为1 d。从图3(c)可知,11号溢流表孔坝段监测点位的压应力总体较平稳,但在观测初期呈现高度的非规律性。因此,在建立应力预测模型时需结合数据预处理方法将异常点进行剔除增补。因观测数据时间跨度大且数据较为完整,因此仅剔除异常点。采用 $3\sigma$ 准则对混凝土坝应力数据进行检验。步骤如下:①构建压应力原始数据数列 $\{y_1, y_2, \dots, y_n\}$ ;②计算平均值 $\bar{y}_i$ 、方差值 $\hat{\sigma}_d$ ,得到 $\bar{y}_i$ 为0.7963,  $\hat{\sigma}_d$ 为0.3118;③计算 $y_i$ 与 $\hat{\sigma}_d$ 差值,判断是否为异常值。经检验,压应力观测数据无异常值,可用于预测模型实测值输出。

2.3 影响因子主成分分析

2.3.1 影响因子选取

混凝土坝的应力主要与水压、温度、自重、湿胀以及时效等因素相关。依据混凝土坝应力监测数据及应力传统统计模型影响因子经验公式<sup>[31]</sup>,确定输出量为坝体压应力 $\sigma$ :

$$\sigma = \sigma_H + \sigma_T + \sigma_G + \sigma_W + \sigma_\theta$$

(1)

式中  $\sigma_H$ 、 $\sigma_T$ 、 $\sigma_G$ 、 $\sigma_W$ 、 $\sigma_\theta$  分别为水压、温度、自重、湿胀及时效分量。其中,自重、湿胀分量在大坝运行后通常不再改变,均视为常量。因此,考虑坝体压应力影响因素为水压分量(影响因子为  $H-H_0$ 、 $H^2-H_0^2$ 、 $H^3-H_0^3$ ,其中  $H$  为上下游水位差,  $H_0$  为基准应力时的水头)、温度分量(影响因子为  $\sin\frac{2\pi t}{365}$ 、 $\sin\frac{4\pi t}{365}$ 、 $\cos\frac{2\pi t}{365}$ 、 $\cos\frac{4\pi t}{365}$ ,其中  $t$  为累计监测天数)、时效分量(影响因子为  $\theta-\theta_0$ 、 $\ln\theta-\ln\theta_0$ ,其中  $\theta=\frac{t}{100}$ )。

2.3.2 主成分提取

根据 PCA 方法原理,基于混凝土坝监测数据,采用 SPSS 软件中 Pearson 相关系数对  $H-H_0$ (水压 1)、 $H^2-H_0^2$ (水压 2)、 $H^3-H_0^3$ (水压 3)、 $\sin\frac{2\pi t}{365}$ (温度 1)、 $\sin\frac{4\pi t}{365}$ (温度 2)、 $\cos\frac{2\pi t}{365}$ (温度 3)、 $\cos\frac{4\pi t}{365}$ (温度 4)、 $\theta-\theta_0$ (时效 1)、 $\ln\theta-\ln\theta_0$ (时效 2) 9 项应力影响因子进行分析,得到应力影响因子相关性矩阵如表 1 所示。Pearson 相关系数可以表达两变量相关程度,相关系数越高表示两变量间的相关程度越密切。

表 1 影响因子相关性矩阵

| Table 1 Influencing factor correlation matrix |         |         |         |         |         |         |         |         |         |
|-----------------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 影响因子                                          | 水压 1    | 水压 2    | 水压 3    | 温度 1    | 温度 2    | 温度 3    | 温度 4    | 时效 1    | 时效 2    |
| 水压 1                                          | 1.0000  | 0.9898  | 0.9743  | -0.0996 | -0.0339 | -0.0083 | -0.0361 | 0.8991  | 0.7201  |
| 水压 2                                          | 0.9898  | 1.0000  | 0.9961  | -0.6676 | -0.0326 | -0.0035 | -0.0250 | 0.8799  | 0.6663  |
| 水压 3                                          | 0.9743  | 0.9961  | 1.0000  | -0.0510 | -0.0354 | -0.0024 | -0.0224 | 0.8626  | 0.6404  |
| 温度 1                                          | -0.0996 | -0.0668 | -0.0510 | 1.0000  | 0.0796  | 0.0572  | 0.0173  | -0.1425 | -0.1558 |
| 温度 2                                          | -0.3389 | -0.0326 | -0.0355 | 0.0796  | 1.0000  | 0.1063  | 0.1065  | -0.0898 | -0.2041 |
| 温度 3                                          | -0.0083 | -0.0035 | -0.0024 | 0.0572  | 0.1063  | 1.0000  | 0.0713  | -0.1056 | -0.1387 |
| 温度 4                                          | -0.0361 | -0.0250 | -0.0224 | 0.0173  | 0.1065  | 0.0713  | 1.0000  | -0.0777 | -0.1656 |
| 时效 1                                          | 0.8991  | 0.8799  | 0.8626  | -0.1425 | -0.0898 | -0.1056 | -0.0773 | 1.0000  | 0.8264  |
| 时效 2                                          | 0.7201  | 0.6663  | 0.6404  | -0.1558 | -0.2041 | -0.1387 | -0.1656 | 0.8264  | 1.0000  |

由表 1 可知,水压、时效分量之间因影响因子选取公式的原因存在高度线性正相关,各水压分量之间相关性达 0.97 以上,各时效分量之间相关性达 0.8264,存在高度相关性影响。3 项水压分量与 2 项时效分量的相关性也达 0.64 以上,属于中度相关范畴。若直接作为裂缝预测模型的输入变量,影响因子间多重共线性将会影响预测模型的预测结果。

进一步基于 PCA 方法计算各成分特征值、贡献率、累计贡献率,如表 2 所示。表 2 中前 4 项主成分特征值大于 1,累计贡献率达 82.027%,可涵盖原数据集的大部分信息,因此选用前 4 项主成分作为模型的输入变量。

2.3.3 主成分挖掘结果

提取 4 项主成分得分系数矩阵,结果见表 3,并将计算结果作为模型的输入变量。

表 2 各成分特征值、贡献率、累计贡献率

| Table 2 Characteristic value, contribution rates, and cumulative contribution rate of each component |                     |                     |         |
|------------------------------------------------------------------------------------------------------|---------------------|---------------------|---------|
| 主成分                                                                                                  | 特征值                 | 初始特征值<br>方差贡献率%     | 累计贡献率%  |
| 1                                                                                                    | 3.293               | 36.586              | 36.586  |
| 2                                                                                                    | 2.032               | 22.574              | 59.160  |
| 3                                                                                                    | 1.076               | 11.735              | 70.895  |
| 4                                                                                                    | 1.002               | 11.132              | 82.027  |
| 5                                                                                                    | 0.883               | 9.815               | 91.843  |
| 6                                                                                                    | 0.709               | 7.881               | 99.723  |
| 7                                                                                                    | 0.022               | 0.247               | 99.971  |
| 8                                                                                                    | 0.003               | 0.029               | 100.000 |
| 9                                                                                                    | $1.31\times10^{-7}$ | $1.46\times10^{-6}$ | 100.000 |

表 3 各主成分得分系数矩阵

| Table 3 Influencing factor score coefficient matrix |        |        |        |        |
|-----------------------------------------------------|--------|--------|--------|--------|
| 影响因子                                                | 主成分 1  | 主成分 2  | 主成分 3  | 主成分 4  |
| 水压 1                                                | 0.309  | -0.001 | -0.017 | -0.029 |
| 水压 2                                                | 0.319  | -0.012 | -0.007 | 0      |
| 水压 3                                                | 0.321  | -0.021 | 0.004  | 0.023  |
| 温度 1                                                | 0.177  | -0.202 | 0.253  | 0.077  |
| 温度 2                                                | 0.001  | 0.095  | 0.967  | -0.071 |
| 温度 3                                                | 0.006  | 0.037  | -0.067 | 1.000  |
| 温度 4                                                | 0      | 0.011  | -0.027 | -0.034 |
| 时效 1                                                | -0.006 | 0.478  | 0.114  | 0.043  |
| 时效 2                                                | 0      | 0.473  | 0.074  | 0.035  |

2.4 模型参数选择

单一模型的参数采用人工手动选取,通过不断调整参数,并结合拟合效果判别模型精度。其主要思路为:每次设定不同参数后运行代码执行程序,经反复试验观察拟合效果的变化,最终从大量试验中筛选出对应拟合效果最优的模型参数。所有测试均在 Python3.8 中运行,试验计算机配置为 16 G 内存,Windows10 运行系统,搭载 2.50 GHz 英特尔 i5-12400F 的 CPU。单一模型参数调整范围及参数取值见表 4。

LightGBM 模型参数众多,除 Boosting 类算法中决策树数量、决策树最大深度、学习率等 3 个主要参数,还包括决策树类型、叶子数量、迭代数据比例和最小子节点权重和。LightGBM 模型参数采用默认值,AVOA-LightGBM 模型参数采用 AOVA 对 LightGBM 模型参数优化后的取值。LightGBM 模型参数取值范围如表 5 所示(表中 DART 为一种防止过拟合的高级正则化方法(dropouts meet multiple additive regression trees),GBTD 为梯度提升决策树(gradient boosting decision tree))。

表 4 单一模型参数调整范围及取值

Table 4 Parameter adjustment range and parameter value of single model

| 预测模型    | 参数        | 参数范围      | 参数取值  |
|---------|-----------|-----------|-------|
| SVR     | 核参数       | 特征尺度或自动或无 | 特征尺度  |
|         | 不敏感损失误差   | 0.001~0.1 | 0.05  |
|         | 惩罚因子      | 0.1~150   | 3.00  |
| RF      | 决策树数量     | 1~1 000   | 80.00 |
|         | 决策树最大生长深度 | ≥1        | 5.00  |
|         | 决策树数量     | 1~1 000   | 65.00 |
| XGBoost | 决策树最大生长深度 | ≥1        | 5.00  |
|         | 学习率       | 0.01~0.5  | 0.10  |

表 5 LightGBM 模型参数调整范围参数取值

Table 5 Parameter adjustment range and parameter value of LightGBM model

| 参数名       | 调参范围                    | 默认值   | 优化后取值 |
|-----------|-------------------------|-------|-------|
| 决策树类型     | RF 或 DART 或 GOSS 或 GBDT | GBDT  | GBDT  |
| 叶子数量      | ≥1                      | 31.0  | 21    |
| 决策树最大生长深度 | ≥1                      | -1.0  | 275   |
| 学习率       | 0.01~0.5                | 0.1   | 0.085 |
| 决策树数量     | 1~1 000                 | 100.0 | 300   |
| 迭代数据比例    | 0~1                     | 0.5   | 0.8   |
| 最小子节点权重和  | ≥0                      | 1.0   | 0.003 |

2.5 预测模型对比

采用 PAC 方法对各模型影响因子主成分进行挖掘,挖掘前后各模型测试集的预测值对比如图 4 所示。

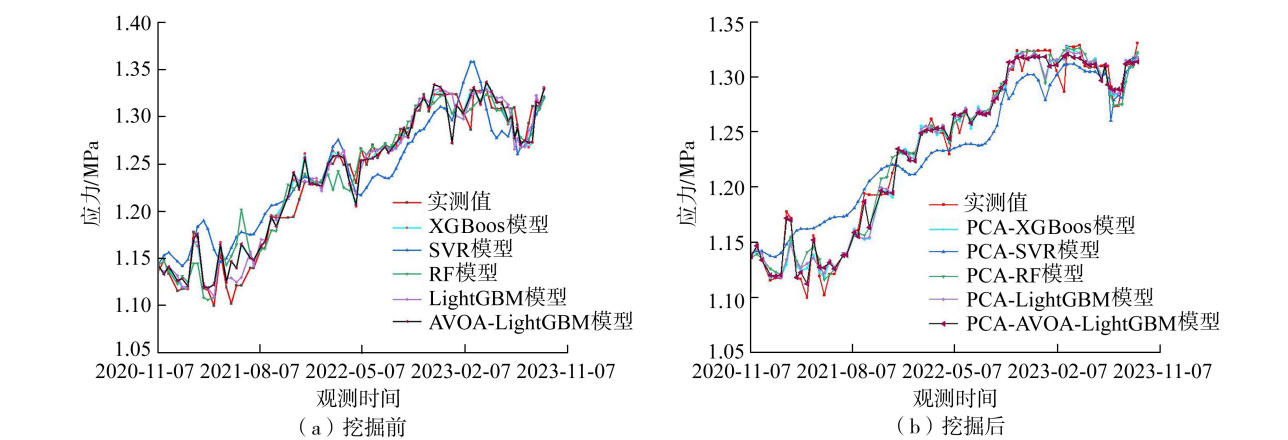


图 4 影响因子主成分挖掘前后各模型测试集对比

Fig.4 Comparison of test sets of each prediction model before and after principal component mining of influencing factors

根据模型测试集的预测值、实际值和运算时间分别计算各模型测试集对应的指标,统计结果如表 6 所示。

由表 6 分析可知,PCA-RF 模型相对于 RF 模型,MAE、MAPE、RMSE、运算时间分别降低了 22%、14%、45%、27%,PCA-SVR 模型相对于 SVR 模型分别降低了 30%、18%、18%、31%,PCA-XGBoost 模型相对于 XGBoost 模型分别降低了 47%、36%、48%、21%,PCA-LightGBM 模型相对于 LightGBM 模型分别降低了 26%、14%、39%、34%,PCA-AVOA-LightGBM 模型相对于 AVOA-LightGBM 模型分别降低了 44%、60%、62%、34%,可见采用 PCA 方法挖掘影响因子主成分后降低了影响因子间多重共线性和模型的计算量,模型预测精度和计算效率均有较大提升。

采用 PCA 方法降维影响因子后,PCA-AVOA-LightGBM 模型预测精度优于 PCA-XGBoost 模型、PCA-LightGBM 模型优于 PCA-RF 模型优于 PCA-SVR 模型,其中 PCA-LightGBM 模型对比 PCA-XGBoost 模型在精

表 6 各模型评价指标对比

| Table 6 Comparison of evaluation indicators for stress prediction models |          |         |          |        |
|--------------------------------------------------------------------------|----------|---------|----------|--------|
| 模型                                                                       | MAE/MPa  | MAPE/%  | RMSE/MPa | 运算时间/s |
| RF                                                                       | 0.125 10 | 2.141 7 | 0.022 5  | 84.5   |
| SVR                                                                      | 0.024 97 | 3.785 0 | 0.027 4  | 62.7   |
| XGBoost                                                                  | 0.015 15 | 2.421 3 | 0.022 3  | 41.5   |
| LightGBM                                                                 | 0.011 59 | 1.859 6 | 0.017 7  | 34.5   |
| AVOA-LightGBM                                                            | 0.011 51 | 1.754 1 | 0.016 9  | 35.1   |
| PCA-RF                                                                   | 0.097 10 | 1.841 8 | 0.012 4  | 61.5   |
| PCA-SVR                                                                  | 0.017 57 | 3.105 9 | 0.022 5  | 43.4   |
| PCA-XGBoost                                                              | 0.008 01 | 1.549 1 | 0.011 5  | 32.7   |
| PCA-LightGBM                                                             | 0.008 52 | 1.606 6 | 0.010 8  | 22.5   |
| PCA-AVOA-LightGBM                                                        | 0.006 42 | 0.715 5 | 0.006 3  | 23.1   |

度评价指标上相差极小,预测精准度基本一致,但其计算效率明显高于 PCA-XGBoost 模型,说明 PCA-LightGBM 模型和其优化后的预测模型在综合性能上具有优势。PCA-AVOA-LightGBM 模型在预测精度上有绝对优势,运算耗时也相对较少,与 PCA-LightGBM 模型相比,MAE 提升了 25%,MAPE 提升了 55%,RMSE 提升了 42%,表明 AVOA 对模型超参数进行了有效的优化,提升了模型预测的精度。

3 结 论

- a. 使用 PCA 方法挖掘提炼各影响因子,组合重构一组具有代表性的综合指标,将 9 项影响因子转化为 4 项主成分因子,降低了影响因子间多重共线性,提高了预测精度和计算效率。
- b. 提出了 AVOA-LightGBM 模型,基于 AVOA 对 LightGBM 模型超参数决策树数量、决策树最大生长深度、学习率、决策树类型、叶子数量、迭代数据比例和最小子节点权重和进行了优化,解决了因 LightGBM 模型超参数设置不当,预测结果无法达到最优的问题。
- c. 依托某混凝土坝应力监测数据,将 PCA 方法应用于 SVR、RF、XGboost、LightGBM 和 AVOA-LightGBM 模型中,结果表明,采用 PCA 方法进行主成分挖掘后各应力预测模型效率和精度均有较大提升,证明了 PCA 方法的可行性;PCA-AVOA-LightGBM 模型相较于 PCA-LightGBM 模型,各项评价指标提升较大,表明 AVOA 有效解决了超参数优化的问题;PCA-AVOA-LightGBM 模型在运算效率和预测精度上表现出了较高的优越性,可在类似混凝土坝的安全监测工程中运用。

参考文献:

[ 1 ] 朱伯芳. 论混凝土坝的使用寿命及实现混凝土坝超长期服役的可能性[J]. 水利学报,2012,43(1):1-9. (ZHU Bofang. On the expected life span of concrete dams and the possibility of endlessly long life of solid concrete dams[J]. Journal of Hydraulic Engineering,2012,43(1):1-9. (in Chinese))

[ 2 ] 周兰庭,邓思源,柳志坤,等. 基于 SSA-LSTM-GF 的混凝土坝变形预测方法[J]. 河海大学学报(自然科学版),2023,51(2):73-80. (ZHOU Lanting, DENG Siyuan, LIU Zhikun, et al. Deformation prediction method of concrete dam based on SSA-LSTM-GF[J]. Journal of Hohai University (Natural Sciences),2023,51(2):73-80. (in Chinese))

[ 3 ] 周秋景,张国新,刘毅. 高拱坝短期变形和应力预测预警方法及工程应用[J]. 水利水电技术,2015,46(1):93-98. (ZHOU Qiujing, ZHANG Guoxin, LIU Yi. Prediction and early-warning method for short-term deformation and stress of high arch dam and its application[J]. Water Resources and Hydropower Engineering,2015,46(1):93-98. (in Chinese))

[ 4 ] LI Bin, YANG Jie, HU Dexiu. Dam monitoring data analysis methods: a literature review[J]. Structural Control Health Monitoring,2019,113:1-4.

[ 5 ] 丛培江,郑东健,仲琳. 大坝应力应变监测的多因素时变分析模型[J]. 武汉大学学报(信息科学版),2008(9):914-917. (CONG Peijiang, ZHENG Dongjian, ZHONG Lin. Time-varying model for dam stress-strain monitoring data based on multiple factors[J]. Geomatics and Information Science of Wuhan University,2008(9):914-917. (in Chinese))

[ 6 ] 田菊飞,苏怀智. 基于随机森林算法的大坝应力预测模型的构建及其应用[J]. 水电能源科学,2018,36(5):54-56. (TIAN Jufei, SU Huaizhi. Development and application of dam stress prediction model based on random forest algorithm[J]. Water Resources and Power,2018,36(5):54-56. (in Chinese))

[ 7 ] 张伟,吴丹,李小奇,等. 基于最小二乘支持向量机的大坝应力预测模型[J]. 水利与建筑工程学报,2011,9(1):26-29. (ZHANG Wei, WU Dan, LI Xiaoqi, et al. Dam's stress prediction model based on LS-SVM method[J]. Journal of Water Resources and Architectural Engineering,2011,9(1):26-29. (in Chinese))

[ 8 ] SU Huaizhi, CHEN Zhixin, WEN Zhiping. Performance improvement method of support vector machine-based model monitoring dam safety[J]. Structural Control and Health Monitoring,2016,23:252-266.

[ 9 ] 崔博,安惠伦,陈文龙,等. 基于 PCA-SSA-XGBoost 算法的拱坝应力预测模型研究[J]. 水力发电,2024,50(5):45-53. (CUI Bo, AN Huilun, CHEN Wenlong, et al. Research on arch dam stress prediction model based on PCA-SSA-XGBoost algorithm[J]. Water Power,2024,50(5):45-53. (in Chinese))

[10] 常留红,李晨玉,曾子彬,等. 基于 WOA-VMD-XGBoost 的混凝土坝变形预测[J]. 水利水运工程学报,2024(3):146-157. (CHANG Lihong, LI Chenyu, ZENG Zibin, et al. Deformation prediction of concrete dams using WOA-VMD-XGBoost methodology[J]. Hydro-Science and Engineering,2024(3):146-157. (in Chinese))

[11] 韩峰,余佳,徐国鑫,等. 基于改进 XGBoost 不平衡围岩超前分类方法的隧洞工程施工仿真研究[J]. 河海大学学报(自然科学版),2023,51(1):150-157. (HAN Feng, YU Jia, XU Guoxin, et al. Study on tunnel construction simulation based on advanced classification of imbalanced surrounding rock using improved XGBoost[J]. Journal of Hohai University (Natural Sciences),2023,51(1):150-157. (in Chinese))

[12] 吴娟,朱跃龙,金松,等. 三种机器学习模型在太湖藻华面积预测中的应用[J]. 河海大学学报(自然科学版),2020,48(6):542-551. (WU Juan, ZHU Yuelong, JIN Song, et al. Area prediction of cyanobacterial blooms based on three machine learning methods in Taihu Lake [J]. Journal of Hohai University (Natural Sciences),2020,48(6):542-551. (in Chinese))

[13] MAHSA H, ABBAS M, REZA G. A novel scheme for mapping of MVT-type Pb-Zn prospectivity: LightGBM, a highly efficient gradient Boosting decision tree machine learning algorithm[J]. Natural Resources Research,2023,32(6):2417-2438.

[14] ZHANG Wengang, WU Chongzhi, TANG Libin, et al. Efficient time-variant reliability analysis of Bazimen landslide in the Three Gorges Reservoir Area using XGBoost and LightGBM algorithms[J]. Gondwana Research,2023,123:41-53.

[15] 高治鑫,包腾飞,李涛涛,等. 基于贝叶斯优化 LightGBM 的大坝变形预测模型[J]. 长江科学院院报,2021,38(7):46-50. (GAO Zhixin, BAO Tengfei, LI Yangtao, et al. Dam deformation prediction model based on bayesian optimization and LightGBM [J]. Journal of Yangtze River Scientific Research Institute,2021,38(7):46-50. (in Chinese))

[16] 张孟昕,陈波,刘伟琪,等. SSA-XGBoost 与时空特征选取的大坝变形预测模型[J]. 水力发电学报,2024,43(1):84-98. (ZHANG Mengxin, CHEN Bo, LIU Weiqi, et al. Dam deformation prediction model selected by SSA-XGBoost with temporal and spatial features[J]. Journal of Hydroelectric Engineering,2024,43(1):84-98. (in Chinese))

[17] 苏燕,付家源,林川,等. 基于时间注意力机制的大坝动态变形预测模型[J]. 水力发电学报,2022,41(7):72-84. (SU Yan, FU Jiayuan, LIN Chuan, et al. Dam dynamic deformation prediction model based on time-attention mechanism[J]. Journal of Hydroelectric Engineering,2022,41(7):72-84. (in Chinese))

[18] 肖渝,孙若辰,王倩,等. 不同降水驱动的 SWAT 模型参数不确定性研究[J]. 河海大学学报(自然科学版),2024,52(2):43-52. (XIAO Yu, SUN Ruochen, WANG Qian, et al. Study on parameter uncertainty in SWAT model driven by different precipitation[J]. Journal of Hohai University (Natural Sciences),2024,52(2):43-52. (in Chinese))

[19] 林川,桂星煜,朱律运,等. 基于因子融合的混凝土面板堆石坝变形预测模型[J]. 水力发电学报,2023,42(10):139-152. (LIN Chuan, GUI Xingyu, ZHU Lyuyun, et al. Deformation prediction model of concrete faced rockfill dams based on factor fusion [J]. Journal of Hydroelectric Engineering,2023,42(10):139-152. (in Chinese))

[20] 刘娣,孙佳倩,余钟波. 基于机器学习模型的多层土壤湿度反演[J]. 河海大学学报(自然科学版),2024,52(3):7-14. (LIU Di, SUN Jiaqian, YU Zhongbo. Multi-layer soil moisture inversion based on machine learning models[J]. Journal of Hohai University (Natural Sciences),2024,52(3):7-14. (in Chinese))

[21] 王仁超,马钰明. 基于 PCA 和 CS-KELM 的重力坝变形预测模型[J]. 水利水运工程学报,2021(4):10-18. (WANG Renchao, MA Yuming. Prediction model of gravity dam deformation based on PCA and CS-KELM [J]. Hydro-Science and Engineering,2021(4):10-18. (in Chinese))

[22] 孙全明,曲志坚,任崇广. 基于粒子群优化和 LightGBM 的情景感知多式联运推荐[J]. 电子学报,2021,49(5):894-903. (SUN Quanming, QU Zhijian, REN Chongguang. Context-aware multi-modal transportation recommendation based on particle swarm optimization and LightGBM [J]. Acta Electronica Sinica,2021,49(5):894-903. (in Chinese))

[23] 赵峰,李妞妞. 基于 VAE-GWO-LightGBM 的信用卡欺诈检测方法[J]. 东北师大学报(自然科学版),2023,55(4):77-84. (ZHAO Feng, LI Niuniu. Credit card fraud detection method based on VAE-GWO-LightGBM [J]. Journal of Northeast Normal University (Natural Science Edition),2023,55(4):77-84. (in Chinese))

[24] 王洁,李金泽,王子瞳,等.改进鲸鱼优化 LightGBM 的可解释性心脏病风险预测模型[J].北京邮电大学学报,2023,46(6):39-45. (WANG Jie,LI Jinze,WANG Zitong,et al. An interpretable prediction model for heart disease risk based on improved whale optimized LightGBM[J]. Journal of Beijing University of Posts and Telecommunications,2023,46(6):39-45. (in Chinese))

[25] 高永涛,朱强,吴顺川,等.基于 AVOA-XGBoost 模型的岩爆预测研究[J].华中科技大学学报(自然科学版),2023,51(12):151-157. (GAO Yongtao,ZHU Qiang,WU Shunchuan,et al. Rockburst prediction study based on AVOA-XGBoost model[J]. Journal of Huazhong University of Science and Technology(Natural Science Edition),2023,51(12):151-157. (in Chinese))

[26] 牛传乐,李芳,任顺,等.基于非洲秃鹫算法优化卷积神经网络的重叠峰解析方法[J].科学技术与工程,2024,24(16):6592-6599. (NIU Chuanle,LI Fang,REN shun,et al. Overapping peak resolution method fox oplimizing convolutional neural network bxased on the Africa vulture oplimization algorithm[J]. Sciencee Technology and Engineering,2024,24(16):6592-6599. (in Chinese))

[27] 罗兴隆,贺兴时,周洁,等.基于非洲秃鹫优化算法改进的密度峰值聚类[J].山东大学学报(理学版),2024,59(1):46-55. (LUO Xinglong,HE Xingshi,ZHOU Jie,et al. Improved density peak clustering approach based on African vultures optimization algorithm[J]. Journal of Shandong University(Natural Science),2024,59(1):46-55. (in Chinese))

[28] GUOLIN K,MENG Q,THOMAS F,et al. LightGBM;a highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural InformationProcessing Systems. New York:Curran Associates Inc. ,2017:3149-3157.

[29] 莫坤,王娜,李恒吉,等.基于 LightGBM 的网络入侵检测系统[J].信息安全研究,2019,5(2):152-156. (MO Kun,WANG Na,LI Hengji,et al. Network intrusion detection system model based on LightGBM[J]. Journal of Information Security Research,2019,5(2):152-156. (in Chinese))

[30] ABDOLLAHZADEH B,GHAREHCHOPOGH F S,MIRJALL S. Africanvultures optimization algorithm;a new natureinspired metaheuristic algorithm for global optimizationproblems[J]. Computers & Industrial Engineering,2021,158:107408.

[31] 吴中如. 水工建筑物安全监控理论及其应用[M]. 北京:高等教育出版社,2003.

(收稿日期:2024-07-30 编辑:胡新宇)

(上接第 33 页)

[28] 余钟波,李敏娟,刘芸辰,等.基于氡同位素的河水与地下水水力交换研究[J].河海大学学报(自然科学版),2020,48(1):8-13. (YU Zhongbo,LI Minjuan,LIU Yunchen,et al. Study on hydraulic exchange of river water and groundwater based on radon isotope[J]. Journal of Hohai University(Natural Sciences),2020,48(1):8-13. (in Chinese))

[29] 骆欢,衣鹏.基于氡同位素的闸塘渗漏检测研究[J].人民长江,2017,48(24):101-105. (LUO Huan,YI Peng. Study on sluice leakage detection based on Rn isotope[J]. Yangtze River,2017,48(24):101-105. (in Chinese))

[30] 焦鹏程,王弭力,刘成林.新疆罗布泊盐湖卤水的氚同位素特征及其地质意义[J].核技术,2004,27(9):710-715. (JIAO Pengcheng,WANG Mili,LIU Chenglin. Characteristics and origin of tritium in the potassium-rich brine in Lop Nur,Xinjiang[J]. Nuclear Techniques,2004,27(9):710-715. (in Chinese))

[31] 巩同梁,田立德,刘东年,等.羊卓雍湖流域湖水稳定同位素循环过程研究[J].冰川冻土,2007,29(6):914-920. (GONG Tongliang,TIAN Lide,LIU Dongnian,et al. A preliminary study of stable isotope cycle processes in lake water in the Yamzho Lake Basin[J]. Journal of Glaciology and Geocryology,2007,29(6):914-920. (in Chinese))

[32] CRAIG H. Isotopic variations in meteoric waters[J]. Science,1961,133(3465):1702-1703.

[33] 郑淑蕙,侯发高,倪葆龄.我国大气降水的氢氧稳定同位素研究[J].科学通报,1983,28(13):801-806. (ZHENG Shuhui,HOU Fagao,NI Baoling. Hydrogen and oxygen stable isotopes of atmospheric precipitation in China[J]. Chinese Science Bulletin,1983,28(13):801-806. (in Chinese))

[34] 郭鑫,李文宝,杜蕾,等.内蒙古夏季大气降水同位素特征及影响因素[J].中国环境科学,2022,42(3):1088-1096. (GUO Xin,LI Wenbao,DU Lei,et al. Characteristics and influence factors for the hydrogen and oxygen isotopic of precipitation in Inner Mongolia[J]. China Environmental Science,2022,42(3):1088-1096. (in Chinese))

[35] 李文鹏,郝爱兵,郑跃军,等.塔里木盆地区域地下水环境同位素特征及其意义[J].地学前缘,2006,13(1):191-198. (LI Wenpeng,HAO Aibing,ZHENG Yuejun,et al. Regional environmental isotopic features of groundwater and their hydrogeological explanation in the Tarim Basin[J]. Earth Science Frontiers,2006,13(1):191-198. (in Chinese))

[36] 甄志磊,李畅游,李文宝,等.内蒙古达里诺尔湖流域地表水和地下水环境同位素特征及补给关系[J].湖泊科学,2014,26(6):916-922. (ZHEN Zhilei,LI Changyou,LI Wenbao,et al. Characteristics of environmental isotopes of surface water and groundwater and their recharge relationships in Lake Dali Basin[J]. Journal of Lake Sciences,2014,26(6):916-922. (in Chinese))

(收稿日期:2024-11-05 编辑:刘晓艳)