

基于 GAMLSS – GLO 及随机森林算法的 土石坝渗流监测模型

赵明哲¹, 樊牧¹, 沈立锋¹, 赵培双¹, 马俊¹, 唐小松¹, 罗航¹, 宋博旭¹, 聂兵兵²

(1. 华能澜沧江水电股份有限公司糯扎渡水电厂, 云南 普洱 665005; 2. 华能澜沧江水电股份有限公司, 云南 昆明 650214)

摘要: 根据土石坝渗流原型观测,厘清水位、降水等因素的影响,模拟大坝渗流的真实状态,合理评估其渗流监测结果,是土石坝安全监控亟待解决的关键问题。基于此,根据数理统计原理,采用随机森林算法构建无降水条件下渗流量与上下游水位的回归模型;考虑渗流量受前期累积降水的综合影响,引入广义可加模型(GAMLSS – GLO),模拟降水影响下土石坝渗流监测值的波动区间,并将其与渗流-水位回归模型叠加,预测土石坝渗流监测的可靠区间;最后,将该方法应用于糯扎渡心墙堆石坝的渗流监测。结果表明:所提模型方法对渗流的水位、降水响应表现出良好的适用性,显著提高了渗流模拟预测质量。同时求解了渗流量置信区间,有利于土石坝的运行工况判断及安全监控。

关键词: 土石坝渗流量; 回归分析; GAMLSS 模型; 随机森林算法; 区间预测

中图分类号:TV641

文献标识码: A

文章编号: 1672-643X(2023)05-0198-09

Seepage monitoring model of earth – rock dam based on GAMLSS – GLO and random forest algorithm

ZHAO Mingzhe¹, FAN Mu¹, SHEN Lifeng¹, ZHAO Peishuang¹, MA Jun¹,
TANG Xiaosong¹, LUO Hang¹, SONG Boxu¹, NIE Bingbing²

(1. Nuozhadu Power Plant of Huaneng Lancang River Hydropower Inc., Puer 665005, China;

2. Huaneng Lancang River Hydropower Inc., Kunming 650214, China)

Abstract: The key problems of the seepage flow monitoring of earth – rock dams are how to clarify the influence of water level, rainfall and other factors, how to simulate the real seepage flow and evaluate the results of the seepage monitoring based on the prototype observation. In view of this, a regression model of seepage flow and upstream and downstream water levels is established under the condition of no rainfall using the random forest algorithm and the principles of mathematical statistics. Regarding to the comprehensive influence of the seepage flow on the accumulated rainfall in the earlier period, the Generalized Additive Models for Location, Scale and Shape of Generalized Logistic Distribution (GAMLSS – GLO) is introduced to simulate the fluctuation intervals of the monitoring values of seepage flow under the influence of rainfall. Then, GAMLSS – GLO is superimposed with the seepage flow – water level regression model to predict the reliable intervals of the seepage flow. Finally, the method is applied to the seepage monitoring of the core wall earth – rock dam in Nuozhadu Reservoir. The results show that the proposed method shows good applicability to the water level and rainfall response of seepage flow, it can significantly improve the quality of seepage simulation prediction, and solve the confidence intervals of seepage flow, which is conducive to judging the operation condition and safety monitoring of earth – rock dams.

Key words: seepage flow of earth – rock dam; regression analysis; GAMLSS model; random forest algorithm; interval prediction

收稿日期:2022-11-29; 修回日期:2023-03-07

基金项目:“十四五”国家重点研发计划项目(2022YFC3005401); 云南省重点研发计划项目(202203AA080009)

作者简介:赵明哲(1992—),男,云南昆明人,博士,工程师,研究方向为水文及水资源。

通讯作者:樊牧(1985—),女,云南曲靖人,硕士,工程师,研究方向为大坝安全监测。

1 研究背景

土石坝在实际运行过程中会经常面临库水位频繁变化以及台风、暴雨等极端天气情况,甚至遭遇多种天气、地质事件叠加的复杂工况,影响土石坝渗流的分布规律及坝坡稳定,众多的影响因素使得土石坝渗流过程具有复杂性、随机性等特点,对土石坝安全运行提出了严峻挑战^[1-4]。由于外界扰动及测量误差的影响,实际工程中的渗流监测数据并不是渗流量、渗透压力、水位等指标的真实值^[5]。因此,为保障土石坝的运行安全,通常构建渗流预测模型,分析实际渗流监测过程的变化规律,建立土石坝渗流监测模型,挖掘渗流序列的随机性、趋势性、周期性特征,实现渗流监测资料的分析、评估与修正^[6]。

目前,大量学者基于回归分析原理建立了考虑上下游水位和降水量等要素的土石坝渗流模型,且已取得丰富的研究成果。基于回归分析的土石坝渗流监测分析模型与方法主要包括多元线性回归模型^[7]、多元非线性回归模型^[2]、时间序列预测模型^[8]、机器学习算法^[9](支持向量机、相关向量机、神经网络)等。研究表明,人工智能和机器学习技术在处理土石坝渗流预测模型等非线性复杂问题中具有明显优势^[10]。Beiranvand 等^[11]对基于人工智能、机器学习技术的独立模型与混合模型预测土石坝渗流及孔隙水压力的相关研究进行了综述,分析了不同大坝渗流建模方法的使用条件和有效性。Li 等^[12]选取 27 个影响土石坝渗流预测模型精度的要素,建立了基于随机森林算法的大坝安全监测模型,并量化了各要素对模型的贡献率。He 等^[13]基于贝叶斯网络(Bayesian network)构建了土石坝安全智能评价框架,评估了土石坝渗流破坏的极端风险。张鹏等^[6]根据土石坝渗流影响要素的滞后效应,探究了上下游水位及降水量共同作用下大坝渗流的变化规律。此外,部分学者还针对土石坝渗流监测的误差问题,采用误差修正模型^[14]以及卡尔曼滤波算法^[15]等方法,对渗流监测结果进行了处理。Rehamnia 等^[16]将卡尔曼滤波器与前馈人工神经网络模型相结合,提出了一种大坝逐日渗流量预测模型。

虽然土石坝渗流监测与统计分析研究已取得一定成果,但现有方法主要将渗流监测值的影响要素作为整体进行考虑,构建多要素影响的渗流预测模型。此类方法难以准确刻画各要素对效应量(渗流监测值)的作用特征及贡献量,也不利于精细化描述各要素对渗流监测值的影响过程。此外,现有研

究主要以确定性方法为主,难以考虑渗流监测结果的随机性和波动性。

因此,本研究针对土石坝渗流监测分析在预测模型、响应解析等方面的不足,提出基于 GAMLSS - GLO 及随机森林算法的土石坝渗流监测模型,以糯扎渡水电站心墙堆石坝为实际工程案例开展相关研究。首先,筛选前 7 d 无降水条件下量水堰监测的大坝渗流量,构成无降水影响的大坝渗流序列,采用随机森林算法,构建无降水条件下渗流量与上下游水位的回归模型;考虑渗流量受前期累积降水的综合影响,提出大坝渗流监测的综合降水因子;在此基础上,引入广义可加模型(generalized additive models for location, scale and shape of generalized logistic distribution, GAMLSS - GLO),模拟降水影响下土石坝渗流监测值的波动特性;将水位影响下渗流回归过程与降水影响下渗流的波动区间相叠加,搭建土石坝渗流监测的区间预测模型。通过上述研究以期提高渗流模拟预测的质量,为土石坝渗流安全监控、合理评价及预警预报提供依据。

2 土石坝渗流分析模型与方法

2.1 基于随机森林算法的水位-渗流量回归模型

近年来,机器学习算法发展迅速,此类方法通过样本训练确定模型的结构和参数,拟合变量间的相依关系,已广泛应用于各个领域,在大坝渗流预测上亦得到了成功应用^[11]。本研究采用参数设置简单、易实现、精度较高的随机森林算法(random forest algorithm),建立土石坝渗流量与上下游水位的相关关系模型,对渗流的水位影响分量进行回归及还原计算,进而对渗流监测序列进行分解。

基于随机森林算法构建水位-渗流量的回归模型时,采用 Bootstrap 方法从总体中抽取样本构成训练集,计算集合特征,再通过分类回归树(classification and regression tree, CART)算法生成回归模型决策树^[17],并且以最小均方差作为决策树节点分裂的依据,如公式(1)所示。

$$m = \min_{A,s} \left[\min_{D_1(A,s)} \sum (y_i - c_1)^2 + \min_{D_2(A,s)} \sum (y_i - c_2)^2 \right] \quad (1)$$

式中: m 为决策树节点分裂的最小均方差; s 为节点包含的所有训练集; A 为节点抽取的特征集; y_i 为子集的输出值; D_1 和 D_2 为节点全集 s 根据特征 A 分裂的子集; c_1 和 c_2 为两子集的均值。对于任意节点,根据公式(1) 计算使得均方差达到最小的分裂方式

及其生长信息。

训练过程中,重复计算各节点的分裂生长信息,当计算达到决策树深度、节点最小样本数、最小均方差阈值等指标的终止条件时,得到 Bootstrap 抽样训练集的回归决策树。对抽样得到的所有训练集进行训练,便形成随机森林回归模型。

2.2 大坝渗流综合降水因子分析方法

大坝渗流量的监测值是一定时效期内库水位和降水等因素动态变化作用下的瞬时效应量^[2],考虑到糯扎渡大坝的规模等特征,结合量水堰实测的渗流量序列与水位、降水序列的定性分析结果,确定其大坝渗流监测值会受到前 7 d 累积降水的影响,并且其影响作用呈现出复杂的非线性关系。因此,本研究采用综合降水因子反映前期累积降水对渗流监测值的扰动作用。影响渗流监测值的综合降水因子 ($Precip_{Factor}$) 计算公式如下:

$$Precip_{Factor} = (P_{day1} \cdot \theta_{day1} + P_{day2} \cdot \theta_{day2} + \cdots + P_{day7} \cdot \theta_{day7}) \theta_{fix} \quad (2)$$

式中: $P_{day1}, P_{day2}, \cdots, P_{day7}$ 分别为渗流监测时间前 1 d 至前 7 d 的降水量, mm; $\theta_{day1}, \theta_{day2}, \cdots, \theta_{day7}$ 分别为逐日降水影响渗流量的效用系数; θ_{fix} 为修正系数。

为量化逐日降水量对渗流监测结果的影响,以综合降水因子与渗流监测降水分量的相关系数最大为目标,采用粒子群优化算法 (particle swarm optimization, PSO), 优化确定逐日降水影响渗流的效用系数,其目标函数如下:

$$Obj. \quad r_{max} = \left(\frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \right) \quad (3)$$

式中: r_{max} 为综合降水因子与渗流监测降水分量序列皮尔逊相关系数 (Pearson correlation coefficient) 的最大值; X_i 为综合降水因子序列样本值, $\bar{X}$ 为其均值; Y_i 为渗流监测降水分量序列样本值, $\bar{Y}$ 为其均值; N 为序列样本容量。

2.3 大坝渗流波动降水影响区间的 GAMLSS - GLO 模型

降水导致渗流监测值的波动特性与降水的大小具有较强的相关性。从统计学角度来看,如果将土石坝渗流监测值的波动作为随机变量,则其概率分布将随综合降水因子的变化而发生变化,即概率分布各阶矩表现出随综合降水因子演化的特征。

传统的概率统计方法 (如固定的概率分布模型) 难以反映渗流监测值的波动随综合降水因子的变化情况,而分位数回归方法则难以全面反映不同降水状态下渗流监测值波动的分布特征。为了更为准确地模拟降水条件下土石坝渗流量监测结果的波动特性,本研究参照非一致水文频率分析中应用的时变矩模型 (time-varying moments model) 和广义可加模型 (generalized additive models for location, scale and shape, GAMLSS) 对趋势性演化的水文序列进行了分析^[18],并提出了渗流降水影响分量的 GAMLSS 模型。

2.3.1 大坝渗流监测波动量的概率分布 考虑到统计分析中常用的正态类 (正态分布和对数正态分布)、 Γ 类 (指数分布、Gamma 分布、P - III 分布、对数 P - III 分布)、极值类 (广义极值分布、Gumbel 分布、Weibull 分布) 等分布类型通常情况下为右偏态 (正偏) 形式,其左偏态 (负偏) 形式下参数取值范围与分布函数、密度函数的形式将发生较大变化,不适合用于构建降水影响下土石坝渗流监测波动量的概率分布。

广义逻辑分布 (generalized logistic distribution) 具有灵活可靠、适应性强等优势^[19],可用于描述土石坝渗流监测波动随机变量的分布情况。为了满足渗流监测波动样本序列的不同偏态特性,本研究采用 Hosking 于 1997 年提出的考虑偏斜系数的 V 型 GLO 分布,构建大坝渗流监测波动随机变量的概率分布。GLO - V 分布的累积分布函数、概率密度函数以及累积分布逆函数如下:

$$F(x) = \frac{1}{(1 + e^{-y})} \quad (4)$$

$$f(x) = \frac{\alpha^{-1} e^{-(1-k)y}}{(1 + e^{-y})^2} \quad (5)$$

$$y = \begin{cases} -k^{-1} \ln \{1 - k(x - \mu)/\alpha\} & (k \neq 0) \\ (x - \mu)/\alpha & (k = 0) \end{cases} \quad (6)$$

$$x(F) = \begin{cases} \mu + \alpha [1 - ((1 - F)/F)^k]/k & (k \neq 0) \\ \mu + \alpha \ln((1 - F)/F) & (k = 0) \end{cases} \quad (7)$$

式中: $F(x)$ 和 $f(x)$ 分别为 GLO - V 分布的累积分布函数和概率密度函数; $x(F)$ 为 GLO - V 分布累积分布的逆函数; μ 为位置参数, $\mu \in (-\infty, \infty)$; α 为尺度参数, $\alpha \in [0, \infty)$; k 为偏斜参数 (系数), $k \in (-\infty, \infty)$ 。

为了反映 GLO - V 分布对于左偏态和右偏态样本序列的适应性,给出了当 $\mu = 0, \alpha = 2, k =$

$\{-0.45, -0.30, -0.15, 0, 0.15, 0.30, 0.45\}$ 时, GLO 分布的正-负偏态示意图,如图 1 所示。

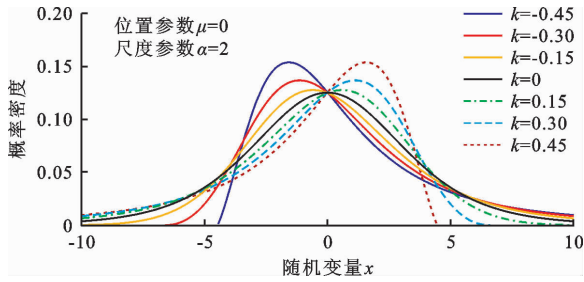


图 1 不同偏斜参数下 GLO 分布正-负偏态示意图

从图 1 可以看出, GLO - V 概率分布对不同样本的适应性较强,特别是对于其概率密度函数的偏态特征,当其偏斜参数 $k < 0$ 时,分布呈左偏态;当偏斜参数 $k > 0$ 时,分布呈右偏态;当偏斜参数 $k = 0$ 时,分布呈现正态化特征。因此,选取 GLO - V 线型构建大坝渗流波动分布具有合理性。

2.3.2 土石坝渗流监测波动的 GAMLSS 模型 综合降水因子作为与渗流监测值物理相关的随机变量,可扩展渗流监测波动概率分布估计的信息量。非一致水文频率分析的相关研究中,为了在估计水文变量概率分布时考虑相关变量的信息,将随机变量概率分布的一阶矩、二阶矩或分布参数解释为随降水、时间、气温、径流等协变量演化的变量,拟合参数变化情况下随机变量的概率分布^[18]。因此,为了充分利用降水信息,尽可能准确地描述大坝渗流监测值的波动特征,以综合降水因子作为大坝渗流监测波动 GLO - V 概率分布参数的协变量,构建大坝渗流波动区间的 GAMLSS - GLO 模型。

设渗流降水影响分量的波动变量服从 3 参数的 GLO - V 概率分布,则大坝渗流波动区间的 GAMLSS - GLO 模型为 $F_{Y, X_{\text{PrecipFactor}}}(Y | \mu_X, \alpha_X, k_X)$, 其中 μ_X, α_X, k_X 为渗流波动量序列的分布参数。通过协变量综合降水因子 $X_{\text{PrecipFactor}}$ 和分布参数演化方程系数 θ_{evolu} , 可以将 GAMLSS 模型的概率分布参数、协变量和演化方程系数之间的关系利用函数关系式 (7) ~ (9) 来表示^[20]:

$$\mu_X = g_\mu(\theta_\mu | X = X_{\text{PrecipFactor}}) \quad (8)$$

$$\alpha_X = g_\alpha(\theta_\alpha | X = X_{\text{PrecipFactor}}) \quad (9)$$

$$k_X = g_k(\theta_k | X = X_{\text{PrecipFactor}}) \quad (10)$$

式中: $g_\mu(\cdot), g_\alpha(\cdot)$ 和 $g_k(\cdot)$ 为 GAMLSS - GLO 模型分布参数的演化方程,常见的演化方程包含线性函数、多项式、幂指数函数、对数函数等; $\theta_\mu, \theta_\alpha, \theta_k$ 为模型各分布参数演化方程的系数, $(\theta_\mu, \theta_\alpha, \theta_k) \subseteq$

θ_{evolu} ; μ_X, α_X 和 k_X 为协变量预测值 $X_{\text{PrecipFactor}} = X$ 时的分布参数。

2.3.3 GAMLSS 模型参数估计与拟合优度检验方法 GAMLSS 模型通常采用极大似然法估计参数。由于 GAMLSS 模型概率分布的样本序列还包含对应的协变量,因此可以通过计算样本序列及其协变量组合的条件对数似然函数值,使全序列条件对数似然函数之和最大的分布参数演化方程系数 θ_{evolu} 成为 GAMLSS 模型的最优参数^[16],其具体计算公式如下:

$$l_{\text{Condition}}(\theta_{\text{evolu}}) = \sum \ln(f_Y(Y | X_{\text{PrecipFactor}}; \theta_{\text{evolu}})) \quad (11)$$

式中: $l_{\text{Condition}}(\theta_{\text{evolu}})$ 表示 GAMLSS 模型深化方程系数为 θ_{evolu} 时的条件对数似然函数之和; $f_Y(Y | X_{\text{PrecipFactor}}; \theta_{\text{evolu}})$ 表示当渗流波动随机变量和综合降水因子协变量分别为 Y 和 $X_{\text{PrecipFactor}}$ 时渗流波动量分布的概率密度函数值。

采用模型残差 Q - Q 图(quantile - quantile plot)、Worm 图,以及 K - S 检验(Kolmogorov - Smirnov)方法对 GAMLSS 模型的拟合优度进行评价。K - S 检验统计量的计算公式如下^[18]:

$$D_{K-S} = \max_{1 \leq t \leq N} |u_t - u_{e,t}| \quad (12)$$

式中: D_{K-S} 为 K - S 检验统计量; t 为样本序列中的样本; N 为序列中的样本容量; $u_t = F(y_t | \theta)$, 为第 t 个样本的理论频率; $u_{e,t} = \text{rank}(y_t) / (1 + N)$, 为第 t 个样本排频的经验频率。当 D_{K-S} 小于置信水平 α 的临界值 D_α 时,所选理论分布通过 K - S 检验。

2.4 计及水位与前期累积降水的土石坝渗流监测区间预测模型

将上述提出的基于随机森林算法的水位-渗流量回归模型与大坝渗流波动的降水影响区间 GAMLSS - GLO 模型相结合,通过渗流水位影响分量与降水影响波动区间的叠加,搭建计及水位与前期累积降水的土石坝渗流监测区间预测模型。其主要步骤如下:

(1) 对渗流量、上下游水位、降水量数据样本序列进行预处理,即审查数据的可靠性,插补或删除缺失值和异常值等,根据日降水数据,将样本总体划分为前 7 d 有降水影响和前 7 d 无降水影响的两个子序列;

(2) 基于随机森林算法,根据前 7 d 无降水影响的水位和渗流子序列,构建上下游水位与大坝渗流量的回归模型;

(3) 针对前 7 d 有降水影响的水位、降水和渗流量子序列,根据上下游水位,采用步骤(2)构建的回归模型计算渗流的水位影响分量,将其与实测渗流序列作差,确定渗流的降水影响分量;

(4) 根据渗流的降水影响分量及其对应的降水量序列,以综合降水因子与渗流监测降水分量的相关性最强为目标,采用粒子群算法求解前期降水影响效用系数,计算综合降水因子;

(5) 根据渗流的降水影响分量和综合降水因子序列,采用 GAMLSS - GLO 模型建立渗流波动量与降水因子的关系,模拟前期累积综合降水影响下大坝渗流监测值的波动区间;

(6) 将步骤(2)和步骤(5)的渗流-水位回归模型与渗流-降水波动区间模拟模型得出的渗流影响分量与降水影响分量叠加,搭建计及水位与前期累积降水的土石坝渗流监测区间预测模型,确定土石坝渗流监测值的可靠区间。

3 实例应用

以糯扎渡水电站心墙堆石坝为例,开展大坝渗流量监测数据的分析预测研究。该坝位于我国云南省普洱市思茅区和澜沧县交界处的澜沧江下游干流,心墙堆石坝最大坝高为 261.5 m,在同类坝型中居亚洲第 1、世界第 3。

糯扎渡大坝渗流监测采用量水堰、渗压计、测压管和水位孔多种方式协同进行。其中,下游围堰处量水堰监测的大坝渗流量为大坝渗流安全监测的首要指标。糯扎渡大坝渗流量采用定时监测(每天 1 次,早上 8:00 自动监测并记录)与手动加密监测(根据需要设定加密监测的时间和密度)相结合的方式监测。本研究采用 2019 年 1 月 1 日—2022 年 10 月 15 日糯扎渡大坝下游围堰处量水堰逐日的大坝渗流监测数据,以及水库上下游水位和日降水量进行实例研究,样本序列如图 2 所示。

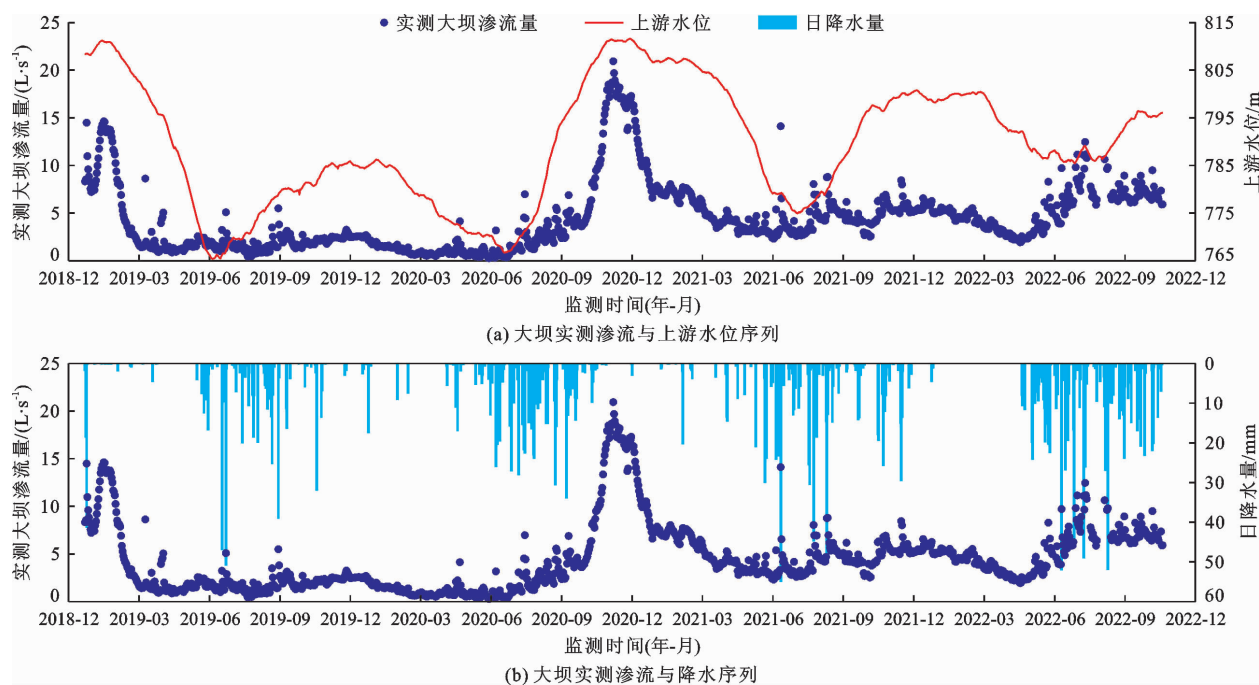


图2 糯扎渡心墙堆石坝渗流实测样本序列

由图2可以看出,大坝渗流量的监测结果与上游水位及降水量具有显著的相关性,符合一般规律。然而,图2所示的渗流监测序列的统计规律在2022年6月前后发生了明显变异(5月底监测的大坝渗流量均值约为4.3 L/s,6月中旬以后监测的大坝渗流量均值超过了7.6 L/s),该时期水库上游水位稳定在780 m附近(属水位影响不敏感段),且刚进入雨季,降水条件类似。经调查电站于2022年5月25日更换过量水堰的流量监测设备,并于当年6月

上旬连续对监测设备基准值进行了调整。因此,初步推断变异点前后测值差异为监测设备更新导致的误差。

为了清楚地反映大坝渗流的真实情况,将样本序列分成两段,其中,2019年1月1日—2022年4月30日的数据用于模型的训练和验证;采用所建立的渗流监测模型对2022年5月1日—2022年10月15日的大坝渗流监测值进行分析与评估,并对异常数据进行修正,以提高监测数据的可靠性。

3.1 大坝渗流量-水位关系回归分析

将上游库水位及下游尾水位作为预测变量,将糯扎渡大坝下游围堰处布置的 DB-WE-11 型量水堰监测的渗流量作为响应变量,选取渗流量监测日期前 7 d 均无降水的数据组成样本序列,并且按 80%

和 20% 的比例从样本序列中随机抽取数据分别作为模型的训练集和验证集,采用随机森林算法,构建无降水条件下糯扎渡大坝渗流量的随机森林回归模型,渗流量模型预测效果及其与水头的相关关系如图 3 所示。

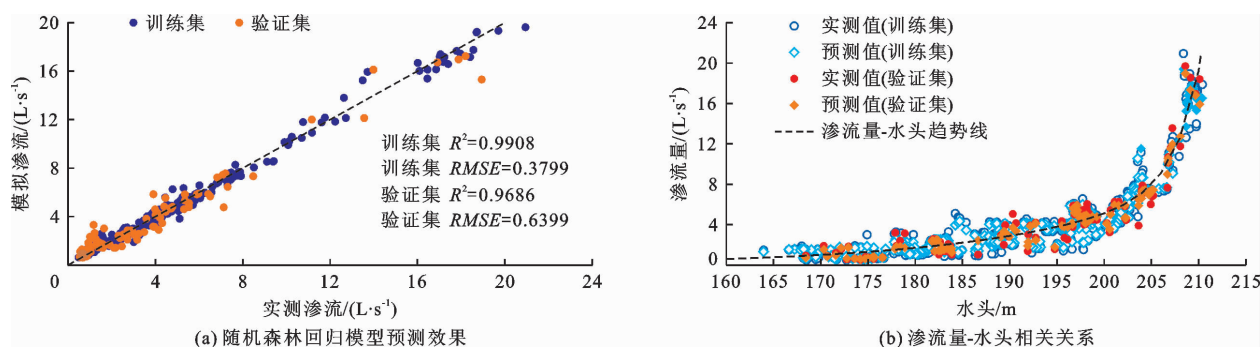


图3 无降水条件下渗流量模型预测效果及其与水头的相关关系

由图 3 可知,随机森林算法构建的渗流量回归模型拟合效果较好,且较为稳定。对比训练集与验证集模型的拟合优度指标可以看出,训练集的 $R^2 = 0.9908$ 、 $RMSE = 0.3799$;验证集的 $R^2 = 0.9686$ 、 $RMSE = 0.6399$,验证集模型的拟合效果仅略低于训练集,表明该模型能够较好地反映无降水条件下大坝渗流量与上下游水位的映射关系。结合图 3 (b) 中大坝渗流量与水头的关系可以看出,不受降水影响的情况下,渗流量与上下游水位的关系非常稳定,与水头大致呈指数关系。当水库的水头从 170 m 增长到 200 m 时,大坝渗流量从 1 L/s 增大到 5 L/s,渗流量增量与水头增量的比值为 0.13:1;而当水库的水头从 200 m 增长到 210 m 时,大坝渗流量从 5 L/s 增大到超过 19 L/s,渗流量增量与水头增量的比值达到 1.4:1,该两个水头段的大坝渗流量对水头的敏感性相差 10 余倍,表明糯扎渡大坝渗

流量对高水头极为敏感,而水库水头较低时大坝渗流量则相对平稳。

3.2 大坝渗流监测响应分解与降水综合影响

为了较为准确地分析大坝渗流量受水位、降水等因素的影响,考虑将各要素影响下的渗流量进行分解,进而分析各要素对渗流监测值的影响效应与程度,实现对土石坝渗流安全重要影响因素的监控。利用构建的水位-渗流量随机森林回归模型,模拟对应上下游水位过程在无降水条件下产生的大坝渗流过程,即渗流监测值的水位影响分量;将渗流水位分量从实测大坝渗流量中扣除,得到大坝渗流监测值的降水影响分量。

通过上述方法,分解糯扎渡大坝下游围堰量水堰监测渗流序列的水位影响分量和降水影响分量,并对各分量的渗流响应关系与效用进行分析,结果如图 4 所示。

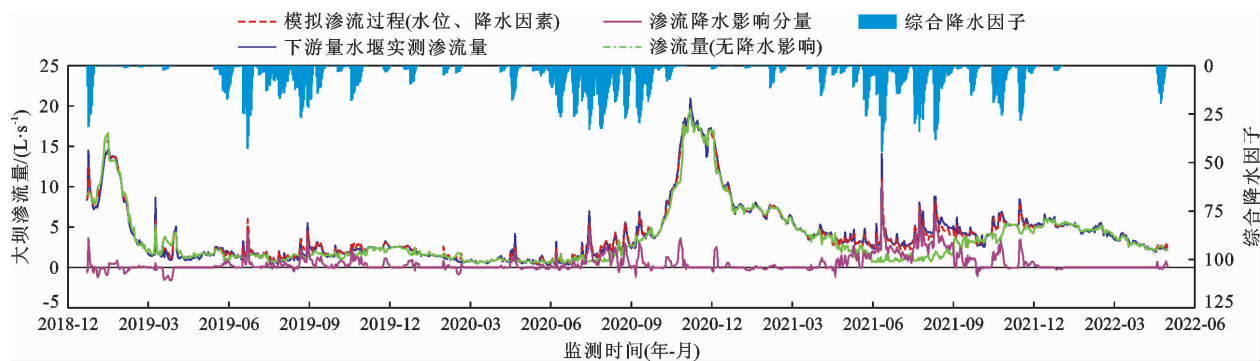


图4 大坝下游量水堰监测渗流量的水位、降水响应关系

图 4 中无降水影响的渗流量为仅受水位影响的大坝渗流量模拟过程,即渗流量-水位随机森林回归

模型的模拟结果,其与下游量水堰实测的大坝渗流过程的差值即为渗流降水影响分量,反映了渗流监

测值受降水影响的情况。对比分析渗流降水影响分量与综合降水因子的关系可以看出,大部分时段的渗流降水影响分量与降水的变化过程一致,两者图形形状大致呈对应关系。表明渗流量监测值对水位、降水因素响应关系的解析结果具有合理性。

在上述分析的基础上,采用多元线性回归方程确定大坝渗流综合降水因子,研究大坝渗流监测值受前期降水的累积影响。根据分解的渗流降水影响分量和日降水量序列,以降水因子与渗流降水分量的相关性为目标,采用 PSO(粒子群)算法优化确定大坝渗流监测值受到降水影响的效用系数 $\theta_{\text{day}} = [1.000, 0.964, 0.772, 0.794, 0.758, 0.402, 0.202]$, 综合降水因子修正系数 $\theta_{\text{fix}} = 0.5198$ 。可以看出,近 2 d 的降水对渗流监测值的影响最为显著;随时间延长,降水对渗流的影响逐渐衰减,监测日期之前第 7 d 的降水对渗流的影响系数仅为 0.202,超过 7 d 的降水则不再影响大坝渗流量的监测结果。

3.3 基于 GAMLSS - GLO 模型的渗流-降水波动模拟

为进一步分析前期降水综合影响下渗流监测值

的波动特征,构建基于 GAMLSS - GLO 的渗流量波动-综合降水因子模拟模型。根据上述得到的渗流降水影响分量与综合降水因子序列,采用 PSO 算法求解最优 GAMLSS - GLO 模型(拟合优度评价指标 $l_{\text{Condition}}(\theta_{\text{evolu}}) = 1019.934$),其分布参数演化方程见方程式(13)~(15);采用 K - S 统计检验、模型残差 Q - Q 图和 Worm 图直观反映 GAMLSS - GLO 模型的拟合效果,如图 5 所示。

$$\mu_X = 6.474 \times 10^{-5} X^3 - 3.794 \times 10^{-4} X^2 + 4.610 \times 10^{-2} X + 3.240 \times 10^{-2} \quad (13)$$

$$\alpha_X = -3.028 \times 10^{-7} X^4 + 2.744 \times 10^{-5} X^3 - 1.226 \times 10^{-3} X^2 - 3.944 \times 10^{-2} X + 3.033 \times 10^{-2} \quad (14)$$

$$k_X = -4.780 \times 10^{-11} X^6 + 1.009 \times 10^{-7} X^5 - 1.260 \times 10^{-5} X^4 + 5.667 \times 10^{-4} X^3 - 1.110 \times 10^{-2} X^2 + 0.1061 X - 0.836 \quad (15)$$

式中: X 为综合降水因子的实际值,根据方程式(13)~(15)可计算该降水条件对大坝渗流监测结果的影响概率。

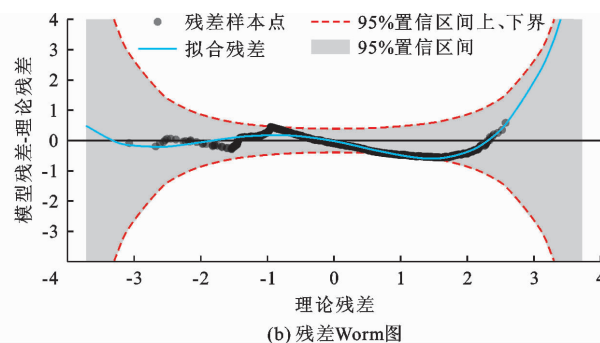
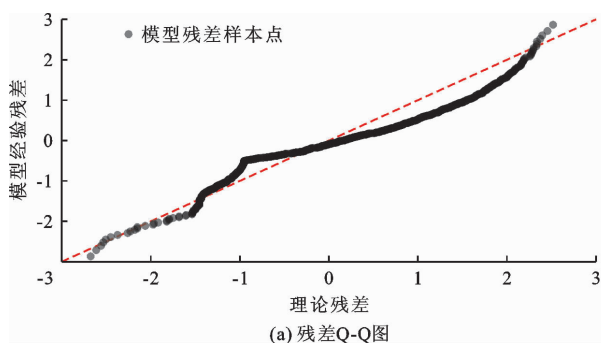


图5 降水影响下大坝下游量水堰监测渗流量波动的 GAMLSS - GLO 模型拟合效果

上述结果表明,GAMLSS - GLO 模型位置参数 μ 、形状参数 α 和偏斜系数 k 演化方程的阶数分别为三阶、四阶和六阶。K - S 检验统计量 $D_{K-S} = 0.0997$ (显著性水平 $\alpha = 0.05$ 时 K - S 检验的临界值 $D_{\alpha} = 0.1025$),通过假设检验,证明 GAMLSS - GLO 模型是合理的,其模拟结果能够反映渗流监测值波动随综合降水因子的变化规律。由图 5 可以看出,模型残差正态 Q - Q 图中残差样本点基本落在 1:1 线附近,模型残差 Worm 图中残差样本点接近于 x 轴($y = 0$) 直线,且均落在 95% 置信区间的灰色区域内呈扁平分布,表明模型的拟合效果良好。为体现 GAMLSS - GLO 模型的优势,将该模型求解的大坝下游量水堰渗流量监测值波动置信区间与分位数回归方法的计算结果进行对照,如图 6 所示。

由图 6 可知,两种方法求解的综合降水因子与量水堰渗流监测值的波动特性规律大致相同,综合降水因子增大也将引起渗流监测值的增大,但不同降水条件下渗流监测值的波动范围无明显变化,总体上表现为渗流监测值相对中位数正偏的波动范围大于负偏,说明渗流监测值的负偏较为集中而正偏相对分散,且相同的偏离程度下负偏的概率更高。不同置信水平下大坝渗流监测值的波动特性方面,随着置信水平的提高,渗流监测值波动区间的宽度大幅增加,当置信水平小于 90% 时,渗流监测值波动区间宽度约为 3 L/s,即相对其中位数的波动范围大致在 $[-1, 2]$ 以内,且该区间基本不随综合降水因子而变化;当置信水平增加到 99% 时,渗流监测值波动区间的宽度约为 8 L/s,且相对其中位数的波动范围随综合降水因

子的变化而变化,具体为正偏先减小后增大,而负偏先增大后减小,其拐点出现在综合降水因子约等于32的点附近。

此外,基于 GAMLSS - GLO 的渗流量波动模型较分位数回归模型更为全面,能够提供任意降水条件下连续的渗流波动量概率分布,有利于土石坝渗流量监测值的区间预测。因此,本研究推荐采用 GAMLSS 模型分析大坝渗流监测值的降水综合影响,并分别给出了降水影响下渗流波动的概率密度分布(probability density function, PDF)和累积概率

分布(cumulative distribution function, CDF),如图7所示,可为后续渗流区间预测及修正研究提供依据。

3.4 大坝渗流监测值的区间预测

根据糯扎渡大坝渗流回归分析、水位与降水要素分解及各要素与渗流响应关系解析的结果,通过水位-渗流量回归模型计算上下游水位影响的渗流分量,利用大坝渗流波动的降水影响区间 GAMLSS - GLO 模型计算给定置信水平下渗流降水分量波动的置信区间,将两分量叠加得到计及水位与前期累积降水的土石坝渗流监测可靠区间,如图8所示。

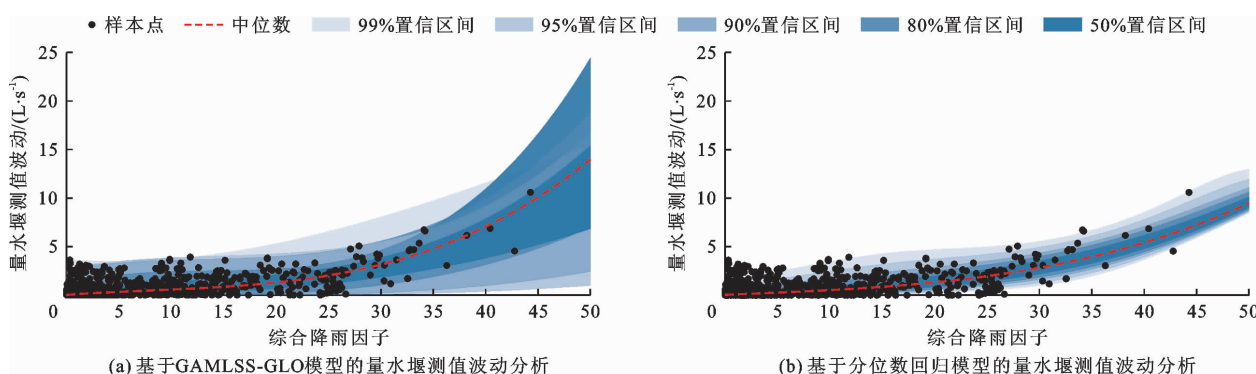


图6 降水影响下 GAMLSS - GLO 模型与分位数回归方法求解的渗流量量水堰测值波动分析

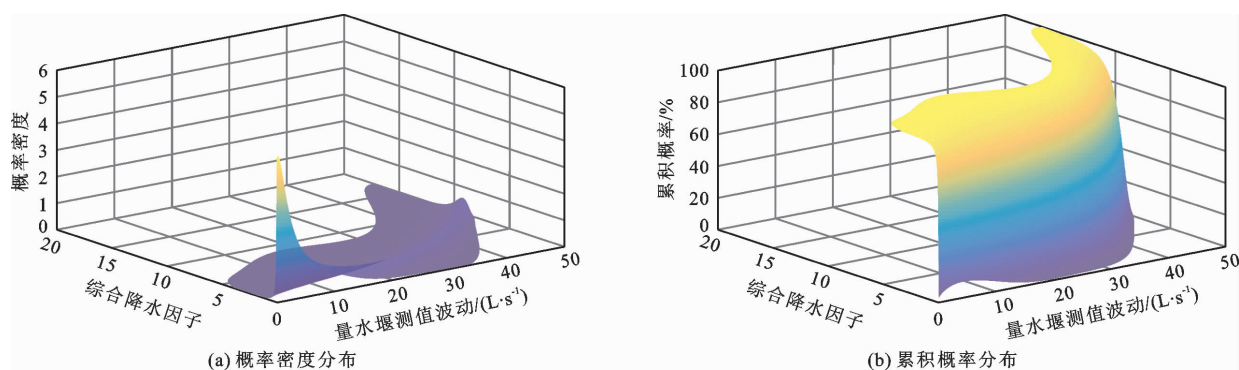


图7 降水影响下基于 GAMLSS - GLO 模型的大坝下游监测渗流量波动概率密度及累积概率分布

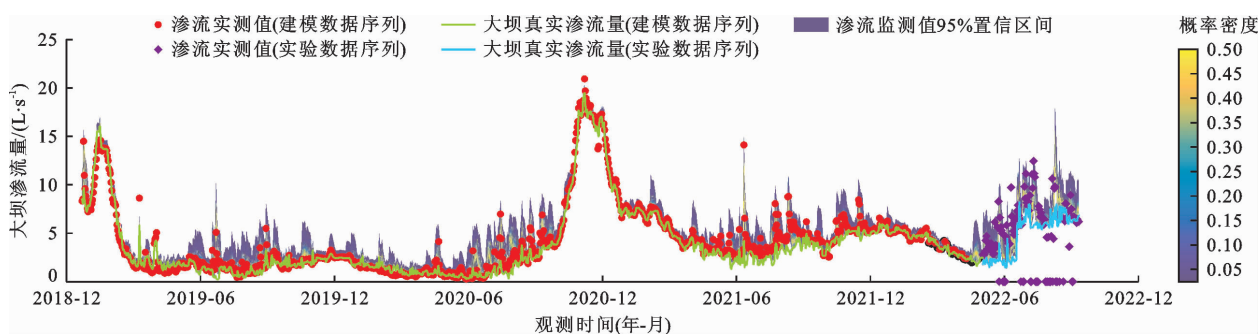


图8 糯扎渡大坝渗流监测值区间预测结果

由图8可以看出,在2022年5月以前(建模数据序列),糯扎渡大坝下游围堰侧量水堰监测的渗流样本点连续,规律性较强,实测点几乎全部落在

95%的置信区间范围内,且大部分样本点紧靠预测区间的中位数,交叉验证了大坝渗流监测值区间预测模型和量水堰渗流监测数据的合理性与可靠性。

对于 2022 年 5 月至 10 月(实验数据序列),5 月份大坝渗流监测值和区间预测结果与建模序列基本一致,但在更换过量水堰的流量监测设备(5 月 25 日)并对新监测设备基准值进行调整(6 月中上旬)之后,为保证模型对变异序列的适应性,根据变异前后样本的均值,将水位-渗流回归模型增加了跳跃变异增加的均值。结果表明,处理后模型对于全部实验序列的模拟效果均良好,变异前后监测值基本都落在置信区间内。此外,该模型还具有评估渗流监测序列的功能,当监测值偏离置信区间时,可向大坝安全监测人员提供预警信息,并对缺测或异常的数据样本进行插补或修正。

4 讨 论

本文针对影响土石坝渗流观测结果的两类关键要素,即水位和降水,基于分解合成的思想,采用随机森林算法构建无降水条件下渗流量与上下游水位的回归模型,同时引入 GAMLSS - GLO 模型量化降水对渗流监测的影响。

在无降水条件下,渗流量回归模型训练期精度达到 99.08%,验证期精度达到 96.86%,上下游水位与渗流量的关系较为稳定且不受降水的影响,这一模拟结果优于已有的大部分同类模型^[21]。例如, Li 等^[12]基于随机森林算法构建的大坝安全监测模型的平均预测精度为 95%;吕鹏^[1]采用自适应差分演化(differential evolution adaptive metropolis, DREAM)算法将渗透系数反演值代入数值模型后得到 97% 的模拟精度。

在降水对渗流监测的影响方面,通过综合降水因子反映了降水对渗流监测的滞后性影响,其结果与其他研究的结论一致^[5]。采用 GAMLSS - GLO 模型计算了降水影响下的渗流量监测的偏差特性,研究结果与吴雄伟等^[14]、冯春燕等^[15]关于土石坝渗流监测偏差修正的研究结论一致。总体来看,本研究提出的基于 GAMLSS - GLO 及随机森林算法的土石坝渗流监测模型可提高渗流模拟精度,降低不确定性的影响。

5 结 论

土石坝的渗流监测是大坝安全监测的重要方面。本研究针对复杂工况下土石坝渗流监测数据失真的问题,依据数理统计原理,提出了基于 GAMLSS - GLO 及随机森林算法的土石坝渗流监测模型,并且以糯扎渡水电站大坝渗流监测为实际工

程案例,开展了量水堰渗流监测分析及修正研究,得到如下结论:

(1)利用分解合成的思想,将渗流监测值的水位和降水影响分解后分别模拟再合成,提出了计及水位与前期累积降水的土石坝渗流监测区间预测模型。该模型能够更为灵活地模拟不同因素对渗流的影响,提高了渗流监测模型的准确性,通过推求大坝的真实渗流量,可为土石坝的运行工况判断及安全监控提供依据。

(2)筛选无降水影响的渗流监测样本,构建了基于随机森林算法的渗流-水位回归模型,以模拟水位变化导致的大坝渗流分量,并提出了降水综合因子量化降水对渗流监测值的持续影响,实现了大坝渗流监测的多要素响应解析。

(3)提出基于 GAMLSS(广义可加模型)的渗流-降水关系分析方法,构建了以综合降水因子为协变量的渗流量监测值波动概率分布模型,求解了大坝渗流监测值随降水因子变化波动的置信区间。该方法能够对前期累积降水影响下的渗流监测值波动区间进行连续模拟,充分考虑了渗流监测的不确定性。

参考文献:

- [1] 吕 鹏. 基于不确定性分析的土石坝渗流性态研究[D]. 天津:天津大学,2019.
- [2] 王 宁. 土石坝渗流影响因素及坝坡稳定性研究[D]. 成都:西华大学,2018.
- [3] 黄侄祥. 降雨入渗及库水位升降对出山店水库土石坝渗流场的影响研究[D]. 武汉:华北水利水电大学,2018.
- [4] XIANG Yan, FU Shuyan, ZHU Kai, et al. Seepage safety monitoring model for an earth rock dam under influence of high-impact typhoons based on particle swarm optimization algorithm[J]. Water Science and Engineering, 2017, 10 (1): 70 - 77.
- [5] 庞 琼,王士军,谷艳昌,等. 基于滞后效应函数的土石坝渗流水位模型应用[J]. 水土保持学报,2016,30(2):225 - 229.
- [6] 张 鹏,李艳玲,李 雪,等. 土石坝渗流监测统计模型的变量分离方法[J]. 水电能源科学,2018,36(1):79 - 81 + 108.
- [7] 钟学梅. 库水位升降条件下土石坝渗流稳定分析及应用研究[D]. 成都:西华大学,2016.
- [8] 盛金保,刘 嘉,李君纯. 土石坝渗流统计数学模型[J]. 水利水运科学研究,1995(4):435 - 443.
- [9] 刘彩花. 汾河水库土石坝渗流特性多模型预警研究[D]. 太原:太原理工大学,2015.
- [10] 刘永涛,郑东健,孙雪莲,等. 基于 EEMD - RVM 的土石坝渗流量时间序列预测模型[J]. 水利水电科技进步, 2021,41(3):89 - 94. (下转第 214 页)

- [8] TAO Hongfei, SHEN Pingping, LI Qiao, et al. Research on head loss of pre-pump micro-pressure filter under clean water conditions[J]. *Water Supply*, 2022, 22(3): 3271–3282.
- [9] ZONG Quanli, ZHENG Tiegang, LIU Huanfang, et al. Development of head loss equations for self-cleaning screen filters in drip irrigation systems using dimensional analysis[J]. *Biosystems Engineering*, 2015, 133: 116–127.
- [10] DE – DEUS F P, MESQUITA M, TESTEZLAF R, et al. Methodology for hydraulic characterisation of the sand filter backwashing processes used in micro irrigation[J]. *Methods X*, 2020, 7: 188–198.
- [11] ZONG Quanli, LIU Zhenji, LIU Huanfang, et al. Backwashing performance of self-cleaning screen filters in drip irrigation systems[J]. *PLoS ONE*, 2019, 14(12): 0226354.
- [12] 喻黎明,曹东亮,李久霖,等. Y 型网式过滤器多目标优化正交试验[J]. *农业机械学报*, 2022, 53(9): 322–333.
- [13] 陶洪飞,滕晓静,马英杰,等. 出水管角度对网式过滤器内部流场影响研究[J]. *节水灌溉*, 2017(8): 21–25+29.
- [14] 谢炎,刘贞姬,李洁,等. 卧式自清洗网式过滤器过滤及排污过程的数值模拟[J]. *浙江大学学报(农业与生命科学版)*, 2022, 48(1): 117–124.
- [15] 杨洪飞,宗全利,刘贞姬,等. 大田滴灌用网式过滤器滤网堵塞成因分析[J]. *节水灌溉*, 2017(2): 94–98.
- [16] 宗全利,杨洪飞,刘贞姬,等. 网式过滤器滤网堵塞成因分析与压降计算[J]. *农业机械学报*, 2017, 48(9): 215–222.
- [17] 朱德兰,王蓉,阮汉铨. Y 型网式过滤器堵塞过程对有机肥浓度的响应研究[J]. *农业机械学报*, 2020, 51(7): 332–337.
- [18] 王睿,王文娥,胡笑涛,等. 微灌用施肥泵施肥比例与肥水比对过滤器堵塞的影响[J]. *农业工程学报*, 2017, 33(23): 117–122.
- [19] 黄玲艳,刘貽雄,唐强,等. 低温过滤器堵塞工况下的压降特性研究[J]. *流体机械*, 2018, 46(7): 9–13.
- [20] 喻黎明,徐洲,杨具瑞,等. CFD–DEM 耦合模拟网式过滤器局部堵塞[J]. *农业工程学报*, 2018, 34(18): 130–137.
- [21] 张凯,喻黎明,刘凯硕,等. 网式过滤器拦截率计算及其影响因素分析[J]. *农业工程学报*, 2021, 37(5): 123–130.
- [22] 陶洪飞,沈萍萍,周洋,等. 微灌用泵前微压过滤器的最佳运行工况研究[J]. *灌溉排水学报*, 2022, 41(6): 72–79.
- [23] 李继霞,姜有忠,黄光迪,等. 网式旋流自清洗泵前过滤器设计与试验[J]. *农机化研究*, 2021, 43(12): 174–180.
- [24] 聂世豪,宗全利,李翠锦,等. 适用恶劣水质泵前无压全自动水力旋喷滚筒网式过滤器: 202120363738. X[P]. 2021-11-09.

(上接第 206 页)

- [11] BEIRANVAND B, RAJAEI T. Application of artificial intelligence-based single and hybrid models in predicting seepage and pore water pressure of dams: a state-of-the-art review[J]. *Advances in Engineering Software*, 2022, 173: 103268.
- [12] LI Xing, WEN Zhiping, SU Huaizhi. An approach using random forest intelligent algorithm to construct a monitoring model for dam safety[J]. *Engineering with Computers*, 2021, 37: 39–56.
- [13] HE Lu, WANG Shijun, GU Yanchang, et al. Seepage behavior assessment of earth-rock dams based on Bayesian network[J]. *International Journal of Distributed Sensor Networks*, 2021, 17(12): 15501477211058672. DOI: 10.1177/15501477211058672
- [14] 吴雄伟,张丽红,唐宏进,等. 基于误差修正模型的土石坝渗流观测资料分析[J]. *人民长江*, 2018, 49(S2): 296–299+320.
- [15] 冯春燕,庞琼,谷艳昌,等. 基于卡尔曼滤波的土石坝渗流安全监控模型[J]. *武汉大学学报(工学版)*, 2019, 52(11): 962–967.
- [16] REHAMNIA I, BENLAOUKLI B, JAMEI M, et al. Simulation of seepage flow through embankment dam by using a novel extended Kalman filter based neural network paradigm: case study of Fontaine Gazelles Dam, Algeria[J]. *Measurement*, 2021, 176: 109219.
- [17] 李焱,贾雅君,李磊,等. 基于随机森林算法的短期电力负荷预测[J]. *电力系统保护与控制*, 2020, 48(21): 117–124.
- [18] 江聪. 考虑水文过程的年径流非一致性分析[D]. 武汉: 武汉大学, 2017.
- [19] 杨冉冉. 三参数 I 型广义 Logistic 分布的参数估计[D]. 北京: 北京工业大学, 2016.
- [20] 熊斌,熊立华. 基于基流退水过程的非一致性枯水频率分析[J]. *水利学报*, 2016, 47(7): 873–883.
- [21] NOURANI V, BEHFAR N, DABROWSKA D, et al. The applications of soft computing methods for seepage modeling: a review[J]. *Water*, 2021, 13(23): 3384.