



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860,CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目: 融合空间关联性的特高拱坝位移概率性预测模型  
作者: 徐丛, 王少伟, 顾冲时, 苏怀智  
DOI: 10.13203/j.whugis20200508  
收稿日期: 2020-09-24  
网络首发日期: 2021-04-19  
引用格式: 徐丛, 王少伟, 顾冲时, 苏怀智. 融合空间关联性的特高拱坝位移概率性预测模型. 武汉大学学报(信息科学版). <https://doi.org/10.13203/j.whugis20200508>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20200508

### 引用格式：

徐丛, 王少伟, 顾冲时, 等. 融合空间关联性的特高拱坝位移概率性预测模型[J]. 武汉大学学报·信息科学版, 2021, DOI: 10.13203/j.whugis20200508 (XU Cong, WANG Shaowei, GU Chongshi, et al. A Probabilistic Prediction Model for Displacement of Super High Arch Dams Considering the Deformation Spatial Association [J]. Geomatics and Information Science of Wuhan University, 2021, DOI: 10.13203/j.whugis20200508)

## 融合空间关联性的特高拱坝位移概率性预测模型

徐丛<sup>1</sup> 王少伟<sup>1,2</sup> 顾冲时<sup>3</sup> 苏怀智<sup>3</sup>

1. 常州大学环境与安全工程学院, 江苏 常州, 213164.

2. 南京水利科学研究院水文水资源与水利工程科学国家重点实验室, 江苏 南京 210029.

3. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098.

**摘要：**基于机器学习语言建模时，传统方法仅以拟合均方误差（mean square error, MSE）最小为单一优化目标，容易引起过拟合问题。对此，基于关联向量机（relevance vector machine, RVM），建立了融合变形空间关联性的双优化目标约束下的概率性预测模型。利用形状相似度指标（shape similarity index, SSI）对拱坝变形的空间关联性进行量化，并将单测点 *MSE* 和区域变形 *SSI* 相融合，共同作为 RVM 模型的训练优化目标，以期实现 *MSE* 尽可能小，而 *SSI* 尽可能大。以锦屏一级拱坝为例，预测均方根误差和最大误差的平均降幅分别为 31.2% 和 24.8%，使用多核函数之后，模型预测性能进一步提升；RVM 模型的预测置信带宽明显小于多元回归模型，平均降幅为 75.1%，由此表明双目标 RVM 模型可有效提升特高拱坝位移预测的精度和稳定性，并降低不确定性。

**关键词：**特高拱坝；位移；空间关联性；概率性预测模型；关联向量机；双目标优化

中图法分类号：P282

文献标志码：A

随着工程技术水平的提高，我国混凝土坝的筑坝高度和建设规模不断加大，截止 2019 年，我国已建、在建和拟建一批 200 m ~ 300 m 级的特高拱坝<sup>[1]</sup>。然而，回顾世界高坝建设和安全管理的发展过程，各国都曾面临诸多重大问题与挑战，缴纳了昂贵的“学费”，甚至付出了惨痛的代价，为确保安全运行，结构生命周期内的安全监控是当前大坝安全管理的主要理念<sup>[2-3]</sup>。

通过对大坝原位监测数据的全面深入分析，应用某种数学方法，建立反映大坝效应集与荷载集关系的监控模型，据此模拟和预测大坝的动态行为和内在作用机理，是实现大坝安全性态监控与健康诊断的最常用手段和方法<sup>[4]</sup>。在众多监测效应量中，变形是混凝土坝真实结构性态的综合反映，河海大学吴中如院士领衔的大坝安全监控团队先后建

收稿日期：2020-09-24

基金项目：国家自然科学基金（51709021）；中国博士后科学基金资助项目（2020M670387）；水文水资源与水利工程科学国家重点实验室“一带一路”水与可持续发展科技基金（2019nkzd03）；中国水利水电科学研究院流域水循环模拟与调控国家重点实验室开放研究基金（IWHR-SKL-KF202002）；江苏省研究生科研与实践创新计划项目（KYCX20\_2560）。

第一作者：徐丛，硕士研究生，主要从事水工结构安全监控研究。18306120727@163.com。

通讯作者：王少伟，博士，副教授，主要从事水工结构安全监控及老化病害研究。shaowei2006nanjing@163.com。

立了统计模型、混合模型和确定性模型等，其中最为经典的是水压-周期性温度-时效 3 因果关系建模因子的 (hydraulic-seasonal-time) HST 模型，其在实际工程变形的变量分离和预测中应用较多，并取得了一定效果<sup>[5-7]</sup>。随着技术的进步和新工程问题的出现，HST 模型也不断被优化和改进。为定量解释部分混凝土坝监测发现的异常变形性状，新的影响因子被添加到 HST 模型中。卢正超等<sup>[8]</sup>在 HST 模型中引入周期性函数表示的冻胀分量，较好地解释了丰满混凝土重力坝坝顶竖向位移的历年“双峰”现象；胡江等<sup>[9]</sup>在 HST 模型中引入了裂缝开度分量，用以解释陈村拱坝坝顶向上游侧的长期趋势性变形；王少伟等<sup>[10]</sup>针对锦屏一级拱坝历年 1880 m 高水位稳定期监测发现的坝体径向位移向下游持续增大的异常现象，建立了考虑变形滞后效应的 (hydraulic-hysteretic-seasonal-time) HHST 模型，定量解释了该异常性状是由粘弹性滞后水压变形和环境温度年周期温降作用共同引起的。

除了优化模型建模因子来兼顾复杂多因素的影响外，如何提高模型对拱坝真实工作性态的解释、预测能力是当前大坝安全监控领域中的另一个热点<sup>[11, 12]</sup>。传统的数学监控模型都是把效应量视为相关影响因素的线性显式函数，并主要采用基于最小二乘法的多元线性回归 (multiple linear regression, MLR) 或逐步回归来确定模型中的系数，而实际上大坝监测效应量与其影响因素之间具有复杂的非线性关系，且包含大量的不确定性。对此，集成人工智能算法的机器学习语言，如人工神经网络、支持向量机、极限学习机、回归树、随机森林、长短时记忆网络等<sup>[13-16]</sup>，已被证明具有强大的非线性数据挖掘能力，因而越来越多地被应用到大坝安全监控领域中。模型参数优化是机器学习建模的关键所在，目前在模型训练时均是以模型输出值和实测值之间的拟合均方差 (mean square error, MSE) 最小化作为参数优化的唯一目标。然而，一味追求训练误差最小并不是总能达到最好的预测效果，某些情况下，训练误差过小反而会导致模型推广能力的下降，即所谓的过拟合<sup>[17]</sup>。此外，部分工程实践结果表明，机器学习模型对因果输入因子的微小波动具有非常强的敏感性，导致模型预测精度和稳定性难以严格保证，而以邻近点变形作为输入因子的机器学习模型则有较好的精度<sup>[18, 19]</sup>。对此，鉴于特高拱坝力学机理和变形性状的空间整体性，可考虑在因果模型的基础上，充分利用坝体多测点位移之间的空间关联信息，以此提高机器学习模型的解释和预测能力。

大坝安全监测的最终目的是将大坝的实际状况与理论模型、历史过程和预测结果进行比较，一旦发现效应量超过监控指标，系统就及时报警，以便立即查找原因，并采取工程或非工程措施。目前国内外普遍使用的监控指标是基于数学模型的置信区间法，然而当前的数学模型主要是点值预测，根据拟合均方差拟定的置信带宽在预测时段内也是定值，没有考虑点值预测不确定性在各时刻的差异性<sup>[20]</sup>。对此，为解决传统机器学习模型的过拟合问题，本文首先对拱坝坝体多测点之间的变形空间关联性进行量化，并将其与拟合均方差相结合，共同作为机器学习模型参数优化时的训练目标，使拟合均方差尽可能小，而空间关联性则尽可能的大；其次，鉴于融合 Bayesian 推理过程的关联向量机 (relevance vector machine, RVM) 能够反映输出结果的概率信息，本文采用构建的双目标函数来优化 RVM 模型中的参数，建立融合变形空间关联性的特高拱坝位移

双目标概率性预测模型, 利用伴随点值预测产生的单点置信带宽来反映复杂影响因素所导致的预测不确定性。

## 1. 融合变形空间关联性的建模双目标优化函数

拱坝在正常运行下的变形性态具有非常强的整体性, 坝体相邻区域具有相似的变形规律, 即就是变形的空间关联性, 且这种关联性随着空间距离的增加而减小。在利用实测数据确定显式型模型中的具体系数或者隐式型机器学习模型的最优参数时, 传统的建模方法仅以实测值和拟合值之间的最小拟合均方差作为单一目标, 只考虑了拱坝变形的单测点因果关系, 没有反映拱坝多测点变形之间的空间关联性, 导致据此建立的机器学习模型容易出现过拟合问题。对此, 可首先对拱坝位移场进行聚类分区, 使同一聚类区域内测点的位移时间序列具有最大的空间关联性, 再通过形状相似度指标对此关联性进行量化, 多测点加权融合得到待研究测点与区域变形的整体相似性后, 再将其与拟合均方差耦合融入到机器学习模型的目标优化函数中, 使拟合均方差尽可能小, 而变形空间关联性则尽可能的大, 以此提高单测点因果模型的预测精度和稳定性。

### 1.1 基于增量距离的拱坝位移场空间聚类分区

位移监测时间序列中的变化量是由库水位和气温的周期性循环所引起, 较绝对位移更能直观反映拱坝的真实结构性态, 因此本文选用基于增量距离的相似性指标进行拱坝位移场的聚类分区。对于任一位移时间序列  $\delta_{it}$  ( $i = 1, 2, \dots, N; t = 1, 2, \dots, T$ ),  $i$  为测点序号,  $t$  为监测时间, 则任意两测点  $i$  和  $j$  之间的增量距离可定义为:

$$d_{ij} = \sum_{t=1}^T \left[ \left( \delta_{it} - \delta_{i(t-1)} \right) - \left( \delta_{jt} - \delta_{j(t-1)} \right) \right]^2 \quad (1)$$

$d_{ij}$  越小表示两测点位移时间序列之间的相似度越大。对此, 使用分层聚类法, 可将具有最小增量距离的测点划分到同一聚类中, 由此产生的聚类偏差平方和,  $F$ , 为:

$$F = \sum_{l=1}^k \sum_{i=1}^{N_l} (X_{il} - \bar{X}_l)^2 \quad (2)$$

式中,  $k$  为聚类的总数,  $l$  为聚类的序号,  $N_l$  为第  $l$  个聚类区中的测点总数,  $X_{il}$  为第  $l$  个聚类中第  $i$  个测点的指标值,  $\bar{X}_l$  为第  $l$  个聚类的指标中心值。

$F$  会随着聚类次数  $p$  的增加而增大, 当  $F_p$  与  $F_{p+1}$  的差值远大于  $F_p$  与  $F_{p-1}$  的差值时, 此时的  $p$  为最佳聚类次数, 聚类的簇数  $k = N - p$ 。

### 1.2 基于时间序列形状相似度的拱坝变形空间关联性量化

拱坝变形的空间关联性主要表现为相同时段内各测点位移具有相似的变化规律, 因此可利用位移监测时间序列的形状信息来量化空间关联性。为研究坝体位移时间序列的形状特征, 需要在保证信息完整性的情况下将总序列划分为一定长度的子序列, 通过滑动窗口法来提取时间序列的波动信息。对于给定的位移时间序列  $\delta_{it}$ , 设定窗口长度和滑动步长分别为  $w$  和  $s$ 。从  $\delta_{i1}$  开始, 由左到右提取滑动窗口中每个子序列的波动趋势, 如下所示:

$$\Delta_{i,t} = \delta_{i,t+w-1} - \delta_{i,t}, \quad 1 \leq t \leq T - w + 1 \quad (3)$$

然后，可定义如下所示形状特征函数：

$$shape = \begin{cases} 1, & \Delta > 0 \\ 0, & \Delta = 0 \\ -1, & \Delta < 0 \end{cases} \quad (4)$$

原位移时间序列  $\delta_{it}$  的形状信息则由包含-1、0、1 的向量  $\mathbf{F}_i$  来描述。对于两个长度相同的时间序列  $\delta_{it}$  和  $\delta_{jt}$ ，两者的形状相似度指数， $SSI_{ij}$ ，可由 Tanimoto 测度表示为<sup>[21]</sup>：

$$SSI_{ij} = \frac{\mathbf{F}_i^T \mathbf{F}_j}{\mathbf{F}_i^T \mathbf{F}_i + \mathbf{F}_j^T \mathbf{F}_j - \mathbf{F}_i^T \mathbf{F}_j} \quad (5)$$

为了进一步得到某测点与同一分区内其它测点的整体相似度  $SSI_i$ ，则需要对单测点的形状相似度进行加权融合。按照单测点相似度越大，融合权重越大的原则，分别计算第  $l$  分区内的第  $i$  个测点和剩余  $N_l - 1$  个测点的形状相似度，则融合权重  $\omega_{ij}$  可计算如下：

$$\omega_{ij} = SSI_{ij} / \sum_{j=1, j \neq i}^{N_l} SSI_{ij} \quad (6)$$

进而， $i$  测点与所在分区的整体形状相似度可表示为：

$$SSI_i = \sum_{j=1, j \neq i}^{N_l} SSI_{ij} \omega_{ij} \quad (7)$$

### 1.3 双目标优化函数的构建

在基于机器学习语言建立特高拱坝位移因果预测模型时，参数优化过程中应使拟合均方差  $MSE$  尽可能的小，而变形空间关联性  $SSI$  则尽可能的大。据此，双目标优化函数的确定过程如下：

(1) 根据初步拟定的模型参数，计算模型拟合值与实测值之间的  $MSE$ ，如下所示：

$$MSE = \frac{1}{T} \sum_{t=1}^T (\delta_{it} - \hat{\delta}_{it})^2 \quad (8)$$

式中， $\delta_{it}$  和  $\hat{\delta}_{it}$  分别为测点  $i$  在  $t$  时刻的位移实测值和拟合值。

(2) 根据式 (7)，计算模型拟合时间序列与区域内测点变形的整体相似度。优化目标函数中， $MSE$  应尽可能小，而空间关联性要尽可能大，因此要对  $SSI$  取倒数处理，即  $S = 1 / SSI$ 。

(3) 两优化目标的量纲不一致，需分别进行归一化处理：

$$MSE' = \frac{MSE - MSE_{\min}}{MSE_{\max} - MSE_{\min}} \quad (9)$$

$$S' = \frac{S - S_{\min}}{S_{\max} - S_{\min}} \quad (10)$$

式中， $MSE_{\max}$ 、 $MSE_{\min}$ 、 $S_{\max}$  和  $S_{\min}$  分别为参数优化范围内  $MSE$  和  $S$  的最大和最小值。

综上，加权融合得到的双目标优化函数如下所示：

$$\begin{aligned} \min L_{\theta}(\gamma, c) &= (1 - \theta) * MSE' + \theta * S' \\ \text{s.t. } 0 &\leq \theta \leq 1 \end{aligned} \quad (11)$$

式中,  $\theta$  为 *SSI* 的权重; 以 Sigmoid 核函数为例,  $\gamma$  和  $c$  为 RVM 模型中需要优化的参数。

## 2. 基于 RVM 的拱坝位移双目标预测模型

### 2.1 RVM 基本原理

RVM 是由 Tipping 提出的一种基于概率统计的机器学习技术, 与支持向量机具有相似的结构形式, 但 RVM 需要的关联向量数目较少, 模型结构更为简化, 且 RVM 融合了 Bayesian 推理, 可据此进行点值的概率性估计。对于训练集  $\{\mathbf{x}_n, t_n\}_{n=1}^N$ , 则 RVM 中输入  $\mathbf{x}_n$  与输出  $t_n$  之间的非线性关系可表示为<sup>[20, 22]</sup>:

$$t_n = \sum_{n=1}^N \mathbf{w}_n K(\mathbf{x}, \mathbf{x}_n) + b + \varepsilon_n \quad (12)$$

式中,  $\varepsilon_n$  为残差余量, 服从  $N(0, \sigma^2)$  的正态分布;  $\mathbf{w}_n$  为权重向量;  $K(\mathbf{x}, \mathbf{x}_n)$  为核函数;  $b$  为常数。

RVM 模型在训练过程中, 要先给定超参数  $\alpha$  和  $\sigma^2$  的初值, 再通过迭代计算, 直到所有参数满足收敛条件。迭代方法已非常成熟, 本文不再详细阐述。迭代收敛后, 超参数更新为  $\alpha_{mp}$  和  $\sigma_{mp}^2$ 。则对于一个新的输入  $\mathbf{x}_*$ , 预测结果为:

$$p(t_* | \mathbf{t}, \alpha_{mp}, \sigma_{mp}^2) = \int p(t_* | \mathbf{w}, \sigma_{mp}^2) p(\mathbf{w} | \mathbf{t}, \alpha_{mp}, \sigma_{mp}^2) d\mathbf{w} = N(t_* | y_*, \sigma_*^2) \quad (13)$$

式中,  $y_* = \mu^T \phi(\mathbf{x}_*)$ ,  $\sigma_*^2 = \sigma_{mp}^2 + \phi(\mathbf{x}_*)^T \Sigma \phi(\mathbf{x}_*)$ ;  $y_*$  表示当模型输入为  $\mathbf{x}_*$  时, RVM 模型的预测输出值; 方差  $\sigma_*^2$  可表示预测值在  $\mathbf{x}_*$  处的随机分布, 则预测结果的 95% 置信区间为  $[y_* - 1.96\sigma_*, y_* + 1.96\sigma_*]$ 。

### 2.2 双目标约束下的 RVM 参数优化方法

以 Sigmoid 核函数  $K(\mathbf{x}, \mathbf{x}_i) = \tanh(\gamma(\mathbf{x} \cdot \mathbf{x}_i) + c)$  为例, 在建立双目标 RVM 模型时, 需要确定的参数有 3 个: *SSI* 的权重  $\theta$ , 核函数中的核宽度  $\gamma$  以及偏差  $c$ , 具体的参数优化步骤如下:

(1) *SSI* 的权重  $\theta$  取值范围在  $[0, 1]$  之间, 取值步长为 0.01, 对于每一个给定的  $\theta_i$ , 可以确定一个相应的目标函数  $L_{\theta_i}$ , 并将其用于优化参数组合  $(\gamma, c)$ 。

(2) 参数  $\gamma$  和  $c$  的初始搜索范围设置为  $[2^{-10}, 2^0]$ , 将它们划分为若干网格, 第  $j$  个网格点的参数组合为  $(\gamma, c)_j$ 。

(3) 对于每个确定的  $\theta_i$  和参数组合  $(\gamma, c)_j$ , 可得相应的目标函数值  $L_{\theta_i, j}$ 。

(4) 比较所有  $L_{\theta_i, j}$  的大小, 得到给定  $\theta_i$  下的最小目标函数值  $L_{\theta_i}^{\min}$ , 并计算出对应的 *SSI* <sub>$\theta_i$</sub>  和 *MSE* <sub>$\theta_i$</sub> 。

(5) 重复步骤 (1) ~ (4), 得到所有  $\theta_i$  下的 *SSI* <sub>$\theta_i$</sub>  和 *MSE* <sub>$\theta_i$</sub> , 并分别进行归一化处理; 为了保证在 *MSE* 不变的情况下, 给予形状相似度 *SSI* 尽可能大的权重, 按照 *SSI*

权重从 1 至 0 的方向，找到两者间距最大的点，所对应的权重  $\theta$  和  $(\gamma, c)$  即为最优参数。

### 2.3 模型评价指标

为定量评价所建模型的预测精度，以复相关系数  $R^2$ 、均方根误差（root mean square error,  $RMSE$ ）和最大绝对误差（maximum absolute error,  $ME$ ）作为量化评价指标，其中  $RMSE$  和  $ME$  越小， $R^2$  越大，表明模型的精度越高；另外，RVM 模型可给出各时刻点值预测的置信区间，结合大坝安全监控领域常用 2S 准则对应的 95% 置信度，本文采用平均置信带宽（average width of confidence interval,  $AWCI$ ）来表示预测时段内置信区间的平均宽度，该值越小，模型预测结果的不确定性越小。各指标的计算方法如下所示：

$$R^2 = \frac{\left[ \sum_{t=1}^T (\hat{\delta}_t - \bar{\hat{\delta}})(\delta_t - \bar{\delta}) \right]^2}{\sum_{t=1}^T (\hat{\delta}_t - \bar{\hat{\delta}})^2 \sum_{t=1}^T (\delta_t - \bar{\delta})^2} \quad (14)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\delta_t - \hat{\delta}_t)^2} \quad (15)$$

$$ME = \max |\delta_t - \hat{\delta}_t| \quad (16)$$

$$AWCI = \frac{1}{T} \sum_{t=1}^T 2 \times 1.96 \times \sqrt{\sigma_t^2} \quad (17)$$

式中， $\bar{\hat{\delta}}$  和  $\bar{\delta}$  分别为时段  $T$  内预测值和实测值的各自平均值； $\sigma_t^2$  为  $t$  时刻的预测方差。对于基于最小二乘法的线性回归模型， $AWCI=2 \times 1.96 \times RMSE$ 。

## 3. 实例分析

锦屏一级拱坝位于四川省凉山州雅砻江干流上，为混凝土双曲拱坝，最大坝高 305 m。坝顶高程 1885 m，水库正常蓄水位 1880 m，死水位 1800 m。2004 年开始前期筹建工作，2009 年 10 月 23 日开始浇筑坝体混凝土，并于 2013 年 12 月 23 日全线浇筑至坝顶高程；2014 年 8 月 24 日，水库首次蓄水至正常蓄水位 1880 m。

### 3.1 HST 和 HHST 因果关系建模因子

对于锦屏一级拱坝，前期研究结果表明<sup>[10]</sup>：该拱坝在每年 1880 m 高水位稳定期（持续时间为 100 至 160 天），坝体存在明显的向下游侧发展的趋势性变形，有限元仿真分析表明该异常变形形态是由水压作用下坝体混凝土徐变引起的粘弹性滞后变形效应和环境温度年周期温降效应共同引起的，进而在 HST 模型的基础上，引入了粘弹性有限元计算得到的滞后水压分量  $\delta_{Hv}$ ，所建 HHST 模型量化得到该趋势性位移增量中两者所引起的比例分别为 30% 和 70%。据此，本文在对该拱坝建立位移预测模型时，分别选用 HST 和 HHST 两组因果关系建模因子作为 RVM 模型的输入。HST 和 HHST 模型的数学表达式如式（18）和（19）所示：

$$\delta = a_0 + \sum_{i=1}^4 a_i H^i + \sum_{i=1}^2 \left( b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) + c_1 \theta + c_2 \ln \theta \quad (18)$$

$$\delta = a_0 + X (\delta_{He}' + \delta_{Hv}') + \sum_{i=1}^2 \left( b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) + c_1 \theta + c_2 \ln \theta \quad (19)$$

式中， $H$  为库水深； $t$  为监测日到起始监测日的累计天数； $\theta = t/100$ ；HHST 模型中  $\delta_{He}' = \sum_{i=1}^4 a_i H^i$  和  $\delta_{Hv}' = d_0 + d_1 \Delta H_{P1} + d_2 \Delta H_{P2} + d_3 \Delta H_{P3} + d_4 \Delta \bar{H}_{P4.5} + d_5 \tau + d_6 \tau^2 + d_7 \tau^3 + d_8 \tau^4$  分别为有限元计算并拟合得到的瞬时水压分量和滞后水压分量， $X$  为总水压分量调整系数，其中滞后水压分量采用约束最小二乘法分段拟合，具体参见文献[10]； $a_0$  为常数项； $a_i$ 、 $b_{1i}$ 、 $b_{2i}$ 、 $c_i$ 、 $d_i$  分别为水压分量、温度分量、时效分量和滞后分量的回归系数。

### 3.2 拱坝位移场空间聚类及形状相似度

根据 2015 年 6 月到 2018 年 12 月的径向位移监测数据，利用基于增量距离的分层聚类法把 34 个测点划分到不同的区域，聚类过程及结果如图 1 所示。由图 1 (a) 可知，总离差平方和随聚类次数的增加而逐渐增大，聚类 30 次之后，总离差平方和开始显著增长，两者之间出现明显的拐点，据此可将 34 个测点划分为 4 类。聚类结果如图 3 所示。从图 1 (b) 可以看出：所有测点被分成 4 个区域，且各区域呈独立分布状态；结合位移监测资料可发现，相同时段内，从 1 区至 4 区，即就是从拱冠梁坝顶到两岸坝肩和河床建基面，坝体径向位移增量呈逐渐减小的趋势，与工程实际情况相符，由此表明了基于增量距离聚类的合理性。

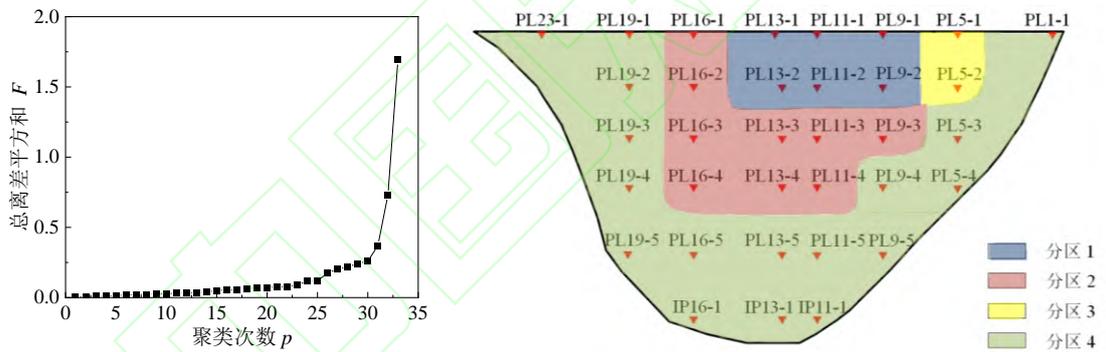


图 1 径向位移场聚类：(a) 聚类总离差平方和，(b) 聚类分区图

Fig. 1 Clusters of the Radial Displacement Field: (a) Clustering Deviations, (b) Distribution of Clusters

该拱坝垂线监测点的采样间隔为 1 天，因此采用滑动窗口法来提取位移时间序列的波动信息时，滑动窗口的步长设置为 1 天。窗口宽度要尽可能小，但同时要避免监测误差的干扰，敏感性分析后将滑动窗口宽度设置为 7 天。以第一分区内的 6 个测点为例，测点之间的形状相似度如图 2 所示。从中可发现，空间位置距离越近的测点，相似度越大，其原因就在于空间距离较小的相邻区域的荷载作用和变形机理也更为接近。

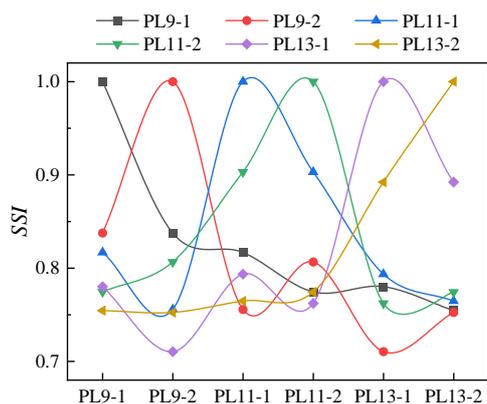


图2 第1分区内测点间形状相似度

Fig. 2 Shape Similarities between Monitoring Points in Cluster 1

### 3.3 单核函数建模结果与分析

对坝体 34 个测点建立预测模型时，所用监测数据时段为 2015 年 6 月 22 日到 2018 年 12 月 31 日。以 2018 年 6 月 13 日为分界点，前面的数据用于模型拟合训练，后面则用于模型预测能力测试。为全面对比分析，本文建立预测模型时采用 HST 和 HHST 两种因果输入因子，且各输入因子组合下分别建立多元回归模型（MLR）、单目标 RVM 模型（RVM1）和双目标 RVM 模型（RVM2），共计 6 种模型，其中 RVM 模型中选用的核函数均为 Sigmoid 核函数。

每个测点的最优权重  $\theta$  的取值是各不相同的，以 PL11-1 和 PL11-2 为例，模型拟合阶段归一化的  $MSE$  和  $SSI$  随  $SSI$  指标权重  $\theta$  的变化规律如图 3 所示。随着  $SSI$  权重的逐渐增大，基于 HST 和 HHST 的 2 种模型的拟合  $MSE$  都会随之增大，这是因为原本的单目标就是使拟合阶段的  $MSE$  最小化，随着新目标  $SSI$  的引入，拟合  $MSE$  必然会增大。另一方面，单目标所建模型的  $SSI$  是比较低的，引入  $SSI$  目标函数后，根据模型输出计算得到的  $SSI$  显著提高。综上，双目标优化建模时的关键在于二者的取舍，要尽量满足  $MSE$  较小而  $SSI$  较大，对于测点 PL11-1 和 PL11-2，HST 模型的  $SSI$  最优权重分别为 0.69 和 0.26，HHST 模型的最优  $SSI$  权重分别为 0.26 和 0.15。

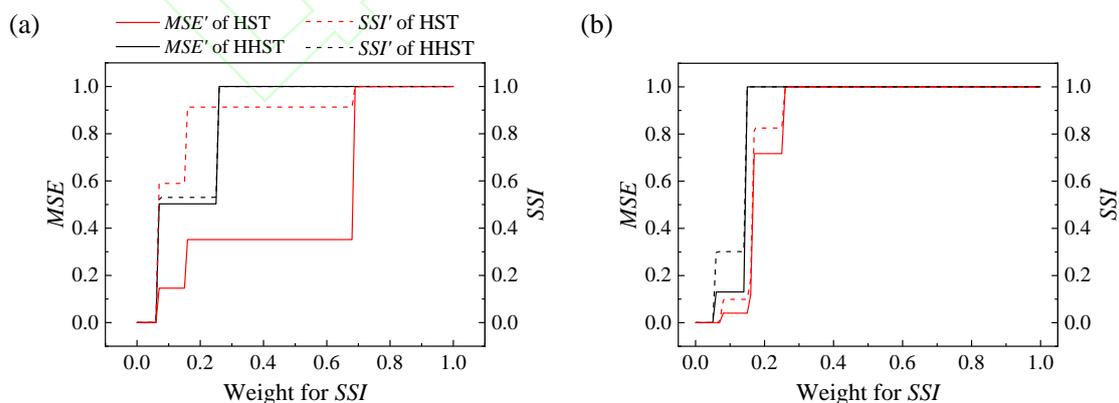


图3 归一化拟合  $MSE$ 、 $SSI$  与权重  $\theta$  的关系：(a) PL11-1, (b) PL11-2

Fig. 3 The Normalized Fitting  $MSE$  and  $SSI$  with respect to the Weight of  $SSI$ : (a) PL11-1, (b) PL11-2

通常而言，基于 MLR 和机器学习语言建立的数学模型均有较好的拟合精度，本文所建 6 种模型亦是如此，因此以下重点对比分析各模型的预测性能。PL11-1 测点的径

向位移预测结果分别如图4所示,同分区内其它5个测点预测结果的整体规律与此类似。从图中可以看出, HST 和 HHST 两种因果因子组合下, 单目标 RVM 模型预测值和实测值的整体偏离程度较大, 绝对残差的平均值和最大值是 3 种建模方法中最大的; 改用双优化目标之后, RVM 模型的残差显著下降, 预测曲线都更接近于实测值, 其与 MLR 模型预测精度的对比需进一步量化分析; 相比与 HST 模型, HHST 模型的预测曲线波动较大, 其原因就在于 HHST 模型中滞后分量的输入因子较多, 且包含库水深的前期变化量, 而水压是最重要的影响因素, 进而导致位移预测值对库水深的微小变化非常敏感, 尽管文献[10]的研究结果表明引入滞后分量能更合理地解释锦屏一级拱坝 1880 m 高水位的异常变形性状, 但从预测角度看, HHST 模型中滞后分量的数学表达式仍需进一步优化, 以便同时保证 HHST 模型的解释能力和预测精度。

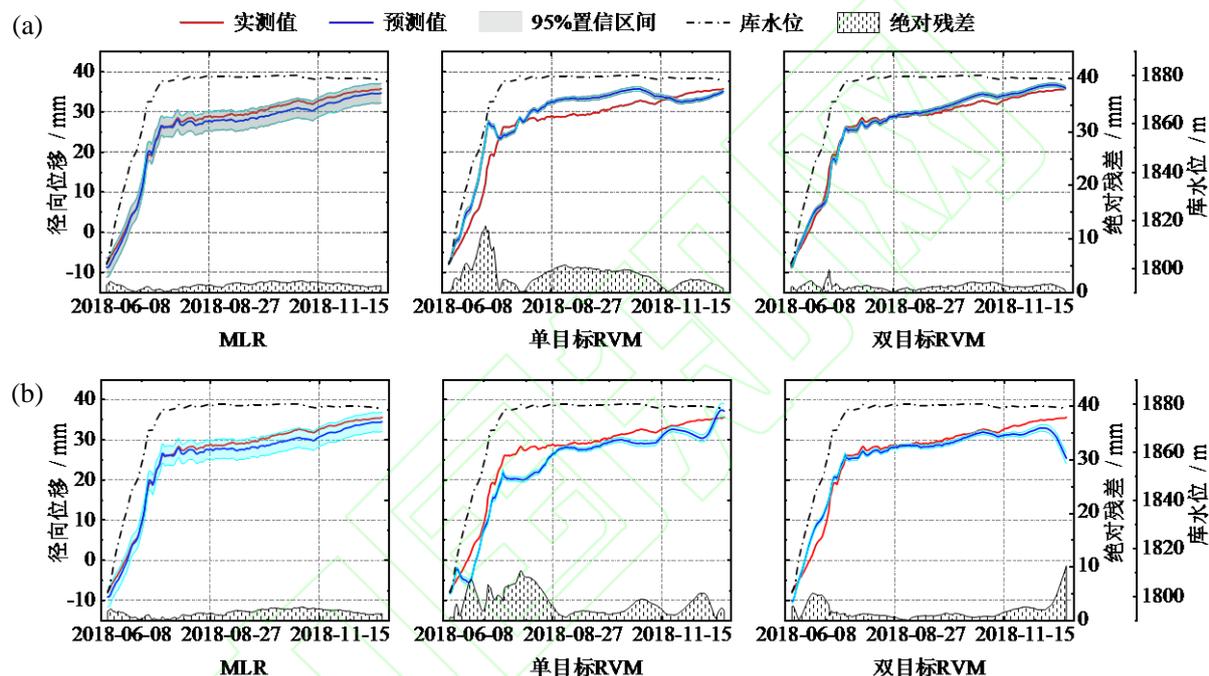


图4 测点 PL11-1 的径向位移实测和预测时间序列: (a) HST, (b) HHST

Fig. 4 Time Series of the Measured and Predicted Radial Displacement of the PL11-1: (a) HST, (b) HHST

2 种因果建模因子下各测点 3 种模型在预测阶段的评价指标对比如图 5 所示。由图可总结出: HST 和 HHST 2 种输入因子下, 相比于单目标 RVM 模型, 仅 PL9-1 和 PL13-2 基于 HHST 模型的预测性能保持不变, 其余测点双目标 RVM 模型的  $RMSE$  和  $ME$  明显减小,  $R^2$  增大,  $RMSE$  和  $ME$  的最大降幅分别为 66.2% 和 71.1%, 平均降幅为 31.2% 和 24.8%, 充分表明引入双目标优化条件能明显提升 RVM 模型的预测精度和稳定性; 与 MLR 模型对比可发现, 几乎所有的单目标 RVM 模型的  $RMSE$  和  $ME$  都明显偏大,  $R^2$  更小, 表明总体预测效果是相对较差的, 引入双目标优化条件后, 预测效果有所提升, 但 12 个模型中仅 4 个模型的预测性能略微优于 MLR 模型, 表明单核函数双目标 RVM 模型的预测性能仍劣于 MLR 模型; 单目标 RVM 模型和双目标 RVM 模型的  $AWCI$  指标值几乎相同, 由此表明引入空间关联性这一约束条件无法降低预测结果的不确定性, 但 RVM 模型在预测阶段的置信带宽明显小于 MLR 模型, 最大降幅和平均降幅分别为 86.9% 和 75.1%。综上所述, RVM 模型能有效减少预测结果的不确定性, 引入双优

化目标可明显提升 RVM 模型的预测性能，但基于单核函数建立的双目标 RVM 模型的预测能力仍然有限，大部分情况下仍然低于 MLR 模型，因此需进一步研究提高双目标 RVM 模型的预测精度和稳定性的方法。

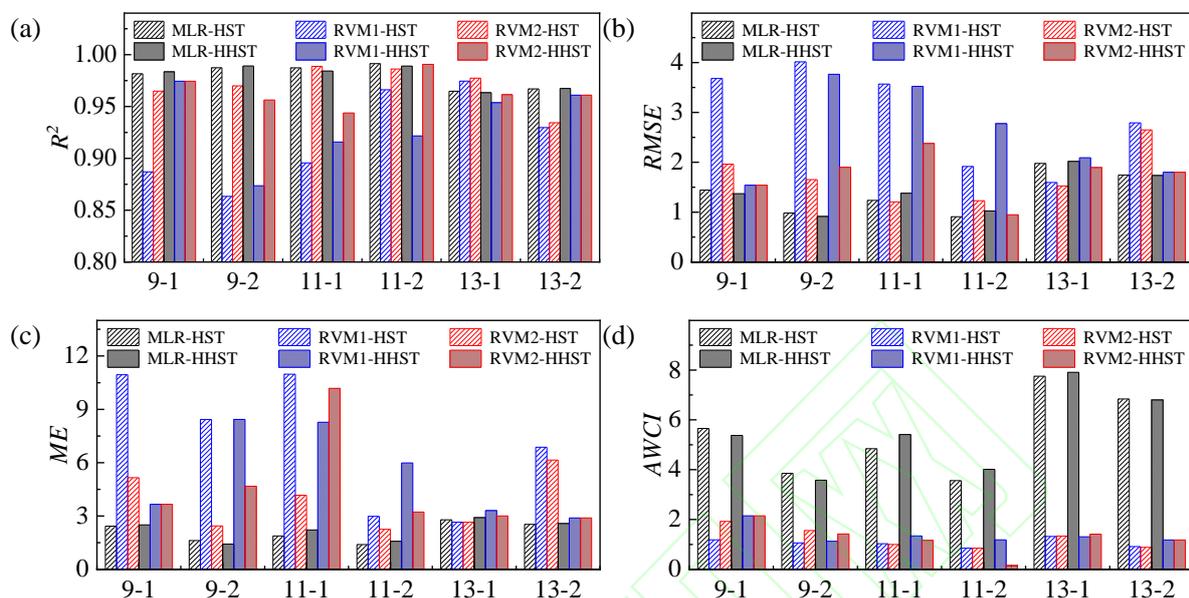
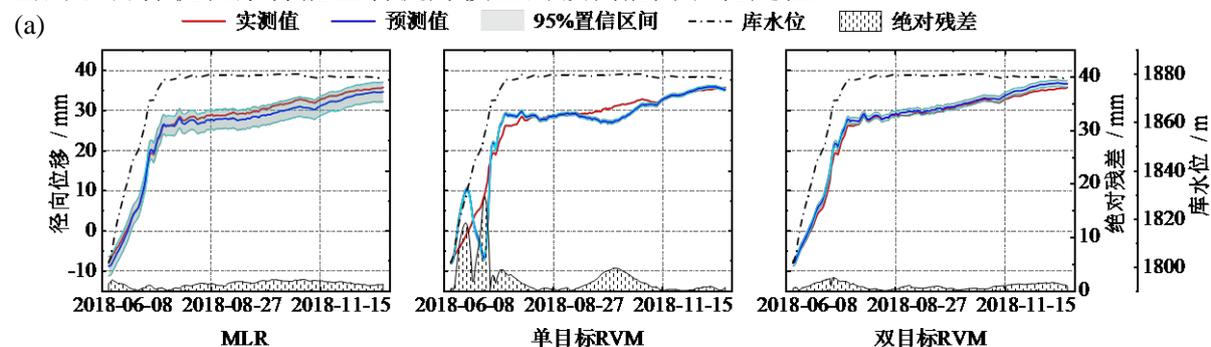


图 5 预测评价指标: (a)  $R^2$ , (b)  $RMSE$ , (c)  $ME$ , (d)  $AWCI$

Fig. 5 Evaluation Index for Model Prediction Performance: (a)  $R^2$ , (b)  $RMSE$ , (c)  $ME$ , (d)  $AWCI$

### 3.4 多核函数建模结果与分析

RVM 模型的性能与相关向量的数量有关，通常而言，采用单一核函数构建的 RVM 模型中的相关向量较少，模型结构较为简单，导致其泛化能力有所欠缺，常用的解决办法是将多个核函数组合使用<sup>[20, 22]</sup>。为进一步提高 RVM 模型的预测精度，本文在原有单一 Sigmoid 核函数的基础上，增加 Gauss 核函数 ( $K(x, x_i) = \exp(-\beta \|x - x_i\|^2)$ ， $\beta$  为核宽度)，进而形成组合核函数，据此得到的 PL11-1 测点的径向位移预测时间序列分别如图 6 所示。从图中可以发现，多核单目标 RVM 模型的预测精度也较差，预测曲线出现了较大的波动，表明其预测稳定性也较差；采用双目标优化后，多核 RVM 模型的预测效果得到显著提高，绝对残差的平均值和最大值均明显降低，进一步验证了本文所提出的双目标优化条件能显著提升模型的预测能力和稳定性。



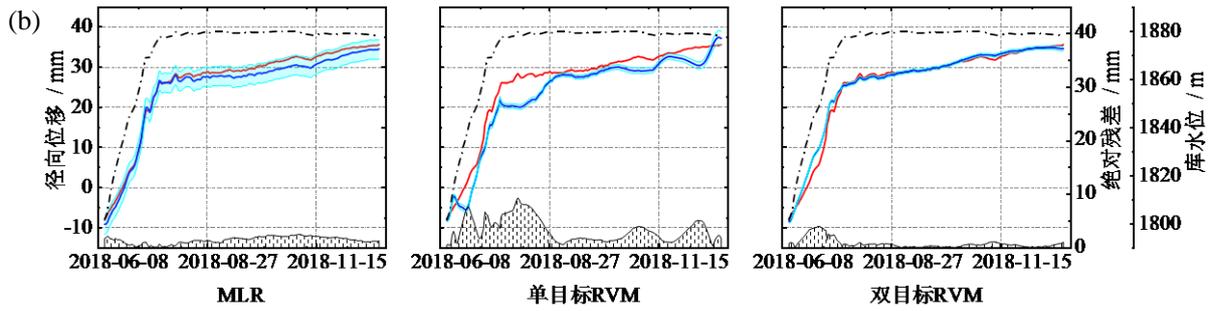


图 6 测点 PL11-1 的径向位移实测和预测时间序列：(a) HST，(b) HHST

Fig. 6 Time Series of the Measured and Predicted Radial Displacement of the PL11-1: (a) HST, (b) HHST

图 7 为多核双目标 RVM 模型与单核双目标 RVM 模型和 MLR 模型的预测  $RMSE$  的对比，降低百分比统计如表 1 所示。从图 7 和表 1 可以看出，使用多核函数之后，双目标 RVM 模型的预测效果进一步提升，6 个测点 HST 和 HHST 模型的预测  $RMSE$  平均分别降低 44.5% 和 29.9%，且除了 PL9-1 和 PL9-2 的 HHST 模型外，其余 10 个多核双目标 RVM 模型的预测精度均高于 MLR 模型，预测  $RMSE$  平均分别降低 30.6% 和 13.4%，充分表明多核函数对 RVM 模型的预测性能有较强的提升效果。

表 1 多核双目标 RVM 模型预测阶段  $RMSE$  降低百分比

Tab. 1 The Decrease Proportion of  $RMSE$  of the Double Objective RVM model with Multi-kernel Function

对比模型	模型因子	9-1	9-2	11-1	11-2	13-1	13-2	平均值
单核双目标 RVM 模型	HST	41.9%	49.4%	27.9%	43.4%	45.4%	58.7%	44.5%
	HHST	6.1%	44.5%	49.9%	33.4%	1.6%	43.6%	29.9%
MLR 模型	HST	21.0%	14.8%	29.1%	23.4%	57.9%	37.3%	30.6%
	HHST	-5.5%	-15.4%	13.8%	38.3%	7.6%	41.5%	13.4%

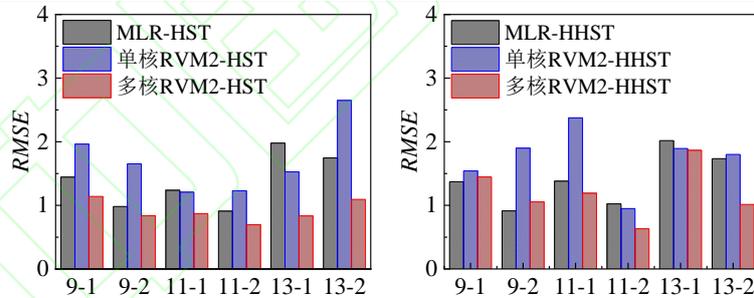


图 7 单、多核函数下双目标 RVM 模型和 MLR 模型的预测  $RMSE$

Fig. 7 Prediction  $RMSE$  of the RVM and MLR models

综上所述，RVM 凭借其强大的数据挖掘能力已成为大坝安全监控领域强有力的建模工具，但针对特高拱坝的位移预测模型，传统建模方法中仅依靠拟合均方差最小的单一优化目标，容易导致模型出现过拟合的问题，所建模型的预测精度和稳定性难以保证，而融合拱坝变形空间关联性的双优化目标则有效地解决了这一问题。以锦屏一级拱坝径向位移增量距离第 1 分区中的 6 个测点为例，相比于单目标 RVM 模型，使用双目标优化之后，模型的  $RMSE$  和  $ME$  平均降幅分别为 31.2% 和 24.8%，采用多核函数可进一步提升模型预测性能，大部分测点的预测效果优于 MLR 模型；RVM 模型的预测置信带宽明显小于 MLR 模型，平均降幅为 75.1%，而且预测带宽对于每个数据点是独立分布的。由此表明多核函数双目标 RVM 模型可有效提升特高拱坝位移预测的精度和稳定性，

并降低预测结果的不确定性，可将其用于拱坝空间变形性态的健康诊断中。

## 4. 结论

基于机器学习语言建立拱坝位移预测模型时，参数优化的传统方法中仅以拟合均方差最小为单一目标，导致所建模型容易出现过拟合问题。对此，通过位移监测时间序列的形状相似度对拱坝变形的空间关联性进行量化，提出了拟合均方差尽可能小且空间关联性尽可能大的双优化目标，据此建立了多核函数双目标 RVM 模型，所建模型能有效提升特高拱坝位移预测的精度和稳定性，同时降低预测结果的不确定性。

引入滞后分量的 HHST 模型，已被证明可以更好地解释锦屏一级拱坝由环境温度年周期温降作用和粘弹性滞后变形效应所共同引起的异常变形性态，但就监控模型的预测精度而言，HHST 模型的预测能力略低于 HST 模型，由此表明在 HST 模型中引入滞后分量对解释锦屏一级拱坝变形性态而言是非常有必要的，但现有 HHST 模型中滞后分量的数学表达式还需进一步优化，以期同时保证监控模型的解释能力和预测精度。此外，在进行双目标约束下的参数优化时，采用的是网格搜索算法，未来研究中可结合遗传算法、粒子群优化算法等智能搜索算法，提升模型优化的精确性。

## 参考文献

- [1] Zhou Jianping, Du Xiaohu, Zhou Xingbo, et al. Research on High Dams and Developing Trends [J]. Journal of Hydroelectric Engineering, 2019, 38(2): 1-14 (周建平, 杜效鹤, 周兴波, 等. 世界高坝研究及其未来发展趋势[J]. 水力发电学报, 2019, 38(2): 1-14)
- [2] Gu Chongshi, Su Huaizhi, Wang Shaowei. Advances in Calculation Models and Monitoring Methods for Long-Term Deformation Behavior of Concrete Dams [J]. Journal of Hydroelectric Engineering, 2016, 35(5): 1-14 (顾冲时, 苏怀智, 王少伟. 高混凝土坝长期变形特性计算模型及监控方法研究进展[J]. 水力发电学报, 2016, 35(5): 1-14)
- [3] Salazar F, Morán R, Toledo M A, et al. Data-Based Models for The Prediction of Dam Behavior: A Review and Some Methodological Consideration [J]. Archives Computational Methods in Engineering, 2017, 24(1): 1-21.
- [4] Wu Zhongru. Safety Monitoring Theory & Its Application of Hydraulic Structures [M]. Beijing: Higher Education Press, 2003 (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)
- [5] Su Huaizhi, Chen Zhexin, Wen Zhiping. Performance Improvement Method of Support Vector Machine-Based Model Monitoring Dam Safety [J]. Structural Control and Health Monitoring, 2016, 23(2): 252-266.
- [6] Li Mingchao, Ren Qiubing, Kong Rui, et al. Dynamic Modeling and Prediction Analysis of Dam Deformation under Multidimensional Complex Relevance [J]. Journal of Hydraulic Engineering, 2019, 50(6): 687-698 (李明超, 任秋兵, 孔锐, 等. 多维复杂关联因素下的大坝变形动态建模与预测分析[J]. 水利学报, 2019, 50(6): 687-698)
- [7] Wei Bowen, Chen Liangjie, Li Huokun, et al. Optimized Prediction Model for Concrete Dam Displacement Based on Signal Residual Amendment [J]. Applied Mathematical Modelling, 2020, 78: 20-36.
- [8] Lu Zhengchao, Zhang Jinping, Li Libing, et al. Analysis on Frost Deformation of The Fengman Concrete Gravity Dam [J]. Journal of China Institute of Water Resources and Hydropower Research, 2006, 4(1): 53-57, 61 (卢正超, 张进平, 黎利兵, 等. 丰满混凝土重力坝的冻胀变形分析[J]. 中国水利水电科学研究院学报, 2006, 4(1): 53-57, 61)
- [9] Hu Jiang, Wu Suhua. Statistical Modeling for Deformation Analysis of Concrete Arch Dams with

- Influential Horizontal Cracks [J]. *Structural Health Monitoring*, 2019, 18(2): 546-562.
- [10] Wang Shaowei, Xu Yingli, Gu Chongshi, et al. Hysteretic Effect Considered Monitoring Model for Interpreting Abnormal Deformation Behavior of Arch Dams: A Case Study [J]. *Structural Control and Health Monitoring*, 2019, 26(10): 1-20.
- [11] Wei Bowen, Xiong We, Li Huokun, et al. Dam Deformation Forecasting of Leapfrog Combined Model Merging Residual Errors of Chaos [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(9): 1072-1078 (魏博文, 熊威, 李火坤, 等. 融合混沌残差的大坝位移蛙跳式组合预报模型 [J]. *武汉大学学报·信息科学版*, 2016, 41(9): 1072-1078)
- [12] He Jinping, Gao Quan. An Improved Algorithm of Cloud Fusion for Dam Health Diagnosis [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(7): 1022-1029 (何金平, 高全. 一种适应大坝健康诊断的改进云合并算法[J]. *武汉大学学报·信息科学版*, 2018, 43(7): 1022-1029)
- [13] Mata J. Interpretation of Concrete Dam Behaviour with Artificial Neural Network and Multiple Linear Regression Models [J]. *Engineering Structures*, 2011, 33(3): 903-910.
- [14] Ranković V, Grujović N, Divac D, et al. Development of Support Vector Regression Identification Model for Prediction of Dam Structural Behaviour [J]. *Structural Safety*, 2014, 48: 33-39.
- [15] Salazar F, Toledo MÁ, Oñate E, et al. Interpretation of Dam Deformation and Leakage with Boosted Regression Trees [J]. *Engineering Structures*, 2016, 119:230-251.
- [16] Liu Wenju, Pan Jianwen, Ren Yisha, et al. Coupling Prediction Model for Long-Term Displacements of Arch Dams Based on Long Short-Term Memory Network [J]. *Structural Control and Health Monitoring*, 2020, 27(3): e2548.
- [17] Salazar F, Toledo MA, Oñate E, et al. An Empirical Comparison of Machine Learning Techniques for Dam Behaviour Modelling [J]. *Structural Safety*, 2015, 56: 9-17.
- [18] Wang Shaowei, Xu Yingli, Gu Chongshi, et al. Two Spatial Association-Considered Mathematical Models for Diagnosing The Long-Term Balanced Relationship and Short-Term Fluctuation of The Deformation Behaviour of High Concrete Arch Dams [J]. *Structural Health Monitoring*, 2020, 19(5): 1421-1439.
- [19] Shao Chenfei, Gu Chongshi, Yang Meng, et al. A novel model of dam displacement based on panel data [J]. *Structural Control and Health Monitoring*, 2018, 25(1): e2037.
- [20] Chen Siyu, Gu Chongshi, Lin Chaoning, et al. Multi-Kernel Optimized Relevance Vector Machine for Probabilistic Prediction of Concrete Dam Displacement [J]. *Engineering with Computers*. 2020 (online first).
- [21] Ren Juan, Chen Qi. Clustering and its empirical study based on shape for multivariable panel data [J]. *Statistics Information Forum*, 2011, 26(10): 28-33. (任娟, 陈圻. 基于形状特征的多指标面板数据聚类方法及其应用[J]. *统计与信息论坛*, 2011, 26(10): 28-33)
- [22] Fan Zhendong, Cui Weijie, Chen Min, et al. Prewarning Model for Dam Safety Based on IPSO -RVM [J]. *Journal of Yangtze River Scientific Research Institute*, 2016, 33(2): 48-51 (范振东, 崔伟杰, 陈敏, 等. 基于 IPSO-RVM 的大坝安全预警模型[J]. *长江科学院院报*, 2016, 33(2): 48-51)

## **A Probabilistic Prediction Model for Displacement of Super High Arch Dams Considering the Deformation Spatial Association**

XU Cong<sup>1</sup>    WANG Shaowei<sup>1,2</sup>    GU Chongshi<sup>3</sup>    SU Huaizhi<sup>3</sup>

1 School of Environmental and Safety Engineering, Changzhou University, Changzhou 213164, China

2 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

3 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

**Abstract:** Machine learning language has become an ideal modeling tool in the field of dam health monitoring with its powerful nonlinear data mining ability. However, the minimum fitting mean square error (MSE) is determined as the only optimization objective in the

traditional modeling process, which is likely to cause over-fitting problems. To overcome this problem, based on the relevance vector machine (RVM), a probabilistic prediction model is established under the constraint of double optimization objectives, which integrates the deformation spatial association and MSE. The deformation spatial association is quantified by the shape similarity index (SSI) at first. The double objective is then established with the combination of the MSE and SSI, and is achieved by making the *MSE* as small as possible, while the *SSI* is as large as possible. Engineering example of the Jinping-I arch dam shows that the average decrease proportion of the root mean square error (RMSE) and maximum absolute error (ME) of the proposed double objective RVM model is 31.2% and 24.8%, respectively, and the prediction performance can be further improved by using the multi-kernel function. The prediction confidence bandwidth of the RVM model is significantly smaller than that of the traditional multiple linear regression model, with an average decrease proportion of 75.1%. Therefore, the multi-kernel double objective RVM model established for the displacement of super high arch dams can effectively improve the prediction performance and reduce the uncertainty.

**Key words:** super high arch dams, displacement, deformation spatial association, probabilistic prediction model, relevance vector machine, double objective optimization

**First author:** XU Cong, master degree candidate, specializes in health monitoring of hydraulic structures. Email: 18306120727@163.com

**Corresponding author:** WANG Shaowei, PhD, associate professor, specializes in health monitoring and aging diseases of hydraulic structures. Email: shaowei2006nanjing@163.com

**Foundation support:** The National Natural Science Foundation of China (51709021), the Project funded by China Postdoctoral Science Foundation (2020M670387), the Belt and Road Special Foundation of the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (2019nkzd03), the Open Research Fund of State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin (China Institute of Water Resources and Hydropower Research) (IWHR-SKL-KF202002), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20\_2560).

## 网络首发:

**标题:** 融合空间关联性的特高拱坝位移概率性预测模型

**作者:** 徐丛, 王少伟, 顾冲时, 苏怀智

**收稿日期:** 2020-12-24

**DOI:**10.13203/j.whugis20200508

## 引用格式:

徐丛, 王少伟, 顾冲时, 等. 融合空间关联性的特高拱坝位移概率性预测模型[J]. 武汉大学学报·信息科学版, 2021, DOI: 10.13203/j.whugis20200508 (XU Cong, WANG Shaowei, GU Chongshi, et al. A Probabilistic Prediction Model for Displacement of Super High Arch Dams Considering the Deformation Spatial Association [J]. Geomatics and Information Science of Wuhan University, 2021, DOI: 10.13203/j.whugis20200508)

---

**网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!**

**您感兴趣的其他相关论文:**

### 利用优化的组合核相关向量机算法构建地表下沉预测模型

罗亦泳, 姚宜斌, 赵庆志, 周世健

武汉大学学报·信息科学版, 2018, 43(9): 1295-1301

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20160368>

### 大坝监测资料的时变 Kalman 预测模型

李子阳, 郭丽, 顾冲时

武汉大学学报·信息科学版, 2010, 35(8): 991-995

<http://ch.whu.edu.cn/article/id/1030>

### 数学形态滤波在大坝安全监控数据粗差检测中的应用

景继, 顾冲时

武汉大学学报·信息科学版, 2009, 34(9): 1126-1129

<http://ch.whu.edu.cn/article/id/1382>