



Multi-kernel optimized relevance vector machine for probabilistic prediction of concrete dam displacement

Siyu Chen^{1,2,3} · Chongshi Gu^{1,2,3} · Chaoning Lin² · Kang Zhang^{1,2,3} · Yantao Zhu^{1,2,3}

Received: 22 October 2019 / Accepted: 23 December 2019 / Published online: 11 January 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

The observation data of dam displacement can reflect the dam's actual service behavior intuitively. Therefore, the establishment of a precise data-driven model to realize accurate and reliable safety monitoring of dam deformation is necessary. This study proposes a novel probabilistic prediction approach for concrete dam displacement based on optimized relevance vector machine (ORVM). A practical optimization framework for parameters estimation using the parallel Jaya algorithm (PJA) is developed, and various simple kernel/multi-kernel functions of relevance vector machine (RVM) are tested to obtain the optimal selection. The proposed model is tested on radial displacement measurements of a concrete arch dam to mine the effect of hydrostatic, seasonal and irreversible time components on dam deformation. Four algorithms, including support vector regression (SVR), radial basis function neural network (RBF-NN), extreme learning machine (ELM) and the HST-based multiple linear regression (HST-MLR), are used for comparison with the ORVM model. The simulation results demonstrate that the proposed multi-kernel ORVM model has the best performance for predicting the displacement out of range of the used measurements dataset. Meanwhile, the ORVM model has the advantages of probabilistic output and can provide reasonable confidence interval (CI) for dam safety monitoring. This study lays the foundation for the application of RVM in the field of dam health monitoring.

Keywords Optimized relevance vector machine · Multi-kernel · Jaya optimization algorithm · Dam health monitoring · Prediction model

Abbreviations

RVM	Relevance vector machine	SR	Stepwise regression
ORVM	Optimized relevance vector machine	PJA	Parallel Jaya algorithm
CI	Confidence interval	ANN	Artificial neural network
HST	Hydrostatic-season-time	MLP	Multilayer perceptron
HTT	Hydrostatic-temperature-time	SLFNs	Single hidden layer feedforward neural networks
MLR	Multiple linear regression	ANFIS	Adaptive neural fuzzy inference system
PLSR	Partial least squares regression	MARS	Multivariate adaptive regression splines
		GPR	Gaussian process regression
		RBF-NN	Radial basis function neural network
		ELM	Extreme learning machine
		SVM	Support vector machine
		SVR	Support vector regression
		MLR-HST	HST-based multiple linear regression
		SumGP	Multi-kernel Gaussian kernel + polynomial kernel
		SumGL	Multi-kernel Gaussian kernel + Laplace kernel
		SumLP	Multi-kernel Laplace kernel + polynomial kernel

✉ Chongshi Gu
csgu@hhu.edu.cn

✉ Chaoning Lin
linchaoning@hhu.edu.cn

¹ State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

² College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China

³ National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, Nanjing 210098, China

G-ORVM	Gaussian kernel-based optimized relevance vector machine
GP-ORVM	SumGP kernel-based optimized relevance vector machine
R^2	Coefficient of determination
RMSE	Root mean square error
MAE	Mean absolute error
ME	Maximum absolute error
AWCI	Average width of confidence interval
AVCI	Average variance of confidence interval

1 Introduction

Dam engineering is an important kind of infrastructure that can bring significant benefits in the economic and social fields, such as flood control, power generation, water supply, and irrigation. The operation state of dams is very complicated for its relations with the water level, ambient temperature dam material properties and geo-mechanical factors [1]. The failure of dam engineering could pose uncontrolled flood and cause disaster to downstream areas. The past century has witnessed many severe accidents of dam engineering failure worldwide such as China (Gouhou CRFD, 1993), France (Malpasset Arch Dam, 1959), Italy (Gleno Multiple-Arch Dam, 1923; Vajont Arch Dam, 1963), Spain (Tous dam, 1982), USA (St. Francis Gravity Dam, 1928; Teton Earth Dam, 1976) [2].

Timely and effective detection of observational data anomalies according to the monitoring model may prevent hidden dangers and the occurrence of accidents. In the daily practice, the prediction models of dam displacement can be categorized into three groups: statistical model, hybrid model, and deterministic model, which are widely used during the construction period, storage period and operation period of dam engineering [1, 3, 4].

The deterministic model, also called numerical models, is established based on the numerical methods, such as the finite element method and discrete element method [5, 6]. The deterministic model can interpret dam displacement in mechanics concept, the modeling is based on many numerical computations and structural simplifications. However, in this approach, modeling and calculations are time-consuming considering various geometries and operation conditions [5]. Moreover, due to the limitation of computational techniques and parameter settings, some special monitoring effects (e.g. seepage, uplift pressure) are difficult to be predicted accurately using the deterministic model. Sometimes a deterministic model cannot be supplied with the thermal effects because of the lack of temperature measurements. In this case, the hybrid model is a good solution where the thermal effect is represented by periodic time functions and the other effects are the same as the deterministic model [1].

The statistical models are based on the previous data and basic mathematical functions. Very well-known statistical models in dam safety procedures are hydrostatic-season-time (HST) and hydrostatic-temperature-time (HTT) model [7], and the former one is widely used in the positive analysis and inverse analysis of dam health monitoring [5, 8, 9]. The unknown coefficients can be obtained by using different regression techniques, such as multiple linear regression (MLR) [10, 11], partial least squares regression (PLSR) [12] and stepwise regression [13].

However, linear regression-based models also contain some disadvantages. On the one hand, they are not well-suited to model nonlinear interactions between input factors and dam displacement [6]. On the other hand, they are easily ill-conditioned [14]. The limitations of conventional statistical models have motivated dam engineers to work on new approaches for dam behavior modeling [6]. With the development of machine learning, a great number of application methods have been proposed recent years and applied in dam engineering, including dam health monitoring [6], reliability analysis [15, 16], seismic evaluation [17], computational cost reduction [18] and uncertainty quantification [19]. Artificial neural network (ANN) [20] and support vector machine (SVM) [21] are the most popular methods, having good computing performance for solving nonlinear problems. There are various types of ANN models and most of the applications are based on the multilayer perceptron (MLP), where the major challenge is learning time requirement and structure selection. Single hidden layer feedforward neural networks (SLFNs), such as radial basis function neural network (RBF-NN) [22, 23] and extreme learning machine (ELM) [24], were tested for modeling in dam health monitoring due to the simple structure and efficient algorithms. RBF-NN is an SLFN that utilizes radial basis functions as activation functions in the hidden layer where the output is a linear combination of the inputs and neuron parameters [20]. ELM was proposed by Huang et al [25]. Compared with other standard SLFNs, ELM decreases the required training time and has fewer parameters settings. Despite the poor stability of the output due to the stochastic untrained weights and biases in the input to the hidden layer, the average performance of ELM is verified to be superior to standard SLFNs, stepwise regression model and MLR model in the application of dam displacement prediction [24]. The SVM, a kernel-based technique, is currently the most popular machine learning method. SVM has a relatively good capability of solving nonlinear problems, especially for the data with fewer samples and high dimensions. In order to enable SVM to solve the regression problem, the insensitive loss coefficient ϵ was introduced into the SVM and support vector regression (SVR) was developed [21]. Some researchers have reported the superior performance of the SVR in dam structural monitoring [26–28]. Except for the

forementioned models, adaptive neural fuzzy inference system (ANFIS) [29], multivariate adaptive regression splines (MARS) [30] and Gaussian process regression (GPR) [31, 32] are also competitive ML models used in dam health monitoring though with high computational cost and complexity. A detailed literature review about ML models used in dam health monitoring can be referred to [5, 6].

Relevance vector machine (RVM) is a predictive machine learning model proposed by Tipping [33]. RVM is a flexible and powerful tool that modifies the principal ideas behind the SVM with Bayesian theory and has a similar form to SVM. The advantage of RVM is its capacity to provide reasonable inferences at low computational cost and improves the inadequacy of SVM in many aspects, including the utilization of non-mercer kernels, reduced sensitivity to hyper-parameter settings and probabilistic output with fewer relevance vectors for a given dataset [33]. RVM is suitable for dealing with complex regression and classification problems and has been verified in many practical problems. Imani et al. [34] examined the capability of relevance vector machine models for predicting sea-level variations and concluded that the RVM approach was superior to ELM in terms of accuracy during the test periods. Zhang et al. [35] utilized the relevance vector machine for stability inference of soil slopes. Wang et al. [36] used the multiclass relevance vector machine approach to classify faulty samples of multi-level inverter system. Kong et al. [37] utilized the relevance vector machine to realize real-time monitoring of the tool wear in machining process. Most of the previous application studies of RVM were based on the Gaussian kernel, where trial-and-error or pilot calculation were often used to determine the hyper-parameter value. It should be noticed that trial-and-error and pilot calculation is time-consuming if the dataset is large and the iteration step is small [35]. In fact, kernel function and hyper-parameters value are important factors that have impacts on the sparsity and generalization performance of RVM.

Currently, there is no general consensus on the appropriate setting of the kernel function and hyper-parameters. The determination of the kernel function and corresponding hyper-parameter value of the RVM model under a given problem can be recognized as a constrained optimization problem. Evolutionary algorithms and swarm intelligence algorithms are two important kinds of population-based heuristic algorithms [38], which have been widely used in a variety of engineering problems. Genetic algorithm [39] and artificial immune algorithm [40] are two typical evolutionary algorithms. Particle swarm optimization [41], artificial fish swarm algorithm [42], artificial bee colony algorithm [43] are popular swarm intelligence algorithms. In addition to the evolutionary and swarm intelligence-based algorithms, there are a variety of other algorithms that work on the principles of different natural phenomena. Jaya algorithm is a recently proposed novel

global optimization method. Compared with other popular algorithms, it does not contain any algorithm-specific parameter to be tuned, which makes the algorithm convenient to implement in the practical application [37, 38, 44, 45].

The purpose of this paper is to develop a novel monitoring model for the probabilistic prediction of concrete dam displacement. An efficient optimization framework for RVM parameters is developed based on the parallel Jaya algorithm (PJA). The proposed optimized relevance vector machine (ORVM) has a good performance on estimating the optimal hyper-parameter values of the RVM real-time and is able to provide the reliable predicted results of concrete dam displacement. In addition, this paper compares the nonlinear mapping capabilities of ORVM models with different kernel functions (e.g. simple kernel functions and multi-kernel functions) and discuss the most suitable choice for given data. The developed ORVM model is performed on a super-high concrete arch dam located in China and a discussion is conducted compared with equivalent SVR, RBF-NN, ELM, and HST-MLR models.

The rest of the paper is organized as follows. In Section 2, related methodologies, such as the statistical monitoring model of concrete dam displacement and description of the proposed ORVM, are illustrated in detail. Data collection, detailed analyses, and comparisons of predicted results are shown in Section 3. The conclusion and future work are summarized in Section 4.

2 Methodologies

2.1 Statistical model for concrete dam displacement monitoring

As a comprehensive response of dam structural behavior, dam displacement is a nonlinear function of hydrostatic pressure, temperature, time effect, and other unexpected unknown factors [1, 5]. According to the current research, the hydrostatic-seasonal-time (HST) model is one of the most popular statistical models for dam deformation monitoring [6]. HST model is based on the analysis of structure and mechanics theory, which can be quantitatively interpreted and approximated by the following expression:

$$\delta = \delta_H(t) + \delta_T(t) + \delta_\theta(t) \quad (1)$$

where water pressure component $\delta_H(t)$ denotes reversible effect of static hydrostatic pressure, temperature component $\delta_T(t)$ denotes reversible effect influenced by seasonal thermal and ambient temperature, $\delta_\theta(t)$ denotes time component (also called aging component).

Under the action of water pressure, $\delta_H(t)$ can be described by a polynomial function consisting of reservoir water levels

H and coefficients a_i , which is given in Eq. (2). The value of h depends on the dam type. For gravity dam, h is set to 3; for arch dam, h is set to 4.

$$\delta_H(t) = \sum_{i=1}^h a_i H^i \tag{2}$$

The temperature component $\delta_T(t)$ describes the displacement caused by the temperature changes in bedrock and dam concrete. The calculation of $\delta_T(t)$ depends on the layouts of the thermometers. If the thermometers equipped in dam engineering is enough and the measured data is sufficient as well as continuous, these measurements can describe the dam temperature field very well and $\delta_T(t)$ can be calculated by Eq. (3). Otherwise, $\delta_T(t)$ can be calculated by a combination of harmonic function given in Eq. (4).

$$\delta_T(t) = \sum_{i=1}^{l_1} b_i T_i \text{ or } \delta_T = \sum_{i=1}^{l_2} b_{1i} \bar{T}_i + \sum_{i=1}^{l_2} b_{2i} \beta_i \tag{3}$$

$$\delta_T(t) = b_1 \sin(d) + b_2 \cos(d) + b_3 \sin(d) \cos(d) + b_4 \sin^2(d) \tag{4}$$

where b_i, b_{1i} and b_{2i} are coefficients; T_i is the observation value of the i th thermometer, l_1 denotes the number of thermometers used for modeling; \bar{T}_i and β_i denote the average value of the measured temperature at i th layer and temperature gradient, respectively; l_2 denotes is the number of layers by where the thermometers are installed; $d = 2\pi t/365$; t is the number of days from the observation date to the beginning of the monitoring sequence.

The time component $\delta_\theta(t)$ reflects the irreversible deformation of the dam body or dam foundation toward a certain direction over time. For a normal concrete dam, $\delta_\theta(t)$ rapidly changes at the initial service life and then stabilizes in the later stage. According to the current research [1], different and strictly monotone functions can be used for modeling the time effect $\delta_\theta(t)$, as shown in Eq. (5)

$$\delta_\theta(t) = c_1 \theta + c_2 \ln(\theta) + c_3(1 - e^{-\theta}) \tag{5}$$

where c_i are coefficients; $\theta = t/100$.

2.2 Optimized relevance vector machine with multi-kernel

2.2.1 Theory of relevance vector machine

The RVM, originally proposed by Tipping [33], is a predictive machine learning model and has the comparable form to SVM as shown in Eq.(6). RVM can be utilized for solving regression problems and provides probabilistic estimates, as opposed to the SVM’s point estimates. Given a set of input

vectors $\{x_n, t_n\}_{n=1}^N$, presume that $t_n = y(\mathbf{x}_n, \mathbf{w}) + \varepsilon_n$, where $\varepsilon_n \sim \mathbf{N}(0, \sigma^2)$ and $\mathbf{N}(0, \sigma^2)$ denotes the normal distribution with mean-zero with variance σ^2 . The output, combined with kernel function $K(x, x_n)$, can be written as

$$\mathbf{y} = f(\mathbf{x}) = \sum_{n=1}^N w_n K(x, x_n) + b \tag{6}$$

where w_n denotes the weights vector which is to be adjusted for the training set, b denotes the bias.

The probabilistic formulation of RVM model can be defined as

$$p(t_n|\mathbf{X}) = \mathcal{N}(t_n|y(\mathbf{x}_n), \sigma^2) \tag{7}$$

where \mathcal{N} denotes the normal distribution over t_n with mean of $y(\mathbf{x}_n)$ and variance σ^2 . $y(\mathbf{x})$ represents a linearly weighted sum of nonlinear fixed basis functions, which contains the same definition as Eq. (6). On account of the assumption of independence of the t_n , the likelihood of the whole dataset can be defined as follows

$$p(\mathbf{t}|\mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{t} - \Phi\mathbf{w}\|^2\right\} \tag{8}$$

where $\mathbf{t} = (t_1 \dots t_N)^T$, $\mathbf{w} = (w_0 \dots w_N)^T$ and $\Phi(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \dots, K(x_n, x_N)]^T$. There are several types of kernel functions could be used in Φ , such as simple kernel and multi-kernel. We will discuss them in detail in Section 2.2.2.

For the purpose of overcoming over-learning from implementation of maximum-likelihood estimation for \mathbf{w} and σ^2 , additional constraint on the parameters can be imposed by adding a complexity penalty to the likelihood or error function. A zero-mean Gaussian prior probability distribution over \mathbf{w} is shown in Eq. (9).

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=0}^N \mathcal{N}(w_i|0, \alpha_i^{-1}) \tag{9}$$

where $\boldsymbol{\alpha}$ is a vector of $N + 1$ hyper-parameters.

By utilizing Bayesian posterior inference, the posterior distribution over \mathbf{w} is given as follows.

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)} \tag{10}$$

Eq. (10) can be written as follows

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-(1+N)/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})\right] \tag{11}$$

Here, the posterior covariance $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$ are given as follows

$$\Sigma = (\sigma^{-2}\Phi^T\Phi + \mathbf{A})^{-1} \tag{12}$$

$$\mu = \sigma^{-2} \Sigma \Phi^T \mathbf{t} \tag{13}$$

where $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$.

For the uniform hyperpriors over σ^2 and α , the term $p(\mathbf{t}|\alpha, \sigma^2)$ needs to be maximized and we can get

$$p(\mathbf{t}|\alpha, \sigma^2) = (2\pi)^{-N/2} \left| \sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T \right|^{-1/2} \exp \left[-\frac{1}{2} \mathbf{t}^T (\sigma^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T)^{-1} \mathbf{t} \right] \tag{14}$$

Values of σ^2 and α that maximize Eq. (14) can be obtained iteratively by using the following updating rules.

$$(\alpha_i)^{\text{New}} = \frac{\gamma_i}{\mu_i^2}, \tag{15}$$

$$(\sigma_i^2)^{\text{New}} = \frac{\|\mathbf{t} - \Phi \boldsymbol{\mu}\|^2}{N - \sum_i \gamma_i} \tag{16}$$

where μ_i is the i th element of the estimated posterior weight w . Define the quantities $\gamma_i \equiv 1 - \sigma_i^2 \Sigma_{ii}$ and Σ_{ii} denotes the i th diagonal element of the posterior covariance matrix Σ from Eq. (12).

The maximization of $p(\mathbf{t}|\alpha, \sigma^2)$ is known as the type-II maximum likelihood method [46] and evidence for hyper-parameter [47]. Once the iterative procedure has converged to the most probable values α_{MP} and σ_{MP}^2 , the predictive results for a new set of inputs x_* , can be written as follows

$$\begin{aligned} p(t_* | \mathbf{t}, \alpha_{MP}, \sigma_{MP}^2) &= \int p(t_* | \mathbf{w}, \sigma_{MP}^2) p(\mathbf{w} | \mathbf{t}, \alpha_{MP}, \sigma_{MP}^2) d\mathbf{w} \\ &= N(t_* | y_*, \sigma_*^2) \end{aligned} \tag{17}$$

where $y_* = \mu^T \phi(x_*)$, $\sigma_*^2 = \sigma_{MP}^2 + \phi(x_*)^T \Sigma \phi(x_*)$. The mean y_* denotes the predicted value of RVM at the test point x_* . The variance σ_*^2 can capture the uncertainty of the predicted distribution at the test point x_* . For example, the 95% confidence interval (CI) of the predicted results can be determined by $[y_* - 1.96\sigma_*, y_* + 1.96\sigma_*]$.

2.2.2 Multi-kernel technique

In RVM modeling, there is no constraint over the type of kernel functions (e.g., the kernels do not have to satisfy Mercer condition) [33]. Meanwhile, it is necessary to select the suitable kernel function empirically and determine the appropriate values of hyper-parameters. However, the construction of new kernel functions with high-performance is very complicated, requiring a lot of trials and computing resources. Therefore, using simple mathematical operations on the simple kernel functions to construct a new kernel is a simple and effective way. In this study, different kernel functions are studied for modeling, including three simple kernel functions and three constructed multi-kernel functions using a weighted combination strategy. Gaussian kernel, Polynomial kernel, and Laplace kernel are commonly-used kernel functions. Through the weighted combination of the simple kernel functions, three multi-kernel functions (SumGL kernel, SumGP kernel, and SumLP kernel) are constructed. The kernel functions aforementioned are summarized in Table 1. Where r denotes the hyper-parameters, and the values of hyper-parameters need to be optimized to make sure that the kernel functions can map the data with high performance.

2.2.3 Parallel Jaya algorithm

Jaya algorithm, a powerful and state-of-art optimization algorithm, was proposed by Rao [38]. The advantage of Jaya algorithm is that it contains only one control parameter. Jaya

Table 1 The used simple kernels and the constructed multi-kernels

Kernel function of RVM	Formula	Hyper-parameters	Number of hyper-parameter
Gaussian	$K_G = \exp \left(-\ x_i - x_j\ ^2 / 2r_G^2 \right)$	r_G	1
Laplace	$K_L = \exp \left(-\ x_i - x_j\ / r_L \right)$	r_L	1
Polynomial	$K_P = (r_{p1}(x_i \cdot x_j) + r_{p2})^{r_{p3}}$	$r_{p_i} (i = 1, 2, 3)$	3
SumGL	$K_{GL} = \lambda K_G + (1 - \lambda) K_L$	r_G, r_L, λ	3
SumGP	$K_{GP} = \lambda K_G + (1 - \lambda) K_P$	$r_G, r_{p_i}, \lambda (i = 1, 2, 3)$	5
SumLP	$K_{LP} = \lambda K_L + (1 - \lambda) K_P$	$r_L, r_{p_i}, \lambda (i = 1, 2, 3)$	5

The multi-kernel function sumGL is constructed by combining Gaussian kernel with Laplace kernel linearly. The multi-kernel function sumGP is constructed by combining Gaussian kernel with Polynomial kernel linearly. The multi-kernel function sumLP is constructed by combining Laplace kernel with Polynomial kernel linearly

algorithm was developed based on the idea that the solution obtained (population) for a given problem should get away the worst solution and move towards the best solution. The description of the Jaya algorithm is as follows.

Let $f(X)$ is the function to be optimized. The number of parameters to be determined is m , and the population size is n (population size $k = 1, \dots, n$). Therefore, the total population can be considered a matrix of dimension (m, n) . Let $f(x)_{best}$ is the best value of the objective function produced by the best candidate. The worst candidate can be defined as the worst objective value (i.e., $f(x)_{worst}$). The solution is updated according to the difference between the best candidate and the existing solution as well as the worst candidate. $X_{j,k,i}$ denotes the value of the j th variable for the k th candidate during the i th iteration, then this value is updated by Eq. (6).

$$X'_{j,k,i} = X_{j,k,i} + r_{1j,i} \left(X_{j,best,i} - |X_{j,k,i}| \right) - r_{2j,i} \left(X_{j,worst,i} - |X_{j,k,i}| \right) \tag{18}$$

where $r_{1j,i}$ and $r_{2j,i}$ are two different random numbers, uniformly distributed in the range $[0, 1]$. $X_{j,best,i}$ denotes the value of the j variable for the best candidate, and $X_{j,worst,i}$ denotes the value of the j variable for the worst candidate. The detailed description of Jaya algorithm can be found in [48]. In order to improve the calculation efficiency, the concept of multi-population is introduced to establish the parallel Jaya algorithm (PJA) based on the static multi-population [49]. The population is divided into several sub-populations, and the sub-population structure is performed to parallelize the sequential algorithm. The flowchart of the multipopulation-based PJA is shown in Fig. 1.

2.2.4 Parameters optimization method for RVM using parallel Jaya algorithm

As mentioned above, the kernel function and its hyper-parameter have significant impacts on the performance and sparsity of the RVM-based model. In general, the kernel function and hyper-parameter value are defined before model implementing. In order to obtain the optimal value of hyper-parameter and prevent over-fitting in the validation dataset, the parameters of RVM to be optimized are encoded in the solutions of Jaya algorithm. The solution can be represented as $s = (s_1, s_2, \dots, s_r)$, where s_r denote the parameters of the kernel function and r is the number of the parameters to be optimized. k -fold cross-validation is a popular way to estimate the model generalization performance. In k -fold cross-validation, the selected dataset is equally partitioned into k subsets. In these k subsets, a single subset is used for validation and the remaining $k - 1$ subsets are used for training. The cross-validation method is then carried out k times, with each of the k subsets used exactly once as the validation dataset.

The target function should be defined in a proper form. In this study, the root mean square error (RMSE) of the solution is chosen as the target function, as shown in Eq. (19).

$$F_{RMSE}(s) = \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{1}{N} \sum_{i=1,k}^N (y_i - y(i)_s)^2} \tag{19}$$

where K is the number of subsets, and N is the number of validation samples. y_i denotes the target value, and $y(i)_s$ denotes the predicted value using RVM model. In this paper,

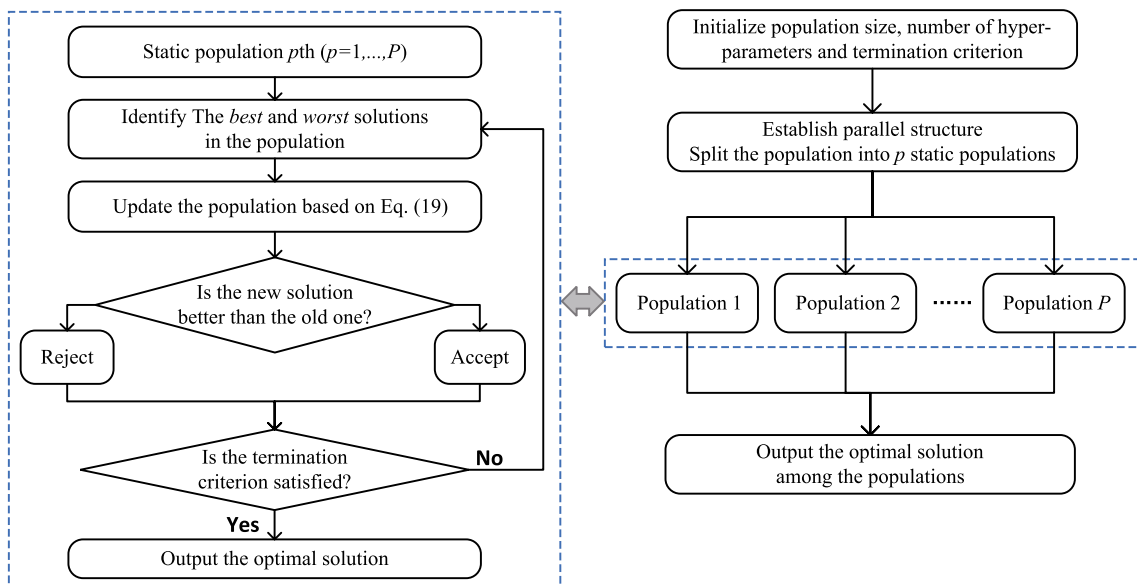


Fig. 1 Flowchart of the multi-population-based parallel Jaya algorithm (PJA)

5-fold cross-validation is carried out for model training. Therefore, K is set to 5.

The adaption of hyper-parameters using PJA contains the following steps:

- (1) Set the population size in the Jaya algorithm, and initialize the kernel function. Calculate the initial solutions by target function shown in Eq. (19).
- (2) Split the population into P sub-populations and build parallel calculation structure. Find the best solution and worst solution in each population.
- (3) In each sub-population, update each solution s_i as a candidate solution c_i by Eq. (19). Calculate the value of the target function by carrying out 5-fold cross-validation.
- (4) For each solution, if $f(c_i) < f(s_i)$, update s_i with c_i ; else, do not update s_i .
- (5) Repeat steps (3) ~ (4), until the maximum number of iterations is reached.
- (6) Record the optimal solution in each sub-population.
- (7) Record the best solution among the optimal solutions obtained in each sub-population. In this manner, the target function value reaches the minimal value.

Combined with the specific monitoring dataset, the optimal hyper-parameter r_{opt} as well as the fitted or predicted outputs can be obtained by performing the above steps.

2.3 Procedure of ORVM for the prediction of concrete dam displacement

As mentioned in Section 2.1, water pressure component, temperature component and time component are selected as independent variables of the model. The displacement is adopted as the dependent variable. It is noted that the initial value should be deducted for the establishment of hydrostatic pressure factors and time factors. Therefore, the input x of the model can be denoted as a vector shown below.

$$x = \left\{ H - H_0, (H - H_0)^2, (H - H_0)^3, (H - H_0)^4, \right. \\ \left. \sin(d), \cos(d), \sin(d)\cos(d), \sin^2(d), t - t_0, \right. \\ \left. (e^{-t_0} - e^{-t}), \ln(t) - \ln(t_0) \right\} \quad (20)$$

where H_0 denotes the water level on the initial monitoring day and t_0 denotes the initial monitoring day. The other symbols have the same meaning as the variables in Eq. (2) ~ Eq. (5).

To eliminate the influence of the dimension on the dataset, the input data are normalized within a range of [0, 1] by

$$f(x_i) = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \quad (21)$$

where x_i represents the value to be normalized. $x_{i\max}$ and $x_{i\min}$ denote the maximum and minimum value of the data to be normalized, respectively.

The flowchart of the proposed ORVM-based probabilistic prediction model for dam displacement is illustrated in Fig. 2, and the main procedure is described as follows.

- (1) Choose the influential components and determinate the displacement.
- (2) Data preparation and normalization. Collect the monitoring data from the dam monitoring system and build the inputs of the model. All the data should be normalized within a range of [0, 1].
- (3) Dataset division. Based on the obtained data, establish the training set and testing test for modeling.
- (4) Optimization of model parameters. Select specific kernel function and determine the value of the hyper-parameters using the adaptive parameters selection method for RVM described in 2.3.3.
- (5) Model establishment. The RVM-based model is built using the training data, the selected kernel function and the optimal value of the hyper-parameters.
- (6) Performance verification. Utilize testing set to verify whether the trained ORVM model has good generalization performance on unknown monitoring data.

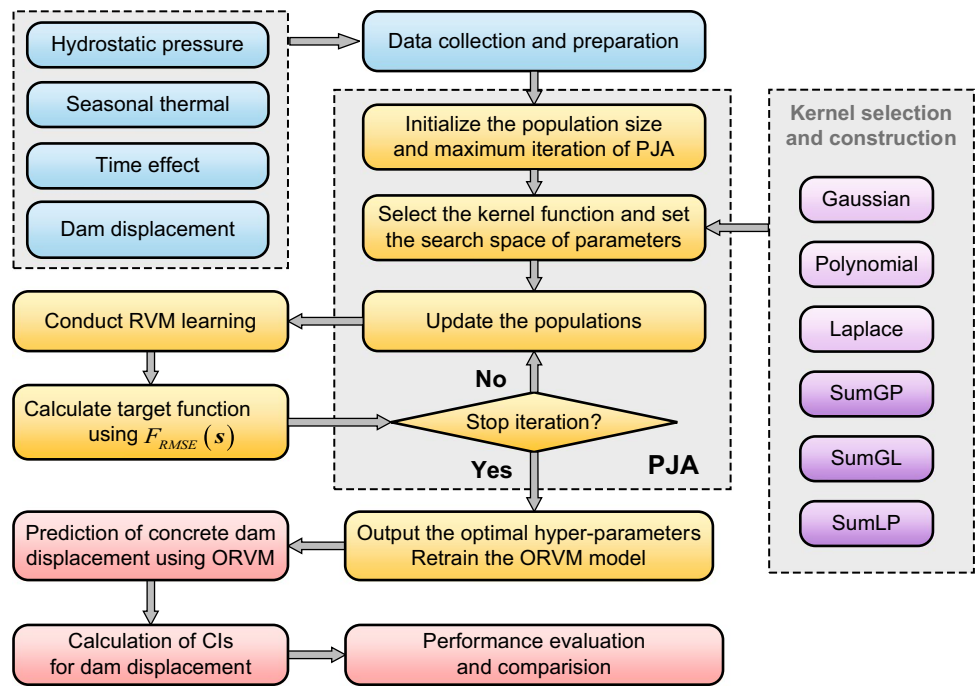
In this study, six statistical metrics are used as the criteria to comprehensively evaluate the predictive performance of the models, including the coefficient of determination (R^2), the root mean square error (RMSE), the mean absolute error (MAE), the maximum absolute error (ME), the average width of confidence interval (AWCI) and the average variance of confidence interval (AVCI), which are expressed in Appendix A. It should be noted that a model is more precise if it contains not only lower RMSE, MAE and ME values, but also the higher R^2 value in both training and testing dataset. AWCI and AVCI reflect the stability and smoothness of the 95% CI obtained by ORVM, respectively. A smaller AWCI represents the more reliable predicted results of ORVM. The smaller AVCI is, the smoother and more stable the 95% CI of ORVM becomes.

3 Application

3.1 Dam engineering profile

Jinping-I hydropower station project is located in Sichuan Province, China. It is mainly composed of a double-curvature concrete arch dam, underground power plant, and water conveyance structures. The dam crest elevation is at 1885m and the maximum height of the arch dam is 305m. The arch dam consists of 26 sections, with a dam crest length

Fig. 2 The flowchart of the ORVM predictive model of concrete dam displacement



of 552m. Jinping-I arch dam is the highest arch dam currently in the world. The dam is equipped with an advanced automatic monitoring system which composed of various instruments, including water level gauges, pendulums, thermometers, strain gauges, osmometers, and piezometers. In this study, the radial displacement measured at the reading station PL13-3 of the central pendulum system is analyzed for modeling. The overlooking of the dam and location of the pendulum in the No. 13 dam section are shown in Fig. 3.

3.2 Data collection and preparation

The radial displacement measurements of monitoring point PL13-3, which are the data implemented in this study, are measured from 1st September 2013 to 7th November 2016 with 380 groups of data samples in total. Figs. 4 and 5 illustrate the time evolution of the measured radial displacement and reservoir water level, respectively. Since there’s no enough continued dam body temperature near PL13-3, the harmonic functions as given in Eq. (4) are used for modeling the temperature effect indirectly. Therefore, a total of 11 factors are selected as independent variables of the models. The measured radial displacement value of monitoring point PL13-3 is selected as the dependent variable.

The first 320 samples that correspond to the period between 3rd September 2013 and 30th July 2016, are used for cross-validation and training. The remaining 60 samples which correspond to a period between 31st July 2016 and 7th October 2016, are utilized as the testing set for evaluating the model performance. The testing set is subdivided into six parts, with the samples of 10, 20, 30, 40, 50, and 60,

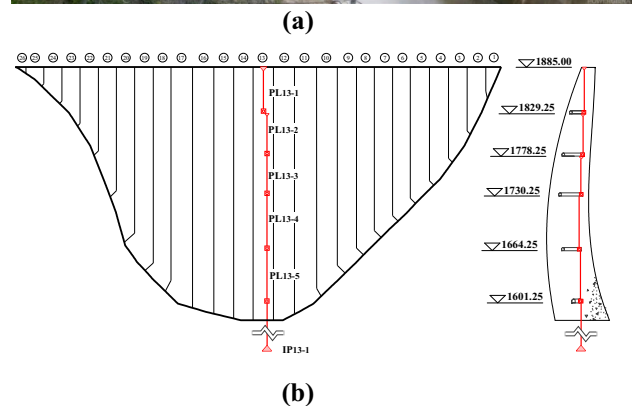


Fig. 3 The concrete arch dam: (a) downstream view; (b) Location of pendulums in No. 13 dam section

Fig. 4 Measured radial displacement

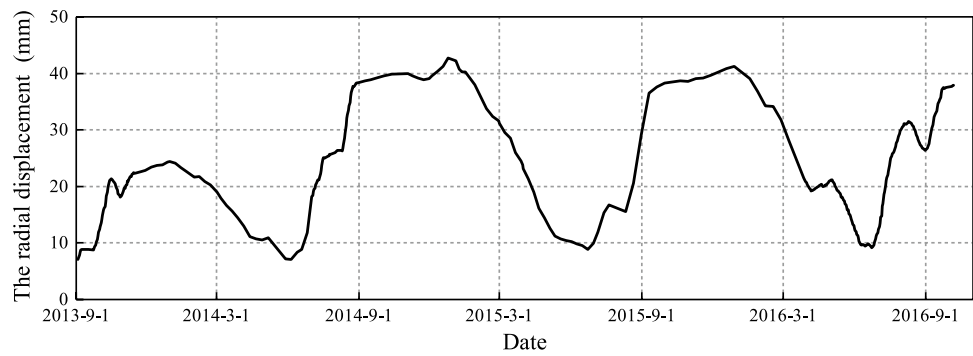


Fig. 5 The reservoir water level

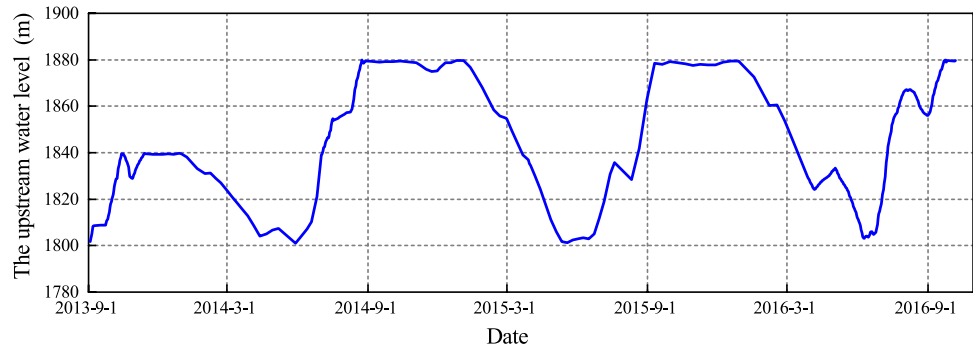


Table 2 Training set and testing test

Dataset no.	Number of samples	Date range
Training set	320	2013.9.3~2016.7.30
Testing set 1	10	2016.7.31~2016.8.9
Testing set 2	20	2016.7.31~2016.8.19
Testing set 3	30	2016.7.31~2016.8.29
Testing set 4	40	2016.7.31~2016.9.8
Testing set 5	50	2016.7.31~2016.9.18
Testing set 6	60	2016.7.31~2016.10.7

respectively. The goal is to test the predictive performance and robustness of RVM for dam displacement. The detailed information on training and testing sets is listed in Table 2. It is noted that the deformation data are weekly recorded rather than daily recorded during the period from September 2014 to March 2016 due to the instrument maintenance and debugging of the monitoring system.

3.3 Performance evaluation of the ORVM models with different kernel functions

For the ORVM model, different simple kernel functions and multi-kernel functions are selected for testing and comparison. The hyper-parameter values of kernel functions are obtained by the optimization method introduced in Section 2.2.4. For the PJA, the population size is set to 40 and the maximum iteration number is set to 150. In order to

establish the parallel calculation structure, the population is divided into 4 parts equally in the multi-population structure and each sub-population has a size of 10.

The search space of the hyper-parameter values and the obtained optimal values of different ORVM models are listed in Table 3. The convergence process curves of simple kernel-based ORVM models and multi-kernel ORVM models are shown in Fig. 6. From the results shown in Fig. 6 and Table 3, it can be observed that PJA has good performance for parameters optimization of RVM as the fitness values of the six models remain stable after 100 iterations. The fitness value of Gaussian kernel-based optimized relevance vector machine model (G-ORVM) is the smallest among three simple kernel-based RVM models, with the fitness value is 0.592. The fitness value of SumGP kernel-based optimized relevance vector machine (GP-ORVM) is 0.566, which is the smallest among the three multi-kernel ORVM models. On the whole, the fitness value of GP-ORVM is the smallest among the six models. Therefore, it could be inferred that the performance of GP-ORVM is the best to some degree.

Take the testing set 3 as an example, four evaluation metrics of ORVM models with different kernel functions are listed in Table 4, and the statistically superior results are shown in boldface. In general, the six ORVM models with different kernel functions achieve satisfactory performance since the coefficients of determination in the training set and testing set are larger than 0.95. The three ORVM models with multi-kernel function also provide reasonably good results especially in the testing set, where the GP-ORVM

Table 3 Search space of the RVM hyper-parameters

Kernel function	Search range of control parameters	Obtained hyper-parameters	Minimum fitness value
Gaussian	$r_G \in [0.01, 50]$	$r_G = 2.87$	0.592
Polynomial	$\{r_{P1}, r_{P2}\} \in [0.01, 50], r_{P3} \in [1, 5]$	$(r_{P1}, r_{P2}, r_{P3}) = (4.37, 25.56, 1.75)$	0.633
Laplace	$r_L \in [0.01, 50]$	$r_L = 1.90$	0.615
SumGP	$\{r_G, r_{P1}, r_{P2}\} \in [0.01, 50], r_{P3} \in [1, 5], \lambda \in (0, 1)$	$(r_G, r_{P1}, r_{P2}, r_{P3}, \lambda) = (2.43, 9.34, 6.16, 1.01, 0.97)$	0.566
SumGL	$\{r_G, r_L\} \in [0.01, 50], \lambda \in (0, 1)$	$(r_G, r_L, \lambda) = (12.81, 1.99, 0.18)$	0.606
SumLP	$\{r_L, r_{P1}, r_{P2}\} \in [0.01, 50], r_{P3} \in [1, 5], \lambda \in (0, 1)$	$(r_L, r_{P1}, r_{P2}, r_{P3}, \lambda) = (1.63, 16.76, 15.21, 2.12, 0.32)$	0.588

Best fitness value is highlighted in boldface

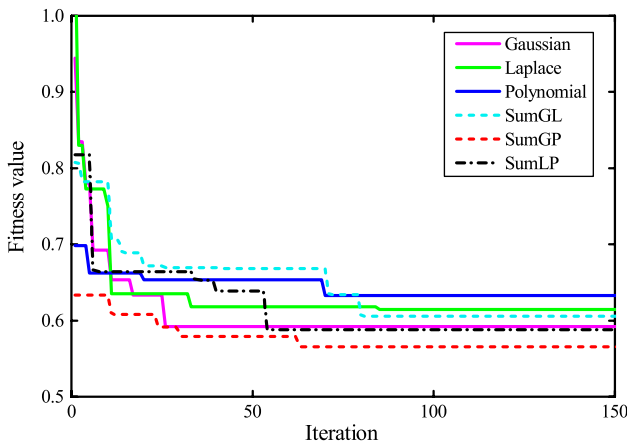


Fig. 6 Convergence characteristics of RVM with different kernels

model has the smallest RMSE, MAE and ME values with the values are 0.2765, 0.2441 and 0.4726, respectively. As for the ORVM models with simple kernel function, the G-ORVM model provides the best predictive performance. Overall, it can be seen from Table 4 that the GP-ORVM model and the G-ORVM model perform better than the other four models considering the generalization and predictive accuracy.

Figs. 7 and 8 depict the advantage of probabilistic prediction for concrete dam displacement using GP-ORVM and G-ORVM models, where the confidence level is set to 95%

and the blue line denotes the measured displacement. In the training sets, the upper and lower bounds of 95% CI get closer to the measured displacement. Except for individual peak points, most of the measured displacement fall into the 95% CIs. In the testing sets, although all the measured displacement falls into the 95% CIs, the GP-ORVM seems to provide a narrower CI than G-ORVM. The relatively narrower CI is significant since it is more sensitive and can capture the abnormal displacement data effectively.

To test the probabilistic prediction performance of GP-ORVM and G-ORVM on different data, six testing sets are adopted for simulation and the calculated values of AWCI and AVCI on different testing sets are shown in Fig. 9. It can be seen that the AWCI values of the GP-ORVM model fluctuate around 1.5mm and are slightly smaller than those of the G-ORVM model, which means the predicted 95% CI using GP-ORVM get a certain degree of compression and the predicted results are more reliable. The AVCI values of both two ORVM models are small with values within the range from 0.08 to 0.24. However, the AVCI values of the GP-RVM model are significantly smaller than those of the G-RVM model, which reflects that the 95% CI of the GP-ORVM is more stable and smoother. The advantage of the smooth CI is that it can improve the reliability of anomaly recognition.

Table 4 The performance of ORVM models based on different kernel functions

Kernel function	Training set				Testing set			
	R^2	RMSE (mm)	MAE (mm)	ME (mm)	R^2	RMSE (mm)	MAE (mm)	ME (mm)
Gaussian	0.9986	0.3328	0.2457	1.1711	0.9874	0.3147	0.2668	0.6173
Polynomial	0.9984	0.3630	0.2661	1.4080	0.9661	0.4927	0.4169	0.8559
Laplace	0.9993	0.2119	0.1537	0.9331	0.9662	0.6238	0.5559	1.0245
SumGP	0.9990	0.2767	0.2042	1.0461	0.9891	0.2765	0.2441	0.4726
SumGL	0.9994	0.2271	0.1669	0.9626	0.9662	0.6048	0.5354	1.0391
SumLP	0.9990	0.1850	0.1327	0.8599	0.9697	0.6549	0.5957	0.9929

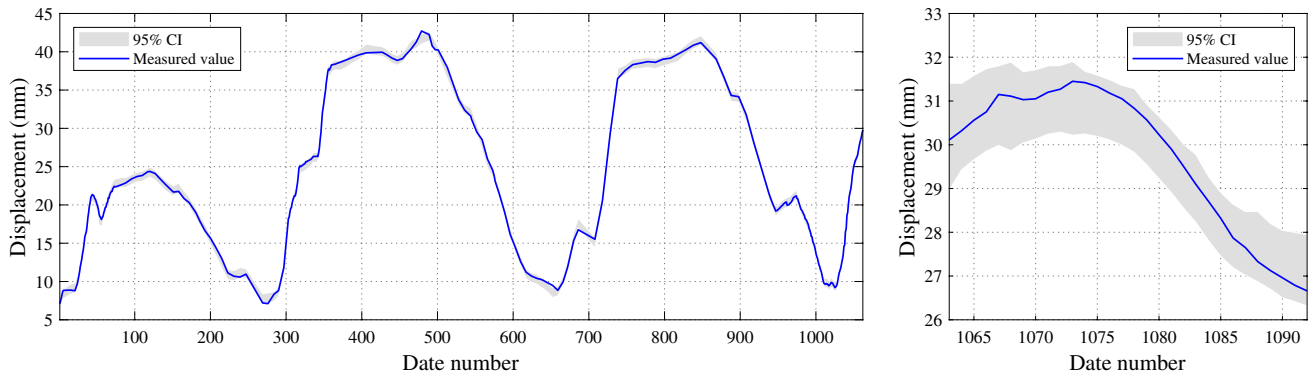


Fig. 7 The output results of the GP-ORVM-based prediction model with 95% CI

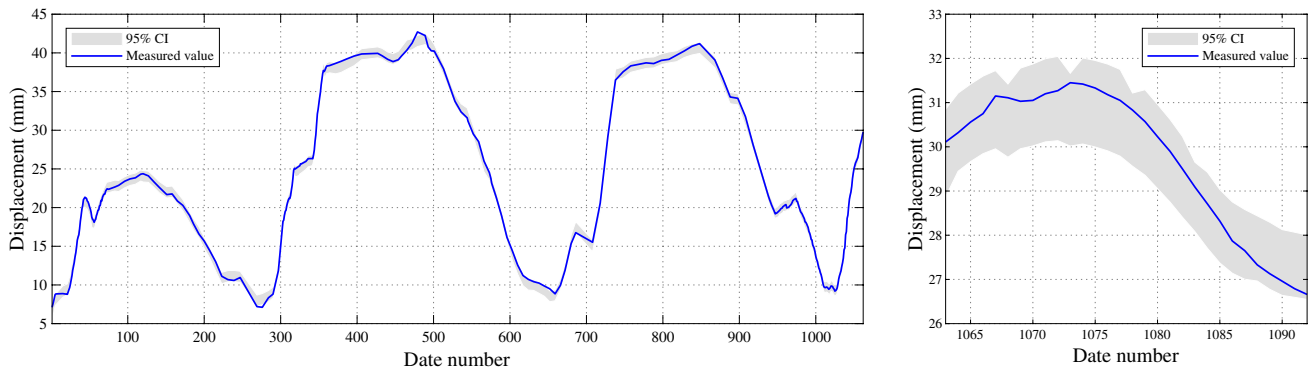


Fig. 8 The output results of the G-ORVM-based prediction model with 95% CI

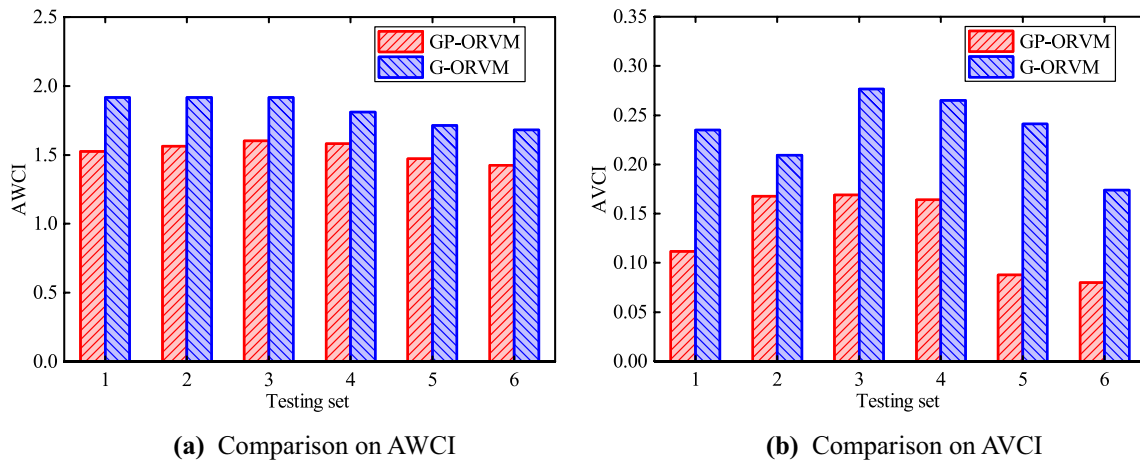


Fig. 9 Performance evaluation of the 95% CIs obtained by GP-ORVM and G-ORVM

3.4 Performance comparison of the existing models

In this section, the RBF-NN, the SVR, ELM and the HST-based multiple linear regression (MLR-HST) models are selected as benchmark models for performance comparison

with the proposed ORVM models. For the MLR-HST model, the regression coefficients are computed by the least square method. For the RBF-NN, SVR and ELM models, the hyper-parameters are determined in an optimal manner rigorously, where 5-fold cross-validation and PJA are performed to

estimate the hyper-parameters, and the target function is the same as Eq. (19).

In the RBF-NN model, the value of spread S_R and the number of neurons N_R in the hidden layer are parameters to be optimized. The training objective of the mean square error is set to 10^{-4} . The value of spread S_R is selected in the interval of [0.01, 100], and the number of hidden layer nodes N_R is selected in the interval of [11, 50]. The optimal values of the spread and number of hidden layer nodes obtained are 7.99 and 14, respectively.

For the ELM model, the sigmoidal function is chosen as activation function. The number of hidden layer nodes N_E is selected in the interval of [11, 50] and the calculated optimal value is 14. Note that the result is obtained based on the average performance from fifty continuous training to reduce the uncertainty.

As for the SVR model, the penalty factor C , kernel parameter σ' and insensitive loss function ϵ are three control parameters. The penalty factor is selected in the interval of [0.01, 100]. The Gaussian kernel is selected as the kernel function of the SVR model, and the value of the kernel parameter is selected in the interval of [0.01, 50]. The insensitive loss function is selected in the interval of [0.001, 0.1]. The optimal values of the penalty factor, the kernel parameter and the insensitive loss function obtained are 2.82, 0.09 and 0.038, respectively.

The search range of the control parameters and the obtained optimal parameter values of the G-ORVM, GP-ORVM, SVR, RBF-NN and ELM models are summarized in Table 5.

In the same manner, take the testing set 3 as an example, a detailed comparison on the prediction performance of the six models is carried out. Evaluation metrics of the fitted and predicted results are listed in Table 6 and the best results are shown in boldface. The fitted and predicted results of these six models are shown in Fig. 10. It can be observed that the R^2 values of the fitted results computed by six models are all approximate to 1.0, which reflects that these six models have satisfactory fitting performance. The RMSE, MAE and ME values of predictive results using GP-ORVM model is the smallest, which indicates the perfect predictive performance of the GP-ORVM model. From Fig. 10 it can be seen that the residual of the predicted results using GP-ORVM and HST-MLR models at peak value (e.g. data on 2016-8-10) is significantly smaller than other models.

In the training period, the GP-ORVM model utilizes 6.56 % of training data as relevance vectors while the G-ORVM model uses 5.94 % of training data as relevance vectors. Compared with the RVM models, the number of support vectors required is near the size of the training set. The developed RVM models can obtain a more sparse solution with very few relevant vectors, namely, most of the α_i tend to infinity and corresponding $w_i = 0$. Therefore, the possibility of overtraining as well as the computational time is minimized.

In order to evaluate the performance of the proposed ORVM models objectively, a detailed comparison using different unknown testing data is made. Bar charts of the predictive performance for these models under six testing sets are shown in Fig. 11, and the evaluation metrics are listed in Table 7. It is noticeable that the performance of

Table 5 Search range of the control parameters and the obtained optimal values of different models

Model	Search range of control parameters	Optimal parameters
G-ORVM	$r_G \in [0.01, 50]$	$r_G = 2.87$
GP-ORVM	$\{r_G, r_{P1}, r_{P2}\} \in [0.01, 50], r_{P3} \in [1, 5], \lambda \in (0, 1)$	$(r_G, r_{P1}, r_{P2}, r_{P3}, \lambda) = (2.43, 9.34, 6.16, 1.01, 0.97)$
SVR	$C \in [1, 100], \sigma' \in [0.01, 50], \epsilon \in [0.001, 0.1]$	$(C, \sigma', \epsilon) = (2.82, 0.09, 0.038)$
RBF-NN	$S_R \in [0.01, 100], N_R \in [11, 50]$	$(S_R, N_R) = (7.99, 14)$
ELM	$N_E \in [11, 50]$	$N_E = 14$

Table 6 Performance of different models using the testing set 3

Model	Training set				Testing set			
	R^2	RMSE	MAE	ME	R^2	RMSE	MAE	ME
GP-ORVM	0.9990	0.2767	0.2042	1.0461	0.9891	0.2765	0.2441	0.4726
G-ORVM	0.9986	0.3328	0.2457	1.1711	0.9874	0.3147	0.2668	0.6173
RBF-NN	0.9970	0.4919	0.3570	1.9685	0.9749	0.4195	0.3608	0.6094
SVR	0.9994	0.2200	0.1248	1.0576	0.9588	0.7718	0.6988	1.0902
ELM	0.9974	0.4669	0.3528	1.9568	0.9816	0.3099	0.2791	0.4814
MLR-HST	0.9964	0.5372	0.4225	1.9611	0.9796	0.2874	0.2515	0.6146

GP-ORVM and G-ORVM models are more robust and reliable than the other models regardless of the size of testing sets. The GP-ORVM model has the smallest RMSE among the models on testing set 1 and testing set 3~6, which are 0.165, 0.277, 0.329, 0.466, 0.501, respectively. In testing set 2, the RMSE value of GP-ORVM is 0.267, which is the second lowest and only slightly higher than the 0.248 and 0.259 values for the MLR-HST and G-ORVM models. The GP-ORVM model has the smallest MAE value on testing set 1 and testing set 3~5 with the values are 0.129, 0.203, 0.244, 0.286, 0.378, respectively. In testing set 2, the MAE value of GP-ORVM is the second lowest with the value is 0.202. It could not be neglected that with the increase of the testing samples, the predictive precision of the MLR-HST model decreases obviously, even if it has good performance in small test samples (e.g. testing set 1~ testing set 3). For SVR and RBF-NN models, the variation trend of the predictive performance does not have the characteristics of monotonous, however, the model performance is not satisfactory in the short-term prediction of concrete dam displacement.

Overall, the predictive performance of the two ORVM-based models is satisfactory. The GP-ORVM model has the minimum values of average RMSE and MAE, which indicates that the GP-ORVM is the most effective model among the listed models. The G-ORVM model has the similar performance to that of the GP-ORVM model as its average RMSE and MAE values of different testing data are only slightly higher than those of the GP-ORVM model.

4 Conclusion and future work

In this study, a novel probabilistic prediction model of concrete dam displacement is presented to build the structural health monitoring framework and mine the effects of hydrostatic, seasonal and irreversible time components on dam deformation. The model is combined with RVM, multi-kernel technique, HST statistical model and PJA using initial service life monitoring measurements collected from a super-high concrete arch dam. The proposed parameters optimization method is verified to be effective for RVM to achieve accurate and robust predictions. Different kernel functions, such as Gaussian kernel, Laplace kernel, Polynomial kernel, and the multi-kernels of their weighted combination are also exploited to build the ORVM to verify their impact on predictive performance. The main conclusions and contributions are summarized as follows:

- The developed ORVM model is suitable for the prediction of non-stationary and nonlinear concrete dam displacement since it provides satisfactory performance both on training and testing sets. The proposed optimization framework for parameters estimation using PJA

can optimize the hyper-parameters of RVM effectively and avoid falling into local optimum, which improves the model predictive performance and robustness.

- The kernel functions and hyper-parameters have significant impacts on the performance of the RVM model. The results suggest that the weighted combination strategy for multi-kernel construction is feasible, and the multi-kernel ORVM models have superior performance than the simple kernel ORVM models.
- Compared with the listed benchmark models, the ORVM is proved to be robust and effective for establishing the dam health monitoring model in predicting the concrete dam displacement. The developed ORVM models with SumGP kernel and Gaussian kernel perform superior to the optimized SVR, RBF-NN, ELM and HST-MLR models in most of the testing sets, reducing the residual. In addition, the developed ORVM is sparser than SVM.
- The developed ORVM model not only obtains the most accurate results in the single point prediction of dam displacement measurements, but also provides probabilistic CIs, which can be used to quantify the uncertainty and identify the abnormal value of dam displacement. In addition, the multi-kernel ORVM can compress and smooth the CI to polish up the reliability of displacement anomaly recognition.

For the future work, the proposed model can be adopted for analysis and prediction of other monitoring measurements in concrete dam engineering, such as tangential displacement, settlement, and seepage. Besides, future studies should involve analyzing and developing the multi-output ORVM model in order to solve the high dimensional regression task and provide more reliable prediction and identification of dam spatial deformation.

Acknowledgements The authors are grateful to the financial sponsorship from National Natural Science Foundation of China (Grant Nos. 51739003, 51779086), National Key R&D Program of China (2018YFC0407104, 2016YFC0401601), Special Project Funded of National Key Laboratory (20165042112) and Key R&D Program of Guangxi (AB17195074).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest regarding the publication of this paper.

Appendix

$$R^2 = \frac{\left[\sum_{i=1}^N (y_S(i) - \bar{y}_S)(y(i) - \bar{y}) \right]^2}{\sum_{i=1}^N (y_S(i) - \bar{y}_S)^2 \sum_{i=1}^N (y(i) - \bar{y})^2} \quad (\text{A.1})$$

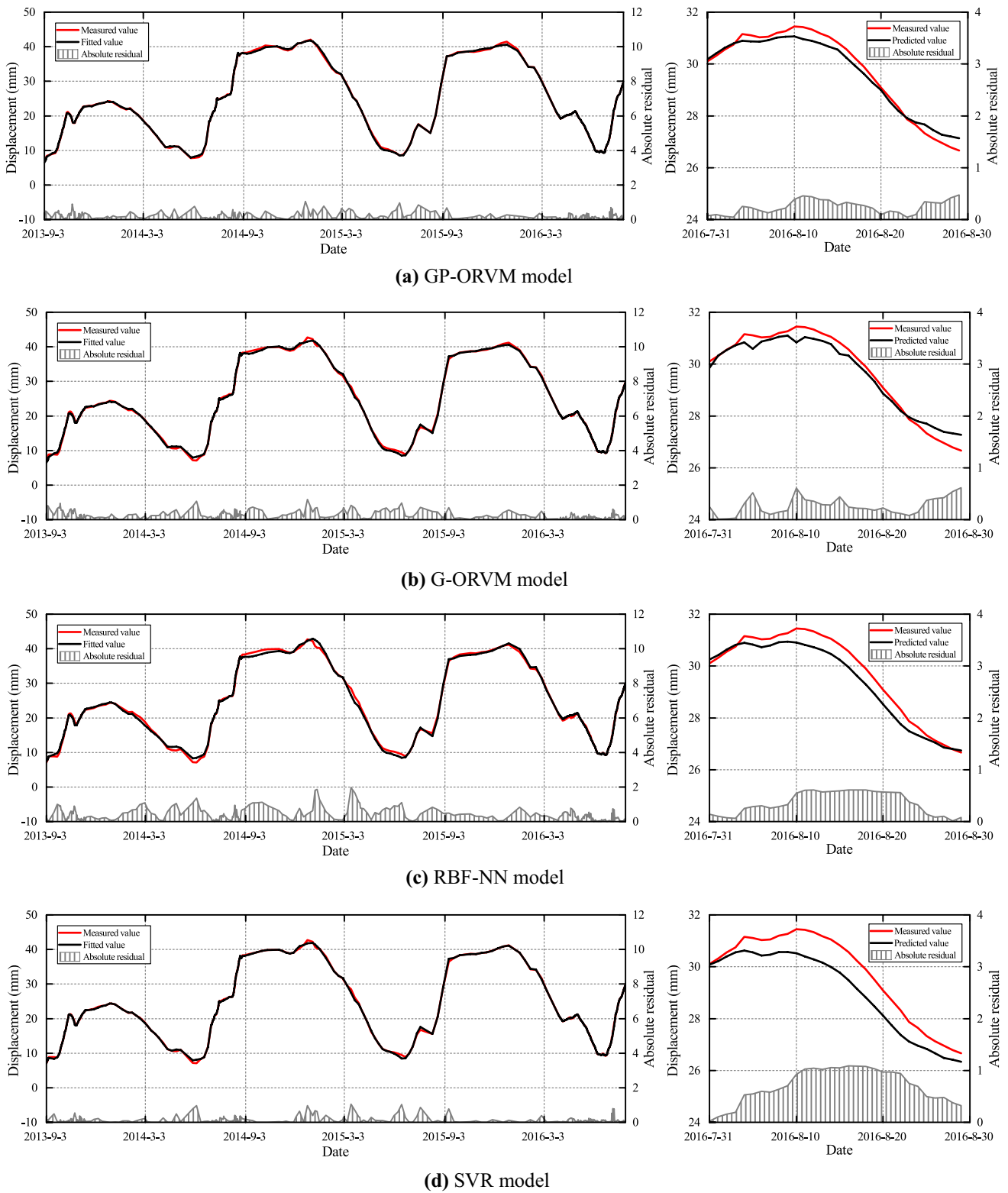
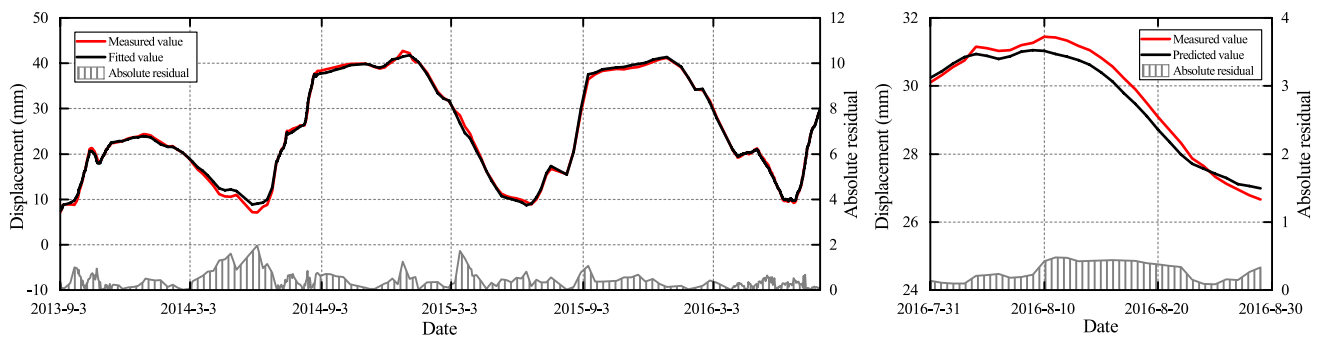
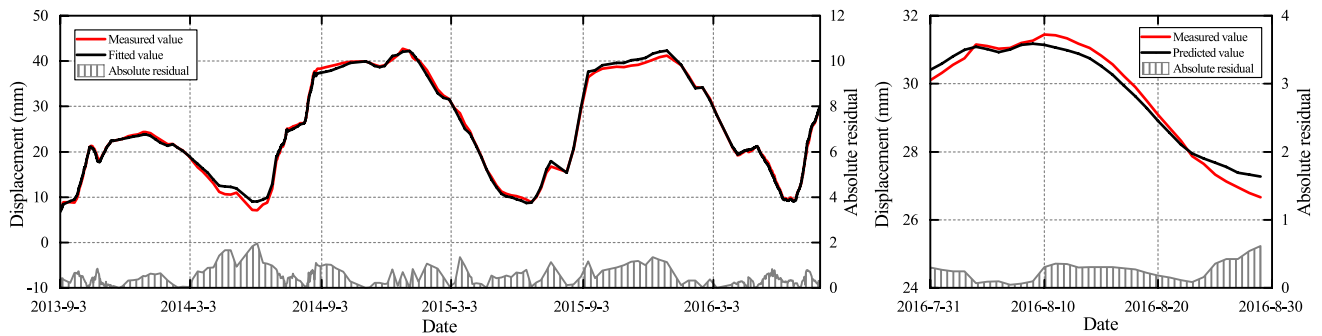


Fig. 10 Performance comparison of the six models



(e) ELM model



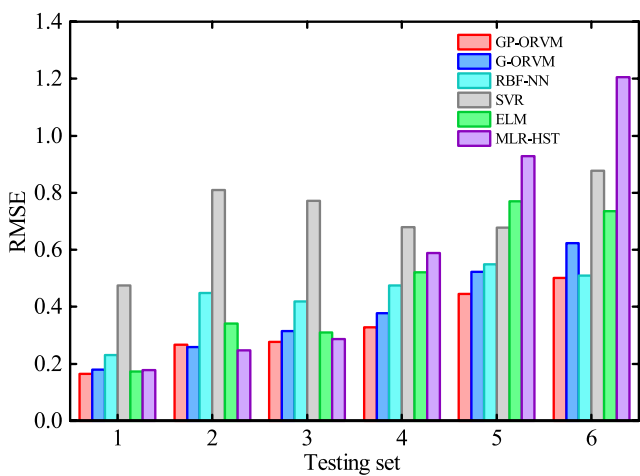
(f) MLR-HST model

Fig. 10 (continued)

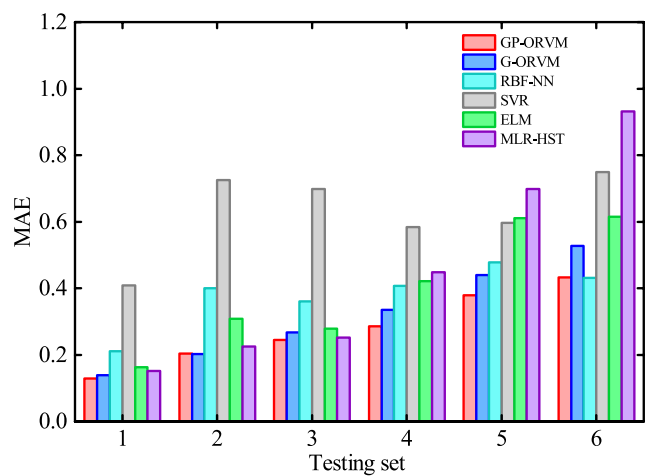
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_S(i) - y(i))^2} \tag{A.2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_S(i) - y(i)| \tag{A.3}$$

$$ME = \max |y_S(i) - y(i)| \tag{A.4}$$



(a) Comparison on RMSE



(b) Comparison on MAE

Fig. 11 Performance evaluation of the six models for dam displacement prediction

Table 7 Comparison of the RMSE and MAE values in six testing sets

Testing set no.	1	2	3	4	5	6	Average
<i>Model</i>							
RMSE							
GP-ORVM	0.165	0.267	0.277	0.329	0.446	0.501	0.331
G-ORVM	0.180	0.259	0.315	0.377	0.523	0.624	0.380
RBF-NN	0.231	0.449	0.419	0.475	0.549	0.509	0.439
SVR	0.475	0.809	0.772	0.680	0.677	0.877	0.715
ELM	0.174	0.342	0.310	0.521	0.771	0.736	0.476
MLR-HST	0.178	0.248	0.287	0.589	0.928	1.206	0.572
MAE							
GP-ORVM	0.129	0.203	0.244	0.286	0.378	0.433	0.279
G-ORVM	0.138	0.202	0.267	0.335	0.439	0.528	0.318
RBF-NN	0.210	0.401	0.361	0.407	0.478	0.431	0.381
SVR	0.409	0.725	0.699	0.584	0.596	0.749	0.627
ELM	0.163	0.308	0.279	0.421	0.611	0.615	0.400
MLR-HST	0.151	0.225	0.251	0.449	0.699	0.932	0.451

$$AWCI = \frac{1}{N} \sum_{i=1}^N 2 \times 1.96 \times \sqrt{\sigma_i^2} \quad (A.5)$$

$$AVCI = \frac{1}{N} \sum_{i=1}^N \left| 2 \times 1.96 \times \sqrt{\sigma_i^2} - AWCI \right| \quad (A.6)$$

where $y_S(i)$ and $y(i)$ denote the model output and measured values of the radial displacement, respectively ($i = 1, 2, \dots, N$); \bar{y}_S and \bar{y} represent the average of the model output and measured values, respectively; N represents the number of observations. σ_i is the predicted variance of ORVM-based prediction model at output point $y_S(i)$.

References

- Wu ZR (2003) Safety monitoring theory and its application of hydraulic structures. Higher Education, Beijing
- Zhao EF (2018) Dam Safety Monitoring Data Analysis Theory & Assessment Methods. Hohai University Press,
- Shi YQ, Yang JJ, Wu JL, He JP (2018) A statistical model of deformation during the construction of a concrete face rock-fill dam. Structural Control & Health Monitoring. <https://doi.org/10.1002/stc.2074>
- Gu CS, Wu ZR (2006) Safety monitoring of dams and dam foundations-theories & methods and their application. Hohai University Press,
- Salazar F, Toledo MA, Onate E, Moran R (2015) An empirical comparison of machine learning techniques for dam behaviour modelling. Struct Saf 56:9–17. <https://doi.org/10.1016/j.strusafe.2015.05.001>
- Salazar F, Morán R, Toledo MA, Oñate E (2015) Data-Based Models for the Prediction of Dam Behaviour: A Review and Some Methodological Considerations. Archives of Computational Methods in Engineering 24(1):1–21. <https://doi.org/10.1007/s11831-015-9157-9>
- Mata J, de Castro AT, da Costa JS (2014) Constructing statistical models for arch dam deformation. Structural Control & Health Monitoring 21(3):423–437. <https://doi.org/10.1002/stc.1575>
- Lin CN, Li TC, Liu XQ, Zhao LH, Chen SY, Qi HJ (2019) A deformation separation method for gravity dam body and foundation based on the observed displacements. Structural Control & Health Monitoring. <https://doi.org/10.1002/stc.2304>
- Sun PM, Bao TF, Gu CS, Jiang M, Wang T, Shi ZW (2016) Parameter sensitivity and inversion analysis of a concrete faced rock-fill dam based on HS-BPNN algorithm. Science China-Technological Sciences 59(9):1442–1451. <https://doi.org/10.1007/s11431-016-0213-y>
- Mata J (2011) Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. Engineering Structures 33(3):903–910. <https://doi.org/10.1016/j.engstruct.2010.12.011>
- Stojanovic B, Milivojevic M, Ivanovic M, Milivojevic N, Divac D (2013) Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. Advances in Engineering Software 65:182–190. <https://doi.org/10.1016/j.advengsoft.2013.06.019>
- Gu CS, Li B, Xu GL, Yu H (2010) Back analysis of mechanical parameters of roller compacted concrete dam. Science China-Technological Sciences 53(3):848–853. <https://doi.org/10.1007/s11431-010-0053-0>
- Xi GY, Yue JP, Zhou BX, Tang P (2011) Application of an artificial immune algorithm on a statistical model of dam displacement. Computers & Mathematics with Applications 62(10):3980–3986. <https://doi.org/10.1016/j.camwa.2011.09.057>
- Gu CS, Wang YC, Peng Y, Xu BS (2011) Ill-conditioned problems of dam safety monitoring models and their processing methods. Science China-Technological Sciences 54(12):3275–3280. <https://doi.org/10.1007/s11431-011-4573-z>
- Hariri-Ardebili MA, Pourkamali-Anaraki F (2018) Simplified reliability analysis of multi hazard risk in gravity dams via machine learning techniques. Arch Civ Mech Eng 18(2):592–610. <https://doi.org/10.1016/j.acme.2017.09.003>
- Hariri-Ardebili MA, Pourkamali-Anaraki F (2018) Support vector machine based reliability analysis of concrete dams. Soil Dynamics and Earthquake Engineering 104:276–295. <https://doi.org/10.1016/j.soildyn.2017.09.016>

17. Hariri-Ardebili MA, Barak S (2019) A series of forecasting models for seismic evaluation of dams based on ground motion meta-features. *Engineering Structures*. <https://doi.org/10.1016/j.engstruct.2019.109657>
18. Hariri-Ardebili MA, Pourkamali-Anaraki F (2019) Matrix completion for cost reduction in finite element simulations under hybrid uncertainties. *Applied Mathematical Modelling* 69:164–180. <https://doi.org/10.1016/j.apm.2018.12.014>
19. Hariri-Ardebili MA, Sudret B (2019) Polynomial chaos expansion for uncertainty quantification of dam engineering problems. *Engineering Structures*. <https://doi.org/10.1016/j.engstruct.2019.109631>
20. Moody J, Darken CJ (1989) Fast Learning in Networks of Locally-Tuned Processing Units. *Neural Computation* 1(2):281–294. <https://doi.org/10.1162/neco.1989.1.2.281>
21. Vapnik V, Golowich SE, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. *Adv Neur In* 9:281–287
22. Chen SY, Gu CS, Lin CN, Zhao EF, Song JT (2018) Safety Monitoring Model of a Super-High Concrete Dam by Using RBF Neural Network Coupled with Kernel Principal Component Analysis. *Mathematical Problems in Engineering* 2018:1–13. <https://doi.org/10.1155/2018/1712653>
23. Kang F, Li JJ, Zhao SZ, Wang YJ (2019) Structural health monitoring of concrete dams using long-term air temperature for thermal effect simulation. *Engineering Structures* 180:642–653. <https://doi.org/10.1016/j.engstruct.2018.11.065>
24. Kang F, Liu J, Li JJ, Li SJ (2017) Concrete dam deformation prediction model for health monitoring based on extreme learning machine. *Structural Control & Health Monitoring*. <https://doi.org/10.1002/stc.1997>
25. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: Theory and applications. *Neurocomputing* 70(1–3):489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
26. Liu CG, Gu CS, Chen B (2017) Zoned elasticity modulus inversion analysis method of a high arch dam based on unconstrained Lagrange support vector regression (support vector regression arch dam). *Engineering with Computers* 33(3):443–456. <https://doi.org/10.1007/s00366-016-0483-9>
27. Su HZ, Chen ZX, Wen ZP (2016) Performance improvement method of support vector machine-based model monitoring dam safety. *Structural Control & Health Monitoring* 23(2):252–266. <https://doi.org/10.1002/stc.1767>
28. Rankovic V, Grujovic N, Divac D, Milivojevic N (2014) Development of support vector regression identification model for prediction of dam structural behaviour. *Struct Saf* 48:33–39. <https://doi.org/10.1016/j.strusafe.2014.02.004>
29. Bui K-TT, Tien Bui D, Zou J, Van Doan C, Revhaug I (2016) A novel hybrid artificial intelligent approach based on neural fuzzy inference model and particle swarm optimization for horizontal displacement modeling of hydropower dam. *Neural Computing and Applications* 29(12):1495–1506. <https://doi.org/10.1007/s00521-016-2666-0>
30. Kang F, Liu X, Li J (2019) Concrete Dam Behavior Prediction Using Multivariate Adaptive Regression Splines with Measured Air Temperature. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-019-04095-z>
31. Lin CN, Li TC, Chen SY, Liu XQ, Lin C, Liang SL (2019) Gaussian process regression-based forecasting model of dam deformation. *Neural Comput Appl* 31(12):8503–8518. <https://doi.org/10.1007/s00521-019-04375-7>
32. Kang F, Li JJ (2019) Displacement Model for Concrete Dam Safety Monitoring via Gaussian Process Regression Considering Extreme Air Temperature. *Journal of Structural Engineering* 146(1):05019001
33. Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 1(3):211–244. <https://doi.org/10.1162/15324430152748236>
34. Imani M, Kao HC, Lan WH, Kuo CY (2018) Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine and relevance vector machine. *Global Planet Change* 161:211–221. <https://doi.org/10.1016/j.gloplacha.2017.12.018>
35. Zhang ZF, Liu ZB, Zheng LF, Zhang Y (2014) Development of an adaptive relevance vector machine approach for slope stability inference. *Neural Comput Appl* 25(7–8):2025–2035. <https://doi.org/10.1007/s00521-014-1690-1>
36. Wang TZ, Xu H, Han JG, Elbouchikhi E, Benbouzid MEH (2015) Cascaded H-Bridge Multilevel Inverter System Fault Diagnosis Using a PCA and Multiclass Relevance Vector Machine Approach. *Ieee T Power Electr* 30(12):7006–7018. <https://doi.org/10.1109/Tpel.2015.2393373>
37. Kong DD, Chen YJ, Li N, Duan CQ, Lu LX, Chen DX (2019) Relevance vector machine for tool wear prediction. *Mechanical Systems and Signal Processing* 127:573–594. <https://doi.org/10.1016/j.ymsp.2019.03.023>
38. Rao R (2016) Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International Journal of Industrial Engineering Computations* 7(1):19–34
39. Holland JH (1975) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, Ann Arbor
40. Farmer JD, Packard NH, Perelson AS (1986) The Immune-System, Adaptation, and Machine Learning. *Physica D* 22(1–3):187–204. [https://doi.org/10.1016/0167-2789\(86\)90240-X](https://doi.org/10.1016/0167-2789(86)90240-X)
41. Eberhart R, Kennedy J A new optimizer using particle swarm theory. In: *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995*. Ieee, pp 39–43
42. Li XL (2003) *A new intelligent optimization-artificial fish swarm algorithm*. PhD Dissertation, Zhejiang University
43. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization* 39(3):459–471
44. Ding ZH, Li J, Hao H (2019) Structural damage identification using improved Jaya algorithm based on sparse regularization and Bayesian inference. *Mechanical Systems and Signal Processing* 132:211–231. <https://doi.org/10.1016/j.ymsp.2019.06.029>
45. Abhishek K, Kumar VR, Datta S, Mahapatra SS (2017) Application of JAYA algorithm for the optimization of machining performance characteristics during the turning of CFRP (epoxy) composites: comparison with TLBO, GA, and ICA. *Engineering with Computers* 33(3):457–475. <https://doi.org/10.1007/s00366-016-0484-8>
46. Berger JO (2013) *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media,
47. MacKay DJJNc (1992) Bayesian interpolation. *4 (3):415–447*
48. Rao R, Waghmare GG (2017) A new optimization algorithm for solving complex constrained design optimization problems. *Engineering Optimization* 49(1):60–83
49. Migallon H, Jimeno-Morenilla A, Sanchez-Romero JL, Rico H, Rao RV (2019) Multipopulation-based multi-level parallel enhanced Jaya algorithms. *J Supercomput* 75(3):1697–1716. <https://doi.org/10.1007/s11227-019-02759-z>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.