



Prediction and interpretation of the deformation behaviour of high arch dams based on a measured temperature field

Shaowei Wang^{1,2,3} · Xupeng Sui¹ · Yi Liu^{2,3} · Hao Gu⁴ · Bo Xu⁵ · Qun Xia¹

Received: 5 July 2022 / Accepted: 2 January 2023 / Published online: 13 January 2023
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

To conduct structural health monitoring, it is important to establish mathematical monitoring models with good prediction and interpretation performance. Generally, the thermal deformation effect interpreted in a displacement monitoring model of concrete dams is represented by the seasonal harmonic factor. The main purpose of this paper is to replace this factor with directly measured dam body temperatures. This approach is conducted by optimally selecting a very small number of the most representative members from hundreds of temperature monitoring points on the dam body, during which the importance of modelling factors is evaluated by a support vector machine (SVM), and automatic and manual criteria are formulated to eliminate the large number of unimportant or similar temperature monitoring points. Then, a hysteretic effect-considered hydraulic, exponential, thermal and time (HETT) model is established, and a causal interpretation of dam deformation behaviour is carried out using a partial dependence diagram to separate displacement components from the SVM model. The world's highest concrete dam, the Jinping I arch dam, is assessed in a case study. Research results demonstrate the efficiency and rationality of the proposed approach. On average, among the 140 total temperature monitoring points in the central cantilever dam section, only 6.3% of them are selected as temperature deformation factors, and these factors can fully characterise the temporal and spatial evolution characteristics of the measured dam temperature field. The prediction accuracy of the HETT model is significantly improved, in which the mean square error, maximum error and correlation coefficient of multiple displacement monitoring points are 60.1%, 86.5% and 101.5% of those of the hydraulic, exponential, seasonal and time (HEST) model, respectively. The thermal deformation effect interpreted by the HETT model is more in accordance with the actual operation condition of the Jinping I arch dam.

Keywords High arch dams · HETT model · Measured temperature deformation factor · Support vector machine · Partial dependence diagram-based causal interpretation

✉ Yi Liu
liuyi@iwhr.com

¹ School of Urban Construction, Changzhou University, Changzhou 213164, China

² Key Laboratory of Construction and Safety of Water Engineering of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

³ State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

⁴ State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

⁵ College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou 225009, China

1 Introduction

Dams are indispensable infrastructure components used in flood control, power generation, irrigation and shipping. With improvements in engineering technology, the scale and maximum dam height of concrete dams have increased rapidly in recent years. The Jinping I arch dam is the highest concrete dam built in the world, with a maximum dam height of 305 m. However, dams bear complex static and dynamic loads and long-term environmental erosion during operation, and the actual load may be larger than the designed scenario, such as the increase in thermal load caused by global climate change in recent years, which is likely to cause potential safety hazards [1]. To ensure dam safety, structural health monitoring plays an important role in dam construction and operation management [2]. In this regard,

benefitting from the rapid development of analysis theory and computer technology, dam safety management has gradually realised a transformation to the ‘digital dam’ and then developed towards the ‘smart dam’, which is characterised by the extensive intelligent real-time analysis of massive monitoring data and the full integration of in-site monitoring, numerical simulations and intelligent control [3].

Among all measured quantities, displacement is the most intuitive reflection of the structural state of concrete dams. The most frequently used mathematical monitoring model is the hydraulic, seasonal and time (HST) model, which represents the thermal deformation effect of concrete dams by the seasonal harmonic factor [4, 5]. To adapt to the development of technology and the emergence of new engineering problems, the HST model has been optimised in two aspects. First, new causal factors have been established to explain the abnormal deformation behaviours of some concrete dams. Hu et al. [6] added a crack opening component into the HST model, and the new component was used to quantitatively evaluate the influence of radial penetrating cracks on the Chencun arch dam. To interpret the measured hysteretic hydraulic deformation behaviour, Wang et al. [7, 8] established a hydraulic, hysteretic, seasonal and time (HHST) model and a hydraulic, exponential, seasonal and time (HEST) model for the Jinping I arch dam, by which the viscoelastic parameters of dam concrete were inversed. Second, with the increase in dam height and the application of thermal insulation measures in severely cold areas, the time lag effects of reservoir water depth and operation mode on the temperature field of concrete dams become more complex, which makes it difficult for the seasonal harmonic factor in the HST model to accurately reflect the actual thermal deformation effect of high-concrete dams. Therefore, to improve the interpretation and prediction accuracy of the displacement monitoring model of concrete dams, it is an effective way to use the measured dam temperature to establish the temperature deformation factor [9]. However, there are hundreds or even thousands of thermometers embedded in each high-concrete dam; thus, it is unrealistic and unreasonable to use all of them as modelling factors. How to extract effective information from the measured massive temperature data is still a key problem that needs to be further solved in the current research. To achieve this goal, Kang et al. [10] took the nonequidistant piecewise average values of the previous air temperature at the dam site as temperature deformation factors, and the length of the previous air temperature period used was determined according to the dam type and dam body thickness. Tatin et al. [11] established the HST-Grad and HST-Layer models using the measured dam temperature in each elevation layer. Mata et al. [12] and Prakash et al. [13] used the main principal components as temperature deformation factors, the former of which were extracted from the measured

temperature time series of multiple monitoring points. Based on the shape similarity of temperature time series, Wang et al. [14] proposed a spatial clustering method for the measured dam temperature field, and principal components were then extracted from the temperature time series of all monitoring points in the same cluster. Belmokre et al. [15] used the concrete temperature of four points as thermal inputs of the random forest-based displacement monitoring model of arch dams. Two temperature points are located on the upstream and downstream faces, and the other two are 2 m inwards of each face. These concrete temperatures can first be calculated through a deterministic thermal model proposed by the same team [16].

Based on optimised causal factors, it is important to improve the prediction and interpretation ability of mathematical monitoring models using different modelling methods. The traditional multilinear regression (MLR) method, including multiple regression and stepwise regression, regards dam response as a linear explicit function of causal factors. In practice, causal mechanisms are nonlinear and dynamic. To solve this problem, artificial intelligence algorithm-integrated machine learning models, such as artificial neural network (ANN), support vector machine (SVM), extreme learning machine (ELM), regression tree (RT), random forest (RF), and long short-term memory (LSTM), have strong nonlinear data mining ability and have been widely used in structural health monitoring [15, 17–21]. With an optimal kernel function, the prediction performance of machine learning models is generally better than that of the MLR model. However, the overfitting problem of the former should be handled carefully, and the Bayesian regularisation method and the deformation spatial association-coupled double objective optimisation method have been proven to be effective for alleviating overfitting [22]. Another disadvantage of machine learning models is that they are usually considered black box models, in which the nonlinear implicit relationship between model input and output is based on network structures. As a result, the application of machine learning models mainly focuses on prediction, while the actual demands for the causal interpretation ability of mathematical models in the field of dam safety monitoring are ignored [23].

The volume of a high arch dam is very large, and the lag influencing mechanisms of air temperature and reservoir water temperature on the dam temperature field are extremely complex [16]. To select a very small number of the most representative members from hundreds or even thousands of dam body temperature monitoring points, by which their measured temperatures are directly used as temperature deformation factors, the importance of modelling factors is evaluated by the SVM, and automatic and manual criteria are formulated to eliminate the large number of unimportant or similar temperature monitoring points.

Then, a hydraulic, exponential, thermal and time (HETT) model is established for the Jinping I arch dam, and the causal mechanisms of the hydraulic and thermal effects on dam displacement are quantitatively interpreted by a partial dependence diagram (PDP).

2 Measured dam temperature-based displacement monitoring model

For the Jinping I arch dam, this paper intends to establish an HETT model for dam displacement based on the SVM and the measured temperatures of the dam body. The modelling process is shown in Fig. 1. The main approach is to select the most representative dam body temperature monitoring points through an importance evaluation of SVM input factors, and the selected dam temperature time series are then denoised and used as temperature deformation factors in the dam displacement monitoring model. To compare with the seasonal harmonic temperature factor-based model, prediction evaluation indices of the mean square error (*MSE*), maximum error (*ME*) and correlation coefficient (R^2) and the hydraulic and temperature components separated by the PDP are used to evaluate the prediction accuracy and causal interpretation ability of the proposed HETT model.

2.1 The HETT model

For the Jinping I arch dam discussed in this paper, previous research results indicate that the measured dam deformation behaviour shows an obvious viscoelastic hysteretic effect, and a hysteretic hydraulic component should be added to the HST model [7]. Therefore, the displacement of this dam needs to be interpreted by four causal components. (1) An instantaneous elastic hydraulic component caused by water pressure, δ_{He} . (2) A viscoelastic creep-induced hysteretic hydraulic component,

δ_{Hv} , is used to characterise the abnormal phenomenon that the measured radial displacement of the dam body continues increasing towards the downstream direction, which mainly appears at the time period that the upstream reservoir water level maintains at the elevation of 1880 m for approximately 100–170 days every year, and it can be represented by a step-type exponential function [8]. (3) A temperature component, δ_T , where the measured temperature of the dam body is used to establish the temperature deformation factor, and it is then compared with the traditional harmonic temperature factor. (4) The irreversible time effect component, δ_θ , is mainly accumulated from creep, plastic deformation, material deterioration, bank slope extrusion and other factors of dam concrete and foundation rock mass. Based on the traditional HST model, the HETT model can then be established as follows:

$$\delta = \sum_{i=1}^4 a_i H^i + \sum_{j=1}^N b_j T_j + c_1 \theta + c_2 \ln \theta + f(\tau) \cdot \{d_1 [1 - \exp(-\alpha\tau)] + d_2 [1 - \exp(-\beta\tau)]\}, \tag{1}$$

where H is the water depth of the upstream reservoir on the displacement monitoring day, T_j is the measured temperature of the j th selected dam body temperature monitoring point, N is the total number of used temperature monitoring points, t is the number of cumulative days from the initial monitoring day, and $\theta = t/100$. $f(\tau)$ is the step-type function; when the upstream reservoir water level is maintained at an elevation of 1880 m, its value is 1; otherwise, it is 0. τ is the duration days of the current water level stable stage, and a_i , b_j , c_1 , c_2 , d_1 and d_2 are regression coefficients. $\alpha = E_{K1}/\eta_{K1}$, and $\beta = E_{K2}/\eta_{K2}$. Here, E_{K1} and E_{K2} are hysteretic elastic moduli in the generalised Kelvin model of dam concrete, and η_{K1} and η_{K2} are viscosity coefficients. For the Jinping I arch dam, $\alpha = 0.5283$, and $\beta = 0.0052$ [8].

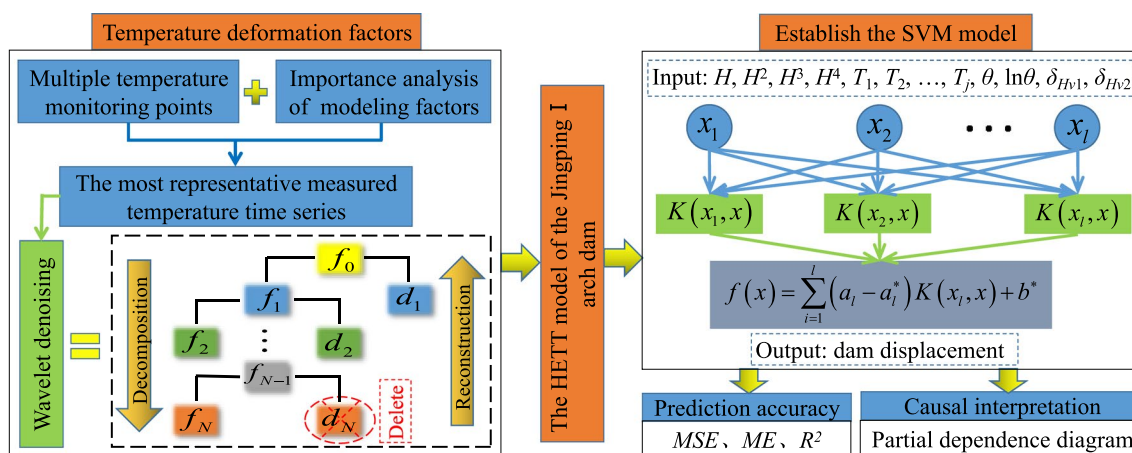


Fig. 1 Modelling process of the measured temperature-based HETT model for the Jinping I arch dam

2.2 SVM model

The essence of SVM regression is to find an optimal classification surface to separate the samples into two groups to minimise the training error. For a training set $\{(x_i, y_i), i = 1, 2, \dots, n\}$, where $x_i = [x_i^1, x_i^2, \dots, x_i^l]^T$ contains l input factors, $y_i \in R$ is the output of the SVM model, and n is the total number of samples. A linear regression model can be established in high-dimensional feature space as follows:

$$f(x) = w\phi(x) + m, \tag{2}$$

where $\phi(x)$ is a nonlinear mapping function, w is the weight, and m is a constant.

ϵ is the linear insensitive loss function and is defined as follows:

$$L(f(x), y, \epsilon) = \begin{cases} 0 & |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & |y - f(x)| > \epsilon \end{cases} \tag{3}$$

Using the Lagrange function, the above regression problem can be transformed into a coupled optimisation problem as follows [4]:

$$\begin{aligned} &\max_{\alpha, \alpha^*} \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \right] \\ &\text{s.t. } \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0; 0 \leq \alpha_i, \alpha_i^* \leq C, \end{aligned} \tag{4}$$

where $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ is the kernel function, C is the penalty factor, and $\alpha_i \geq 0$ is the Lagrange multiplier.

The optimal solution of the Lagrange multiplier can be obtained as $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]$ and $\alpha^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*]$, and thus, the SVM regression function can then be expressed as follows:

$$f(x) = w^* \phi(x) + b^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b^*. \tag{5}$$

2.3 Measured temperature deformation factors

2.3.1 Importance analysis of SVM modelling factors

On the premise of ensuring accuracy, to reduce the number of modelling factors used in the dam displacement SVM model, the input factors of the initial SVM model can be optimised by eliminating some unimportant or similar modelling factors. The importance of an input factor can

be expressed as the partial derivative of the model output $f(x_i)$ to the input factor, so the importance degree of the r th modelling factor can be quantified as follows [24]:

$$S_r = \frac{1}{n} \sum_{i=1}^n \left| \frac{\partial f(x_i)}{\partial x_{ir}} \right| \quad r = 1, 2, \dots, l, \tag{6}$$

where n is the total number of model training samples.

For the frequently used radial basis kernel function, the partial derivative in Eq. (6) can be calculated as follows:

$$\begin{aligned} \frac{\partial f(x)}{\partial x_r} &= \frac{\partial}{\partial x_r} \left(\sum_{i=1}^{n_s} (-\hat{\alpha}_i + \hat{\alpha}_i^*) K(x_i, x) + \hat{b} \right) \\ &= \frac{\partial}{\partial x_r} \left(\sum_{i=1}^{n_s} (-\hat{\alpha}_i + \hat{\alpha}_i^*) \exp \left(-\gamma \sum_{j=1}^n (s_{ij} - x_{rj})^2 \right) + \hat{b} \right) \\ &= 2\gamma \sum_{i=1}^{n_s} (-\hat{\alpha}_i + \hat{\alpha}_i^*) \exp \left(-\gamma \sum_{j=1}^n (s_{ij} - x_{rj})^2 \right) (s_{ij} - x_{rj}), \end{aligned} \tag{7}$$

where n_s is the total number of support vectors, s is the support vector, and γ is the parameter optimised in the kernel function.

2.3.2 Optimisation of the most representative dam temperature monitoring points

In practice, there are hundreds or even thousands of temperature monitoring points arranged on the dam body of a high-concrete arch dam, and the temperature evolution laws of different monitoring points have both similarities and differences, which depend on the spatial distance between them. Therefore, the main issue of establishing a measured temperature-based displacement monitoring model is to select a very small number of the most representative dam temperature monitoring points. To achieve this goal, based on the importance analysis of the input factors of the SVM model, some temperature monitoring points, the measured temperature of which has the lowest effect on the performance of the displacement SVM model, can be eliminated in turn. The optimisation method can be implemented as follows:

Step 1: Take all effective temperature monitoring points of the dam body as temperature deformation

factors, and an initial SVM model including all input factors can be established.

Step 2: According to Eq. (6), calculate the importance degree of each input factor used in the current SVM model.

Step 3: Eliminate the least important input factor and create a new input factor set with all remaining factors.

Step 4: Based on the optimised input factor set, a new SVM model can be established.

Step 5: Calculate the mean square error (MSE) of the new SVM model, which is used to evaluate the performance of the SVM model, and return to Step 2 until only one modelling factor is reserved.

Step 6: Find the minimum MSE among all the above SVM models and mark it as MSE_{min} . On this basis, recur forwards from the last SVM model, and the best SVM model in the automatic elimination process can be determined according to the first exceeding criterion of the MSE, namely, the MSE exceeds the value of $(1 + \text{preset threshold}) * MSE_{min}$ for the first time.

Step 7: Implement the manual elimination process to further optimise the temperature deformation factors automatically retained in Step 6.

The implementation process of establishing the measured dam temperature-based HETT model is shown in Fig. 2. In Step 6, the preset threshold is the maximum allowable increase ratio of the MSE of the best SVM model with respect to the MSE_{min} , and it can be determined as 5%, according to the conventional accuracy requirement of engineering projects. If the reduction in the number of modelling factors of the best SVM model, compared with the SVM model with respect to the MSE_{min} , does not exceed 5% of the total number of all initial modelling factors, the latter can then be adjusted as the best SVM model to improve the accuracy. With the increase in elimination order, the remaining modelling factor has a greater effect on the prediction performance of the SVM model, and its elimination will cause a larger increase in the MSE; thus, the selection of the best SVM model should be conducted from the last towards the first SVM model in turn. Although the automatic process in Step 6 can effectively eliminate all unimportant factors before the MSE is exceeded, there are still a small number of similar temperature factors. Therefore, if the temperature time series of two automatically retained monitoring points and their importance degree are both similar, the lower importance point in each duplicate pair can then be manually eliminated on the premise that the MSE increase ratio of the new SVM model does not exceed the threshold.

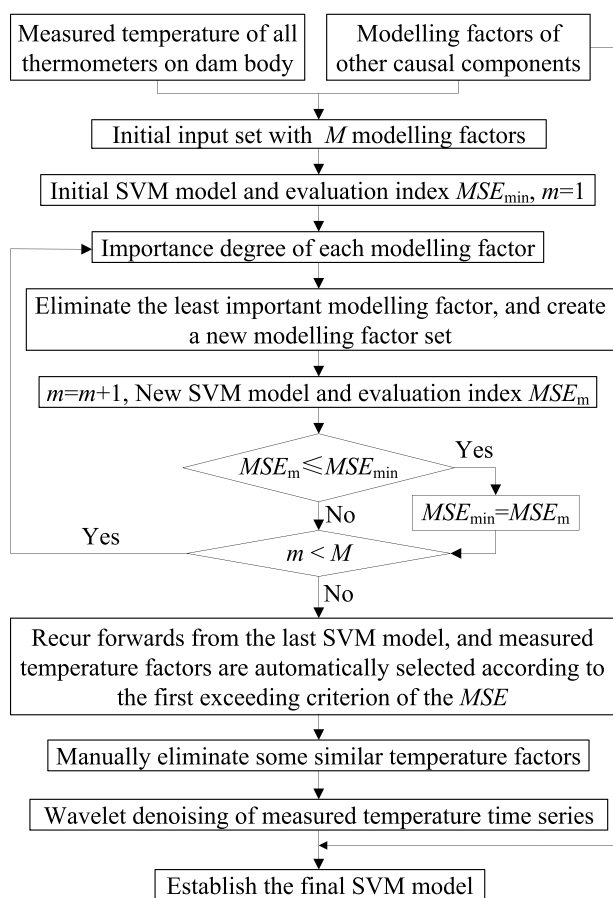


Fig. 2 Implementation process of establishing the measured dam temperature-based displacement monitoring model

2.3.3 Wavelet denoising of measured temperature time series

Measured temperatures of a dam body are usually affected by solar radiation, reservoir water level fluctuation and monitoring errors, especially because the monitoring points arranged near the dam surface are directly radiated by sunlight; thus, temperature fluctuations are more severe. If these temperatures are directly used to establish a displacement monitoring model, it will lead to the overmining of fluctuation data in the machine learning model, which will ultimately affect the prediction accuracy of the SVM model. Therefore, it is necessary to denoise any temperature time series with large fluctuations.

Wavelet multiple-resolution analysis is a signal analysis method in the time–frequency domain. It can decompose a temperature time series into subcomponents with different frequency characteristics, and some noise components with high frequency can then be eliminated. In the decomposition process, only the low-frequency component

obtained in the previous step is decomposed again, and the decomposition process can be expressed as follows:

$$f_0 = f_1 + d_1 = f_2 + d_2 + d_1 = \dots = f_N + d_N + d_{N-1} + \dots + d_2 + d_1 = f_N + \sum_{i=1}^N d_i, \tag{8}$$

where f_0 is the original signal, and f_i and d_i are the low- and high-frequency subcomponents.

2.4 Causal interpretation of dam deformation behaviour based on the SVM model

The nonlinear relationship between the input and output of a machine learning model is implicit and based on network structure, so it is difficult to interpret the causal mechanism of dam deformation behaviour. To solve this problem, the PDP can be used to mine the implicit relationship of machine learning models. For an established machine learning model, when using the PDP to quantify the influence of an input factor on model output, only this factor is determined as the independent variable, and its input data are all replaced by a fixed value, which continuously increases from the actual minimum value to the maximum value, while all other input factors are used as control variables with their actual values. Through the evolution law of the model output, a characteristic function that only depends on the studied input factor can then be obtained. Finally, the causal interpretation of a machine learning model can be represented by calculating the relative increment of each input factor to the model output. Therefore, in this paper, the PDP can be used to separate the hydraulic and temperature components in machine learning models of dam displacement. However, it is not completely reasonable to use the PDP to quantify the influences of the two time effect factors because their input values in the training set increase sequentially, and each value appears only once.

If the model input factor X is divided into x_s and its supplement $x_c = X/x_s$, the partial dependence of the model output on the response of x_s can be defined as follows:

$$f_s(x_s) = E_{x_c} \left[\hat{f}(x_s, x_c) \right] = \int \hat{f}(x_s, x_c) p_c(x_c) dx_c, \tag{9}$$

where $p_c(x_c)$ is the marginal probability density of x_c , namely $p_c(x_c) = \int p(x) dx$.

This can be further estimated from a set of discrete training data as follows:

$$f_s(x_s) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x_s, x_{i,c}), \tag{10}$$

where $x_{i,c} (i = 1, 2, \dots, n)$ is the value of training sample x_c .

2.5 Model performance evaluation

The fitting and prediction accuracy of mathematical monitoring models are frequently evaluated by the MSE , ME and R^2 , as shown in Eqs. (11) to (13), in which a smaller MSE and ME and a larger R^2 indicate better performance of the model:

$$MSE = \frac{1}{n} \sum_{t=1}^n (\delta_t - \hat{\delta}_t)^2, \tag{11}$$

$$ME = \max \left| \delta_t - \hat{\delta}_t \right|, \tag{12}$$

$$R^2 = \frac{\left[\sum_{t=1}^n (\hat{\delta}_t - \bar{\delta})(\delta_t - \bar{\delta}) \right]^2}{\sum_{t=1}^n (\hat{\delta}_t - \bar{\delta})^2 \sum_{t=1}^n (\delta_t - \bar{\delta})^2}, \tag{13}$$

where $\bar{\delta}$ and $\bar{\hat{\delta}}$ are average values of the fitted (predicted) displacement $\hat{\delta}_t$ and measured displacement δ_t , respectively.

3 Case study

The Jinping I arch dam, located on the main stream of the Yalong River in Liangshan Prefecture, Sichuan Province, China, is currently the highest constructed concrete dam in the world. The maximum dam height is 305 m. It is a double curvature arch dam and consists of 26 dam sections, of which the top and bottom thicknesses of the No. 13 central cantilever dam section are 16 and 63 m, respectively. Dam construction started in 2005, and on December 23, 2013, the dam body was fully poured to the dam crest elevation of 1885 m. The upstream reservoir water level reached the designed normal elevation of 1880 m on Aug. 24, 2014 for the first time, and it then cycled with an annual period between elevations of 1800 and 1880 m.

The downstream view of the Jinping I arch dam and layout of the plumb line monitoring system are shown in Fig. 3a and b, respectively. To monitor the temperature field of the dam body, thermometer monitoring systems are arranged in the No. 9, 13 and 19 dam sections. Generally, the temperature field of the central cantilever dam section is the most representative for concrete arch dams. Therefore, all temperature monitoring points of the No. 13 dam section, as shown in Fig. 3c, are preliminarily selected as temperature deformation factors. Except for the three bottom elevation layers, along the centreline of the dam section, there are five thermometers embedded on each elevation layer; two are 5 cm inside the upstream and downstream surface, respectively, and the other three are arranged with equal intervals between these two.

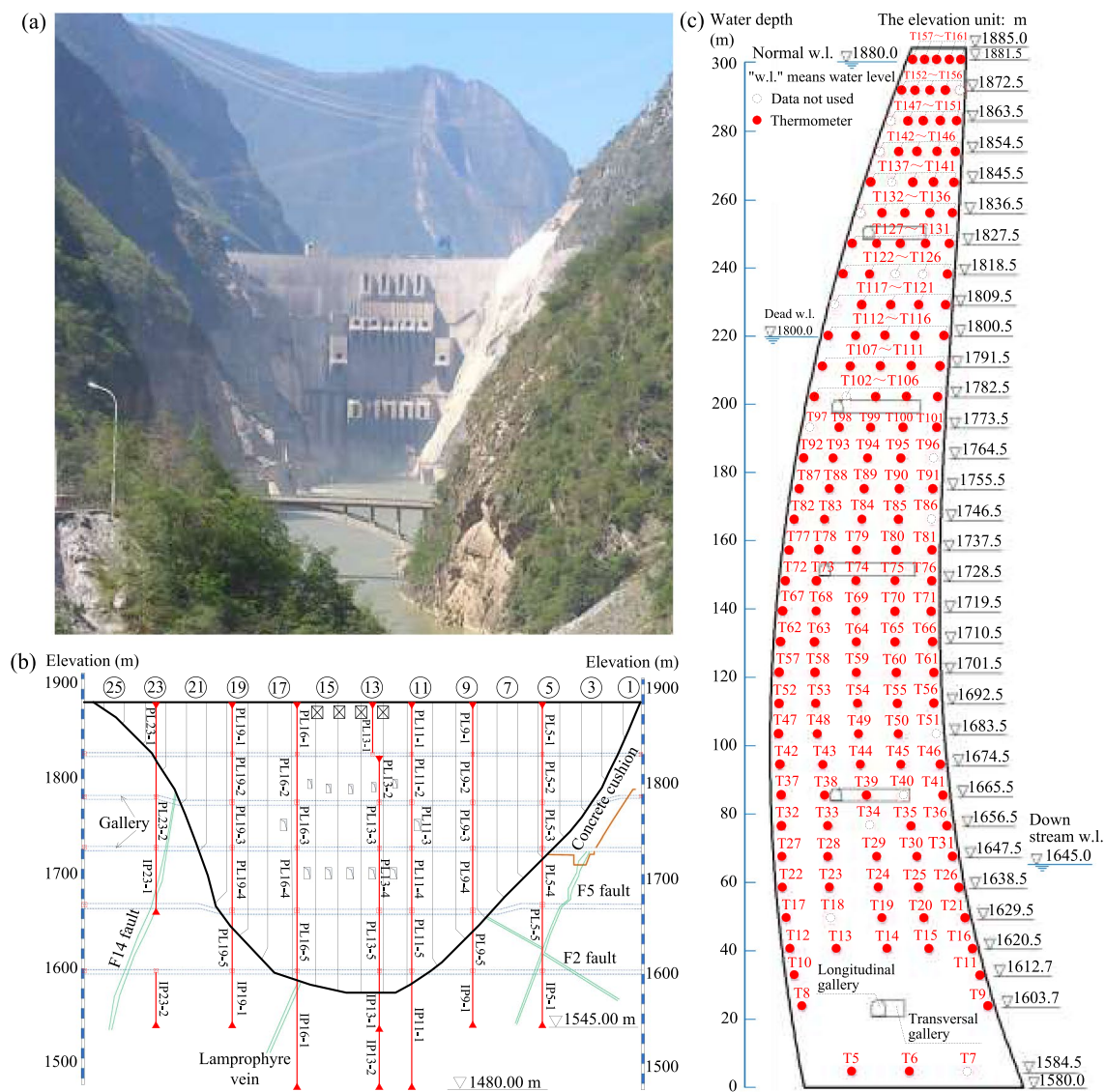


Fig. 3 a Downstream view of the Jinping I arch dam, b layout of the plumb line monitoring system, and c layout of thermometers in the No. 13 dam section

The time period of the measured radial dam displacement and temperature used in this paper is from September 1, 2015 to December 31, 2018, and the sampling frequency is once a day, of which the observation data after May 28, 2018 are used for model prediction performance testing. To avoid the accidental influence of modelling results evaluated by a single monitoring point, a total of seven dam displacement monitoring points, which are all normal plumb line monitoring points on the No. 13 central cantilever dam section and the PL9-1 and PL11-1 monitoring points on the dam crest, are modelled by the proposed approach.

4 Results and discussion

4.1 Measured temperature-based temperature deformation factors

There are 148 modelling factors in the initial SVM model, including 4 hydraulic factors, 140 measured temperature factors, 2 time effect factors and 2 hysteretic hydraulic factors. The main objective of the importance analysis of modelling factors is to reduce the total number of modelling factors, during which a large number of unimportant or similar members from the total 140 initial temperature factors are eliminated.

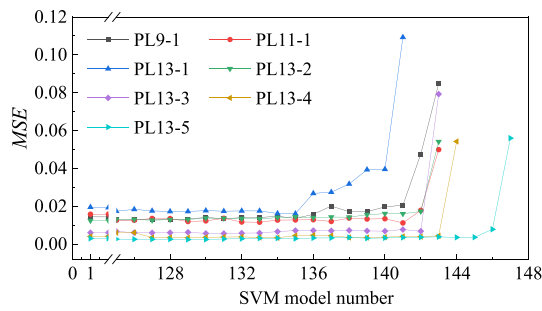


Fig. 4 Evolution of the *MSE* during the successive elimination process of modelling factors

4.1.1 Optimisation of temperature monitoring points

The evolution of the *MSE* during the successive elimination process of modelling factors is shown in Fig. 4. Overall, before the 130th SVM model, after removing the least important modelling factors, the *MSE* of the SVM model decreases slightly or remains basically unchanged, and thus, the elimination of these modelling factors will not weaken the prediction performance of the SVM model. In contrast, in most cases, this elimination plays a role in improvement. In addition, modelling factors eliminated in an early order are all measured temperature factors, and only a small number of non-temperature factors are eliminated a little earlier before the first exceeding of the *MSE*. To compare with the seasonal harmonic factor-based HEST model, these non-temperature factors are still retained in the final HETT model. In this study, the threshold of the *MSE* increase ratio, compared with the MSE_{min} , is set to 5%. According to Fig. 4,

the best SVM model can be determined and is shown in Table 1.

Taking the PL13-3 monitoring point as an example, in Fig. 5a and Table 2, it can be seen that among these automatically selected temperature monitoring points, the temperature time series of T13-146 and T13-139, T13-160 and T13-158, T3-65 and T13-70, and T13-15 and T13-38 are similar in pairs, in which the former in each pair is eliminated in an early order. If T13-146, T13-160, T13-65 and T13-15 are eliminated in manual mode, the *MSE* of the new SVM model, compared with the previous SVM model before eliminating these factors in the overall elimination process, increases by no more than 5% or even decreases, so these four temperature deformation factors can be effectively eliminated.

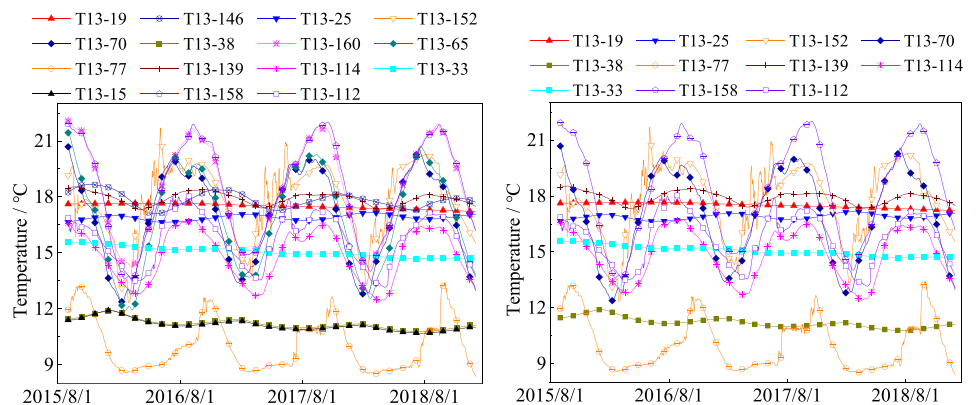
Manual elimination processes are also conducted for the other six displacement monitoring points, and the final retained modelling factors are given in Table 3. The measured temperature time series of the selected dam body temperature monitoring points for each displacement monitoring point are shown in Fig. 6. In addition, as seen in Table 3, the two time effect modelling factors, θ and $\ln \theta$ in Eq. (1), have both been removed for all seven analysed displacement points in the automatic elimination process. The reasons are that these two factors are used to model the trend component of dam displacement, but their evolution laws are repeated with some of the measured temperature factors shown in Fig. 6.

The numbers of finally selected measured temperature factors for PL9-1, PL11-1, PL13-1, PL13-2, PL13-3, PL13-4 and PL13-5 are 9, 4, 8, 7, 11, 12 and 11, respectively. On average, among the 140 temperature monitoring points in the

Table 1 Sequence number of the best SVM model in the automatic elimination process

Displacement point	PL9-1	PL11-1	PL13-1	PL13-2	PL13-3	PL13-4	PL13-5
Sequence number	135	141	135	138	130	129	129

Fig. 5 Measured temperature series selected for PL13-3



(a) Before manual elimination

(b) After manual elimination

Table 2 Modelling factors of PL13-3 retained by automatic elimination

Model number	Model <i>MSE</i>	Least important factor	Model number	Model <i>MSE</i>	Least important factor	Model number	Model <i>MSE</i>	Least important factor
130	0.00554	T13-19	137	0.00703	T13-65	144	0.20133	T13-112
131	0.00558	T13-146	138	0.00725	T13-77	145	3.06870	H ⁴
132	0.00579	T13-25	139	0.00695	T13-139	146	3.06471	H ³
133	0.00586	T13-152	140	0.00684	T13-114	147	3.04936	H ²
134	0.00656	T13-70	141	0.00765	T13-33	148	3.03707	H
135	0.00712	T13-38	142	0.00679	T13-15	/	/	/
136	0.00715	T13-160	143	0.00793	T13-158	/	/	/

Table 3 Retained modelling factors after automatic and manual elimination

Displacement point	Modelling factors (the importance degree increases from left to right)
PL9-1	T13-131, T13-81, T13-79, T13-152, T13-47, T13-148, T13-20, T13-109, H, H ² , T13-65, H ⁴ , H ³
PL11-1	T13-153, T13-133, T13-145, H, T13-111, H ⁴ , H ³ , H ²
PL13-1	T13-131, T13-94, T13-105, T13-158, T13-149, T13-107, T13-77, H, T13-82, H ⁴ , H ³ , H ²
PL13-2	T13-149, H, T13-151, T13-148, T13-120, T13-158, T13-82, T13-121, H ⁴ , H ³ , H ²
PL13-3	T13-19, T13-25, T13-152, T13-70, T13-38, T13-77, T13-139, T13-114, T13-33, T13-158, T13-112, H ⁴ , H ³ , H ² , H
PL13-4	T13-152, T13-65, T13-119, T13-43, T13-116, T13-158, T13-33, T13-15, T13-102, T13-80, T13-154, T13-114, H ³ , H ⁴ , H ² , H
PL13-5	T13-60, T13-37, T13-114, T13-152, T13-35, T13-153, δ_{Hv2} , T13-140, T13-119, H, H ² , H ³ , T13-135, T13-39, H ⁴ , T13-154

Note: δ_{Hv2} is the second modelling factor of the hysteretic hydraulic component. In this paper, the purpose of eliminating modelling factors is to optimise the measured temperature deformation factors; thus, the eliminated non-temperature factors are still retained in the final HETT model

central cantilever dam section, only 6.3% need to be used to represent the thermal deformation effect of the dam, which indicates the high efficiency of the proposed optimisation approach.

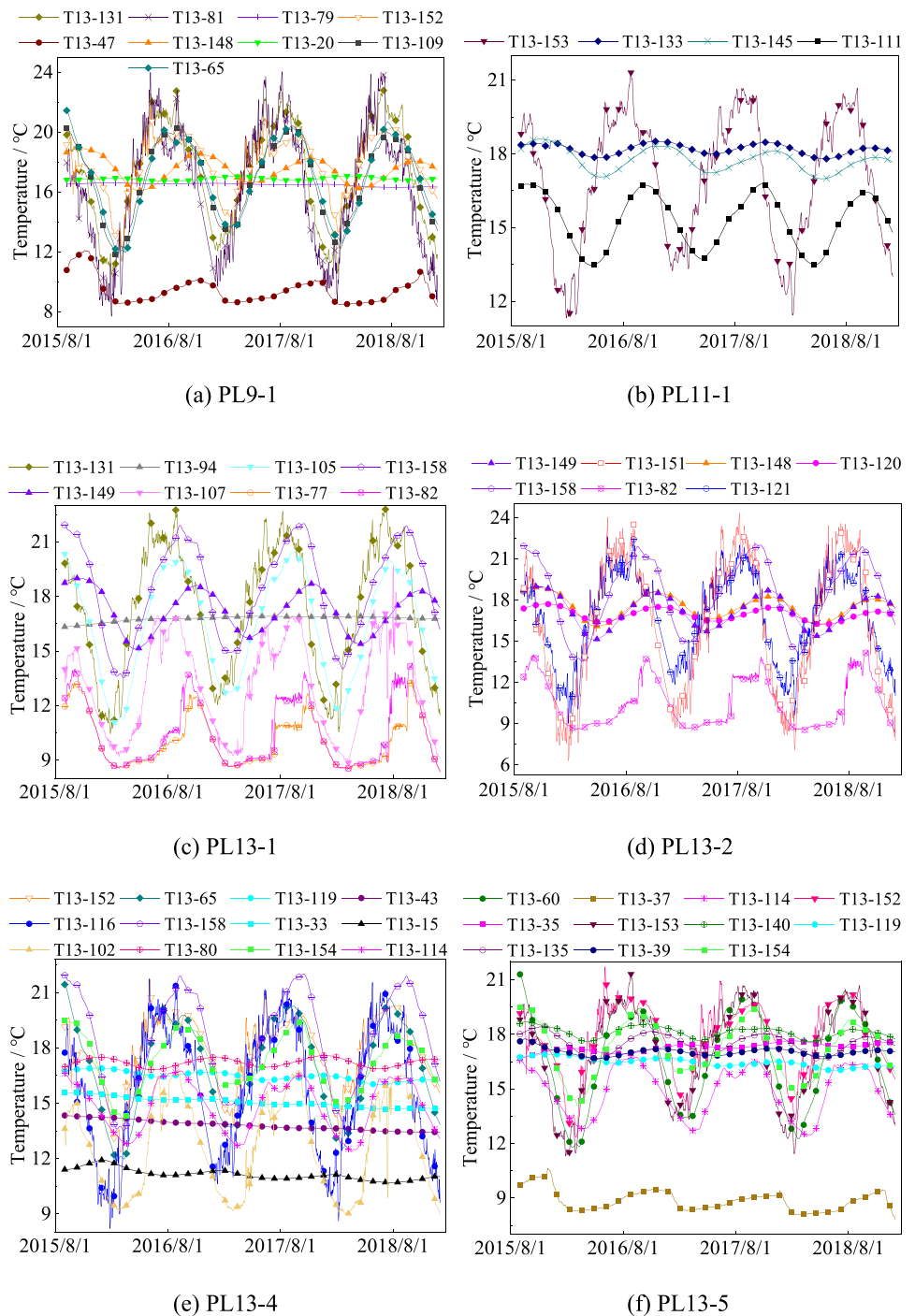
4.1.2 Rationality of the selected temperature deformation factors

In the displacement monitoring model of concrete dams, the influence of a temperature deformation factor is mainly reflected in two aspects: shape characteristics of the used time series and its time lag effect with respect to the air temperature. Therefore, the measured temperature time series shown in Fig. 6 can be divided into six categories according to their shape characteristics and the lag effect mechanism of the dam temperature field. The measured temperature time series of some typical temperature monitoring points in each category are shown in Fig. 7, and their evolution characteristics and distributions are summarised as follows:

(1) Category I: The downstream surface and crest of the dam body are directly affected by solar radiation, and thus their measured temperature time series are very

- similar to the air temperature of the dam site and have great fluctuations, such as T13-81, T3-121 and T13-153. To reduce the interference of data fluctuation on machine learning models, temperature time series in Category I need to be denoised.
- (2) Category II: Temperature monitoring points in Category II are mainly distributed in the upstream surface and middle of the dam body that are affected by the beneficial regulation of the reservoir water level, such as T13-65, T13-102 and T13-112. In the annual rising stage of air temperature, the measured temperatures of these monitoring points rise rapidly due to the low reservoir water level. However, when the air temperature annually drops, the reservoir water level at this stage is maintained at the highest elevation of 1880 m; thus, the reservoir water-induced heat conduction leads to an obvious time lag phenomenon between the temperature change at these monitoring points and the air temperature.
- (3) Category III: The annual periodic evolution law of measured temperatures is basically in accordance with the seasonal harmonic factor, such as T13-47 and T13-149, and they are mainly distributed in the upstream surface of the middle and lower elevation parts of the

Fig. 6 Measured temperature time series selected for each displacement monitoring point



- dam body and the interior of the dam crest. The influence of the annual cycle change in the upstream reservoir water level on these areas has been gradually weakened.
- (4) Category IV: This category is similar to Category III, and the temperature evolution law is also basically consistent with the seasonal harmonic factor, but the annual variation amplitude of measured temperatures has significantly decreased, such as T13-80, T13-119

- and T13-135. Category IV is mainly distributed in the middle and upper elevation parts of the interior dam body.
- (5) Category V: The measured temperature time series shows a linear downwards trend, such as T13-19, T13-43 and T13-79, which are mainly distributed inside the dam heel and the middle and lower elevation parts of the interior dam body. The upstream reservoir water depth and dam body thickness in these areas are both

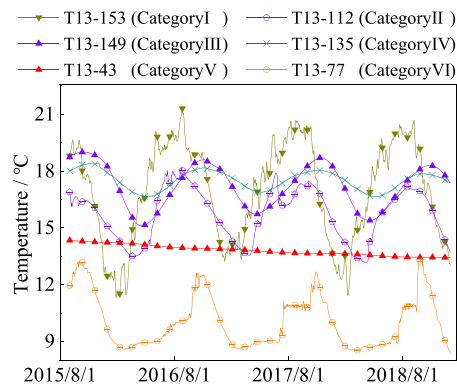


Fig. 7 Measured temperature time series of typical temperature deformation factors

very large, which makes them less affected by the ambient temperature. In addition, the influence of concrete hydration heat in these areas is significantly reduced, and the dam temperature will gradually drop to the joint closure temperature field and form a stable temperature field.

- (6) Category VI: This category is obviously affected by the water storage process of the upstream reservoir, and the measured temperatures rise sharply during this period, such as T13-77, T13-82 and T13-109. The upstream surface of the dam body near the dead water level elevation of 1800 m, with a height range of approximately 100 m, belongs to Category VI.

In summary, for the modelled seven displacement monitoring points in this case study, the evolution characteristics of the measured dam temperature time series used for PL9-1, PL13-1, PL13-2 and PL13-3 are the most complete. The main reason is that these displacement monitoring points are located in the middle and upper elevation parts of the dam body and are widely affected by temperature changes in the whole dam body. Other displacement monitoring points have also selected these representative temperature time series, which verifies the rationality of the proposed approach for optimising the measured temperature deformation factors.

4.1.3 Wavelet denoising of measured temperature time series

Temperature monitoring points in Category I are distributed on the downstream surface of the dam body and the dam crest; thus, the measured temperature is directly affected by solar radiation and fluctuates greatly. To ensure the prediction accuracy of the SVM model, it should be denoised first. The temperature time series of T13-121 before and after wavelet denoising are shown in Fig. 8. By removing some high-frequency components with large fluctuations, the

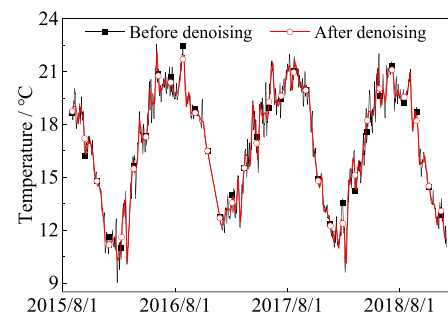


Fig. 8 Temperature time series of T13-121

denoised temperature time series is smooth and still maintains the original evolution characteristics. Temperature time series of T13-11, T13-77, T13-81, T13-102, T13-116, T13-131 and T13-153 are also denoised by wavelet multiresolution analysis.

4.2 Prediction performance of the HETT-SVM model

Combined with the instantaneous hydraulic component, hysteretic hydraulic component and time effect component, the HETT and HEST models are established using the measured temperature factor and seasonal harmonic temperature factor, respectively, and an HETT-o model is also established, in which the selected dam temperature time series have not been denoised. The SVM is used to conduct the nonlinear modelling. The radial displacement time series of the measured and fitted (predicted) values of PL11-1 and PL13-4 are shown in Fig. 9, in which the positive and negative values represent the radial displacement towards the downstream and upstream directions, respectively. The results of other displacement monitoring points are similar to these two. For all seven analysed displacement monitoring points, in the fitting stage, the displacement time series of the HETT, HETT-o and HEST models are basically consistent and very close to the measured value. In the prediction stage, PL11-1 has the best accuracy, while PL13-4 has the largest deviation. With the duration extension of the reservoir water level maintained at an elevation of 1880 m, the prediction deviations of some displacement monitoring points increase gradually. In general, the displacement predicted by the HETT model is closer to the measured value.

The prediction performance evaluation indices of the three models are shown in Fig. 10. As seen in the figure, the prediction performance of the HETT-o model fluctuates greatly. The reason is that this model is established without conducting the wavelet denoising process, whereas the measured temperatures of some monitoring points are seriously affected by solar radiation and have obvious fluctuations. After wavelet denoising, the prediction *MSEs* of the HETT models of PL9-1, PL13-1, PL13-2, PL13-3,

Fig. 9 Radial displacement time series of the measured and fitted (predicted) values

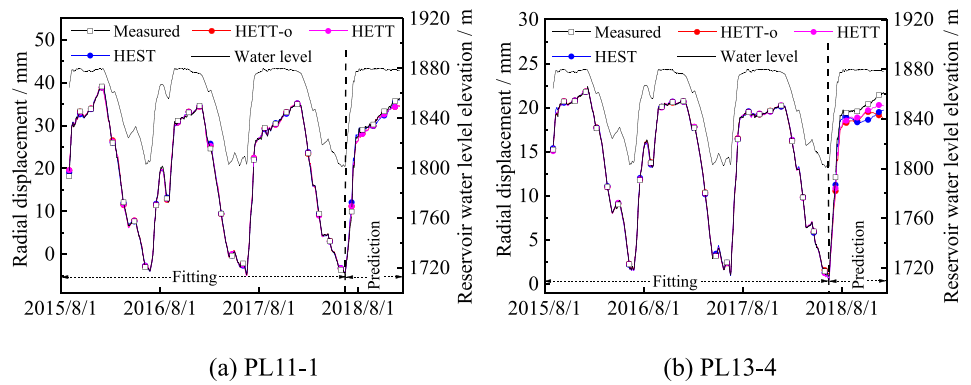
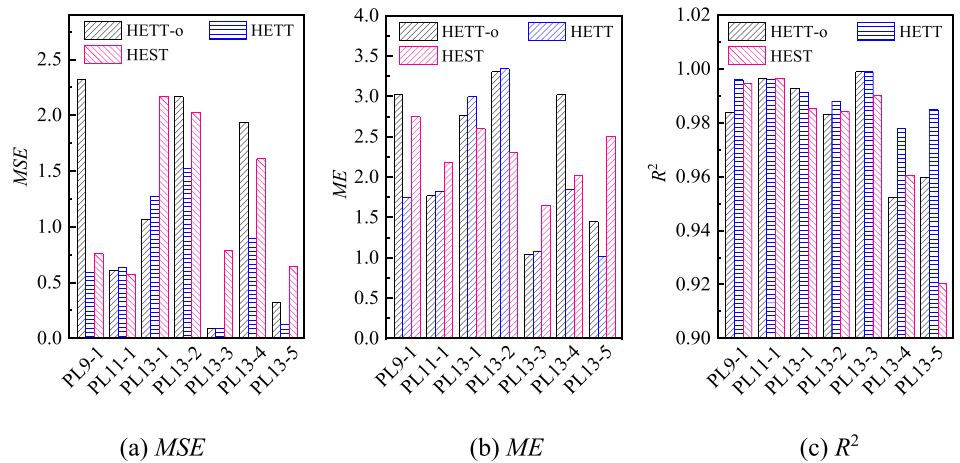


Fig. 10 Evaluation indices of model prediction performance



PL13-4 and PL13-5 are significantly better than those of the HEST models. Overall, for the seven analysed displacement monitoring points, the average prediction *MSE* of the HETT model is only 60.1% of that of the HEST model. The prediction *MEs* of the HETT models of PL9-1, PL11-1, PL13-3, PL13-4 and PL13-5 are also significantly smaller than those of the HEST models, which indicates that the HETT model can better describe the actual thermal deformation effect of concrete arch dams. The average *ME* and R^2 of multiple monitoring points are 86.5% and 101.5% of those of the HEST model, respectively. In conclusion, compared with the traditional seasonal harmonic temperature factor, the measured temperature deformation factor, optimally selected through the importance analysis of SVM modelling factors, can better characterise the thermal deformation effect of arch dams, and the established displacement monitoring model has higher prediction accuracy. However, wavelet denoising must be conducted for the temperature time series of these monitoring points that are directly affected by solar radiation.

4.3 Causal interpretation of the HETT-SVM model

Based on the PDP, the evolution law of hydraulic displacement separated from the SVM model with respect to reservoir water depth is shown in Fig. 11, and the temperature component is shown in Fig. 12. The results of other displacement monitoring points are similar. On the whole, using the measured temperature factor, the hydraulic and temperature components of displacement monitoring points distributed at the middle and upper elevation parts of the dam body are almost unchanged, but they have changed obviously for displacement monitoring points with lower elevation. Figure 10 shows that the influence laws of reservoir water depth on the radial displacement of the dam body obtained by the three models are basically the same. The radial hydraulic displacement increases with reservoir water depth and the elevation of the displacement monitoring point, and the nonlinear relationship is more obvious for the dam crest area. Except for PL13-5 with lower elevation, the relationship between hydraulic

Fig. 11 Relationship between the SVM model-separated hydraulic radial displacement and upstream reservoir water depth

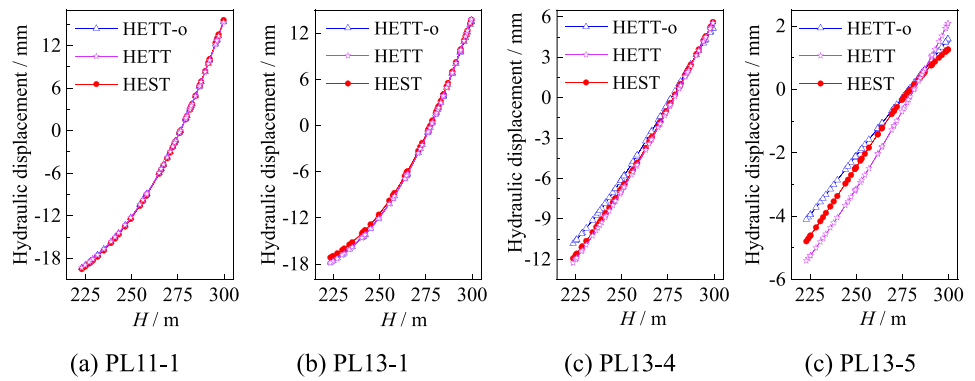
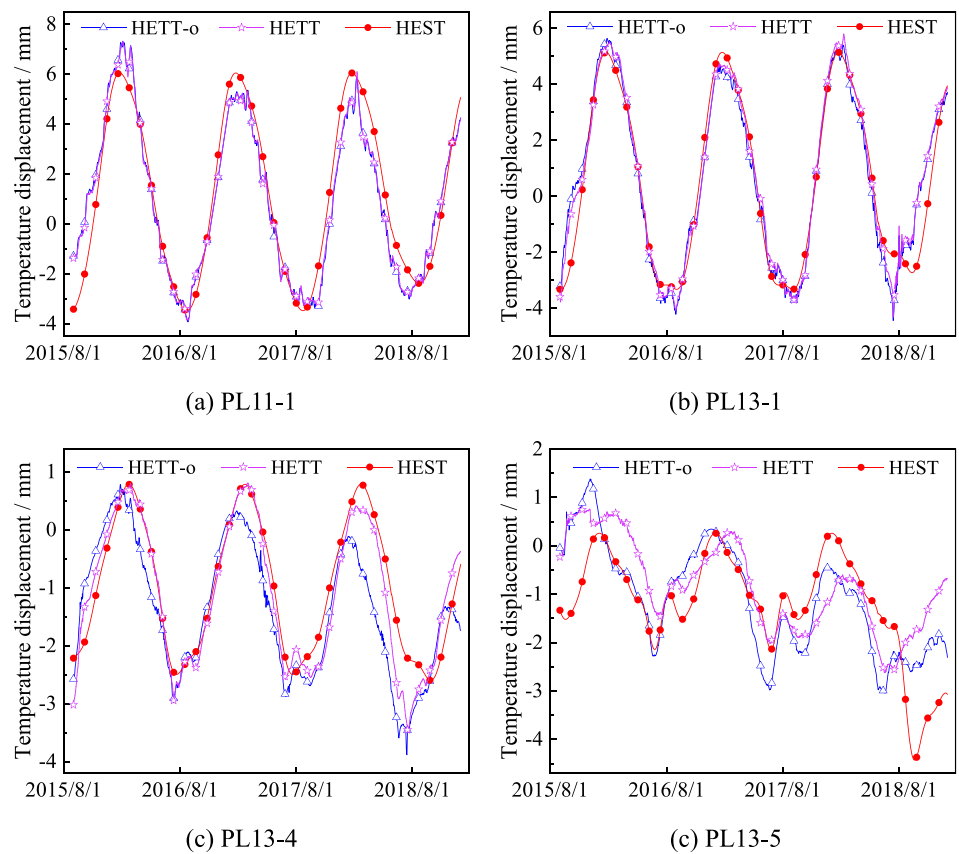


Fig. 12 Radial displacement time series of the SVM model-separated temperature component



displacement and water depth can be expressed by the quartic polynomial function, which is consistent with the deformation theory of concrete arch dams.

The overall evolution laws of the temperature components in Fig. 11 are also basically the same, and they change with an annual periodicity, which is also in accordance with the actual situation of arch dams. The temperature component of PL13-5 shows an increasing trend towards the upstream direction. This phenomenon is caused by the lower elevation of this displacement monitoring point, whereas the measured temperature near the dam heel shows that this area is currently in an overall temperature drop state, as shown in

Category V in Fig. 6. Therefore, the local temperature drop effect in the dam heel area causes the upstream direction temperature deformation trend of the low and medium elevation parts of the dam body. Although the HEST model also interprets this temperature deformation trend, its evolution process is discontinuous, and the trend component of the temperature displacement suddenly increases in the fourth year. However, the trend temperature displacements obtained by the HETT and HETT-o models are gentler and in accordance with the actual situation of the Jinping I arch dam, the measured temperature field of which slowly drops to the joint closure temperature field at the current operation stage.

The comparison results show that temperature monitoring points selected through the importance analysis of SVM modelling factors can better represent the measured temperature field and the thermal deformation effect of high arch dams, and thus the proposed approach can be used to extract effective information from the massive temperature monitoring data of the dam body. Compared with the traditional seasonal harmonic temperature factor-based HEST model, the newly established HETT model has better prediction and interpretation performance.

5 Conclusion

To use the measured dam temperature as the temperature deformation factor of high arch dams, the key issue is to extract effective information from the massive temperature monitoring data of the dam body. In this paper, the importance analysis of SVM modelling factors is used to select the most representative dam temperature monitoring points, and the proposed HETT model has good prediction performance and causal interpretation ability for the Jinping I arch dam. The following conclusions can be drawn:

- (1) The importance analysis of SVM modelling factors can effectively optimise the measured temperature deformation factors used in the dam displacement monitoring model, and the temperature time series of the selected temperature monitoring points can comprehensively describe the temporal and spatial evolution characteristics of the measured temperature field of high arch dams. For the seven analysed displacement monitoring points of the Jinping I arch dam, the selected maximum and minimum numbers of the measured temperature deformation factors are 12 and 4, respectively, with an average number of 8.9 and an average rate of 6.4% from the total 140 effective temperature monitoring points in the central cantilever dam section.
- (2) The prediction performance of the measured temperature factor-based HETT model is better than that of the seasonal harmonic temperature factor-based HEST model, in which the average prediction MSE , ME and R^2 of multiple displacement monitoring points of the former are 60.1%, 86.5% and 101.5% of those of the latter, respectively. However, the temperature time series of the solar radiation-affected temperature monitoring points on the dam surface must be denoised.
- (3) The PDP can be used to explore the causal interpretation of dam deformation behaviour modelled by machine learning models. The hydraulic and temperature components separated from the SVM model agree with the deformation mechanism of high arch dams. The temperature at the middle and lower elevation parts

of a high arch dam is mainly affected by the stable temperature field of the upstream reservoir water, and thus the measured temperature evolution laws of these dam areas obviously deviate from the seasonal harmonic temperature factor. The measured temperature factor has a better interpretation ability for the thermal deformation behaviour of high arch dams.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 51709021, 52079120, 52278240), the Project funded by China Postdoctoral Science Foundation (Grant No. 2020M670387), the Open Research Fund of Key Laboratory of Construction and Safety of Water Engineering of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research (Grant Nos. 202009, 202107), 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) (Grant No. IWHR-SKL-KF202002), and Self-Topic Fund of State Key Laboratory of Simulation & Regulation of Water Cycle in River Basin (Grant No. SKL2020ZY10).

Data availability The data presented in this study are available on request from the first or corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to this work.

References

1. Santillán D, Salet E, Toledo MÁ (2015) A methodology for the assessment of the effect of climate change on the thermal-strain-stress behaviour of structures. *Eng Struct* 92:123–141
2. Zhao EF, Wu CQ (2020) Unified egg ellipse critical threshold estimation for the deformation behavior of ultrahigh arch dams. *Eng Struct* 214:110598
3. Zhong DH, Wang F, Wu BP, Cui B, Liu YX (2015) From digital dam toward smart dam. *J Hydroelectric Eng* 34(10):1–13 (in Chinese)
4. Su HZ, Chen ZX, Wen ZP (2016) Performance improvement method of support vector machine-based model monitoring dam safety. *Struct Control Health Monit* 23(2):252–266
5. Wei BW, Chen LJ, Li HK, Yuan DY, Wang G (2020) Optimized prediction model for concrete dam displacement based on signal residual amendment. *Appl Math Model* 78:20–36
6. Hu J, Wu SH (2019) Statistical modeling for deformation analysis of concrete arch dams with influential horizontal cracks. *Struct Health Monit* 18(2):546–562
7. Wang SW, Xu YL, Gu CS, Bao TF, Xia Q, Hu K (2019) Hysteretic effect considered monitoring model for interpreting abnormal deformation behavior of arch dams: a case study. *Struct Control Health Monit* 26(10):1–20
8. Wang SW, Xu C, Liu Y, Xu B (2022) Zonal intelligent inversion of viscoelastic parameters of high arch dams using an HEST statistical model. *J Civil Struct Health Monit* 12(1):207–223
9. Salazar F, Morán R, Toledo MÁ, Oñate E (2017) Data-based models for the prediction of dam behaviour: a review and some methodological considerations. *Arch Computat Methods Eng* 24(1):1–21

10. Kang F, Li JJ, Zhao SZ, Wang YJ (2019) Structural health monitoring of concrete dams using long-term air temperature for thermal effect simulation. *Eng Struct* 180:642–653
11. Tatin M, Briffaut M, Dufour F, Simon A, Fabre J-P (2018) Statistical modeling of thermal displacements for concrete dams: influence of water temperature profile and dam thickness profile. *Eng Struct* 165:63–75
12. Mata J, Castro AT, Costa JS (2014) Constructing statistical models for arch dam deformation. *Struct Control Health Monit* 21(3):423–437
13. Prakash G, Sadhu A, Narasimhan S, Brehe J-M (2018) Initial service life data towards structural health monitoring of a concrete arch dam. *Struct Control Health Monit* 25(1):e2036
14. Wang SW, Xu C, Gu CS, Su HZ, Hu K, Xia Q (2020) Displacement monitoring model of concrete dams using the shape feature clustering-based temperature principal component factor. *Struct Control Health Monit* 27(10):e2603
15. Belmokre A, Mihoubi MK, Santillán D (2019) Analysis of dam behavior by statistical models: application of the random forest approach. *KSCE J Civ Eng* 23(11):4800–4811
16. Santillán D, Salet E, Vicente DJ, Toledo MÁ (2014) Treatment of solar radiation by spatial and temporal discretization for modeling the thermal response of arch dams. *J Eng Mech-Asce* 140(11):05014001 (1–18)
17. Mata J (2011) Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. *Eng Struct* 33(3):903–910
18. Ranković V, Grujović N, Divac D, Milivojević N (2014) Development of support vector regression identification model for prediction of dam structural behaviour. *Struct Saf* 48:33–39
19. Salazar F, Toledo MÁ, Oñate E, Suárez B (2016) Interpretation of dam deformation and leakage with boosted regression trees. *Eng Struct* 119:230–251
20. Liu WJ, Pan JW, Ren YS, Wu ZG, Wang JT (2020) Coupling prediction model for long-term displacements of arch dams based on long short-term memory network. *Struct Control Health Monit* 27(3):e2548
21. Li MC, Ren QB, Kong R, Du SL, Si W (2019) Dynamic modeling and prediction analysis of dam deformation under multidimensional complex relevance. *J Hydraul Eng* 50(6):687–698 (in Chinese)
22. Wang SW, Xu C, Liu Y, Wu BB (2022) A spatial association-coupled double objective support vector machine prediction model for diagnosing the deformation behaviour of high arch dams. *Struct Health Monit* 21(3):945–964
23. Hu J, Ma FH (2021) Comparison of hierarchical clustering based deformation prediction models for high arch dams during the initial operation period. *J Civil Struct Health Monit* 11(4):897–914
24. Wang SW, Xu YL, Gu CS, Bao TF (2018) Monitoring models for base flow effect and daily variation of dam seepage elements considering time lag effect. *Water Sci Eng* 11(4):344–354

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.