WILEY | Hindawi

Research Article

Multipoint Deformation Safety Monitoring Model for Concrete Arch Dams Based on Bayesian Model Selection and Averaging

Lin Cheng,[1,2] **Jiamin Chen** ,[1,2] **Chunhui Ma** ,[1,2] **Jie Yang,**[1,2] **Xiaoyan Xu,**[1,2] **and Shuai Yuan** [1,2]

[1]*State Key Laboratory of Eco-Hydraulics in Northwest Arid Region, Xi'an University of Technology, Xi'an 710048, China*
[2]*Faculty of Water Resources and Hydro-Electric Engineering, Xi'an University of Technology, Xi'an 710048, China*

Correspondence should be addressed to Chunhui Ma; shanximachunhui@foxmail.com

The deformation properties of concrete arch dams are affected by the synergistic effects of multiple factors, featuring strong, multidimensional spatialtemporal evolution and distribution characteristics. This paper proposes a zoned safety monitoring model for arch dam deformation based on spatialtemporal similarity and model optimization to evaluate the deformation safety state of arch dam structures. First, zoned clustering of the deformation monitoring points at different locations of an arch dam was performed using a panel data multi-index clustering method to determine the deformation laws at different positions of the dam. Next, multipoint comprehensive displacements of the deformation properties of each zone were extracted using principal component analysis to extract the uniform deformation law of the monitoring point in each zone. Finally, we adopted Bayesian model selection (BMS) and Bayesian model averaging (BMA) for the regression model set, considering the uncertainty of the model. The engineering case study showed that BMA yielded robust and effective prediction results for the deformation of the arch dam. The analysis of the zoned deformation mechanism indicated that the deformation of the arch dam followed the general rule. The temperature component of the arch dam was mainly reflected in the middle with a hysteresis effect, and the time-dependent component was evident in both sides of the dam shoulder. The arch dam deformation safety monitoring model proposed in this study has strong robustness and interpretability, which can provide valuable technical support for analyzing the evolution of arch dam deformation properties.

1. Introduction

The construction of dams, an important engineering measure to transform nature and exploit water resources, also brings huge economic and social benefits. Recently, several 200–300 m class extrahigh arch dams have been planned, constructed, and implemented in China. These projects are completed under complex topographic and geological conditions. During their long-term services, the body and base of arch dams suffer various levels of aging and damage, resulting in reduced structural load-bearing capacities and increased risk of significant engineering accidents [1]. The concept of "structural health monitoring" has been put forward since the 1990s for which the measurements of the displacement. The study of the evolution of deformation

properties based on monitoring data is critical for safe arch dam operation [2].

A commonly used method to monitor dam safety involves a mathematical monitoring model based on the prototype measurement data from a monitoring instrument to analyze the mechanical deformation of the dam structure [3, 4]. Mata et al. [5] used the three-factor hydraulic-seasonal-time (HST) model to describe the displacement monitoring data of concrete dams. In another study, Hu and Ma [6] proposed two improved statistical models to address the nonstationary thermal and nonmonotonic time-dependent effects and analyze the deformation mechanism of high-arch dams at the initial impoundment. Relevant research of Wang et al. [7] introduced a hysteresis water-pressure component into the HST model and

proposed a hydraulic-hysteretic-seasonal-time (HHST) model to explain the anomalous deformation behaviors caused by hysteresis water-pressure deformation and ambient temperature drop effects. The above studies focus on single monitoring points while ignoring the deformation similarity among various monitoring points.

The general structure of arch dams has unique deformation characteristics, so a deformation analysis using a traditional monotonous method is not appropriate. Instead, a comprehensive evaluation with multiple measuring points and factors should be developed to replace the independent single-factor evaluation based on the complex spatialtemporal characteristics of arch dam deformation [8]. In developing a safety monitoring model for arch dam deformation at multiple monitoring points, Wei et al. [9] introduced spatial coordinates and calculated the water-pressure component using the finite element method. In addition, they proposed a hybrid model of a particle swarm algorithm-optimized support vector machine (SVM) for concrete arch dam deformation monitoring at multiple monitoring points.

Furthermore, Xu et al. [10] used principal component analysis (PCA) to extract the comprehensive displacement of multiple monitoring points. They then decomposed the data into seasonal and trend displacements using the Bayesian estimator of abrupt change, seasonal change, and trend (BEAST) methodology, followed by quantifying and analyzing the changing laws of displacement. Similarly, Wang et al. [11–15] introduced the spatialtemporal clustering concept into the multipoint analysis of arch dam deformation and clustered or divided the objects in space according to the spatialtemporal data.

These mathematical models usually employ spatial coordinates, PCA dimensional reduction, and panel data clustering to establish the dependence relationship between multipoint deformation and environmental quantities. Regression analysis or optimization methods follow to establish the model's coefficients for each factor and physical interpretation of the components to further analyze the dam's working state [3, 16].

With the development of computing technology and artificial intelligence, algorithms such as random forest (RF) [17, 18], neural network (NN) [19, 20], long- and short-term memory (LSTM) [21, 22], SVM [23–26], extreme learning machine (ELM) [27], and Gaussian process regression (GPR) [28, 29] have been applied to security monitoring successively. However, such models are usually aimed at the fitting accuracy to test their validity, while mathematical model optimization based on optimization theory has attracted significant attention.

For instance, Huang et al. [3] proposed the optimal statistical model selection method for dam monitoring from the perspective of system theory with the optimization principles of good model fit, test validity, and model simplicity. Although such a method considered the fitting accuracy, model factors, and verification effect, most variables in the model had to be known while ignoring the model's uncertainty. Bayesian class selection has been well studied in structural and geotechnical engineering [30–32]. Thus,

Gamse et al. [33] applied the Bayesian model selection (BMS) to the HST model updating, which could rapidly acquire new information embedded in additional measurements to improve the dam's safety and reliability. Similarly, Prakash and Balomenos [34] used the Bayesian method to estimate and update model parameters, proposing the following two approaches: the naive Bayesian model (NBM) and the Bayesian model averaging (BMA). BMS and BMA can solve the uncertainty in dam modeling and yield the most reasonable HST model from a set of model class candidates based on the model criterion provided by the data.

This paper proposes a multipoint deformation safety monitoring model for concrete arch dams based on BMS and BMA. In this work, a comprehensive list of similarity indicators was established using a panel data multi-index clustering method, and zoned clustering was performed on the deformation of the dam measuring point using an analytic hierarchy process. In addition, PCA was carried out on the deformation of each zone to extract multipoint displacements containing the deformation state of each zone, and a regression model was established to form a model set. Next, BMS was performed on the model set, and BMA was conducted considering the model uncertainty, followed by an analysis of the period prediction results and deformation mechanism.

This paper is composed of three parts. First, an introduction to the theories and principles of spatialtemporal clustering is presented along with multipoint comprehensive displacement extraction and BMS and BMA based on panel data. Then, the building process of the arch dam multipoint BMS is described. Finally, a verification of the robustness and strong interpretability of the multipoint deformation safety monitoring model of concrete arch dams based on BMS and BMA is performed through engineering cases.

2. Multipoint Monitoring Model for Arch Dams Based on BMS

2.1. Spatialtemporal Clustering Method Using Panel Data. Traditional analyses of concrete arch dam deformation monitoring data focus on the time-domain characteristics of single monitoring points without considering spatial information. We can achieve more dynamic information and empirical results using comprehensive panel data modeling rather than single-point modeling, improving the model's validity and accuracy. Two-dimensional panel data include timeseries and cross-sectional information, [12] as shown in Figure 1(c). Cross-sectional data indicate the locations of the observation points, as shown in Figure 1(a), and time-series data reflect the time effects of the prototype observation data, as shown in Figure 1(b).

One common method for panel data modeling is to cluster or divide the monitoring points in space according to the spatialtemporal characteristics of the panel data, through which monitoring points with similar characteristics can be classified together. When performing the cluster analysis, the following two fundamental issues should be considered: (i) the method to measure the similarity between monitoring
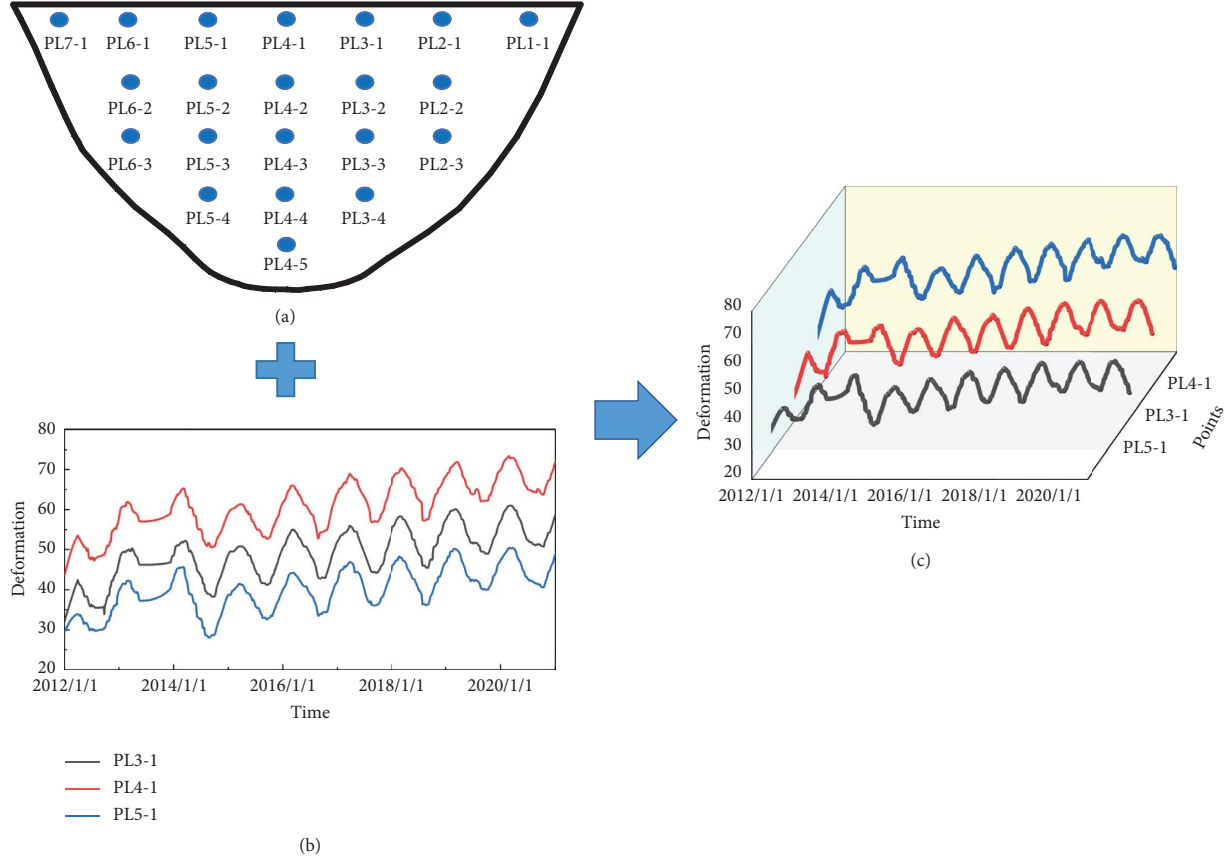
Figure 1: Panel data schematic: (a) cross-sectional data; (b) time-series data; (c) panel data.

points and (ii) the systematic clustering method to be used. In this paper, similarity measurement based on the comprehensive index of panel data was adopted, and the Ward-based hierarchical clustering method was used for the clustering.

To reflect the correlation and dynamic change information between the displacements at various areas of the arch dam, the time series of measured displacements at multiple monitoring points can be expressed in the form of panel data as follows:

$$\delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1T} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \cdots & \delta_{NT} \end{bmatrix}, \quad (1)$$

where δ_{it} $(I = 1, 2, \ldots, N; t = 1, 2, \ldots, T)$ is the actual measured data of dam displacement, N is the number of monitoring points, and T is the number of observations of each monitoring point time series.

We can divide panel data cluster analysis into single-indicator and multi-indicator forms [35]. Most current studies use the single-indicator panel data clustering method, which is simple to use with a wide application. However, the prototype dam monitoring data present a more complex pattern, and using a single-indicator cannot fully reflect the changing process and trend of each

monitoring point. Therefore, this study performed the clustering of dam deformation monitoring data using the following three basic similarity indicators: the absolute distance d_{ij} (AD), incremental distance d_{ij} (KD), and speed increase distance d_{ij} (GRD). Thus, the deformation monitoring values, such as absolute initial value, incremental range, growth rate, and acceleration, are thoroughly evaluated. The definitions of the three basic similarity indicators are found in the literature [12]. Their integration can help obtain a general similarity indicator to measure the overall similarity of the deformation monitoring points. This comprehensive index is as follows:

$$d_{ij}(CD) = \beta_1 d_{ij}(AD) + \beta_2 d_{ij}(KD) + \beta_3 d_{ij}(GRD), \quad (2)$$

where d_{ij} (CD), d_{ij} (AD), d_{ij} (KD), and d_{ij} (GRD) are the comprehensive, absolute, incremental, and speed increase distances between point i and point j, respectively, and β_1, β_2, and β_3 are the weight coefficients of d_{ij} (AD), d_{ij} (KD), and d_{ij} (GRD), respectively. The weight coefficients can be assigned subjectively or objectively based on the actual situation, and the weight coefficients in equation (2) were determined using the variation coefficient method [36].

Common cluster analysis methods mainly consist of partition and hierarchical clustering. Comparatively, hierarchical clustering is more suitable for dam monitoring points because the number of clusters of concrete arch dam deformation data cannot be determined in advance. This

paper used the Ward method as the clustering method for concrete dam deformation monitoring data [37].

Assuming that the N monitoring points of the dam can be divided into k regions, denoted as $G_1, G_2, \ldots, G_k$, within the deformation data, Y_i refers to the indicator value of the monitoring point $i\,(i = 1, 2, \ldots, N)$ in H, and $\overline{Y}$ is the indicator center of H_u. The sum of deviation squares W_u of the monitoring points in H_u is calculated based on the following equation:

$$W_u = \sum_{i=1}^{N_u} \left(Y_i - \overline{Y}\right)^2. \tag{3}$$

The pedigree clustering dendrogram and the deformation status of the monitoring points should be considered before the final determination of the number of clusters k. In practice, the number of hierarchical clusters can be defined by the abrupt change of the sum of deviation squares.

2.2. Bayesian Model Selection and Averaging Principle

2.2.1. Bayesian Model Selection.
BMS is a statistical analysis method that considers the uncertainty of the model based on the Bayesian theory. This modeling technology is commonly used to deal with model uncertainty problems [35–38]. This study used two cases for each factor of the statistical model, i.e., selected model and unselected model, resulting in a model space $\mathbf{M} = \{M_1, M_2, \ldots, M_v\}$ (composed of all

possible models) to be generated. The posterior probabilities of its parameters can be derived from equation (4):

$$P(\boldsymbol{\theta}|\mathbf{D}) = \sum_{j=1}^{v} P\left(M_j|\mathbf{D}\right) P\left(\boldsymbol{\theta}\big|M_j, \mathbf{D}\right), \tag{4}$$

where $\boldsymbol{\theta}$ is the parameter vector to be estimated, i.e., the coefficients of each factor of the statistical model; $\mathbf{D}$ is the observed data sample; and M_j is the jth model in the model space. The posterior density distribution of the parameter vector $\boldsymbol{\theta}$ is the weighted average of the posterior density distribution of $\boldsymbol{\theta}$ under the model space condition, i.e., the posterior probability $P(M_j|\mathbf{D})$ of the model, which the following equations can calculate:

$$P\left(M_j|\mathbf{D}\right) = \frac{P\left(\mathbf{D}\big|M_j\right) P(M_v)}{\sum_{h=1}^{v} P\left(\mathbf{D}\big|M_h\right) P(M_h)}, \tag{5}$$

$$P\left(\mathbf{D}\big|M_j\right) = \int P\left(\mathbf{D}\big|\theta_j, M_j\right) P\left(\boldsymbol{\theta}_j\big|M_j\right) d\boldsymbol{\theta}_j. \tag{6}$$

Equation (6) denotes the likelihood function integral corresponding to model M_j, where θ_j is to the parameter vector corresponding to model M_j, $P\left(\theta_j \mid M_j\right)$ is the prior probability distribution of parameters in model M_j, $P\left(\mathbf{D} \mid \theta_j, M_j\right)$ is the likelihood function of model M_j, and $P\left(M_j\right)$ is the prior distribution of model M_j.

We calculate the posterior mean and posterior variance of the parameter vector $\boldsymbol{\theta}$ using equations (4)–(6), which is expressed as the follows:

$$
\begin{aligned}
\mathbf{E}[\boldsymbol{\theta}|\mathbf{D}] &= \sum_{j=1}^{v} \mathbf{E}\left[\boldsymbol{\theta}\big|\mathbf{D}, M_j\right] P\left(M_j|\mathbf{D}\right), \\
\mathbf{Var}[\boldsymbol{\theta}|\mathbf{D}] &= \sum_{j=1}^{v} \left(\mathbf{Var}\left[\boldsymbol{\theta}\big|\mathbf{D}, M_j\right] + \left(\mathbf{E}\left[\boldsymbol{\theta}\big|\mathbf{D}, M_j\right]\right)^2\right) - \mathbf{E}[\boldsymbol{\theta}|\mathbf{D}]^2,
\end{aligned}
\tag{7}
$$

where $\mathbf{E}[\theta|\mathbf{D}]$ and $\mathbf{Var}[\theta|\mathbf{D}]$ are the posterior mean and posterior variance of the parameter vector $\boldsymbol{\theta}$, respectively, and $\mathbf{E}[\theta|\mathbf{D}, M_j]$ and $\mathbf{Var}[\theta|\mathbf{D}, M_j]$ are the posterior mean and posterior variance of $\boldsymbol{\theta}$ in model M_j, respectively.

2.2.2. Bayesian Model Averaging.
The posterior probabilities of each model can be used as the weights for inference, and the weighted average of the models of interest can be obtained to determine the BMA [39–43]. The BMA modeling approach is based on a large model space, but performing scientific sampling in the model space to identify models with high posterior probabilities remains a challenge in applying BMA.

Madigan and Raftery [44] proposed a highest probability model that used Occam's window approach to delete models with low posterior probabilities in the model space. In addition, Chib [45] recommended the Markov chain Monte Carlo (MCMC) method for

sampling in the model space, which is effective at dealing with high-dimensional space problems. The BMA method needs to calculate the marginal likelihood function of the posterior probability distribution of models in the model space, and the calculation process often includes high-dimensional and complex integrals of all parameters and hidden variables. The MCMC method can calculate the posterior distribution of the weights and variances of models and parameters quickly. Therefore, this study adopted the MCMC method for BMA.

3. Construction Process of Multipoint BMS and BMA Models for Arch Dams

By combining the panel data clustering, PCA dimensional reduction, BMS, and BMA, we inferred the model M_{mop} that could best reflect the properties of "real" dam engineering systems from model set $\mathbf{M}$ before analyzing the

spatialtemporal deformation properties of the arch dam. As shown in Figure 2, the arch dam multipoint model based on BMS and BMA included the following four main steps:

3.1. Data Preprocessing.

We extracted the data from the dam monitoring database. As the prototype dam observation data contain unequal time interval observations, coarse error processing and a missing value complement task should be conducted before the clustering analysis of observation data. The preprocessed data comprised a factor set and an effect-quantity set.

3.2. Panel Data Clustering.

We applied the Ward method for the dam body deformation and panel data clustering. First, based on the effect-quantity set obtained by preprocessing, we analyzed the deformation similarity characteristics, and then calculated the absolute distance, incremental distance, and speed increase distance between the monitoring points. Next, according to the three distance indicators, we calculated the weight coefficients using the variation coefficient method to obtain the comprehensive similarity indicator. Finally, based on the comprehensive similarity indicator, we conducted the spatialtemporal monitoring point clustering using the Ward hierarchical clustering to obtain the arch dam zoned deformation family, followed by a qualitative analysis.

3.3. Comprehensive Displacement Extraction.

The comprehensive displacement extraction of the deformation in the dam zones was realized through PCA dimension reduction analysis [9]. First, we performed the equalization treatment of the data before conducting the PCA analysis of the deformation data for each zone. Next, we calculated the co-variance matrix of each zone displacement matrix and obtained the eigenvalues and eigenvectors. In addition, the cumulative contribution information was calculated based on the eigenvalues, and the comprehensive displacement best reflecting the zoned deformation properties was selected from the principal components.

3.4. BMS and BMA

(a) We preselected the factor sets in line with the actual operation of the dam based on the qualitative analysis of the measured point deformation monitoring volume and the rigorous physical derivation theories of previous studies.

(b) We then constructed the BMS and BMA zoned data samples with comprehensive displacements, followed by the elimination of the influence of the factor set dimension and standardization of the dataset.

(c) The posterior inclusion probability (PIP) of each factor and model was calculated using the MCMC sampler, followed by the uncertainty analysis.

(d) We applied the probabilities to determine a weighted average of the models of interest. Models with higher posterior probabilities were assigned higher weights, and models with lower posterior probabilities were given lower weights to quantify the uncertainty of the models.

(e) The prediction results of BMA were compared to analyze the robustness of the model.

(f) Finally, we conducted the quantitative analysis and interpretation of the mechanistic causes of each zone based on the BMA results.

4. Case Study

4.1. Project Overview.

The water-retaining structure of a hydropower station proposed in this study was a concrete double-curvature thin arch dam arranged with seven sets of vertical lines used to monitor the dam body's horizontal displacement. Each line set included a combination of positive and inverted vertical lines, totaling 28 positive vertical line monitoring points and seven inverted vertical line monitoring points. Table 1 defines the radial and tangential displacements of the vertical line system in the positive direction, and the monitoring layout is shown in Figure 3.

Figures 4(a)–4(f) show the radial displacement data of each monitoring point on the dam. The time series of the data analyzed in this paper was from January 2014 to December 2020, with a sampling frequency of once every seven days. A total of 365 datasets were collected at each monitoring point. Figure 5 shows the water level and air temperature at the upstream reservoir as functions of time. Figures 4 and 5 yielded the following findings:

(i) The radial displacement of the arch dam shows strong temporal and spatial regularity. From Figures 4(b)–4(f), it can be seen that the periodicity of the measurement point of the top part of the 2–6 dam of the perpendicular dam section is obvious, and the periodicity is weaker the closer to the dam foundation, so it is necessary to cluster the measurement point data

(ii) The water level of the upstream reservoir had two short-lived lifts during 2015, followed by small water level changes.

(iii) A six-month gap in air temperature data was present in 2018.

4.2. Cluster Analysis of Concrete Arch Dam Deformation Based on Comprehensive Similarity Indicator.

This study adopted the deformation monitoring data during the dam's operation period. We calculated d_{ij} (AD), d_{ij} (KD), and d_{ij} (GRD) according to the absolute initial value, incremental range, growth rate, and acceleration of the deformation. Additionally, we determined the weight coefficients using the variation coefficient method, as shown in Table 2. Based on the weight coefficients, the indexes had influences on the clusters, and the absolute distance accounted for a large proportion.
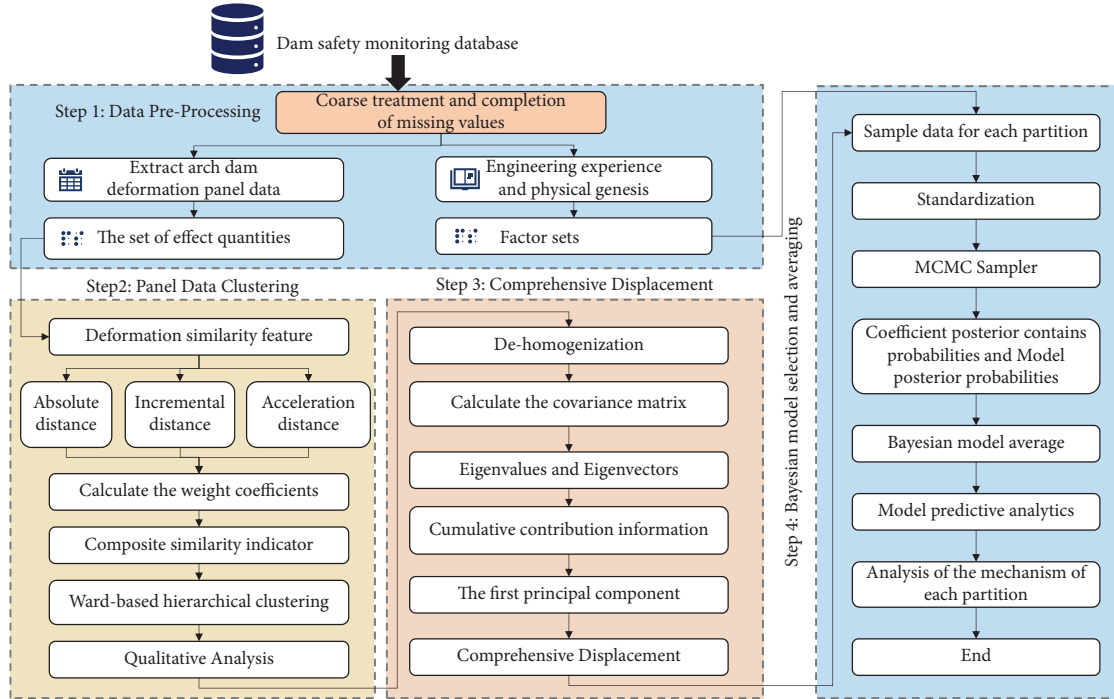
FIGURE 2: Flow chart of the multimeasurement point modeling method of arch dam based on Bayesian model selection and averaging.

TABLE 1: Radial and tangential displacement positive direction of the perpendicular system.

Perpendicular number	Dam section (#)	Radial positive direction (°)	Tangentially to the positive direction (°)
No. 1	1	NE 40.5	NE 130.5
No. 2	4	NE 42.1	NE 132.1
No. 3	7	NE 49.4	NE 139.4
No. 4	11	NE 81.0	NE 171.0
No. 5	16	NE 112.6	NE 202.6
No. 6	19	NE 119.8	NE 119.8
No. 7	22	NE 121.3	NE 211.3

Then, we calculated the comprehensive similarity indicator and performed the spatialtemporal multipoint clustering accordingly to obtain the pedigree clustering dendrogram. In Figure 6, the selected monitoring points are represented by the horizontal coordinates, and the vertical coordinates are the sum of deviation squares of the indexes between the monitoring points. The spectrum diagram Figure 6(b) shows that smaller link distances between clusters yielded more clusters. As shown in Figure 6(a), the sum of deviation squares had two obvious points of abrupt change, so we chose three clusters. Figure 7 shows the clustering results of the spatial monitoring points, and Figure 8 illustrates the measurement process hydrograph of the monitoring points in each zone.

Figures 7 and 8 yield the following findings:

(i) The radial deformation of the arch dam presented a gradual, downward trend from the middle of the dam to the shoulders on both sides, with the largest deformation in the middle of the dam roof, and good symmetry of the deformation distribution.

(ii) The monitoring points in each zone showed high spatialtemporal correlation. For example, Zone I, located in the middle of the arch dam top, was easily affected by temperature fluctuations. Zone II, in the middle of the arch dam, was similar, especially at a high water level, while Zone III, located at the dam base, was less affected by temperature fluctuations.

(iii) According to Zone I, the deformation peak fluctuations and temperature changes showed a certain negative correlation and hysteresis. For instance, the air was in the high-temperature peak stage during August, while the deformation showed a lower peak in September.

4.3. Comprehensive Displacement Extraction for Each Deformation Zone of Concrete Arch Dam. The above results reveal that the clustering in Zones I–III was more concentrated and Zone IV on both sides of the dam shoulder exhibited dispersed clustering. Using the PCA to obtain the
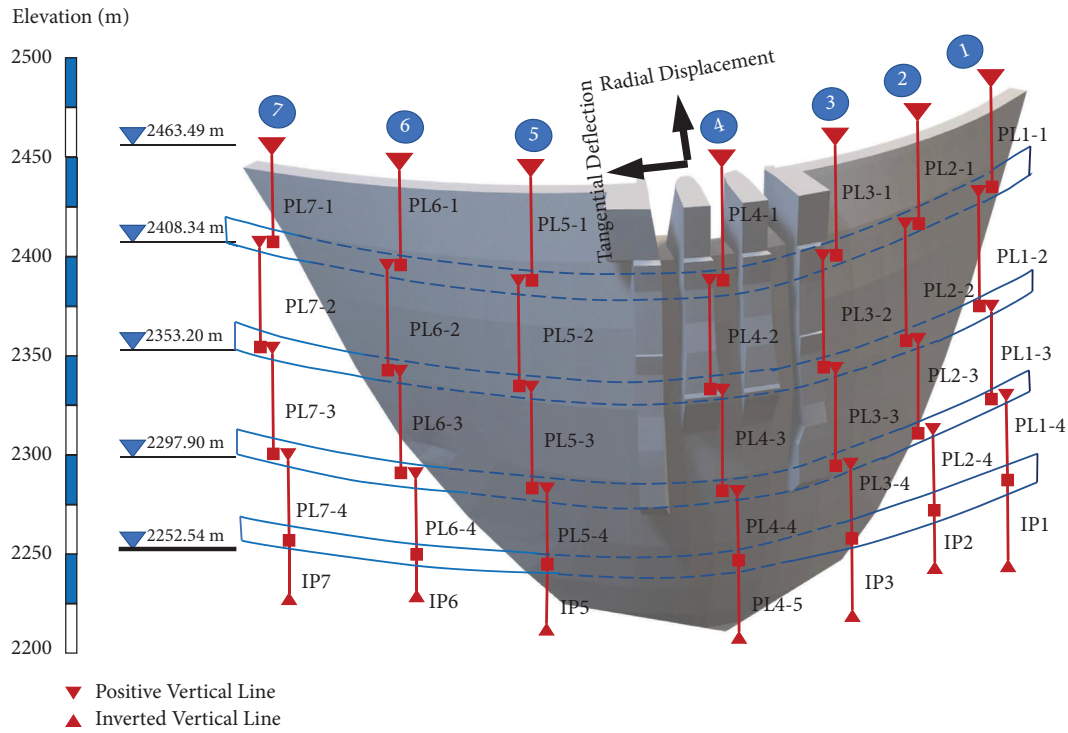
Figure 3: Schematic diagram of the monitoring layout of the positive inverted line of an arch dam.
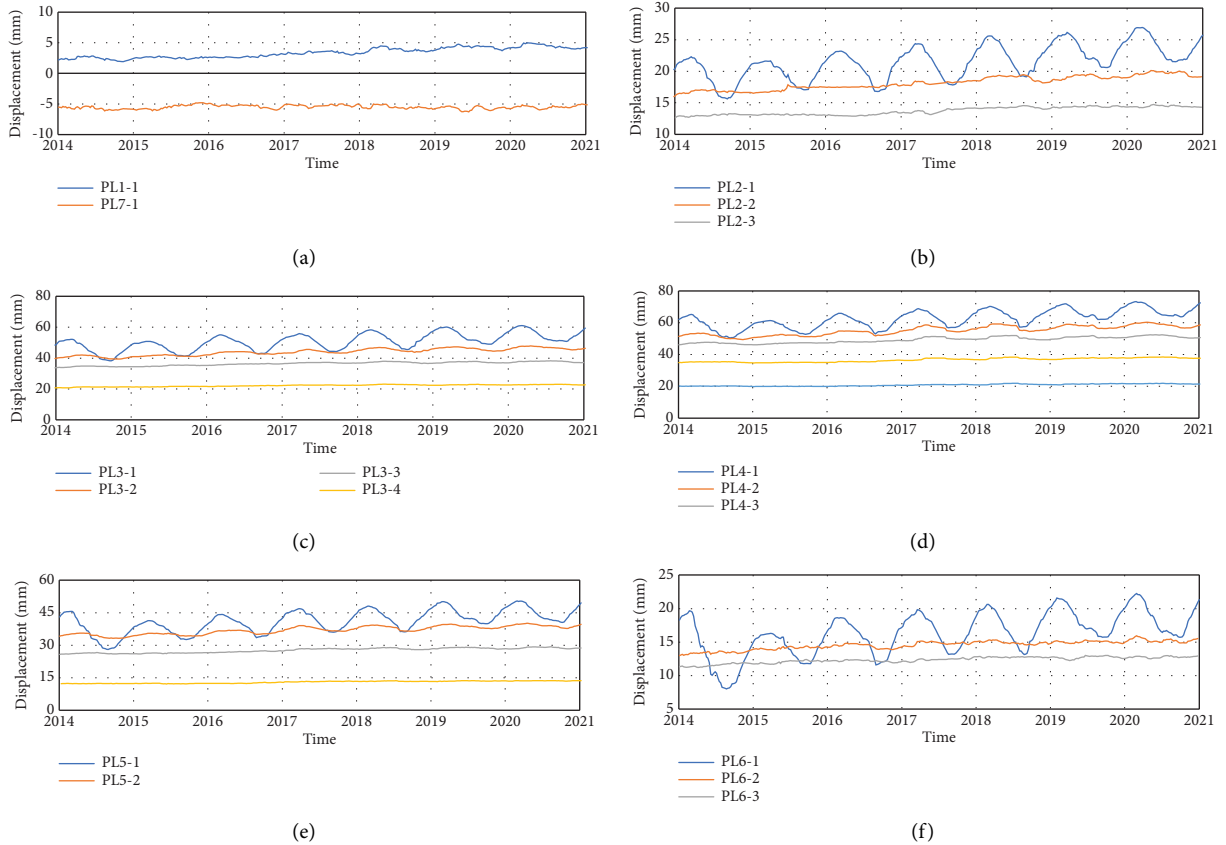

Figure 4: Measured radial displacement: (a) perpendicular number no. 1 and no. 7, (b) perpendicular number no. 2, (c) perpendicular number no. 3, (d) perpendicular number no. 4, (e) perpendicular number no. 5, and (f) perpendicular number no. 6.
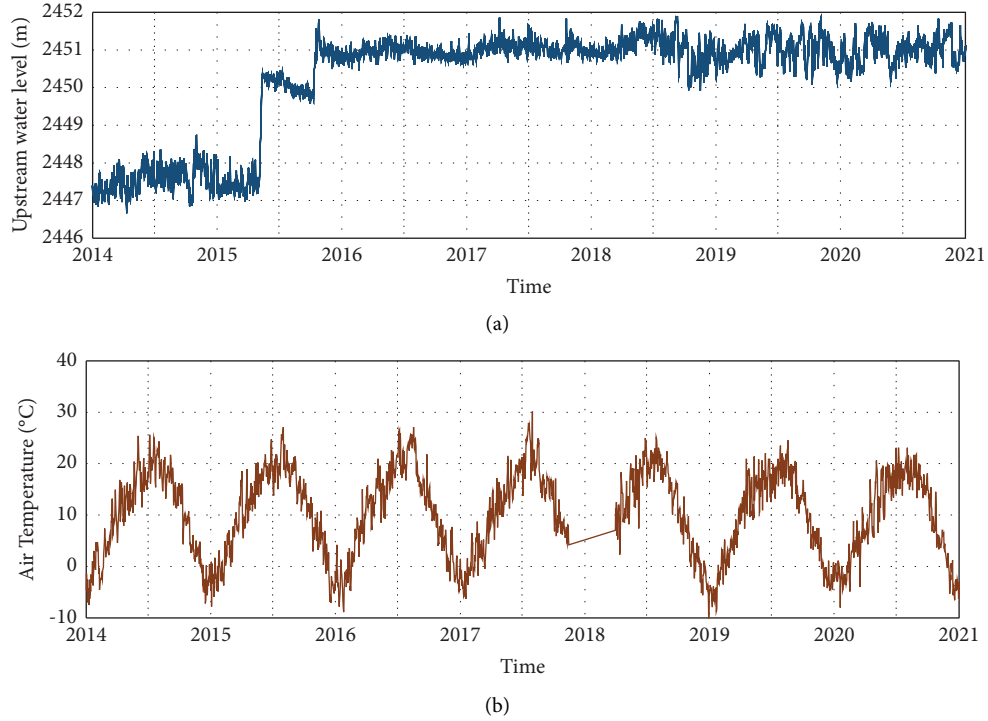
(a)

(b)

FIGURE 5: Time series of environmental quantities: (a) upstream water level; (b) air temperature.

TABLE 2: The weight coefficient is obtained by using the coefficient of the variation method.

Distance indicator	Absolute distance (d_{ij} (AD))	Incremental distance (d_{ij} (KD))	Growth distance (d_{ij} (GRD))
Weight coefficient	0.373	0.283	0.344

(a)

(b)
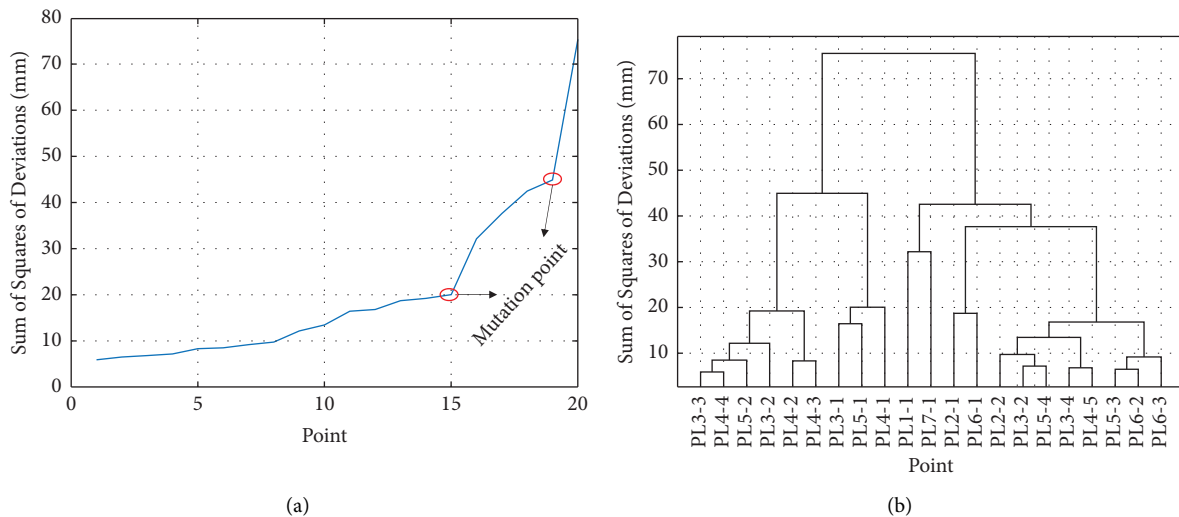
FIGURE 6: Arch dam deformation sequence division lineage cluster tree: (a) sum of squares of deviations; (b) genealogy chart.

comprehensive displacement law of multiple monitoring points is a means to extract the deformation characteristics of multiple monitoring points [9]. This research extracted the comprehensive displacement (PC) for multiple monitoring points of Zones I–III, as shown in Figure 9. Figure 9(a) displays the eigenvalues and the cumulative

contribution rate, and Figure 9(b) compares the comprehensive displacement PC of each zone with the deaveraging information of the monitoring points.

Figure 9 provides the following information:

(i) The first principal component of Zones I–III contained more than 95% of the information on the
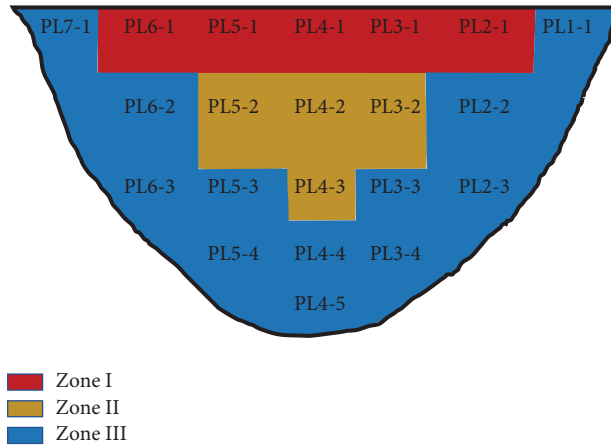
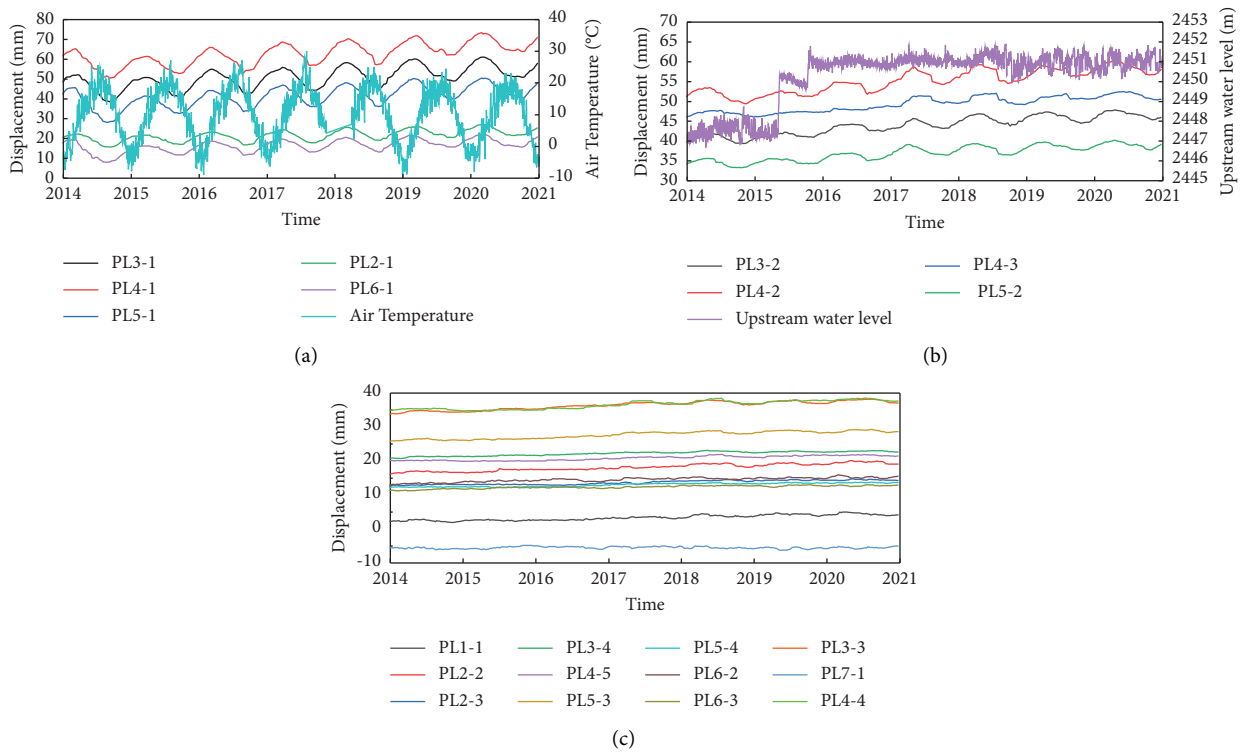FIGURE 7: Arch dam deformation cluster zoning.

FIGURE 8: Clustering result process line analysis diagram: (a) Zone I; (b) Zone II; (c) Zone III.

sequence of measured values of the zoned monitoring points. In addition, the eigenvalues showed a downward trend, as shown in Figure 9(a), indicating that the first principal component meets the requirement for analyzing the comprehensive information of multiple monitoring points in the zone.

(ii) According to Figure 9(b), the upward or downward trend of the comprehensive displacement was more evident than that of the monitoring points. Additionally, in Zone II from 2016 to 2017, the comprehensive displacement mainly occurred at the

mean value of several data in the zone if the change of each monitoring point was stable. Furthermore, the periodicities of the comprehensive displacement of each zone and each monitoring point were nearly equal, and the zones' periodic variations were larger than that of each monitoring point of the zone.

(iii) The comprehensive displacement retained the main information of several monitoring points and amplified the small changes in the overall law of zones, which was conducive to the analysis of the deformation properties in different zones of the arch dam.
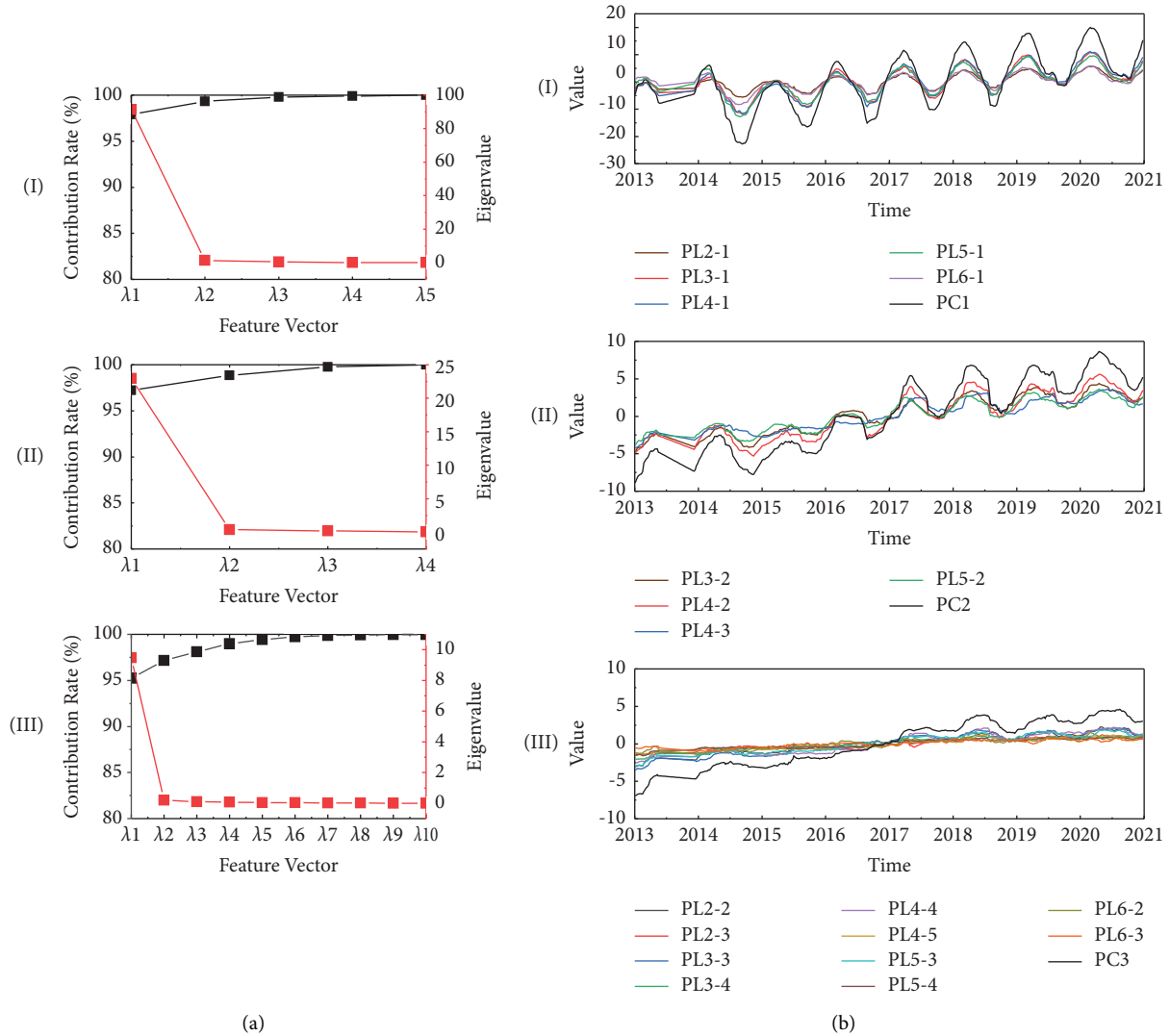
Figure 9: Comprehensive displacement analyses.

4.4. Arch Dam Multipoint BMS and BMA Analysis. The above analysis showed that Zone I had significant periodic variations. In this section, we first used BMS for the model set selection based on the engineering experience and the physical causes of the monitoring volume and then adopted BMA for uncertainty interpretation.

4.4.1. Determination of Model Class. According to the working principle of concrete arch dams, the main statistical model factors of arch dam deformation include water pressure, temperature, and time dependence. A priori knowledge of dam system indicates that different components have specific optional factor sets.

  (i) For the water-pressure component, a priori information shows that its factor set comprises the primary, secondary, tertiary, and quadratic upstream water depths. However, the lift pressure of the arch dam has a negligible impact on the displacement and is therefore not considered.

  (ii) For the temperature component, the dam has been in service for over ten years, and the dam temperature field is in a quasi-stable temperature field state, so the multi-period (one year and six months) harmonic combination can be used as the factor set of temperature component.

  (iii) For the time-dependent component, after more than ten years in operation, the dam's time-dependent deformation of the concrete and rock mass is mainly irreversible creep with slight restorable creep. Therefore, the creep factor of the concrete and rock mass caused by water level ebb and flow is not considered. In addition, the optional domain of the time-dependent component is time polynomial (primary to quadratic), time logarithm, and *e*-exponential of time. As a result, the optional factor sets for each component of the arch dam are shown in Table 3, with the statistical model expressed as the following equation:

TABLE 3: Factor sets selected by the Bayesian models.

Ordinal	The component name	The parameter value	Component symbols
X1			H^1
X2	Water-pressure component	Upstream water depth H	H^2
X3			H^3
X4			H^4
X5			$\sin(1 \times \alpha)$
X6			$\cos(1 \times \alpha)$
X7	Temperature component	$\alpha = 2\pi t/365$	$\sin(2 \times \alpha)$
X8			$\cos(2 \times \alpha)$
X9			Air temperature T
X10			τ
X11			τ^2
X12	Aging components	$\tau = (t - t_0)/100$	τ^3
X13			τ^4
X14			$\ln \tau$
X15			$e^{-\tau}$

$$Y_t = a_0 + \sum_{i=1}^{4} a_i H^i + \sum_{i=1}^{2} \left[b_{1i} \sin(i\alpha) + b_{2i} \cos(i\alpha) \right] + b_3 T + \sum_{i=1}^{4} c_i \tau^i + c_5 \ln \tau + c_6 e^{-\tau}, \tag{8}$$

where Y_t is the zone's comprehensive displacement; a_i, b_{1i}, b_{2i}, b_3, and $c_1–c_6$ re to the regression coefficients of each component, respectively; H is the water depth in front of the dam, $\alpha = 2\pi t/365$, and τ is the number of days from the monitoring date to the starting date divided by 100.

4.4.2. Model Selection Strategy Analysis. The comprehensive displacement data (PC1, PC2, and PC3) were used for the BMS of Zones I–III. We employed 15 factors in the full model, and Table 4 lists their statistics (mean, median, minimum, and maximum). Large numerical differences were evident between the comprehensive displacements and the factor sets. Therefore, data standardization was required before model selection to eliminate the magnitude and numerical differences.

(1) BMS. Each factor has two possibilities during model selection, i.e., selected or not selected. So the best model is one of 32,768 (or 2^{15}) possible models. This study aims to find the best parsimonious model for future prediction and dam health monitoring. For model selection in this study, we explored options through the MCMC sampler based on the data from Zone I. A total of 10 [6] sample models were simulated using the MCMC algorithm, while also investigating the convergences. The coefficient PIP convergence plot of Zone I is shown in Figure 10, and Figure 11 shows the model posterior probability convergence plot. Notably, most points remained on the 45° diagonal, and most coefficient PIP and model posterior probabilities of the MCMC method converged to the theoretical posterior probabilities, indicating strong reliability.

For a clearer view of the model comparison, we used the best model space image (Figure 12) to depict the model's uncertainty. Figure 12 shows the predictors, including the intercept on the Y-axis, while different models are plotted on the X-axis, i.e., each vertical column corresponding to a model and each horizontal row representing a variable. The black rectangles in Figure 12 indicate the variables that were not included, and the color of each column is proportional to the logarithm of the posterior probability. Models with the same color had similar posterior probabilities and could be clustered together. The analysis of Figure 12 shows that at least three factors were not selected in the models with the top 18 posterior probabilities, and $\cos(2\alpha)$, T, τ^2, τ^3 and τ^4 were not selected in most models. This phenomenon is consistent with the results of the PIPs of the model coefficients in Figure 13, in which the PIPs of all three factors were less than 0.5, while those of the other factors were greater than 0.5.

(2) BMA. After treatment with the MCMC sampler, we performed BMA on all possible models and estimated the parameters of BMA models. Table 5 shows the posterior mean, posterior standard deviation, and PIP of each coefficient estimated using BMA. In the model, the PIPs of the intercept and $\cos(\alpha)$ were 1, and the PIPs of other factors, except $\cos(2\alpha)$, T, τ^2, τ^3, and τ^4, were all greater than 0.5, indicating that other factors were most likely to be included in the model compared with $\cos(2\alpha)$, T, τ^2, τ^3, and $\tau^4 w$. The posterior probability distribution of each factor is shown in Figure 14.

After obtaining the model posterior probabilities, we performed BMA and plotted the comparison diagram between fitted and measured values (Figure 15). The monitoring points of Zones I–III and PL1-1 were all close to the 45° diagonal, implying the BMA fitting was effective.

TABLE 4: Statistics of the comprehensive displacement, PL1-1 measured value and factor set of zone I–zone III.

| Component	Mean	Median	Min	Max
H^1	$1.737E + 02$	$1.743E + 02$	$1.700E + 02$	$1.752E + 02$
H^2	$3.016E + 04$	$3.036E + 04$	$2.892E + 04$	$3.071E + 04$
H^2	$5.238E + 06$	$5.291E + 06$	$4.917E + 06$	$5.381E + 06$
H^4	$9.099E + 08$	$9.220E + 08$	$8.361E + 08$	$9.430E + 08$
$\sin(\alpha)$	$4.447E - 04$	$4.452E - 03$	$-1.000E + 00$	$1.000E + 00$
$\cos(\alpha)$	$-2.446E - 04$	$-6.455E - 04$	$-1.000E + 00$	$1.000E + 00$
$\sin(2\alpha)$	$4.293E - 04$	$-1.291E - 03$	$-1.000E + 00$	$1.000E + 00$
$\cos(2\alpha)$	$2.723E - 04$	$-1.900E - 03$	$-1.000E + 00$	$1.000E + 00$
T	$9.861E + 00$	$1.030E + 01$	$-6.600E + 00$	$2.710E + 01$
τ	$9.861E + 00$	$1.030E + 01$	$3.640E + 00$	$2.912E + 01$
τ^2	$3.227E + 02$	$2.683E + 02$	$1.325E + 01$	$8.480E + 02$
τ^3	$7.068E + 03$	$4.395E + 03$	$4.823E + 01$	$2.469E + 04$
τ^4	$1.649E + 05$	$7.199E + 04$	$1.756E + 02$	$7.191E + 05$
$\ln \tau$	$6.177E + 00$	$6.304E + 00$	$4.807E + 00$	$6.879E + 00$
$e^{-\tau}$	$3.530E - 02$	$7.000E - 07$	$4.410E - 13$	$9.700E - 01$
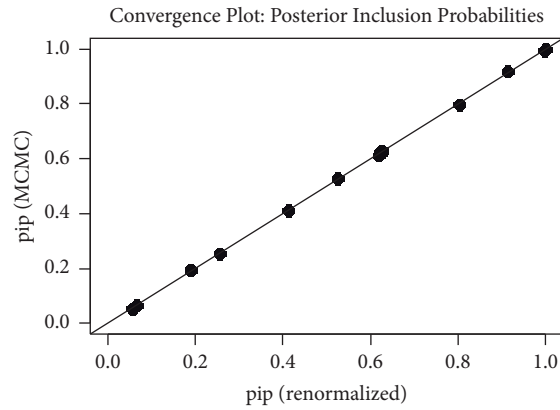
FIGURE 10: Coefficient posterior contains probability (pip) convergence plots.
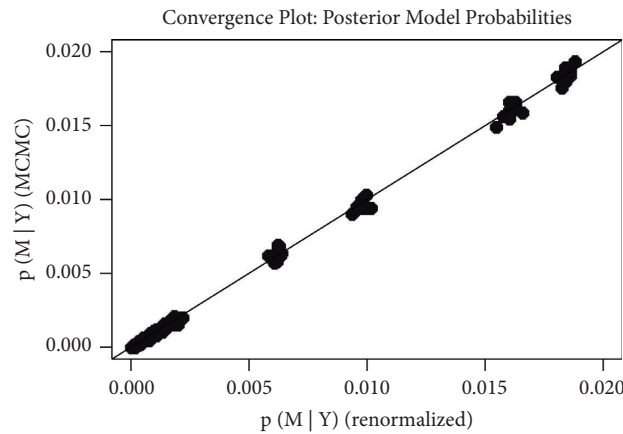

FIGURE 11: Model posttest probability convergence diagram.

(3) Predictive Analysis. To verify the model's predictive ability, we chose various test samples to compare the average prediction results of the model and the commonly used ordinary least-squares (OLS) and stepwise regression (SR) methods. The test sample sizes were 10, 40, and 70, and the mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and root-mean-square error (RMSE) were the indicators for model evaluation. The specific calculations are shown in Table 6.

Figure 16 shows the prediction results, in which Figures. 16(a), (a2), and (a3) are the radar plots of the prediction results of Zone I with different test sample sizes and
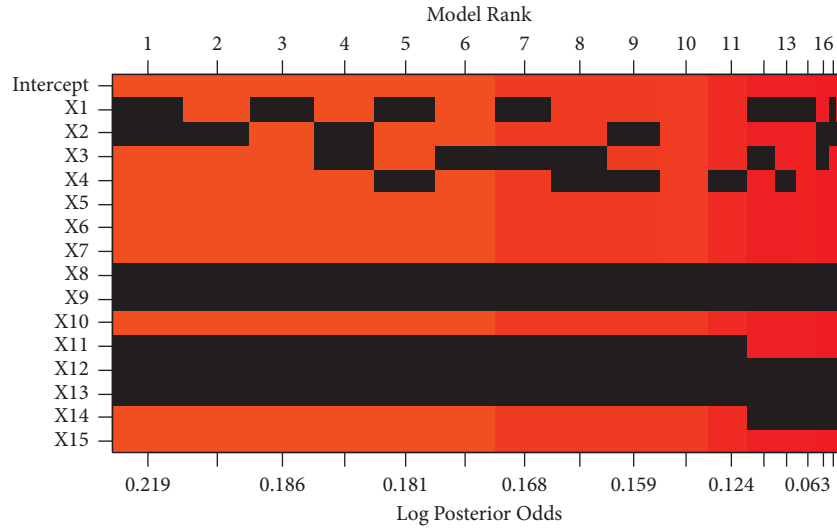
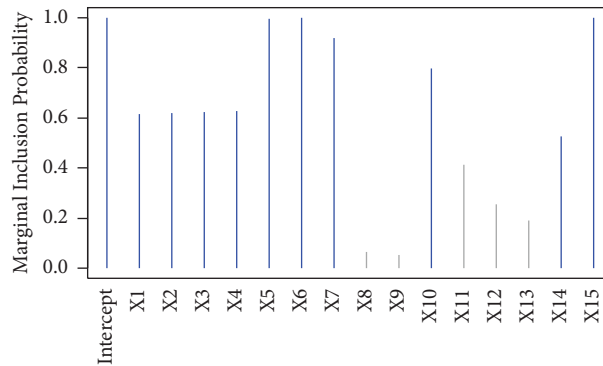FIGURE 12: Description of the uncertainty of the model.

FIGURE 13: Postindex inclusion probabilities for each coefficient.

TABLE 5: Estimation of posterior mean, posterior standard deviation, and posterior inclusion probability (PIP) of partition I coefficients using the BMA model.

Ordinal	Post mean	Post SD	PIP = P (B! = 0)
Intercept	$-1.206E + 00$	$8.981E - 02$	$1.000E + 00$
X1	$4.189E + 01$	$4.338E + 01$	$6.299E - 01$
X2	$1.356E + 01$	$6.737E + 01$	$6.304E - 01$
X3	$-1.206E + 01$	$6.870E + 01$	$6.375E - 01$
X4	$-4.292E + 01$	$4.415E + 01$	$6.397E - 01$
X5	$7.356E - 01$	$1.107E - 01$	$9.988E - 01$
X6	$7.639E + 00$	$1.116E - 01$	$1.000E + 00$
X7	$-2.649E - 01$	$1.192E - 01$	$9.135E - 01$
X8	$-1.189E - 01$	$1.303E - 01$	$5.329E - 01$
X9	$-8.225E - 04$	$7.249E - 02$	$6.129E - 02$
X10	$5.125E + 00$	$2.593E + 01$	$2.927E - 01$
X11	$-1.206E + 00$	$4.585E + 01$	$3.290E - 01$
X12	$-4.235E + 00$	$4.400E + 01$	$4.232E - 01$
X13	$4.926E + 00$	$1.800E + 01$	$4.936E - 01$
X14	$3.924E + 00$	$6.931E + 00$	$7.588E - 01$
X15	$2.414E + 00$	$3.164E - 01$	$9.967E - 01$

Figures 16(b) and 16(c) are the radar plots of the prediction results of different zones with a test sample quantity of ten. Figure 16 reveals the following:

(i) The results of different test samples of Zone I showed that the prediction performance was best when the test sample size was ten and worst with a test sample
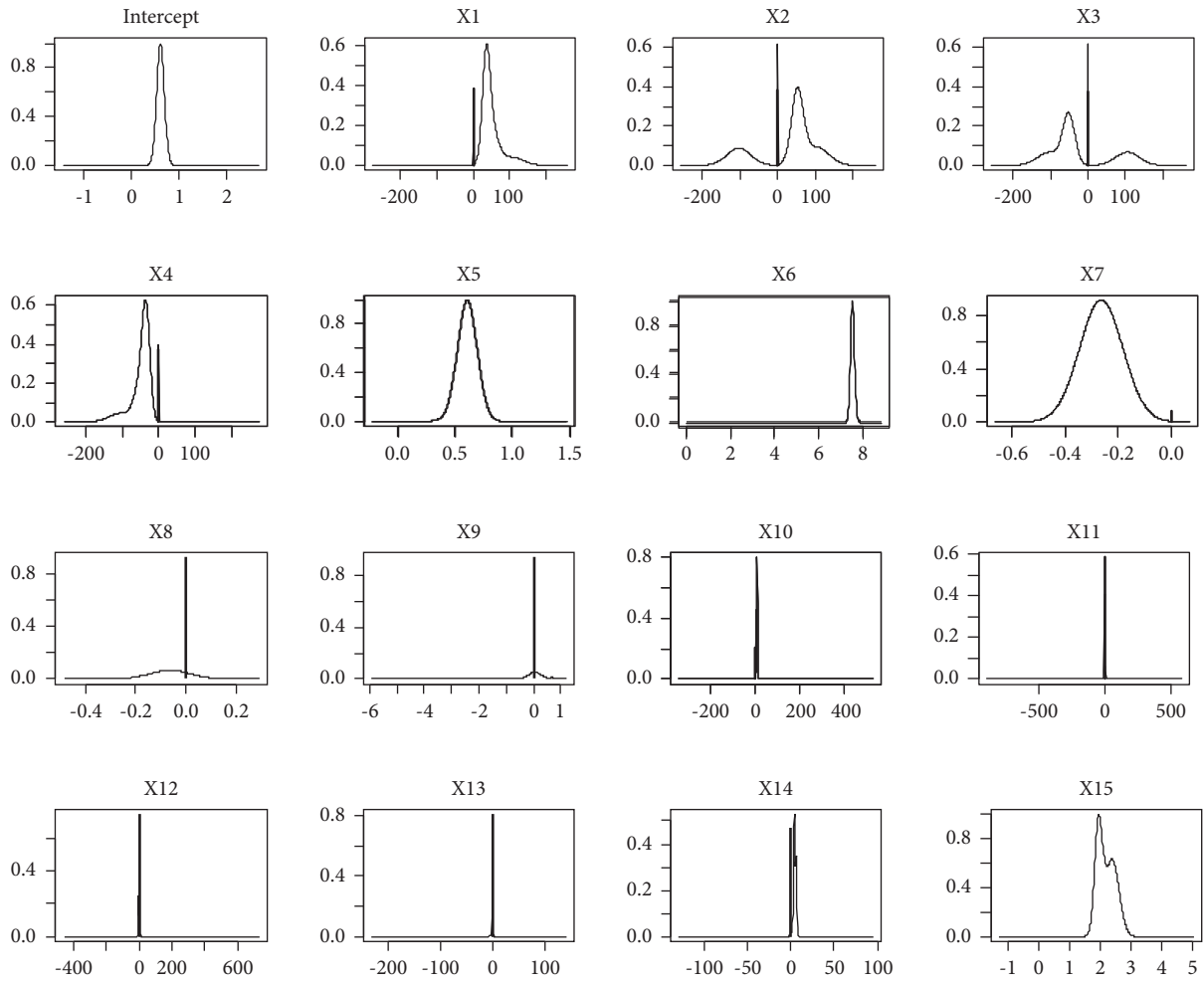
FIGURE 14: Schematic diagram of the posterior probability distribution of each coefficient in the BMA model.
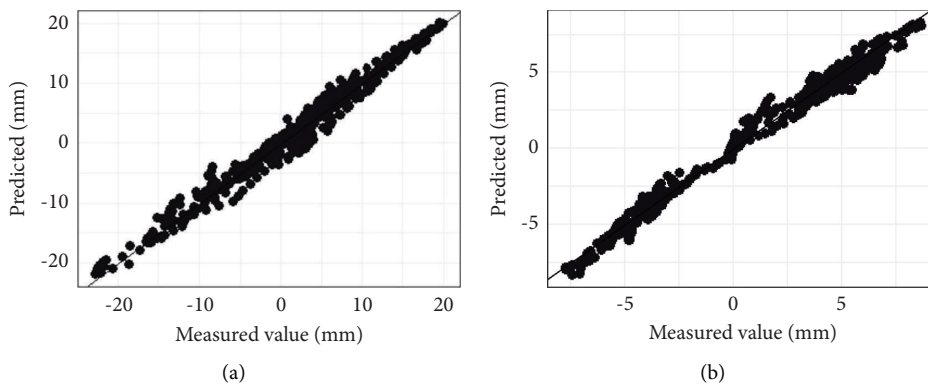

(a)

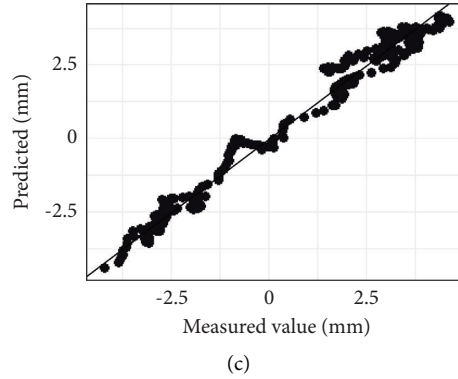(b)

FIGURE 15: Continued.

(c)

FIGURE 15: Comparison of the fitted values of each partition with the observed values: (a) Zone I; (b) Zone II; (c) Zone III.

TABLE 6: Model prediction evaluation indicators.

Evaluation indicators	Calculation formula
Mean absolute error	$\text{MAE} = 1/N\sum_{i=1}^{N}|y_i - F(x_i)|$
Mean square error	$\text{MSE} = \sum_{i-1}^{N}(y_i - F(x_i))^2/N$
Mean absolute percentage error	$\text{MAPE} = 100/N\sum_{i=1}^{N}|(y_i - F(x))/y_i|$
Root-mean-square error	$\text{RMSE} = \sqrt{\sum_{i-1}^{N}(yi - F(x_i))^2/N} = \sqrt{\text{MSE}}$

size of 40. The BMA had the best prediction performance and better prediction evaluation indicators of MSE, RMSE, MAE, and MAPE than the OLS and SR methods.

(ii) Comparing the results of different zones in the case of a test sample size of 10, the prediction results of BMA in all zones were better than those of other models, except for the BMA in Zone III, which had a worse prediction performance than the SR method.

In summary, the BMA method could assign more weights to the models containing more information with better accurate predictions, thus realizing the robustness of the model and achieving the best predictions. In addition, this result verified the scientificity of the BMA method.

4.5. Analysis of Arch Dam Body Zoned Deformation Mechanism. A conventional statistical model can only evaluate the impact of deformation components for a single monitoring point. To solve this problem, this paper proposes a model featured with regionality and integrity in evaluating the impact of deformation components to better reflect the deformation mechanism of the arch dam. Figures 17 and 18 show the contributions of the components of each zone to the zoned deformation, as described as follows:

(i) The contributions of the temperature and time-dependent components were dominant, while water pressure had little influence on the deformation, consistent with the characteristics of the daily regulation capacity of the arch dam. From the spatial perspective, the water-pressure component in Zones I and II was more significant and concentrated in 2013–2016, with little influence in other years. The main reason is that both zones are located in the middle of the dam and the water-pressure contribution mainly occurs in the reservoir storage period.

(ii) The temperature component's contribution showed a spatial trend of gradual decrease from Zone I to Zone III (Figure 18). When the temperature decreased, all the zones presented a downstream deformation trend, and a temperature increase yielded an upstream deformation trend. In addition, the temperature component effect showed a certain hysteresis compared with the measured temperature, which was more significant closer to the shoulder of the dam.

(iii) The contribution of the time-dependent component presented a spatial trend of gradual increase from Zone I to Zone III, in which the middle of the dam body was less affected than both sides of the dam shoulder.

(a)   (b)   (c)
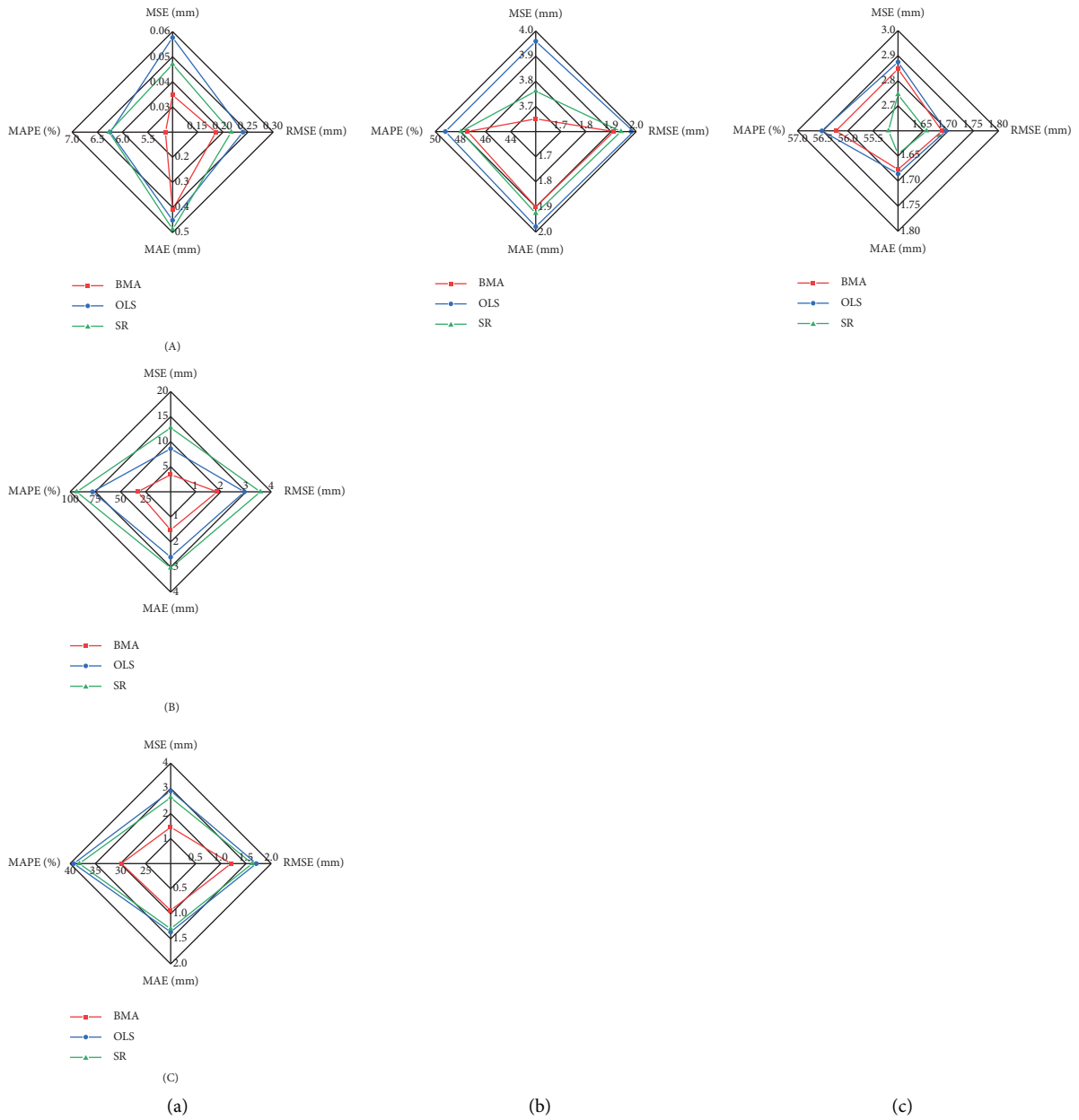
Figure 16: Radar charts of predictors for different test samples in each division: (A), (B), (C) comparison of different sample sizes; (A), (b), (c) comparison of different partitions. (a) Zone I: (A) the test sample is 10, (B) the test sample is 40, and (C) the test sample is 70; (b) Zone II: the test sample is 10; (c) Zone III: the test sample is 10.
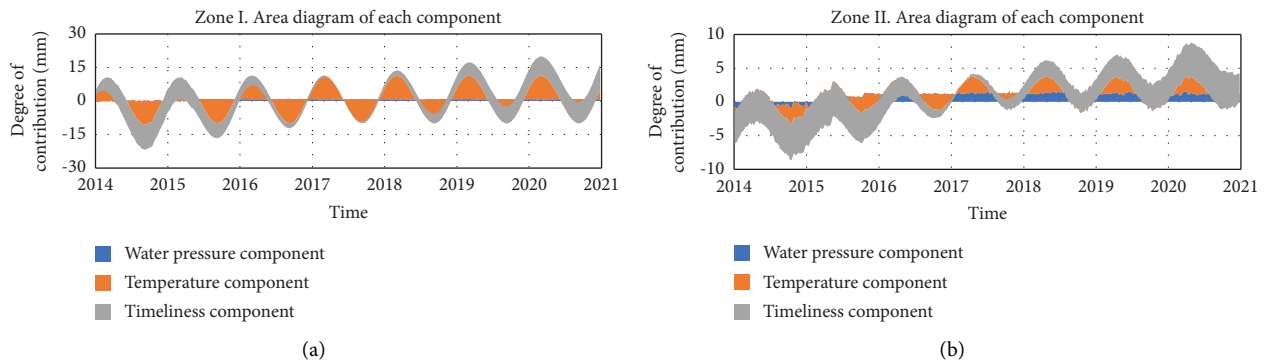

(a)   (b)
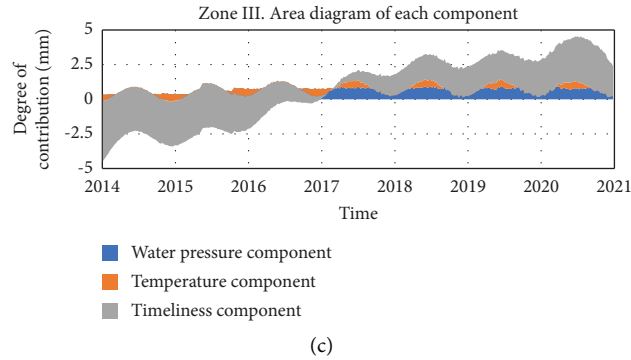
Figure 17: Continued.

(c)

Figure 17: The contribution of the components of each partition of the dam to the deformation: (a) Zone I; (b) Zone II; (c) Zone III.
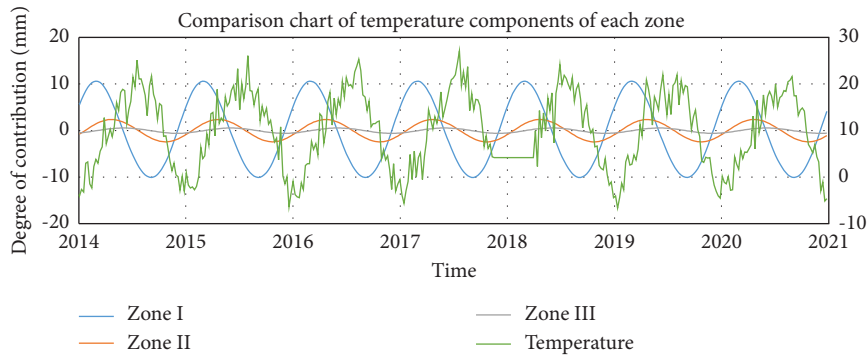

Figure 18: Comparison of the contribution of each partition of the temperature component.

5. Conclusions

Based on a limited deformation sequence, this paper distinguishes the deformation laws of different locations of a dam through cluster analysis, and comprehensive displacements were extracted through PCA, BMS, and BMA. In the full factor set, we used BMS to identify the influential factors at different dam locations to analyze the zone mechanism. The usability of the model was verified with the arch dam, and the results showed that the model could effectively identify the deformation characteristics of each zone. The main conclusions of this paper include the following:

(i) The deformation laws of different locations of the arch dam are distinguished based on the panel data clustering of similarity indicators, considering various deformation characteristics, such as the absolute initial value of deformation, incremental range, growth rate, and acceleration. The zone results are consistent with the deformation laws of the dam.

(ii) The comprehensive displacement based on PCA amplified the common small changes in deformation. The first principal component contains more than 95% of the information at the zone monitoring points, enabling a better analysis of the evolution of the safety properties of all the zones of the observed arch dam.

(iii) BMS and BMA can analyze the deformation properties of arch dams effectively by considering the model uncertainty for optimal models as well as integrated models. The results showed that BMA performs well in dam performance monitoring.

(iv) The analysis of the BMA mechanism showed that, under the effect of daily storage, water-pressure contributes little to the deformation of each zone. In addition, the temperature component is mainly reflected in the middle of the dam body, showing a certain hysteresis effect. The time-dependent component is evident in both sides of the dam shoulder.

Data Availability

The data used to support the finding of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] B. F. Zhu, "On the expected life span of concrete dams and the possibility of endlessly long life of solid concrete dams," *Journal of Hydraulic Engineering*, vol. 43, pp. 1–9, 2012.

[2] C. S. Gu, H. Z. Su, and S. W. Wang, "Advances in calculation models and monitoring methods for long-term deformation behavior of concrete dams," *Journal of Hydroelectric Engineering*, vol. 35, pp. 1–14, 2016.

[3] Y. Y. Huang, Y. Y. He, and Z. Z. Shen, "Optimization method of optimal statistical model of dam monitoring data," *Journal of Hydraulic Engineering*, vol. 53, pp. 154–164, 2022.

[4] J. Hu and S. H. Wu, "Statistical modeling for deformation analysis of concrete arch dams with influential horizontal cracks," *Structural Health Monitoring*, vol. 18, no. 2, pp. 546–562, 2018.

[5] J. Mata, A. Tavares de Castro, and J. Sá da Costa, "Constructing statistical models for arch dam deformation," *Structural Control and Health Monitoring*, vol. 21, no. 3, pp. 423–437, 2014.

[6] J. Hu and F. Ma, "Statistical modelling for high arch dam deformation during the initial impoundment period," *Structural Control and Health Monitoring*, vol. 27, no. 12, p. 2638, 2020.

[7] S. W. Wang, Y. L. Xu, C. S. Gu, T. Bao, Q. Xia, and K. Hu, "Hysteretic effect considered monitoring model for interpreting abnormal deformation behavior of arch dams: a case study," *Structural Control and Health Monitoring*, vol. 26, no. 10, p. 2417, 2019.

[8] Z. R. Wu and B. Chen, "A review on development of dam safety monitoring models," *Moder Surveying and Mapping*, vol. 39, pp. 1–11, 2016.

[9] B. W. Wei, B. Liu, and F. G. Xu, *Multi-point Hybrid Model Based on PSO-SVM for concrete Arch Dam Deformation Monitoring*, Geomatics and Information Science of Wuhan University, Wuhan, China, 2021.

[10] X. Y. Xu, J. Yang, C. H. Ma, X. Qu, J. Chen, and L. Cheng, "Segmented modeling method of dam displacement based on BEAST time series decomposition," *Measurement*, vol. 202, Article ID 111811, 2022.

[11] S. W. Wang, C. Xu, Y. Liu, and B. Wu, "Mixed-coefficient panel model for evaluating the overall deformation behavior of high arch dams using the spatial clustering," *Structural Control and Health Monitoring*, vol. 28, no. 10, p. 2809, 2021.

[12] T. Y. Hu, "Spatial and temporal clustering model of concrete arch dam deformation data based on panel data analysis method," *Journal of Yangtze River Scientific Research Institute*, vol. 38, pp. 39–45, 2021.

[13] J. M. Wang, C. S. Gu, and C. Zhang, "Deformation behavior analysis of Jinping arch dam based on spatiotemporal model of variable intercept panel data," *Journal of Hydroelectric Engineering*, vol. 39, no. 11, pp. 21–30, 2020.

[14] J. Hu and F. Ma, "Zoned deformation prediction model for super high arch dams using hierarchical clustering and panel data," *Engineering Computations*, vol. 37, no. 9, pp. 2999–3021, 2020.

[15] S. W. Wang, T. F. Bao, and K. Hu, "PCA -based monitoring model for spatial integration of deformation of high concrete dam," *Water Resources and Hydropower Engineering*, vol. 49, pp. 123–127, 2018.

[16] B. Chen, T. Y. Hu, Z. S. Huang, and C. Fang, "A spatio-temporal clustering and diagnosis method for concrete arch dams using deformation monitoring data," *Structural Health Monitoring*, vol. 18, no. 5-6, pp. 1355–1371, 2019.

[17] X. Li, Z. P. Wen, and H. Z. Su, "An approach using random forest intelligent algorithm to construct a monitoring model for dam safety," *Engineering with Computers*, vol. 37, no. 1, pp. 39–56, 2021.

[18] B. Dai, C. Gu, E. F. Zhao, and X. Qin, "Statistical model optimized random forest regression model for concrete dam deformation monitoring," *Structural Control and Health Monitoring*, vol. 25, no. 6, p. 2170, 2018.

[19] X. Y. Miao, J. K. Chu, and X. W. Du, "Application of LM-BP neural network in predicting dam deformation," *Computer Engineering and Applications*, vol. 47, no. 1, pp. 220–222, 2011.

[20] G. Prakash, R. Dugalam, M. Barbosh, and A. Sadhu, "Recent advancement of concrete dam health monitoring technology: a systematic literature review," *Structures*, vol. 44, pp. 766–784, 2022.

[21] W. J. Liu, J. W. Pan, Y. S. Ren, Z. Wu, and J. Wang, "Coupling prediction model for long-term displacements of arch dams based on long short-term memory network," *Structural Control and Health Monitoring*, vol. 27, no. 7, p. 2548, 2020.

[22] X. D. Qu, J. Yang, and M. Chang, "A deep learning model for concrete dam deformation prediction based on RS-LSTM," *Journal of Sensors*, vol. 2019, Article ID 4581672, 14 pages, 2019.

[23] S. W. Wang, C. Xu, Y. Liu, and B. Wu, "A spatial association-coupled double objective support vector machine prediction model for diagnosing the deformation behaviour of high arch dams," *Structural Health Monitoring*, vol. 21, no. 3, pp. 945–964, 2022.

[24] E. F. Zhao and C. Q. Wu, "Centroid deformation-based nonlinear safety monitoring model for arch dam performance evaluation," *Engineering Structures*, vol. 243, Article ID 112652, 2021.

[25] V. Ranković, N. Grujović, D. Divac, and N. Milivojevic, "Development of support vector regression identification model for prediction of dam structural behaviour," *Structural Safety*, vol. 48, pp. 33–39, 2014.

[26] B. Chen, X. Fu, X. Y. Guo, C. Gu, C. Shao, and X. Qin, "Zoning elastic modulus inversion for high arch dams based on the PSOGSA-SVM method," *Advances in Civil Engineering*, vol. 2019, Article ID 7936513, 13 pages, 2019.

[27] S. W. Wang, C. Xu, Y. Liu, and B. Xu, "Zonal intelligent inversion of viscoelastic parameters of high arch dams using an HEST statistical model," *Journal of Civil Structural Health Monitoring*, vol. 12, no. 1, pp. 207–223, 2022.

[28] Y. T. Li, T. F. Bao, X. S. Shu, Z. Chen, Z. Gao, and K. Zhang, "A hybrid model integrating principal component analysis, fuzzy C-means, and Gaussian process regression for dam deformation prediction," *Arabian Journal for Science and Engineering*, vol. 46, no. 5, pp. 4293–4306, 2021.

[29] C. N. Lin, T. C. Li, S. Y. Chen, X. Liu, C. Lin, and S. Liang, "Gaussian process regression-based forecasting model of dam deformation," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8503–8518, 2019.

[30] K.-V. Yuen, "Recent developments of Bayesian model class selection and applications in civil engineering," *Structural Safety*, vol. 32, no. 5, pp. 338–346, 2010.

[31] L. Fatahi and S. Moradi, "Multiple crack identification in frame structures using a hybrid Bayesian model class selection and swarm-based optimization methods," *Structural Health Monitoring*, vol. 17, no. 1, pp. 39–58, 2018.

[32] K.-V. Yuen, J. Ching, and K.-K. Phoon, "Bayesian learning methods for geotechnical data," *ASCE-ASME J Risk Uncertainty Eng Syst, Part A: Civil Engineering*, vol. 7, no. 1, Article ID 02020002, 2021.

[33] S. Gamse, W.-H. Zhou, F. Tan, K. V. Yuen, and M. Oberguggenberger, "Hydrostatic-season-time model updating using Bayesian model class selection," *Reliability Engineering and System Safety*, vol. 169, pp. 40–50, 2018.

[34] G. Prakash and G. P. Balomenos, "A Bayesian approach to model selection and averaging of hydrostatic-season-temperature-time model," *Structures*, vol. 33, pp. 4359–4370, 2021.

[35] Z. Li and Y. X. Liu, "Methods and empirical evidence of multi-index clustering and variable coefficient models for panel data," *Statistics and Decisions*, vol. 34, pp. 11–14, 2014.

[36] Y. Yang, "Evaluation of empowerment methods in multi-index comprehensive evaluation," *Statistics and Decisions*, vol. 14, pp. 17–19, 2006.

[37] Y. L. Li, K. Y. Min, Y. Zhang, and L. Wen, "Prediction of the failure point settlement in rockfill dams based on spatial-temporal data and multiple-monitoring-point models," *Engineering Structures*, vol. 243, Article ID 112658, 2021.

[38] L. Wang, "Overview and prospect of bayesian model averaging," *Journal of Technical Economics and Management*, vol. 3, pp. 19–23, 2016.

[39] M. A. Clyde, J. Ghosh, and M. L. Littman, "Bayesian adaptive sampling for variable selection and model averaging," *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, pp. 80–101, 2011.

[40] M. M. Barbieri, J. O. Berger, E. I. George, and V. Rockova, "The median probability model and correlated variables," *Bayesian Analysis*, vol. 16, no. 4, p. 16, 2021.

[41] S. Reuschen, W. Nowak, and A. Guthke, "The four ways to consider measurement noise in bayesian model selection-and which one to choose," *Water Resources Research*, vol. 57, no. 11, 2021.

[42] M. A. Clyde, "Bayesian model averaging and model search strategies," *Bayesian Statistics*, vol. 30, 1999.

[43] J. Piironen, *Bayesian Predictive Methods for Model Selection*, Aalto University, Espoo, Finland, 2013.

[44] D. Madigan and A. E. Raftery, "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1535–1546, 1994.

[45] S. Chib, "Marginal likelihood from the gibbs output," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1313–1321, 1995.