RESEARCH ARTICLE

Development of an optimization model for a monitoring point in tunnel stress deduction using a machine learning algorithm

Xuyan Tan[1,2] | Weizhong Chen[1,2] | Luyu Wang[3] | Wei Ye[1,2]

[1]State Key Laboratory of Geomechanics and Geotechnical Engineering, Chinese Academy of Sciences, Wuhan, China

[2]Institute of Rock and Soil Mechanics, University of Chinese Academy of Sciences, Beijing, China

[3]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

Correspondence

Weizhong Chen, State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China.
Email: wzchen@whrsm.ac.cn

Luyu Wang, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China.
Email: luyu.wang@polyu.edu.hk

Abstract

Monitoring of the mechanical behavior of underwater shield tunnels is vital for ensuring their long-term structural stability. Typically determined by empirical or semi-empirical methods, the limited number of monitoring points and coarse monitoring schemes pose huge challenges in terms of capturing the complete mechanical state of the entire structure. Therefore, with the aim of optimizing the monitoring scheme, this study introduces a spatial deduction model for the stress distribution of the overall structure using a machine learning algorithm. Initially, clustering experiments were performed on a numerical data set to determine the typical positions of structural mechanical responses. Subsequently, supervised learning methods were applied to derive the data information across the entire surface by using the data from these typical positions, which allows flexibility in the number and combinations of these points. According to the evaluation results of the model under various conditions, the optimized number of monitoring points and their locations are determined. Experimental findings suggest that an excessive number of monitoring points results in information redundancy, thus diminishing the deduction capability. The primary positions for monitoring points are determined as the spandrel and hance of the tunnel structure, with the arch crown and inch arch serving as additional positions to enhance the monitoring network. Compared with common methods, the proposed model shows significantly improved characterization abilities, establishing its reliability for optimizing the monitoring scheme.

KEYWORDS

machine learning, monitoring, optimization, simulation, tunnel

Highlights

- Scientific guidance for monitoring scheme optimization is provided.
- The stress distribution of the overall section is characterized using limited numbers of monitoring points.
- Interdisciplinary application of machine learning helps in solving practical engineering problems.

1 | INTRODUCTION

Assessments of stress patterns in tunnel engineering are invaluable, as they directly impact the safety, efficiency, and success of tunnel projects, making rigorous stress analysis an indispensable component of tunnel engineering practice. However, the challenges of accurate stress distribution deduction in tunnel engineering are equally enormous. In addition to the complex nature of subsurface conditions and the intricate interplay of geology, the geometric intricacies of tunnel design add complexity. In recent years, structural health monitoring (SHM) has become an essential technology to determine the stress distribution and ensure the long-term stability of underwater shield tunnels (Cunha, 2022; Memisoglu Apaydin et al., 2022). Even so, it is difficult to capture the mechanical behaviors of the whole tunnel structure. On the one hand, the number of monitoring points and sensors used in practical engineering is limited, making it difficult to monitor the whole section of the tunnel (Tan et al., 2021). Moreover, some data recorded by sensors are invalid, so they cannot be used for analysis (He

et al., 2022). On the other hand, an optimized monitoring scheme is essential to obtain useful information of the structure. However, in practical engineering, the locations of monitoring sections, and the number and distribution of monitoring points are mostly determined based on empirical or semi-emperical methods (Ariznavarreta-Fernández et al., 2016; Mghazli et al., 2023; Wei et al., 2023). A common problem is that the abnormal response of the structure cannot be captured by sensors in case it occurs exactly at the point without sensors (Li et al., 2020). To solve this problem, an optimal sensor installation scheme is needed. Therefore, this study focused on developing an optimization model for monitoring point selection by using a machine learning algorithm and characterizing the mechanical behavior of the whole section based on data from limited sensors.

The purpose of SHM is to obtain the dynamic variation information of the structure, identify the abnormal behavior in advance, and prevent the occurrence of disasters (Gómez et al., 2020; Tan et al., 2023). The SHM system has already been installed in many well-known tunnels at home and abroad, such as the Channel Tunnel in Europe, the Seikan Tonneru in Japan, and the Botlek tunnel in the Netherlands and the Wuhan Yangtze River tunnel, the Nanjing Yangtze River tunnel, and the Xiang'an tunnel in China (Ikuma, 2003; Li et al., 2020; Tan, Chen, Wu, et al., 2020). Although these systems provide important data information for the analysis of structural mechanical behavior, it is impossible to evaluate whether the location of the monitoring points is reasonable and the number of monitoring sensors really meets the engineering needs (Soleimani-Babakamali et al., 2023; Zhang et al., 2022). In addition, there are few reports about optimization of the tunnel monitoring scheme in previous studies, and no unified specifications and standards have been published as guidance for monitoring scheme optimization.

Previous engineering cases show that complex topographic conditions and emergencies may trigger structural abnormal behaviors at the location without any monitoring points (Xu et al., 2022). Accordingly, experts and scholars are increasingly realizing that SHM is insufficient in that it only relies on limited points to determine the mechanical behaviors of tunnels (Du, Zhou, et al., 2021; Maes et al., 2022). Currently, the studies on sensing the mechanical behavior of the entire tunnel surface can be divided into two categories: physics-guided models and data-driven models. Numerical simulation is the most widely used physics-guided method (Yang et al., 2023), such as the finite element method, discrete element method, and their corresponding variations (Liao et al., 2023; Wang et al., 2022). Numerical results are of great importance and can be used to determine structurally sensitive positions of abnormalities and structural variation (overall changing trends), but the absolute value or amplitude of the simulation results often deviates from the actual situation, which makes it difficult to apply in practical engineering. In recent years, with the rapid development of computer science and big data technology, the data-driven based model used to deduce the mechanical behavior of the whole section of a tunnel has attracted

considerable attention (Salimi et al., 2019; Sun et al., 2022; Zhu et al., 2020). For example, based on the compressive sensing method, Bao proposed a wireless sensing technique that can recover lost data (Bao et al., 2015). Wan proposed a Bayesian multitask learning methodology to reconstruct the lost temperature data (Wan & Ni, 2019). Compared with other areas, the application of a data-driven model is quite limited in civil engineering, especially in tunnel engineering. Use of data-driven models to perform minimal unbiased estimations has become an important solution for reconstructing structural responses (Xie et al., 2022; Zhang & Wu, 2019). However, the validity of a data-driven model and the accuracy of reconstructing results are considerably influenced by the monitoring data used in the data experiment (Ángela et al., 2022).

Based on available research and practical engineering needs, this study proposes a spatial deduction model for monitoring scheme optimization and stress distribution characterization using a machine learning algorithm. The Nanjing Yangtze river tunnel was selected as a case study. First, a simulation model was developed to generate a data set that reflected the variation characteristics of tunnel mechanical behaviors. Then, clustering analysis was performed on the data set to determine the key points for monitoring. Lastly, on the basis of the monitoring data, a spatial deduction model was developed and data at the positions without sensors were derived. As an important application, the presented model was applied in the case study to characterize the stress distribution of the whole section.

2 | METHODOLOGY

In this section, the framework of monitoring scheme optimization is introduced, including the methodology to determine the typical monitoring points and derive the mechanical behaviors of overall structures based on sparse monitoring data.

2.1 | The framework of monitoring scheme optimization

The framework of monitoring scheme optimization is composed of three modules: (1) numerical simulation, (2) clustering analysis to determine typical positions, and (3) deduction of the overall stress distribution. In fact, the clustering and deduction analysis are processes of data clustering–reconstruction. The flowchart of the proposed model is shown in Figure 1.

The typical positions are defined as the points that can be used to describe the mechanical states of any position; thus, it is essential to obtain the information of the whole section. In order to realize this goal, a numerical model is developed first to constitute the data set of the mechanical behavior of the overall tunnel section. The tunnel section is divided into a number of elements, whose mechanical information is recorded for feature analysis of machine learning. Then, the constituted data set is fed into the clustering model to determine the optimized positions of
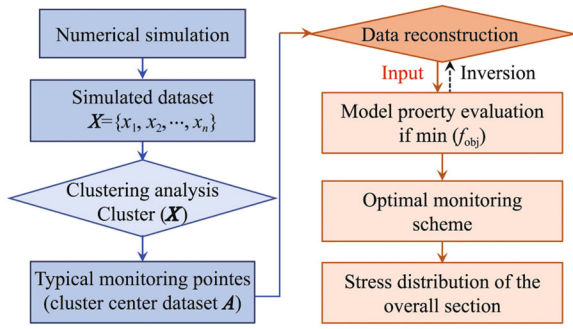
FIGURE 1 Flowchart of monitoring scheme optimization, where f_{obj} represents the objective function.

sensors. The clustering algorithm is superior in terms of determining the key points of the mechanical behavior of the entire data set. By learning the variation features of the time series of each element, the clustering model divides all of the elements into different clusters. The mechanical behavior of elements in one cluster can be characterized by only one element, that is, a clustering centroid. Accordingly, it is determined as the typical positions of monitoring points. Subsequently, the data reconstruction analysis is conducted using supervised learning models to derive the mechanical behavior of the whole tunnel section based on the limited data from typical monitoring positions. By evaluating the characterization results derived by the data from different points, the optimal positions and numbers of points can be obtained. Finally, the optimal monitoring scheme is determined.

2.2 | Data clustering–reconstruction method

The spectral algorithm is a clustering algorithm developed on the basis of graph theory. This algorithm has been widely used as it does not require many assumptions about the data structure (Alshammari et al., 2021). In addition, considering that the responses of the tunnel structure have strong spatial and temporal correlations, and linear regression (LR) is a classical model used to describe the relationship between time series (Tan, Chen, Wang, et al., 2020), the spectral algorithm and the LR model are used in this study for data clustering–reconstruction analysis. Specifically, the spectral algorithm is used to determine the typical positions for monitoring and LR mode is used to derive the information of those points without monitoring.

In the spectral model, all data are regarded as nodes in a graph and the similarities between every two nodes are represented by edge weights, where a higher value of edge weight indicates stronger similarity. By properly partitioning the graph into different groups, a partition of the graph can be obtained, where the edges between different groups have very low weights and the edges within a group have high weights. On this basis, data clustering can be achieved (Alshammari et al., 2021). Essentially, spectral clustering makes use of the spectrum (eigenvalues) of the similarity matrix of the data to perform dimension reduction and then clusters the data in a lower dimension using conventional

clustering methods like K-means. More specifically, for data set $X = (x_1, x_2, \cdots, x_n)$, x_i denotes the simulated data of element i, w_{ij} are the edge weights between x_i and x_j, and d_i is defined as the sum of weights of all the edges that are connected to x_i and can be expressed as follows:

$$d_i = \sum_{i=1}^{n} w_{ij}. \tag{1}$$

For points connected by an edge, $w_{ij} > 0$, or otherwise $w_{ij} < 0$. The K-means algorithm is usually adopted to calculate the weight w_{ij}, and radial basis function (RBF) is selected as the kernel function; thus, the edge weight can be defined as follows:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{2}$$

where σ is the kernel function of RBF.

Based on Equations (1) and (2), all clustering centers can be obtained. Assuming that $A(x_{c1}, x_{c2}, \cdots, x_{cc})$, $A \in X$ is the clustering center set of simulated data, and only the data at the central points are known. Then, these data are fed into LR mode for reconstruction. In the LR model, the data information at the location of the cluster center and no measurement point is assumed to have a linear relationship to train the model and minimize the error between the reconstruction results and the data of the cluster center. The information at any position can be expressed by the data of central points as follows (Yang et al., 2018):

$$x_i' = \alpha_0 + \alpha_1 x_{c1} + \cdots + \alpha_c x_{cc}, \tag{3}$$

where x_i' is the reconstruction result of element i; c denotes the number of clustering centers; and $\alpha_j (j = 0, 1, 2, \cdots, c)$ is the regression parameter.

In addition, the objective function of LR model training is expressed as

$$f_{obj} = \sqrt{\frac{1}{c}\sum_{i=1}^{c}\left(x_i - x_i'\right)^2}. \tag{4}$$

2.3 | Evaluation indicators for model performance

The performance of the monitoring scheme can be evaluated by calculating the characterization error of a model on the overall section. Three evaluation indicators are adopted in this study, namely, root mean square error ($RMSE$), mean absolute error (MAE), and mean absolute percentage error ($MAPE$) (Du, Li, et al., 2021; Du, Zhou, et al., 2021). The detailed descriptions of these three indicators are as follows.

$RMSE$ is a widely used indicator in machine learning to describe the deviation between the deduction result and ground truth, which is considered as a numerical result in this study. Given the deduction result $\hat{y} = \{\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n\}$ and the numerical simulation result

$y = \{y_1, y_2, \cdots, y_n\}$, the deviation between the deduction value and the numerical simulation data is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \quad (5)$$

MAE is an indicator that describes the absolute error between the deduction result and the numerical result, which can directly reflect the error. It can be calculated as follows:

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \quad (6)$$

MAPE is a percentage indicator to describe the absolute error between the deduction result and the numerical result. This indicator can avoid the effect of data size, and the recommended form is expressed as follows:

$$MAPE(y, \hat{y}) = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \quad (7)$$

3 | NUMERICAL SIMULATION OF AN UNDERWATER SHIELD TUNNEL

According to the flowchart of monitoring scheme optimization, the numerical simulation was conducted first. The Nanjing Yangtze River tunnel was selected as the case study, and a numerical model was developed according to the geological condition and tunnel structure to obtain the data of mechanical behaviors.

3.1 | Background for the case study

The Nanjing Yangtze River tunnel is located in Jiangsu, China, and became operational in January 2016. This tunnel is 7014 m long, including the shield-crossing

portion of 3537 m. This tunnel is constructed with fabricated lining, and each lining ring consists of 10 segments, namely, one key segment, two adjacent segments, and seven standard segments, as shown in Figure 2a. The external diagram of tunnle segment is 14.5 m, and the width and thickness of the segment are 2.0 and 0.6 m, respectively. Consisting of two lines, the tunnel is designed in the form of two-tubing dual eight-lane expressways. The tunnel internal structure is shown in Figure 2b. The main geological layers to which the tunnel boring machine (TBM) is exposed are mucky silty clay, silty clay with silty sand, silty-fine sand, medium-coarse sand, gravelly sand, siltstone, and pebble. The highest water pressure that this tunnel subjected to is nearly 0.7 MPa. It is one of the biggest underwater shield tunnels in the world. In addition, the geological conditions to which this tunnel is exposed are also complicated compared to other such tunnels worldwide. More detailed information about this case study can be found in our previous studies (Tan, Chen, Wang, et al., 2020).

3.2 | Numerical modeling

To determine the mechanical behaviors of a certain section of the tunnel structure, the finite element method performed on the second development was used for numerical modeling. The geological layers of this section are shown in Figure 3a, and the corresponding mechanical property parameters are presented in Table 1. Along this line, a numerical model is developed and is shown in Figure 3b. The numerical model of the tunnel segment was divided into 100 elements in the circumferential direction and eight elements in the longitudinal direction, namely, 800 elements in total.

More specifically, the beam-spring model was adopted to simulate the interaction between the tunnel lining and surrounding rock (Koizumi et al., 1988), where a number of nonlinear springs were arranged around the external surface of the tunnel lining to simulate the ground resistances. The stiffness properties of nonlinear springs are shown in Figure 4 and can be
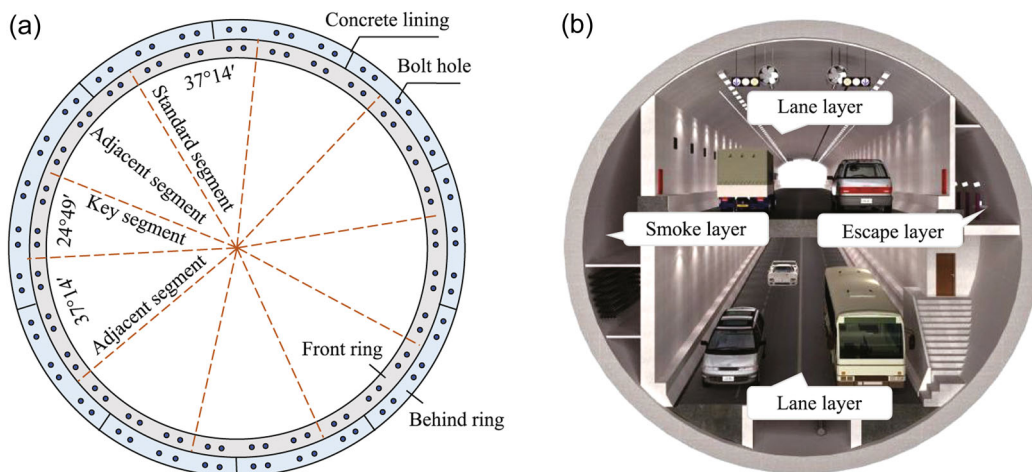

FIGURE 2 Structure of the Nanjing Yangtze tunnel: (a) the assembly form of the shield segment and (b) internal structure.
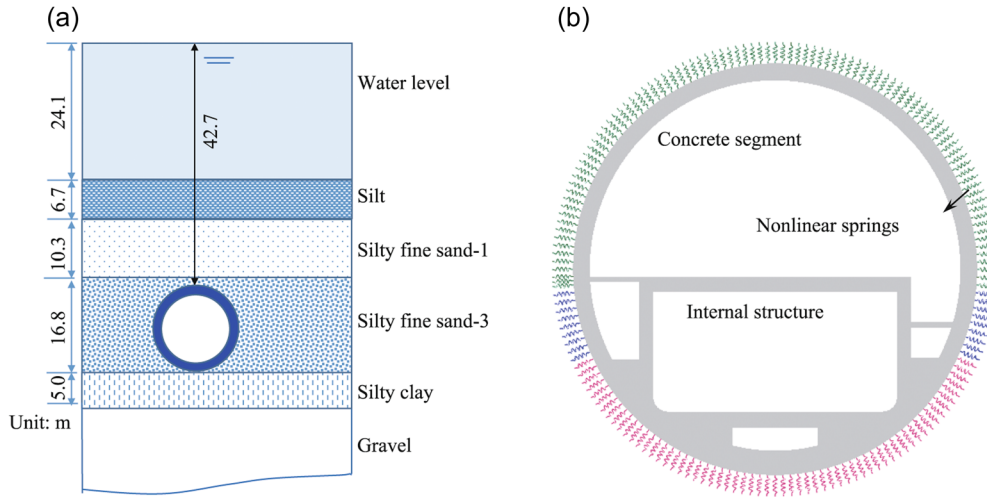
FIGURE 3 Stratum and numerical model of the case study. (a) The stratigraphic diagram and (b) the developed numerical model, where the color of springs represents the stiffness difference.

TABLE 1 Parameters of different geological layers.

Geological layers	Unit weight (kN/m³)	Ground resistance coefficient (MPa/m)	Lateral pressure coefficient
Silt	19.4	5	0.43
Silty fine sand-1	19.3	50	0.40
Silty fine sand -3	20.2	35	0.37
Silty clay	18.6	12	0.65
Gravel	20.3	45	0.30

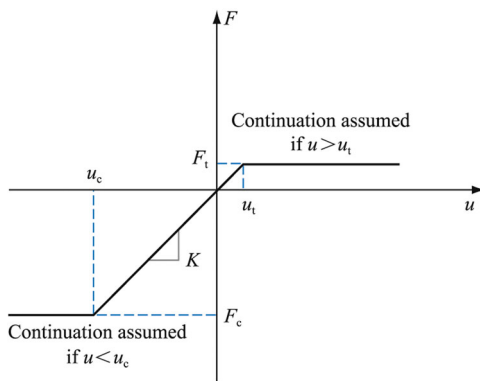**FIGURE 5** Boundary conditions applied on the numerical model.

FIGURE 4 Stiffness properties of nonlinear springs, where u is the relative displacement of the nonlinear spring and u_t and u_c are maximum displacements under tensile and compression conditions, respectively. F_t and F_c are maximum forces (Tan, Chen, Wang, et al., 2020).

calculated using Equation (8). The stiffness of the nonlinear spring is almost 0 when the spring is under tension and related to the mechanical parameters of the geological layers when the spring is compressed.

$$K = \frac{L}{N}\varphi, \tag{8}$$

where L denotes the length of the tunnel edge surrounded by springs; N is the number of springs around the tunnel edge; and φ is the ground resistance.

The material of the tunnel segment is concrete (Type: C60), whose elastic modulus is 36 GPa, possion's ration is 0.2, and unit weight is 24 kN/m³. Considering the influence of segment joints on the stiffness of the shield tunnel, the elastic modulus of concrete is multiplied by 0.8 (Chen et al., 2015). In addition, the boundary conditions of the numerical model are shown in Figure 5. The soil pressure and water pressure are separately applied on the tunnel because sand is the main geological layer that the tunnel is exposed to. The water pressure W_i can be calculated as follows:

$$W_i = \gamma_w h_i, \tag{9}$$

where γ_w denotes the unit weight of water.

Soil pressure applied to the tunnel structure includes overlying soil pressure S_i^{up}, the foundation soil pressure S_i^{bot}, and the lateral soil pressure S_i^{side}. If the center point of the tunnel is considered as the origin of coordinates and the vertical direction as the y-axis, the soil pressure can be calculated as follows:

$$S_i^{up} = \sum_{j=1}^{n} h_j\left(\gamma_j^s - \gamma_w\right), \tag{10}$$

$$S_i^{bot} = S_i^{up} + (G - F)/d, \tag{11}$$

$$S_i^{\text{side}} = \sum_{j=1}^{n} \left(h_j + d/2 - y\right)\left(\gamma_j^{\text{s}} - \gamma_{\text{w}}\right)\psi_j, \qquad (12)$$

where n is the number of overlying soil layers; h_j and γ_j^{s} are the thickness and unit weight of soil layer j, respectively; ψ_j is the earth pressure coefficient; d represents the external diameter of the tunnel segment; and G and F are the gravity and buoyancy that the tunnel is subjected to.

3.3 | Numerical results of tunnel lining stress

The water pressure applied to the tunnel varies periodically from year to year. It consists of two components: initial pressure and dynamic pressure. The former is determined by construction investigations, and is 0.437 MPa for this section, calculated according to Figure 3a. The latter varies with the season and is recorded by the Changjiang Maritime Safety Administration (specific water level data are available from the website https://cj.msa.gov.cn/). As an example, the variation of the water level of the Yangtze River tunnel over 1 year (366 days) was determined to calculate the external load applied on the tunnel structure, as shown in Figure 6. Therefore, the actual water pressure applied on the model is calculated by the sum of the dynamic pressure and the initial water pressure.

Pressure applied to the tunnel can be calculated using Equations (9)–(12), so the corresponding mechanical behavior of the tunnel can also be obtained. Each element records the stress variation in 1 year, so a time series containing 366 data included one element, expressed as $\boldsymbol{x}_i = \{x_i^1, x_i^2, \cdots, x_i^{366}\}$, and a data set $X \in R^{800 \times 366}$ containing 800 time series of all elements was also prepared.
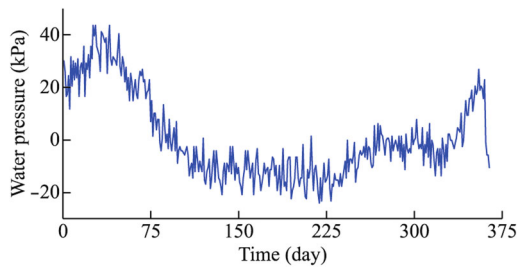
For example, Figure 7 shows some numerical results of stress variation at the locations of arch crown, hance, and spandrel.

4 | DETERMINATION FOR THE MONITORING SCHEME OF A TUNNEL SEGMENT

Based on the data set prepared by numerical simulation, data clustering and deduction model are instantiated in this section to obtain the optimal monitoring scheme. In addition, a series of comparison experiments are conducted to verify the optimization results.

4.1 | Model instantiation on data set

First, the clustering experiment was conducted. The number of clustering centroids was set to 4, 8, 10, and 20 in sequence to determine the reasonable number of monitoring points and the typical positions of segment stress variation. Based on the numerical results obtained for the tunnel, the cluster centroids were obtained using spectral clustering algorithms, as shown in Figure 8. For
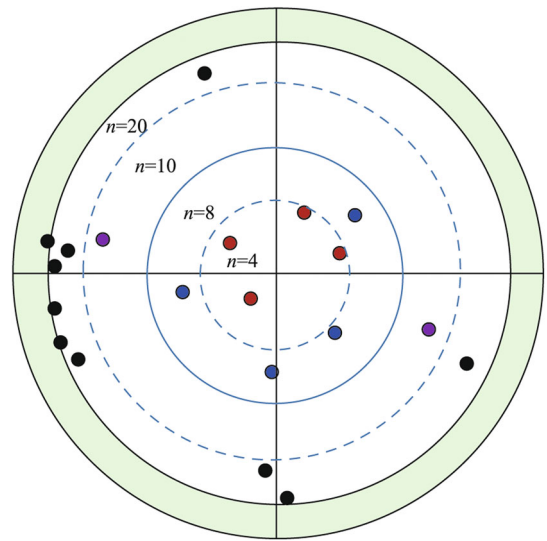

FIGURE 8 Distribution of typical monitoring points obtained from the Spectral model.

FIGURE 6 Water pressure applied in the numerical model.
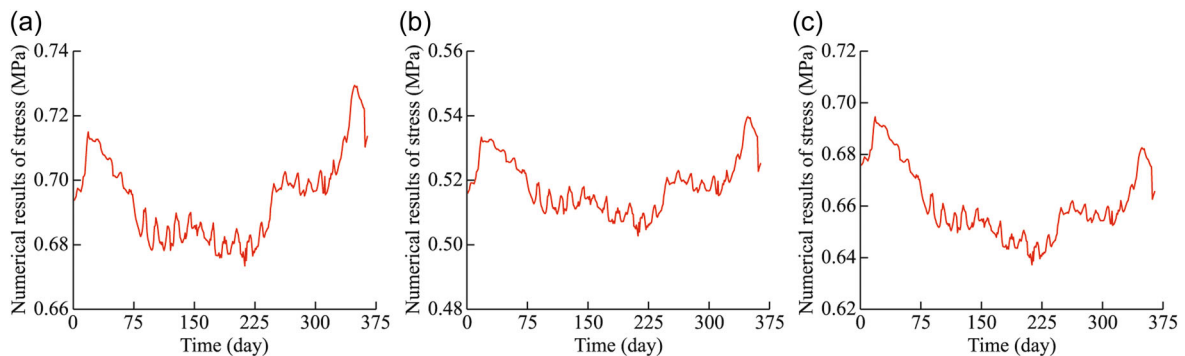

FIGURE 7 Numerical results of segment stress at some typical positions: (a) arch crown, (b) spandrel, and (c) hance.

comparison, the monitoring scheme with different numbers of points was uniformly distributed on a tunnel surface. The red points represent the positions of clustering centroids when $n = 4$, while the blue points represent the positions of the four added clustering centers relative to the first four clustering centers when $n = 8$; the purple and black points also have similar meanings. Newly added clustering centroids and previous ones are separated by a circle. The place of the monitoring points in the segment is represented by the distance from the clustering centroids to the center of the circle. Specifically, the smaller the distance, the closer it is to the inside of the segment.

Based on the clustering analysis results, the data reconstruction experiment was conducted. The data at the clustering centroids are regarded as the input to be fed into the LR model to derive the stress information of all sections. As a result, the characterization ability of the model under different monitoring conditions was evaluated by three indicators, as shown in Figure 9.

4.2 | Discussion of experimental results

According to the monitoring scheme shown in Figure 8, it can be found that when $n = 4$, the clustering centroids are mainly distributed in the spandrel and hance of the tunnel. With increasing monitoring points, some clustering centroids are added without changing previous ones and distributed close to previous ones in the space. Furthermore, if there are too many additional measurement points, the monitoring points show a clustering distribution feature, such as the distribution when $n = 20$. Based on the analysis of model reconstruction results shown in Figure 9, it can be seen that the characterization ability of the model is not directly proportional to the increase in the clustering center number. For example, in contrast to the condition with four monitoring points, although the number of clustering centers under the condition with eight monitoring points is doubled, the characterization ability of the model is not significantly enhanced. In addition, the

characterization ability of the LR model when $n = 8$ is stronger than when $n = 10$. Therefore, a larger number of monitoring points in practical engineering does not mean a better choice because too many monitoring points can result in information redundancy and resource wastage.

To this end, for the case study in this paper, it was suggested to set up eight monitoring points in the field to capture the stress variations of the whole tunnel section after considering the influence of equipment life and other factors on the data quality. In general, spandrel and hance are critically important locations for sensing the whole surface behavior of the tunnel. Then, some monitoring points can be added to the arch crown and inch arch to monitor more mechanical behaviors. Based on the optimized monitoring scheme, the elements randomly selected from the arch crown, spandrel, and hance are derived as examples, which are shown in Figure 10. As shown, the deduction results are in good agreement with the numerical results. This flow can deduce the mechanical behavior of the overall section with limited sensors and provide a deduction accuracy of up to 98% (100%–1.091%). Therefore, the optimal monitoring scheme is reasonable.

4.3 | Verification of the proposed model

In order to verify the reliability of the proposed monitoring scheme optimization model, other widely used methods are adopted to conduct data clustering and reconstruction experiments. Clustering algorithms used in this study included density peaks clustering algorithm (DPCA), K-means, and fuzzy C-means (FCM). Also, the support vector machine (SVM) model was adopted for data reconstruction experiments (Alex & Alessandro, 2014; Gao et al., 2017; Tan et al., 2022). These clustering models are instantiated on the prepared numerical data set and the key points of segment stress variation when $n = 8$ are determined, as shown in Figure 11.

As shown in Figure 11, although there are some differences in the distribution of clustering centers determined by different methods, the distribution
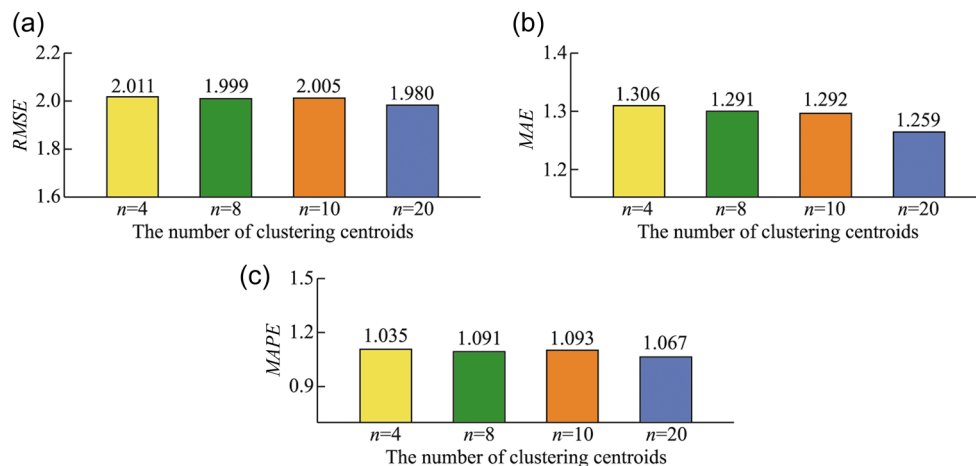

FIGURE 9 Characterization ability of a model with different numbers of monitoring points. (a) Root mean square error (RMSE), (b) mean absolute error (MAE), and (c) mean absolute percentage error (MAPE).
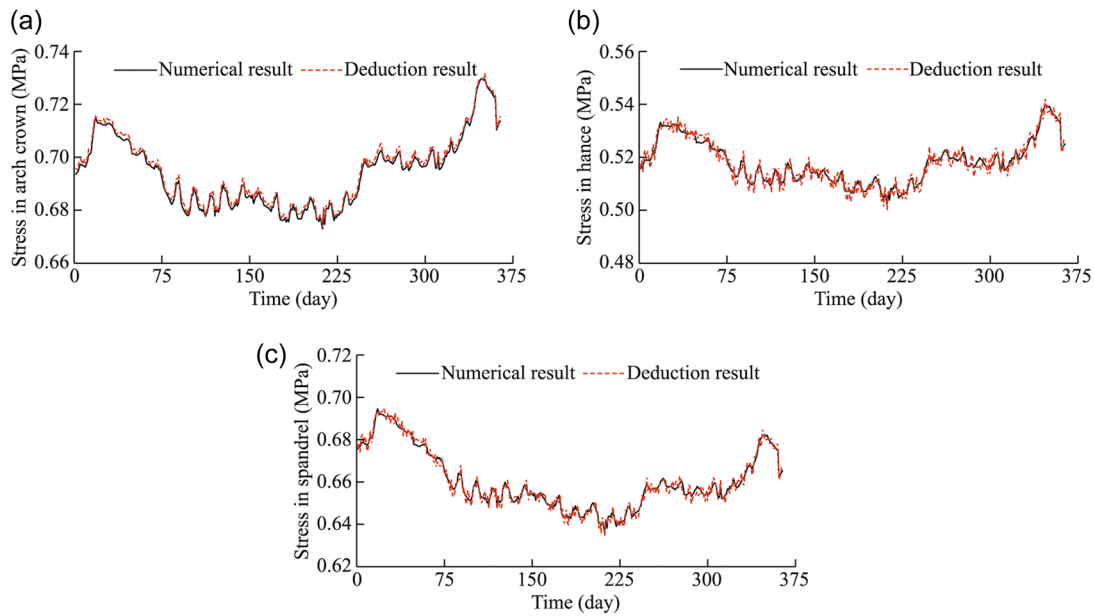
FIGURE 10 Comparison of numerical results with deduction results. (a) Arch crown, (b) hance, and (c) spandrel.
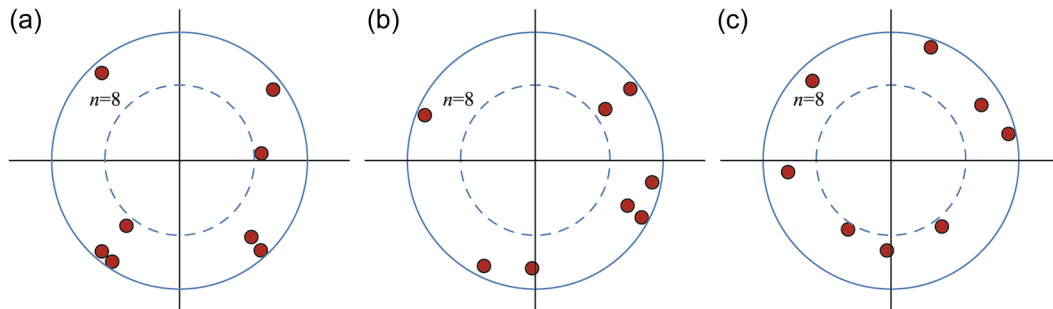

FIGURE 11 Key points of segment stress determined by other baselines: (a) Density peaks clustering algorithm, (b) K-means, and (c) fuzzy C-means.

positions are similar, that is, spandrel, hance, inch arch, and arch crown are key positions of stress variation. Furthermore, the data at key positions determined by different clustering models were fed into SVM model and the LR model to characterize the stress distribution of the overall section. The deduction ability of different models is evaluated, as shown in Figure 12.

The performances of the LR model and the SVM model are compared in terms of three indicators: *RMSE*, *MAE*, and *MAPE*. Comparison results show that the values of all three evaluation indicators of the LR model are smaller than those of the SVM model, indicating that the LR model is superior to the SVM model in terms of tunnel mechanical behavior deduction. In addition, among all clustering algorithms, the spectral algorithm is the most suitable one to cluster the data of tunnel mechanical behaviors. The error between the numerical simulation results and the deduction results obtained from the combined application of spectral and LR models is the smallest among all models. To this end, it can be concluded that the combined application of spectral and LR models is reliable to characterize the stress distribution of the overall tunnel section. The proposed model of monitoring scheme optimization is reliable.

5 | APPLICATION IN THE NANJING YANGTZE RIVER TUNNEL

In order to ensure the safe operation of the Nanjing Yangtze River tunnel, some sensors have been installed in the field to determine the mechanical behaviors of tunnel segments. For each monitoring section, a total of 20 stress sensors have been installed, where each segment is equipped with two sensors, one inside and the other outside, as shown in Figure 13. In order to characterize the stress distribution of the overall section, the proposed model was applied in this case study, and the characterization results of one section were derived as an example.

Since the sensor is embedded inside the segment, its position cannot be changed. Accordingly, the data of the eight sensors closest to the optimization positions determined by the spectral model were selected for application. At certain time points, the stress distribution of this section was derived according to the monitoring data recorded by these 8 sensors, as shown in Figure 14. It can be found that at a specific time point, the maximum stress is located at the arch crown, and the
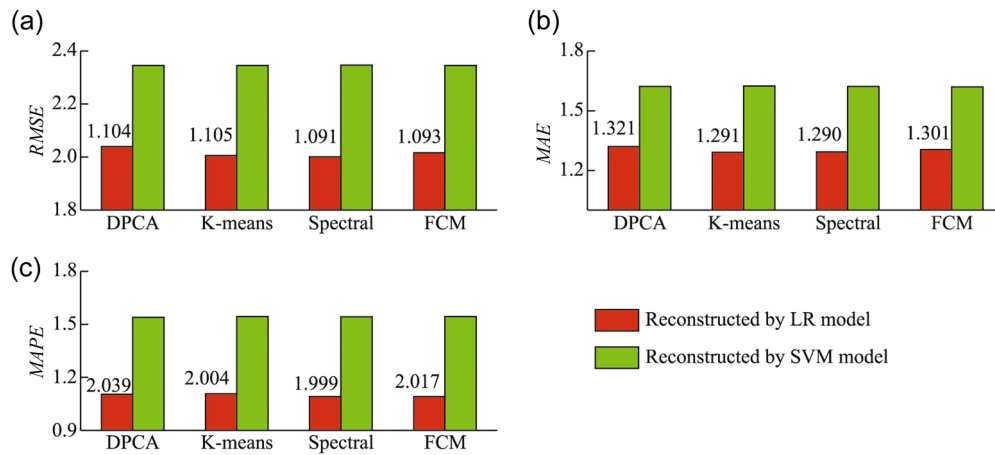
FIGURE 12 Comparison of the deduction capability of different models. (a) Root mean square error (*RMSE*), (b) mean absolute error (*MAE*), and (c) mean absolute percentage error (*MAPE*).
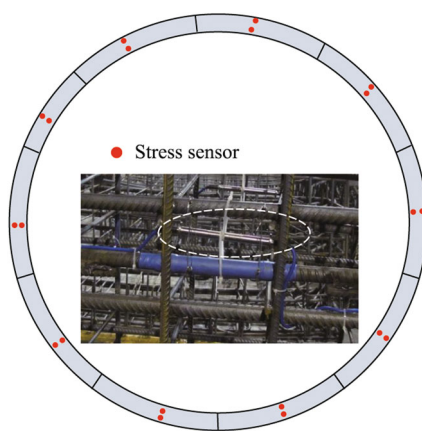

FIGURE 13 Monitoring scheme of stress sensors at the site.
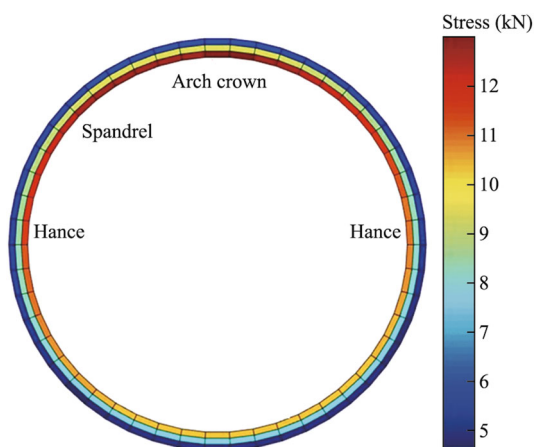

FIGURE 14 Spatial deduction results for the real-time variation of segment stress.

stress inside the segment is significantly greater than that outside the segment.

In addition, the stress evolution trend of some typical positions, such as the arch crown, spandrel, and hance, was also determined using the proposed model, as shown in Figures 15 and 16. The results show that the tunnel segment stress varied periodically in this period, and the

evolution trend is basically the same, but there are certain differences in the variation amplitude between different positions. For example, the variation amplitude of the arch crown is significantly higher than that of other positions. The deduction results for the real-time distribution of the overall section and its evolution trend can provide an important scientific basis for the prevention of disasters.

6 | CONCLUSIONS

In this study, a monitoring scheme optimization method for a tunnel structure was proposed to characterize the stress distribution of the overall tunnel section. This provided scientific guidance for monitoring scheme optimization and solved the problem of insufficient monitoring for the overall tunnel structure using limited sensors. Some specific conclusions were drawn as follows.

The monitoring scheme optimization model was developed on the basis of the combined application of spectral and LR models. The spectral model was applied to determine the typical positions of monitoring points, and the LR model was used to reconstruct the data information of the overall section using the data at the typical positions determined by the spectral model. The results show that sensors should be placed at the spandrel and hance of the tunnel, and can be placed at the inch arch and arch crown if necessary. However, too many monitoring points can result in information redundancy, which consequently decreases model characterization ability.

The reliability of the presented model was verified by comparative analysis with some widely used models. Three indicators, namely, *RMSE*, *MAE*, and *MAPE*, were used to evaluate the characterization ability of the proposed model. It was found that the combined application of spectral and LP models outperforms all other models, and the optimized monitoring scheme is reliable. The accuracy of the presented deduction model is up to 98%, which means that the deduction model is reasonable. As an important application, the proposed
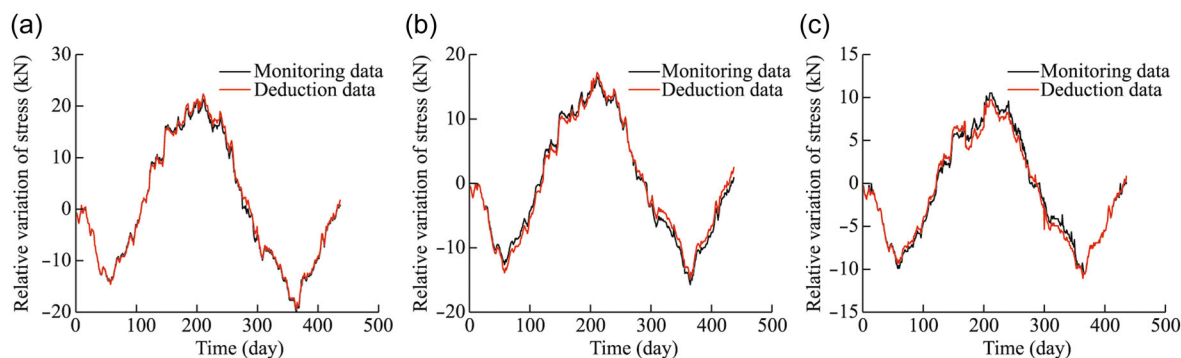
FIGURE 15 Comparison of the spatial deduction model and monitoring data at some positions: (a) arch crown, (b) hance, (c) and spandrel.
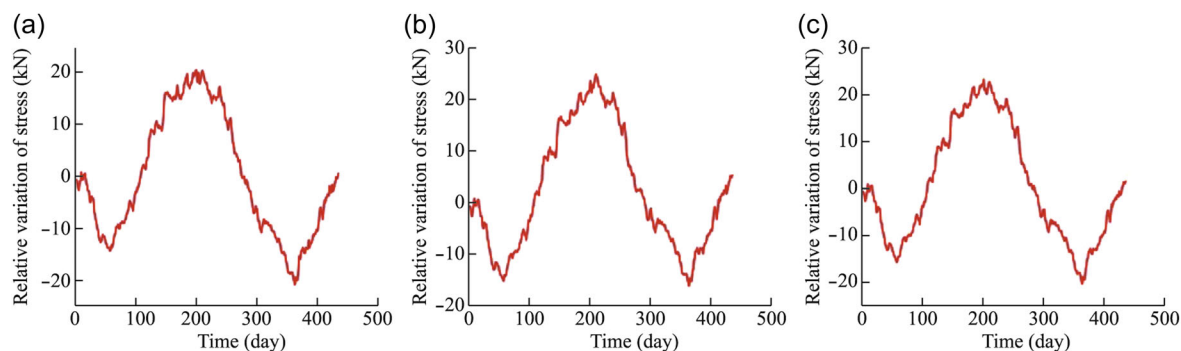

FIGURE 16 Spatial deduction results for segment stress at some positions without sensors: (a) hance, (b) spandrel, and (c) arch crown.

model was applied in the Nanjing Yangtze River tunnel to derive the stress distribution of the overall section and its evolution trend, which provides an important scientific basis for disaster prevention.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data will be made available on request.

ORCID
Xuyan Tan http://orcid.org/0000-0002-4919-7241

REFERENCES
Alshammari M, Stavrakakis J, Takatsuka M. Refining a k-nearest neighbor graph for a computationally efficient spectral clustering. *Pattern Recog*. 2021;114:107869.

Ángela F, Juan B, José RD. Supervised outlier detection for classification and regression. *Neurocomputing*. 2022;486(16):77-92.

Ariznavarreta-Fernández F, González-Palacio C, Menéndez-Díaz A, Ordoñez C. Measurement system with angular encoders for continuous monitoring of tunnel convergence. *Tunnel Undergr Space Technol*. 2016;56:176-185.

Bao Y, Yu Y, Li H, et al. Compressive sensing based lost data recovery of fast moving wireless sensing for structural health monitoring. *Struct Control Health Monitor*. 2015;22(3):433-448.

Chen WZ, Li CJ, Zeng CJ, Yang JP, Liu JQ. Establishment and application of structural health monitoring system for large shield tunnel. *Chinese J Rock Mechs Eng*. 2015;37(1):1-13.

Cunha A. Special issue on structural health monitoring of intelligent infrastructure: advanced research and real-world applications. *J Civil Struct Health Monitor*. 2022;12:1265.

Du B, Li W, Tan X, Ye J, Chen W, Sun L. Development of load-temporal model to predict the further mechanical behaviors of tunnel structure under various boundary conditions. *Tunnel Undergr Space Technol*. 2021;116:104077.

Du Z, Zhou C, Yao J, et al. CogKR: cognitive graph for multi-hop knowledge reasoning. *IEEE Trans Knowl Data Eng*. 2021;99:1.

Gao T, Li A, Meng F. Research on data stream clustering based on FCM algorithm. *Proc Comp Sci*. 2017;122:595-602.

Gómez J, Casas JR, Villalba S. Structural health monitoring with distributed optical fiber sensors of tunnel lining affected by nearby construction activity. *Autom Constr*. 2020;117:103261.

He R, Xiao T, Qiu S, Gu J, Wei M, Xu P. A rule-based data preprocessing framework for chiller rooms inspired by the analysis of engineering big data. *Energy Build*. 2022;273(15):112372.

Ikuma M. Maintenance of the undersea section of the Seikan Tunnel. *Tunnel Undergr Space Technol*. 2005;20:143-149.

Koizumi A, Murakami H, Nishino K. Study on the analytical model of shield tunnel in longitudinal direction. *Doboku Gakkai Ronbunshu*. 1988;1988:79-88.

Li C, Hou S, Liu Y, Qin P, Jin F, Yang Q. Analysis on the crown convergence deformation of surrounding rock for double-shield TBM tunnel based on advance borehole monitoring and inversion analysis. *Tunnel Undergr Space Technol*. 2020;103:103513.

Liao P, Ni P, Ma B, Sun W, Wang A. Centrifuge and numerical modeling of brittle damage of buried pipelines subjected to tunneling induced ground settlements. *Tunnel Undergr Space Technol*. 2023;138:105158.

Maes K, Salens W, Feremans G, Segher K, François S. Anomaly detection in long-term tunnel deformation monitoring. *Eng Struct*. 2022;250:113383.

Memisoglu Apaydin N, Zulfikar AC, Cetindemir O. Structural health monitoring systems of long-span bridges in Turkey and lessons learned from experienced extreme events. *J Civil Struct Health Monitor*. 2022;12:1375-1412.

Mghazli MO, Zoubir Z, Nait-Taour A, Cherif S, Lamdouar N, El Mankibi M. Optimal sensor placement methodology of triaxial accelerometers using combined metaheuristic algorithms for structural health monitoring applications. *Structures*. 2023;51:1959-1971.

Rodriguez A, Laio A. Clustering by fast search and find of density peaks. *Science*. 2014;344(6191):1492-1496.

Salimi A, Rostami J, Moormann C. Application of rock mass classification systems for performance estimation of rock TBMs using regression tree and artificial intelligence algorithms. *Tunnel Undergr Space Technol*. 2019;92:103046.

Soleimani-Babakamali MH, Soleimani-Babakamali R, Sarlo R, Farghally MF, Lourentzou I. On the effectiveness of dimensionality reduction for unsupervised structural health monitoring anomaly detection. *Mech Syst Signal Process*. 2023;187:109910.

Sun C, Xu R, Wang C, Ma T, Chen J. Coal rock image recognition method based on improved CLBP and receptive field theory. *Deep Undergr Sci Eng*. 2022;1:165-173.

Tan X, Chen W, Wang L, Tan X, Yang J. Integrated approach for structural stability evaluation using real-time monitoring and statistical analysis: underwater shield tunnel case study. *J Perform Constr Facil*. 2020;34(2):04019118.

Tan X, Chen W, Wu G, Wang L, Yang J. A structural health monitoring system for data analysis of segment joint opening in an underwater shield tunnel. *Struct Health Monitor*. 2020;19(4):1032-1050.

Tan X, Chen W, Yang J, Tan X. Temporal–spatial coupled model for multi-prediction of tunnel structure: using deep attention-based temporal convolutional network. *J Civil Struct Health Monitor*. 2022;12:675-687.

Tan X, Chen W, Yang J, Tan X. Application of a data-driven intelligent information system in infrastructure: underwater tunnel case study. *J Perform Constr Facil*. 2023;37:1.

Tan XY, Wang YH, Du BW, et al. Analysis for full face mechanical behaviors through spatial deduction algorithm with real-time monitoring data. *Struct Health Monit*. 2021;21(4):147592172110448.

Wan HP, Ni YQ. Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Struct Health Monitor*. 2019;18(4):1282-1309.

Wang L, Wang Y, Vuik C, Hajibeygi H. Accurate modeling and simulation of seepage in 3D heterogeneous fractured porous media with complex structures. *Comput Geotech*. 2022;150(3):104923.

Wei B, Luo S, Yuan D. Optimized combined forecasting model for hybrid signals in the displacement monitoring data of concrete dams. *Structures*. 2023;48:1989-2002.

Xie H, Leung C, Wang J, Li X. Advancing deep underground research through integration of engineering and science. *Deep Undergr Sci Eng*. 2022;1(1):1-2.

Xu ZD, Liu X, Xu W, Sun B, Liu X, Xu D. Analysis on the disaster chain evolution from gas leak to explosion in urban utility tunnels. *Eng Failure Anal*. 2022;140:106609.

Yang JP, Chen WZ, Li M, Tan XJ, Yu J. Structural health monitoring and analysis of an underwater TBM tunnel. *Tunnel Undergr Space Technol*. 2018;82:235-247.

Yang ZY, Qi WQ, Ding YJ, et al. Numerical investigation on the spewing mechanism of earth pressure balance shield in a high-pressure water-rich sand stratum. *Deep Undergr Sci Eng*. 2023;2:74-87.

Zhang XH, Wu ZB. Dual type structural response reconstruction based on moving window Kalman filter with unknown measurement noise. *J Aerospace Eng*. 2019;32(4):04019029.

Zhang Y, Tang Z, Yang R. Data anomaly detection for structural health monitoring by multi-view representation based on local binary patterns. *Measurement*. 2022;202:111804.

Zhu H, Wang X, Chen X, Zhang L. Similarity search and performance prediction of shield tunnels in operation through time series data mining. *Automat Constr*. 2020;114:103178.

AUTHOR BIOGRAPHY

Weizhong Chen is a professor and chief engineer at the Institute of Rock and Soil Mechanics. With a background in large-scale hydropower engineering, transportation, and energy projects in China, his research is focused on (1) underground engineering stability analysis and control, (2) technology for disaster control of underground engineering with large deformations of soft rock, and (3) online monitoring, big data analysis, and early warning system for major underground engineering.