Research on prediction performance of multiple monitoring points model based on support vector machine

Kaiyi Min¹, Yanlong Li¹*, Qiaogang Yin¹ and Lifeng Wen¹

¹State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi’an University of Technology, Xi’an, Shaanxi, 710048, China

*Corresponding author’s e-mail: liyanlong@xaut.edu.cn

Abstract: The multiple monitoring points model is an important means of dam structure health monitoring. Combined with the strong nonlinear mapping capability of support vector machine, the fitting and prediction accuracy of the model is further improved. According to the point selection of the multiple monitoring points SVM model, five kinds of testing schemes are designed. The influence of points selection on the prediction performance of the model is verified by case analysis. The results show that the predictive ability of multiple monitoring points model based on SVM is greatly affected by the correlation degree of the monitoring points. Therefore, it is very important to reasonably select the point data with high similarity as the training samples for the effective prediction of deformation monitoring model.

1. Introduction

Dam safety monitoring is an important part of dam safety management, and it is also the main means to master the dam's operational behavior. Establishing a reliable dam deformation monitoring model is of great significance for timely detecting safety hazards and ensuring stable operation of dams[1]. In recent years, some experts and scholars have analyzed many monitoring points from the whole, and established a multiple monitoring points model (MMP) which can reflect the displacement relationship between each monitoring point of the dam, so that the analysis and evaluation of the dam working performance move forward in a more in-depth direction[2-4]. The MMP model breaks through the limitation of single monitoring point modeling, and transforms the previous "single-point" analysis method into "cross-section" analysis method, so as to better reflect the whole deformation state of the dam. However, there are no uniform rules for determining the sample selection of the monitoring points range, and the selection criteria are relatively vague. In some papers [5-6], all the monitoring points in the dam section are directly selected as the training samples. Such selection method has large randomness and human factors. Due to the diversity of influencing factors such as load method, constraint conditions and material properties, the dam deformation distribution is regional, and the heterogeneity between different parts is large. The unified analysis of the deformation values of all monitoring points is easy to ignore the heterogeneity between the points, leading to large errors in the prediction results.

In view of the advantages of the support vector machine (SVM) in dealing with small samples, nonlinear data classification and regression problems[7], this paper combines SVM with MMP displacement models. Compared with the traditional statistical model, the SVM-based monitoring model can effectively improve the fitting and prediction accuracy[8]. In order to improve the predictive ability of multiple points model based on SVM, this paper designs different test schemes, and analyzes
the selection range of suitable sample points by using case analysis. It has certain reference value for constructing MMP model and analyzing the working behavior of dam.

2. Multiple monitoring point statistical model
At present, the single monitoring point model is widely used in the dam displacement prediction mathematical model, that is to establish a model for the certain monitoring point displacement in a certain direction. In the traditional statistical model, the cause of settlement deformation at a single point of the dam body is composed of three parts: water pressure component ($\delta_H$), temperature component ($\delta_T$) and time effect component ($\delta_\theta$)\(^9\). The general expression of the model is as follows:

$$\delta = \delta_H + \delta_T + \delta_\theta$$

$$= \sum_{i=1}^{1} H + \sum_{i=1}^{i} (\cos \frac{2\pi t \theta}{365} + \sin \frac{2\pi t \theta}{365}) + \theta + \ln \theta$$

where $H$ is the water level difference between the upstream and the downstream on the date of observation; $\bar{H}$ represents the average air temperatures in the 3 days before the observation date; $t$ is the cumulative number of days from the monitoring date to the starting date; $\theta = t/100$.

For the deformation distribution of MMP model, the position parameters $x, y$ and $z$ of monitoring points are usually introduced to expression, and the control range is expanded from one-dimensional to three-dimensional plane through the coordinate position of the settlement point as the input quantity. The spatial–temporal distribution model of the displacement field is established, so as to show the effect of the spatial distribution on settlement\(^10\). The MMP model combines each environmental impact factor and pure coordinate factor to form a new independent variable factor set containing the water pressure component, temperature component and time effect component of the statistical model. In summary, the model can be expressed as followed:

$$\hat{\delta} = \delta(H, T, \theta, x, y, z) = \begin{bmatrix} u(H, T, \theta, x, y, z) \\ v(H, T, \theta, x, y, z) \\ w(H, T, \theta, x, y, z) \end{bmatrix}$$

where $\hat{\delta}$ is displacement vector; $u, v, w$ are three coordinate axis directions of space rectangular coordinate system.

The three-dimensional, three-direction MMP model represented by the above formula can degenerate into two-dimensional, one-dimensional and two-direction, single-direction. The specific expression of the two-dimensional MMP settlement model is:

$$\tilde{\delta} = f(H, T, \theta, x, y)$$

$$= \delta_{H,x,y}(t) + \delta_{T,x,y}(t) + \delta_{\theta,x,y}(t)$$

where $x$ is the distance from the location of the monitoring point to the center line of cross-section, and $y$ is the distance from the location of the monitoring point to the dam crest. In addition, the one-dimensional MMP model in this study is similar to the two-dimensional model, so it will not be discussed here.

3. Support vector machine for multiple monitoring points model
Support vector machine is a supervised learning model proposed by Vapnik based on statistical learning theory in the mid-1990s. SVM is widely used in pattern classification, optimal control, regression analysis, nonlinear modeling and prediction. For the complex nonlinear problem of dam deformation prediction, the SVM regression prediction model can achieve better results.

In the two-dimensional MMP statistical model, due to the amount of independent variable data is large (43 items), the workload of direct calculation is huge and the uncertainty of factors will lead to low fitting accuracy. In this study, four factor sets are becomed as the preset factor set of regression
analysis, and stepwise linear regression analysis is carried out. The impact factors of higher correlation are retained, and the factors with poor correlation are eliminated, so that the model can better fit the measured data. The final filtered impact factor is brought into the model as an input variable. In the one-dimensional MMP statistical model, the monitoring points participating in the modeling are located on the same cross-section and the same straight line of the dam, and the number of independent variable factors is significantly reduced compared with the two-dimensional MMP model. Since the SVM has strong nonlinear mapping ability, the factors which are related to deformation can be directly selected as the input variable in the one-dimensional model. The measured values of the input and output variables are taken as the training samples of the model, the output variables are the dam deformation observation values, and the settlement value of the face rockfill dam is selected as the output value.

All simulations are carried out in the MATA LB R2016b environment. Before running the models, all data for SVM algorithm have been normalized to [-1,1]. In the process of model training, the values of penalty function C and kernel function parameter g have a great influence on model performance. In this study, cross validation (CV) method is used to select parameters, which can get the optimal parameters in a certain sense, effectively avoid the occurrence of over learning and under learning States.

4. Case study
The monitoring data of a concrete face rockfill dam are selected for analysis. The dam started to be filled in October 2008, and the main body of the dam was completed in September 2009. The construction period of face slab concrete is from March 15, 2010 to May 14, 2010, and the water storage began on October 14, 2010. The project reservoir is a daily regulating reservoir with a normal water storage level of 1856 m. The total storage capacity of this reservoir is about 294 million m$^3$. In order to effectively monitor the displacement of face rockfill dam, three straight lines of settlement displacement monitoring points are embedded in the dam body for monitoring. The ES4 and ES1 series are located in the primary rockfill, and the ES2 upper monitoring points are located in the downstream rockfill area, as shown in Figure 1. In this study, the 36 monitoring points on the dam cross-section will be analyzed. The data set is built over a period from December 2010 to June 2015, which is from the completion of the first impoundment to the completion of the third impoundment. Each monitoring point has totally 200 groups of monitoring data, the first 190 groups of data are used for fitting and training, and the last 10 groups of data are used for prediction. The large range of monitoring points can reflect the deformation of the whole rockfill dam under the external load, and its wide range of variation is helpful to study the prediction performance of the model.

Figure 1 Distribution map of the monitoring points.
In order to accurately explore the influence of the selection of monitoring points on the prediction ability in the multiple monitoring points model, five different schemes are designed for comparison. Scheme 1 is the traditional single monitoring points model, scheme 2-4 is the multiple monitoring points model, the monitoring point selection and independent variable parameters of each scheme are different, and the specific content is shown in Table 1. Among them, the selection range of scheme 3 and 4 is indicated by a dotted line frame and a solid line frame in Figure 1, and the measurement points of scheme 5 are marked with red punctuation.

Hu and Shao\textsuperscript{[11,12]} pointed out that the specific monitoring points in the dam have similar deformation laws. Ji\textsuperscript{[13]} called the monitoring points groups with the same straight line, the same monitoring instrument and the same monitoring direction as the homologous monitoring points series. In the four schemes of MMP model, the monitoring points range is gradually reduced from the whole dam cross section to homologous monitoring points series, and the number of monitoring points is gradually reduced from 36 to 6, but the correlation degree of monitoring points is gradually increasing.

Table 1 The description of schemes.

| Scheme  | Number of points (Remark) | Impact factors |
|---------|---------------------------|----------------|
| Scheme 1 | One (ES1-9) | \{ $h, h^2, h^3, \bar{h}, \theta, \ln \theta, \sin \frac{2\pi t}{365}, \sin \frac{4\pi t}{365}$ \} |
| Scheme 2 | Thirty-six (Cross-section of dam) | \{ $x, y, \ln \theta, \theta x, \bar{h}, \theta y, h^2, \bar{h}y, \ln \theta, \cos \frac{2\pi t}{365}, \sin \frac{2\pi t}{365}, \sin \frac{4\pi t}{365}$ \} |
| Scheme 3 | Fifteen (Maximum settlement area of dam body) | Same as scheme 2 |
| Scheme 4 | Thirteen (The series of ES1) | \{ $x, h, hx, h^2, \bar{h}, \bar{h}x, \theta, \theta x, \ln \theta, \sin \frac{2\pi t}{365}, \sin \frac{4\pi t}{365}, \cos \frac{2\pi t}{365}, \cos \frac{4\pi t}{365}$ \} |
| Scheme 5 | Six (ES1-7~ES1-12) | Same as scheme 4 |

As the ES1-9 monitoring point is close to the actual maximum settlement point of the dam, the prediction result of this point has important analytical value. The performance of each scheme will be compared by their prediction results of ES1-9 monitoring point. In this study, correlation coefficient ($R$), mean absolute error (MAE) and mean absolute percentage error (MAPE) are used as evaluation criteria for model performance. Their expressions are as follows:

\[
R = \frac{\sum_{i=1}^{n} (y_D(i) - \bar{y}_D)(y(i) - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_D(i) - \bar{y}_D)^2} \sqrt{\sum_{i=1}^{n} (y(i) - \bar{y})^2}} \quad (4)
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_D(i) - y(i)| \quad (5)
\]
\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y(i) - y_D(i)}{y(i)} \right| \]  

(6)

where \( y_D \) and \( \bar{y} \) represent the simulated values and average value, \( y \) and \( \bar{y} \) denote the measured values and the average value. and \( n \) is the number of observation.

| Scheme   | MAE   | MAPE  | \( R \) |
|----------|-------|-------|--------|
| Scheme 1 | 10.31 | 2.42\%| 0.816  |
| Scheme 2 | 9.97  | 2.28\%| 0.829  |
| Scheme 3 | 8.70  | 1.81\%| 0.831  |
| Scheme 4 | 6.39  | 1.12\%| 0.907  |
| Scheme 5 | 5.78  | 1.00\%| 0.913  |

The prediction performance of each scheme is shown in Table 2, and the specific prediction values are shown in Figure 2. The settlement prediction curves of scheme 2 and scheme 3 fluctuate have large fluctuations up and down, which are not well consistent with the measured curves. The guess reason is that the selected monitoring points are distributed in a large range. The heterogeneity among the monitoring points leads to the low fitting degree of deformation law. In general, the predicted displacement of scheme 4 and scheme 5 is in good agreement with the change trend of measured displacement, and the residual value of scheme 5 is the smallest, indicating that the prediction accuracy of scheme 5 model is the highest. It can be seen that the selection of monitoring point samples has a great impact on the prediction accuracy.

Figure 2 Predicted values and monitoring values in different schemes.

As shown in Table 2, we find that the prediction performances of the multiple monitoring points models are better than that of the traditional single point model, which shows that the MMP model can
better reflect the dam deformation law. In the four schemes of MMP model, the three evaluation indexes of MAE, MAPE and R are basically inversely proportional to the number of monitoring points. Comparing scheme 2 with scheme 5, we can see that MAE is reduced from 9.97 to 5.78, MAPE is reduced from 2.28% to 1.00%, and correlation coefficient R is increased from 0.829 to 0.913, which shows that the prediction performance is significantly improved with the reduction of the point selection range. From the range of three evaluation indexes, when the selection range transforms from the maximum settlement area to the ES1 series straight line, the prediction accuracy improves the most, but it should be noted that the number of monitoring points does not change much at this time (The number of points changed from 15 to 13). In the process of scheme 2-3 and scheme 4-5, the number of selected monitoring points is almost reduced by half, but the prediction accuracy of the model is only slightly improved, and the change range is small, indicating that the correlation between the number of monitoring points and the prediction ability is low. However, in the above several cases, the model prediction performance shows a relatively obvious correspondence with the selection of the monitoring points in the scheme. As the correlation degree of the selected points is gradually improved, the prediction accuracy is also significantly improved. It can be seen that the simple increasing or decreasing number of monitoring points has little effect on the prediction performance of the model. The essential reason for the improvement of the model performance is to eliminate the weakly correlated monitoring points data of "misleading" model, because the retained monitoring point samples have higher similarity, so it can better fit the actual deformation law.

5. Conclusion
This paper selected different points to establish MMP deformation prediction models based on SVM. The prediction ability is studied by analyzing the error between the prediction value and the monitoring value under different schemes. The case analysis shows that:

(1) The MMP model has good predictive performance. Compared with the single point model, it can better reflect the overall deformation law, and has higher application value for dam structure safety evaluation and prediction.

(2) When too many points are directly selected and brought into the training model, the prediction performance is poor. The reason is that there are some weakly correlated and uncorrelated monitoring points, which will disturb the fitting of multiple monitoring points model to the deformation law of the target monitoring point, resulting in the decline of the fitting and prediction accuracy of the model.

(3) It is the number of high correlation points that affects the prediction performance of the multiple monitoring points model. Since the points in the homologous series have high correlation, in the case of a certain monitoring point deformation prediction in the actual engineering, points on the same straight line of the target point can be directly selected to construct a one-dimensional MMP model, so as to obtain a stable and reasonable prediction model.

(4) In the design of model schemes, there is a certain degree of subjective selectivity and randomness in the points selection. Therefore, how to objectively select suitable and highly correlated monitoring points to construction model needs further research.

Acknowledgments
The research described in this paper was funded by the National Natural Science Foundation for Excellent Young Scientists of China (Grant No. 51722907); the National Natural Science Foundation of China (Grant No. 51979224) and the Project 2019JLZ-14 supported by Joint Foundation of Shaanxi.

References
[1] Salazar, F., Morán, Rafael, Toledo, Miguel á., Oñate, E. (2017) Data-based models for the prediction of dam behaviour: a review and some methodological considerations. Archives of Computational Methods in Engineering, 24(1), 1-21.
[2] He, J.P., Li, Z.Z.(1994) Research on Mathematical Model of Multi-measurement Point of Dam
Structural Behavior. Journal of Wuhan University, (02), 134-136+1378+138-142. (in Chinese)

[3] He, J.P., Tu, Y.Y., Shi, Y.Q. (2011) Fusion model of multi monitoring points on dam based on bayes theory. Procedia Engineering, 15(none), 2133-2138.

[4] Stoian, I., Kovacs, I., Ignat, S., Ghiran, O., Oltean, G., Gordan, M. (2018) Multisensorial Data Fusion and Correlation for Dams Structural Health Monitoring. In: IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). Cluj-Napoca, Romania.

[5] Chai, L.S., Qi, D., Wu, H. (2014) Application of multi-point multi-direction BP network model in dam deformation monitoring. Water Resources and Power, (3), 94-97. (in Chinese)

[6] Huang, D.W., Liu, B., Dai, W.J. (2017) Multi-measurement displacement model of Wuqiangxi dam based on neural network model. Geotechnical Geotech Invest Survey, (5). (in Chinese)

[7] Su, H., Chen, Z., Wen, Z. (2016) Performance improvement method of support vector machine-based model monitoring dam safety. Structural Control and Health Monitoring, 23(2), 252-266.

[8] Rezapour, T.M.M., Zarif, S.H.R. (2018) Prediction of the intermediate block displacement of the dam crest using artificial neural network and support vector regression models. Soft Computing.

[9] Shi, Y., Yang, J., Wu, J., He, J. (2018) A statistical model of deformation during the construction of a concrete face rockfill dam. Structural Control & Health Monitoring, 25(2), e2074.

[10] He, J.P. (2010) Dam Safety Monitoring Theory and Application. China Water & Power Press, Beijing. (in Chinese)

[11] Shao, C., Gu, C., Yang, M., Xu, Y., Su, H. (2017) A novel model of dam displacement based on panel data. Structural Control and Health Monitoring, e2037.

[12] Hu, Y.T., Shao, C.F., Gu, C.S., Meng, Z.Z. (2019) Concrete Dam Displacement Prediction Based on an ISODATA-GMM Clustering and Random Coefficient Model.11, 714.

[13] Ji, W., Liu, X.Q., Lin, C.N. (2019) New multi-measurement displacement model of dam based on BP neural network. Water Resources and Hydropower Engineering, 50(02), 115-120. (in Chinese)