doi:10.3233/ATDE220301

Deformation Prediction of Gravity Dam Based on EMD-SARIMA Model

Bin LI^{a,b,1}, Xingping BAI^b, Qun ZHANG^b and Jun LI^b

^aNorthwest Engineering Corporation Limited, Power China Water Conservancy & Hydropower Engineering Div./Pumped Storage Engineering Div., Xi'an, Shannxi,

China

^bXi'an University of Technology State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an, Shannxi, China

> Abstract. Hydraulic engineering plays an important role in energy construction in China. As the most important water retaining structure, the deformation trend and safety state of dam is undoubtedly the most concerned problem in engineering. Dam deformation monitoring data is the most critical information to understand dam deformation. So, the analysis and prediction of dam deformation monitoring data is an important measure to master dam safety state. However, the monitoring data of dam generally contains noise components. In order to reduce the noise influence and improve the stability and accuracy of dam monitoring data. EMD-SARIMA model was established in this paper. The monitoring data was decomposed into several Intrinsic Mode Function (IMF) from high to low frequency by using Empirical Mode Decomposition (EMD). Then, the data was reconstructed after eliminating the IMF mainly containing noise based on the Continuous Mean Square Error (CMSE) criterion. Finally, a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model was established for the reconstructed data. The results show that EMD can effectively reduce the noise in dam monitoring data. The reconstructed data is more stable than the original data, and closer to the actual displacement process of the dam. Compared with SARIMA model, the prediction accuracy of EMD-SARIMA model meets the requirements, and is more accurate and less noise effect. It can be applied to denoise data and prediction analysis of gravity dam.

> **Keywords.** Gravity dam, deformation prediction, empirical mode decomposition, time series analysis, EMD-SARIMA model

1. Introduction

The main environmental factors affecting the deformation of gravity dams include aging, water pressure and temperature [1]. Under the action of water pressure, gravity dam will deform with time. This slow trend deformation is usually presented as a low frequency component in dam monitoring data. The water pressure factor usually has certain periodicity due to the regulation form of reservoir. The temperature factor usually also has a certain periodicity due to the alternation of seasons. Therefore, gravity dam will produce periodic deformation under the action of periodic factor. Periodic deformation

¹ Corresponding Author, Northwest Engineering Corporation Limited, Power China Water Conservancy & Hydropower Engineering Div./Pumped Storage Engineering Div., Shannxi, China; Xi'an University of Technology State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an, Shannxi, China; Email: libin@nwh.cn.

345

is usually presented as intermediate frequency component in dam monitoring data. During the data acquisition process, the monitoring data often contains certain noise components. This is usually due to the defects of instrument design or installation, as well as the interference in the signal transmission process. The noise is usually presented as high frequency components in dam monitoring data [2]. Feature extraction of gravity dam deformation monitoring data can well understand the variation trend, scope and cause. It is very important for guaranteed dam safety [3, 4]. However, the variation range of dam deformation monitoring data is small, and it is close to the high frequency noise. The noise will drown the real data features and affect the accuracy of monitoring data. In order to analyze the monitoring data more accurately and reasonably, the influence of high frequency noise should be reduced as much as possible. In this way, it can better obtain accurate and real data and to better monitor dam safety.

At present, noise reduction methods for dam monitoring data mainly include Wavelet Analysis and Empirical Mode Decomposition (EMD). Wavelet analysis has good time-frequency local characteristics and multi-resolution analysis functions. It has been widely used in data noise reduction, signal analysis, image processing and other fields [5, 6]. Wu [7] added wavelet analysis to GM (1,1) model, and the results showed that wavelet threshold denoising can obviously remove the noise in original data. Li [8] also used wavelet analysis to denoise the dam deformation data, and then reconstructed the extracted comprehensive components to obtain a hybrid model to predict the dam deformation. However, wavelet analysis is not adaptive. The denoising effect depends on the threshold selected by the expert. If the threshold is too small, part of the noise may be retained, while if the threshold is too large, part of the useful data may be deleted. EMD is an adaptive decomposition method for nonlinear and non-stationary data. It was proposed by Huang [9]. It does not need to determine any function and can directly and effectively decompose the data. Compared with wavelet analysis, EMD has the characteristics of simple calculation, intuitive and adaptive. EMD is also widely used in data feature extraction and noise reduction [10, 11]. Liu [12] and Jin [13] both have combined EMD and some prediction methods to fit and predict dam displacement, and obtained good results.

Dam deformation monitoring data is a set of one-dimensional time series. The Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model can find the change characteristics, trends and development laws by analyzing the historical change process of a time series. SARIMA adds a seasonal difference step to the Auto-regressive Integrated Moving Average (ARIMA) model [14]. It is suitable for analyzing data series with certain tendency and periodicity, and is not limited by missing data [15, 16]. In this paper, EMD method is used to denoise and extract features for dam deformation data. Then, SARIMA is used to build a dam displacement prediction model, so as to fit and predict dam displacement.

2. Empirical Mode Decomposition

2.1. EMD Theory

The effect of different influencing factors in dam displacement data is reflected as different frequencies. EMD uses frequency as the judgment criterion to calculate the data series. Then, the Intrinsic Mode Function (IMF) with frequency from high to low can be obtained [17]. The mid and low frequency IMF components usually reflect the basic

characteristics of the data, while the high frequency IMF components generally are noise. EMD calculation steps are as follows:

- (1) Find the extreme points of the data sequence Y(t), connect all the maximum points and minimum points with a cubic spline function respectively, then obtain the upper envelope $V_{\max}(t)$ and lower envelope $V_{\min}(t)$.
- (2) Calculate the mean value $\overline{V}_1(t)$ of the envelopes.

$$\overline{V}_{1}(t) = [V_{\max}(t) + V_{\min}(t)]/2$$
(1)

• (3) Calculate the difference D(t) between Y(t) and $\overline{V}_1(t)$.

$$D(t) = Y(t) - \overline{V}_1(t) \tag{2}$$

- (4) Judge whether D(t) satisfies the two conditions: (a) The number M with 0 in D(t) and the number N with extreme points must satisfy the inequality |M-N|≤1;
 (b) The local mean value of V_{max}(t) and V_{min}(t) both are 0. If D(t) meets the two conditions, D(t) is the first IMF component of Y(t) and is expressed as L₁. Otherwise, D(t) is used as a new original sequence and the above steps are repeated until the conditions are satisfied.
- (5) Calculate residual sequence B_1 .

$$B_1 = Y(t) - L_1 \tag{3}$$

• (6) Take B_1 as a new data sequence and repeat the above steps until all components are decomposed. So far, Y(t) can be expressed as the sum of $n L_i$ sequences and one residual sequence *res*.

$$Y(t_i) = \sum_{i=1}^n L_i + res.$$
(4)

Each IMF has a different frequency. The residual sequence is the trend component and represents the average trend.

2.2. Continuous Mean Square Error Criterion

The original monitoring data can be decomposed to multiple IMFs. The high frequency IMFs mainly contains noise information. The mid and low frequency IMFs mainly contains real information. How to determine which high-frequency IMFs contain noise is the key to data denoising. Continuous Mean Square Error (CMSE) criterion [18] can find the global minimum point of noise energy mutation, so as to find the IMFs mainly containing noise. The calculation formula is as follows:

$$CMSE(L_k, L_{k+1}) = \frac{1}{n} \sum_{i=1}^{n} [L_k(t_i) - L_{k+1}(t_{i+1})]^2 \qquad (k=1, 2, \cdots, n-1)$$
(5)

$$k=argmin[CMES(L_k, L_{k+1})]$$
(1 $\leq k \leq n$) (6)

where: L_k is the *k*th IMF component, t_i is the *i*th data of IMF component, *n* is the number of IMF components.

After the k value is determined, the original data can be reconstructed to obtain the data after denoising.

$$L_t(t) = \sum_{i=k+1}^n \text{IMF}_i(t) + res.$$
(7)

CMSE criterion is to find the kth IMF, so as to determine that the first k high-frequency IMFs are mainly noisy information. This method has the advantages of simple calculation and strong adaptive, and does not need to set the threshold manually.

3. SARIMA Model

3.1. SARIMA Model Definition

Dam displacement monitoring data usually have tendency and periodicity. SARIMA (p, d, q) (P, D, Q) model can analyze and predict this kind of data well. It removes nonstationarity and periodicity of time series by ordinary differencing and seasonal differencing. The original time series is transformed into a stationary and aperiodic series. Then build an ARMA model for analysis. The model is defined as follows [14]:

$$\Phi_p(B)U_P(B^S)(1-B)^d(1-B^S)^D Y_t = \Theta_q(B)V_Q(B^S)\alpha_t$$
(8)

where: p, d, q, P, D, Q are all integer parameters. d is the ordinary differencing order $(d \le 2)$, D is the seasonal difference order. B is a hysteresis operator. S is season length. a_t is a white noise sequence with a mean of 0 and variance of $\sigma^2 \cdot \Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, this is an autoregressive model with p order. $U_P(B^S) = 1 - \Gamma_1 B^S - \Gamma_2 B^{S2} - \cdots - \Gamma_p B^{SP}$, this is a seasonal autoregressive model with P order. $\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, this is a moving average model with q order. $V_Q(B^S) = 1 - H_1 B^S - H_2 B^{S2} - \cdots - H_Q B^{SQ}$, this is a seasonal moving average model with Q order.

3.2. Modeling Steps

The modeling steps of SARIMA are as follows:

- (1) Ordinary difference and periodic difference are performed on the data sequence to determine the parameters *d* and *D*.
- (2) Autocorrelation and partial autocorrelation analysis are performed on the sequence after difference, so as to determine *p*, *q*, *P* and *Q*.
- (3) The least square method is used to estimate the parameters of *p*, *q*, *P* and Q of the primary several models.
- (4) The models are screened by AIC, BIC and DW criteria [19].
- (5) Test the residual of the model. If the residual has the white noise feature, the model meets requirements. Otherwise, the model needs to be re-established.
- (6) The established model is used for prediction analysis. Root mean square error (RMSE), mean absolute error percentage (MAPE) and adjusted mean absolute error percentage (AMAPE) indexes are used to evaluate prediction performance [20].

4. Application Example

The maximum dam height of a RCC gravity dam is 113.00 m, the total length of dam crest is 308.50 m. It is an annual regulation reservoir. In this paper, the data of horizontal displacement monitoring point P8 is selected for analysis. One displacement data is measured daily at P8. The data from 2003 to 2007 is used as training data, and the data from 2008 is used as test data. P8 is placed in No. 5 dam section, in the middle of the riverbed. The data can reflect the general law of dam horizontal displacement. The original displacement data of point P8 is shown in Figure 1.



Figure 1. The original displacement data of point P8.

The data has a slow growth trend and annual periodicity. The trend displacement is mainly due to the creep of concrete. The concrete dam has a large rigidity, so its trend displacement is small. The periodic displacement is mainly caused by temperature and water pressure. In addition, the data has a characteristic with tiny high-frequency fluctuations. Considering that the data has the characteristics of different frequencies, EMD method is used for data denoising calculation.

4.1. EMD Denoising Calculation

The P8 data is decomposed into IMF_1 - IMF_7 and a *res.* sequence by EMD. The results are shown in Figure 2. The frequency of IMF_1 - IMF_4 is high. They contain the main noise information. The IMF_5 - IMF_7 have a certain periodicity and are regarded as the intermediate frequency components. The *res.* sequence is a monotone increasing trend and is regarded as low frequency. In order to reasonably reduce the noise information from the high frequency IMFs, the CMES value should be calculated. The CMES value of high frequency IMFs is shown in Figure 3.

The CMES values of the high frequency IMFs have a minimum point 3. Therefore, the first 3 IMFs are considered as noise. After the first 3 IMFs are removed, the remaining components are used to data reconstruction. The original data and the reconstructed data of P8 are shown in Figure 4.

The reconstructed data has a high degree of consistency with the original data. Compared with the original data, the reconstructed data is smoother. This is because the reconstructed data removes the high frequency noise of the original data. The reconstructed is closer to the real displacement of P8. Therefore, the reconstructed data is used to replace the original data for SARIMA modeling analysis in this paper.





Figure 2. The EMD decomposition of original data of point P8.

Figure 4. The original data and the reconstructed data of point P8.

4.2. SARIMA Modeling

The reconstructed data also has a certain periodicity and tendency. After a 1-order conventional difference and periodic difference, the reconstructed data is converted into a stationary time series. So the parameters d and D of SARIMA model are both 1.After difference, the autocorrelation coefficient and partial autocorrelation coefficient of the reconstructed data end at order 1 and order 5, respectively. Therefore, it can be preliminarily determined that P, Q, p and q are 5, 1, 1 and 1 respectively. Several SARIMA models are selected and the evaluation indexes of each model are calculated. The results are shown in Table 1.

Model					AIC	BIC	DW	R^2
SARIMA (1,	1,	1)(1,	1,	1)	-11.93	-11.91	0.51	0.9910
SARIMA (2,	1,	1)(1,	1,	1)	-13.79	-13.77	1.66	0.9986
SARIMA (3,	1,	1)(1,	1,	1)	-14.00	-13.97	2.13	0.9989
SARIMA (4,	1,	1)(1,	1,	1)	-14.04	-14.00	1.99	0.9989
SARIMA (5,	1,	1) (1,	1,	1)	-14.11	-14.08	1.97	0.9990

Table 1. Evaluation indexes of SARIMA models.

Only the DW statistics of the last two models are between 1.8 and 2.1. It indicates that their fitting residual is stationary white noise. So, the two models are meet the requirements. The R^2 indexes of all 5 models are close to 1. This indicates that there is a great correlation between fitting data and reconstructed data. The R^2 value of the SARIMA (5, 1, 1) (1, 1, 1) model is the largest, indicating that the model has the best correlation. In addition, the SARIMA (5, 1, 1) (1, 1, 1) model has the smallest AIC and BIC values, which also indicates that the model is optimal.

To sum up, EMD denoising is performed for original data first. Then, SARIMA model is established for the reconstructed data. This is the EMD-SARIMA model. Finally, EMD-SARIMA model and SARIMA model are respectively established for reconstructed data. The fitting results of the two models are shown in Figure 5.



Figure 5. The fitting results of EMD-SARIMA and SARIMA to reconstructed data.

There is little difference between the fitting results of the two models. The correlation coefficients are both 0.96, and the errors are stationary white noise. That shows that the fitting effect of the two models are both good. However, the fitting result of EMD-SARIMA model is relatively smooth and there is no high-frequency noise. It is more consistent with the real process of dam displacement. The fitting result of SARIMA model is affected by the high frequency noise of original data.

In this paper, EMD-SARIMA model and SARIMA model are used to predict the displacement data of P8 in 2008. The results are shown in Figure 6, and the prediction indexes are shown in Table 2.

The two models have high prediction accuracy. Compared with SARIMA model, the prediction results of EMD-SARIMA model are more stable, and without high frequency noise. This is because the EMD-SARIMA can denoise the original data. SARIMA model is trained with data containing high frequency noise, so its prediction results also contain high frequency noise. In Table 2, RMSE, MAPE% and AMAPE% values of EMD-SARIMA model are all smaller than SARIMA model. It shows that

EMD-SARIMA model reduces the influence of noise and improves the prediction accuracy.



Figure 6. The predicted results of EMD-SARIMA and SARIMA to original data of point P8.

Table 2. Prediction indexes of EMD-SARIMA model and SARIMA model.

Model	RMSE	MAPE%	AMAPE%
EMD-SARIMA	0.0639	5.5395	5.2857
SARIMA	0.0688	6.1058	5.5924

5. Conclusion

Data noise reduction is of great significance to obtain stable, reliable dam monitoring data. It is helpful to improve the accuracy of data analysis. In this paper, EMD-SARIMA model and SARIMA model are used to fit and predict the original data of concrete dam displacement. The results show that the EMD-SARIMA model can effectively reduce the noise of the monitoring data, and the prediction accuracy is improved. While the SARIMA model cannot exclude the influence of noise in data, and its prediction results still contain noise

There are many kinds of dam monitoring data, and the variation law is not limited to trend and periodicity. Therefore, in order to adapt to a variety of data analysis, it is necessary to study a prediction model with universal applicability. In addition, the noise frequency of different instruments is different, and the calculation results of different noise reduction algorithms are also different. In order to effectively denoise the monitoring data, it is necessary to further study the denoising standards and methods.

References

- Gu CS, Wu ZR. Safety monitoring of dams and dam foundations-theories & methods and their application. Nanjing: Hohai University Press, 2006.8-41p.
- [2] Jiang ZX, Xu ZK, Wei BW. A monitoring model of dam displacement based on wavelet decomposition and support vector machine. Journal of Yangtze River Scientific Research Institute. 2016;33(1):43-47.
- [3] Wu ZR, Chen B. A review on development of dam safety monitoring models. Modern Surveying and Mapping. 2016;39(5):1-8.
- [4] Cheng JT, Xiong Y. Application of extreme learning machine combination model for dam displacement prediction. Procedia Computer Science.2017;107:373-378.
- [5] Lin CN, Li TC, Chen SY, Liu XQ, Lin C, Liang SL. Gaussian process regression-based forecasting model of dam deformation. Neural Computing and Applications. 2019;31(6): 8503-18.

- [6] Liu Q, Xu Y, Zhang LJ, Liang CY. Research on heart sound signal denoising algorithm based on variational mode decomposition and wavelet threshold. Journal of Computer and Communications. 2021;9(10):110-21.
- [7] Wu Y, Gu Y. Metabolic grey early warning model for dam deformation based on wavelet denoising. MATEC Web of Conferences. 2018;246(9):02033.
- [8] Li JM, Bao TF, Gao JJ, Lu YF, Yang G. Wavelet EGM-ISFLA-SVR-based hybrid prediction model for dam deformation. Water Resources and Hydropower Engineering. 2018;49(5):57-62.
- [9] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N, Tung CC, Liu HH. The empirical mode decomposition and the Hibert spectrum for nonlinear and nonstationary time series analysis. Proceedings of Roval Society of London. Series A: Mathematical, Physical and Engineering Sciences. 1998;454(3):903-995.
- [10] Xue F, Sun XW, Dong ZX, Yang HR, Wang HY. Research on data denoising algorithm based on EEMD. Mechanical Engineering & Automation. 2021;228(05):9-11.
- [11] He KJ, Chen YH, Tso GKF. Price forecasting in the precious metal market: A multivariate EMD denoising approach. Resources Policy. 2017;54:9-24.
- [12] Liu SM, Xu JT, Ju BX. Dam deformation prediction based on EMD and RBF neural network. Bulletin of Surveying and Mapping. 2019;509(8):88-91+95.
- [13] Jin SJ, Bao TF, Chen DH, Qian Q. EMD decomposition method-based dam deformation prediction model and its application. Water Resources and Hydropower Engineering. 2017;48(12):41-44.
- [14] Bui KTT, Xiao YX, Doan CV, Do XK, Tran TV, Mai DS. Application of statistic model and backpropagation neural network to analyzing and forecasting hydropower dam displacement. VNU Journal of Science Earth and Environmental Sciences. 2021;37(1):44-51.
- [15] Li B, Hu DX, Yang J, Cheng L. Displacement prediction of earth dam based on MLR-SARIMA model. Advanced Engineering Sciences. 2019;51(02):108-14.
- [16] Fang TT, Lahdelma R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Applied Energy. 2017;179(1):544-552
- [17] Cao EH, Bao TF, Liu YT, Li H. EMD-RVM-ARIMA based prediction model of dam deformation and its application. Water Resources and Hydropower Engineering. 2018;49(12):59-64.
- [18] Li L, Zhang YX, Liu SY. Denoising of rotor vibration signals based on improved EMD-wavelet analysis. Noise and Vibration Control. 2015;35(2):170-4.
- [19] Zhang J. A Case Study of Time Series Analysis Applied for Construction Monitoring of Long-Span Bridge. (Guangzhou: South China University of Technology) Master's Thesis. 2010.
- [20] Maria DCB, Ortiz J, Ballesteros L, Martorell S. Evaluation of a multiple linear regression model and SARIMA model in Forecasting 7Be air concentrations. Chemosphere. 2017;177:326-33.