DAM DEFORMATION PREDICTION BASED ON EEMD-SARIMA MODE

CHEN Tianwei [1,2] , DUAN Qingda[1,2], ZHENG Xudong[1,2] , GAN Ruo[1,2] , PAN Mei[1,2]

[1]College of Geomatic Engineering and Geoinformatics, Guilin University of Technology, 12 Jiangan Road, Guilin 541004, China;

[2]Guangxi Key Laboratory of Spatial Information and Geomatics, 12 Jiangan Road, Guilin 541004, China

KEY WORDS: Dam deformation; Deformation prediction; EEMD; SARIMA; Seasonality; Periodicity

ABSTRACT:

There are many factors affecting dam deformation, and the time series of deformation data is directly modeled without considering the seasonality and periodicity of each influencing factor, the Ensemble Empirical Mode Decomposition (EEMD) and the Seasonal Autoregressive Integrated Moving Average (SARIMA) is proposed for prediction in this paper. Firstly, the time series of deformation data is decomposed by EEMD, which weakens its volatility to some extent, and decomposes various factors affecting dam deformation, so as to obtain a series of Intrinsic Mode Function (IMF) with different frequencies; secondly, according to the seasonal characteristics and periodic characteristics of each IMF, the SARIMA model was established respectively for rolling prediction; thirdly, the final forecast results can be obtained by superimposing the forecast results of each IMF. It is verified by experiments and compared with Gray Model, Kalman Filter Model and SARIMA model that EEMD-SARIMA model has higher prediction accuracy, and it has better fitting degree, which means that it is an effective method for dam deformation prediction.

INTRODUCTION

Dam will be affected by many factors during the process of construction and use. Under the joint action of various factors, the series of deformation data observed during the process of construction and use of dam tend to be highly volatile. It is of great significance for dam safety to grasp dam deformation law and make accurate forecasts in time(JIANG .et al,2006). At present, the prediction methods commonly used in dam prediction research include time series method, grey theory, neural network, regression analysis and wavelet analysis. For dam deformation data with strong volatility, various methods have certain limitations(WEN .et al.2000). The accuracy of time series method and grey theory is lower for the data with high volatility; neural network has strong dependence on the selection of the influence factors, and it is prone to over-fitting; however, regression analysis only analyzes from the perspective of data, lacking physical significance and objectivity; wavelet analysis has strong dependence on the selection of basis function, and it lacks of adaptability(RONG .et al.2018). The key of prediction is the fitting of deformation trend. Only by accurately fitting the deformation trend of settlement data can the prediction accuracy be fundamentally improved. The main influencing factors of dam deformation include aging, temperature, water pressure, etc.(LIU .et al.2009). These factors often have obvious periodicity and seasonality, which makes deformation data have certain periodicity and seasonality. However, the periodicity and seasonality of various factors inevitably have slight differences. If we can distinguish the influencing factors, it is a reasonable method to establish a reasonable model for the periodicity and seasonality of single or a small number of influencing factors. Based on this, this paper establishes a model based on EEMD and SARIMA. Because EEMD decomposition has adaptability and decomposed each IMF have certain physical meaning(PAN .et al.2018), EEMD is used

to decompose the original observation data of the horizontal displacement of dam crest, and separate the influencing factors, so as to weaken the volatility to a certain extent, then, SARIMA model is established to better grasp the periodic and seasonal influences and fit the trend of the dam deformation, so as to improve the prediction accuracy.

1. MODEL INTRODUCTION

1.1 EEMD

EEMD is an improved model of Empirical Mode Decomposition (EMD)(ZHAO .et al.2015). Huang believed that any original signal is composed of several overlaps of eigenmode functions(WANG .et al.2010). In essence, EMD stabilizes the signal to obtain a series of IMF with gradually stable frequency and a margin B. However, when the time scale of the signal has a jump change, IMF will contain different time scale characteristic components, and EMD itself is prone to modal aliasing. In view of these situations, Huang, Wu proposed EEMD(Wu Z H .et al.2009). By adding white noise in the signal to be decomposed, EEMD utilizes the uniform distribution of the white noise spectrum to make the original signal spread over the white noise background with uniform distribution throughout the time-frequency space. Signals of different time scales will be automatically distributed on the appropriate reference scale, so that the signal has continuity on different scales, so as to achieve the purpose of suppressing mode aliasing. Because the mean value of white noise is zero, after adding white noise multiple times to obtain an average value, the additional noise will be eliminated, and the result is the signal itself. The decomposition steps of EEMD have been studied in great detail by predecessors, so this paper will not go into details.

When performing EEMD, the added white noise amplitude of normal distribution is α and the number of repetitions n needs to be determined through multiple experiments. Chen Zhong(CHEN .et al.2013)pointed out that α is usually takes

0.01~0.5, and the repetition number n is generally 100~200. Through repeated experiments, the author selects α =0.15, n =100.

1.2 SARIMA

SARIMA is an improved Model of Autoregressive Integrated Moving Average Model(ARIMA)(CHEN .et al.2018). ARIMA model is also written as ARIMA(p,d,q) (p is the autoregressive term, d is the number of differential, and q is the moving average term). It is obtained by applying the dth order differential of ARMA model. Through such operation, a non-stationary time series is transformed into a stationary time series, and then (p,q) parameters are determined by observing the truncation and trailing characteristics of correlation functions and analyzing autocorrelation and partial correlation, so as to determine the model parameters and conduct modeling of time series. However, when there are obvious cyclical trends and seasonal changes in the time series, the results obtained are often not ideal if directly analyzing it by ARIMA . At this time, the seasonal product model SARIMA(p,d,q)(P,D,Q)(p,d,q is the same as the above representatives, P is the seasonal autoregressive order, Q is the seasonal sliding average order, and D is the order of the seasonal difference)is appropriately introduced, so as to get a better analysis of the result. The time series affecting the dam settlement displacement value is often accompanied by periodicity and seasonality. Therefore, the author chooses to use SARIMA to conduct experimental analysis.

Modeling steps of SARIMA:

Step 1　Establish model ARMA(p,q)　for the time series Y_t of equidistant sampling:

$$Y_t = \zeta_q(B)\varepsilon_t / \psi_p(B) \qquad (1)$$

Among them, q is the lag operator, $\zeta_q(B)$ is the polynomial of sliding average coefficient of order q, ε_t is the white noise, and $\psi_p(B)$ is the polynomial of order p.

Step 2　When the original time series Y_t is unstable, carry out d-order differential processing according to the following formula:

$$u_t = \nabla^d X_t = (1-B)^d X_t \qquad (2)$$

Among them, u is the differential operator, d is the differential order, and ∇ is the differential symbol.

Step 3　After the d order differential, the unstable time series Y_t becomes a stationary time series. After the correlation and partial correlation analysis and observing its trailing and truncating properties, ARIMA(p,d,q) modeling can be carried out for the post-processing time series. The general form of the model is as follows:

$$\psi_p(B)\nabla^d Y_t = \zeta_q(B)\varepsilon_t \qquad (3)$$

Among them, ∇ is the differential symbol, $\nabla^d = (1-B)^d$.

Step 4　If seasonal factors exist in the original time series, SARIMA with a period of c is used to extract its seasonal periodicity, and ARIMA model is used to extract its short-term correlation(FAN .et al.2009). The two are combined to establish the SARIMA model. The general form is as follows:

$$\nabla^d \nabla_C^D Y_t = \frac{\zeta_q(B)T_m(B^C)}{\psi_p(B)\Pi_n(B^C)}\varepsilon_t \qquad (4)$$

Among them, D is the seasonal differential order, C is the seasonal period, $\nabla_C^D = (1-B^C)^d$, $T_m(B^C)$ is the seasonal m-order sliding average polynomial, and $\Pi_n(B^C)$ is an autoregressive polynomial of n-order of seasonality.

1.3　Establish EEMD-SARIMA Model

Due to the complexity of the dam project and the uncertainty of the external environment, as well as all kinds of uncertain factors exist in the process of observation, the volatility of the time series of dam observations tend to be stronger, and the original data are decomposed by EEMD, the frequencies of different IMF is obtained, and the more suitable parameters of the IMFs with different frequencies are selected to establish the SARIMA model, which can better fit the IMF of each order and is more practical. Therefore, EEMD-SARIMA model is adopted to analyze settlement data. The modeling steps are as follows:

Step 1　Perform EEMD decomposition on the original observation data to obtain a series of IMF with different frequencies;

Step 2　Parry out unit root test for each IMF order, and perform differential treatment on unstable IMF;

Step 3　Preliminarily determine the model through correlation analysis and partial correlation analysis of each IMF order;

Step 4　Adjustment model to be optimal by Akichi Information Criterion (AIC) and Schwartz Criterion (SC);

Step 5　Predict the IMF values of each order using the tested model;

Step 6　Overlaps the IMF forecast results of each order to obtain the final result.
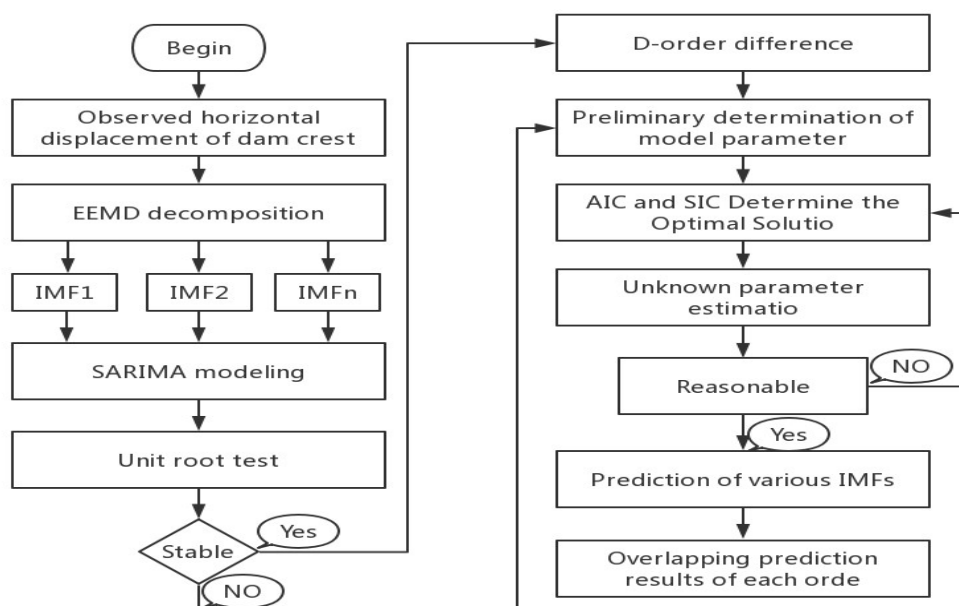
The basic flow chart of the model is as follows:

Figure.1   EEMD-SARIMA model flow chart

EXPERIMENTAL ANALYSIS

2.1   Data Analysis

This paper selects the horizontal displacement monitoring quantity of a monitoring point of Xiaolangdi Dam from 2006 to 2009(LU .et al.2012) and Fengman Dam from 1985 to 1989(JIANG .et al.2009) as the original data of this experiment. Since the sampling interval of the original observation data is extremely uneven, the author first interpolates the original data to obtain the time resolution which is the settlement data of one day, and analyzes whether it satisfies the experimental conditions. Correlation analysis and significance test were carried out on the data of each dam over the years by Statistic Package for Social Science (SPSS). (Due to the limited number of data, correlation analysis is only performed on the year in which the data is complete) .

As shown in the following table:

	Correlation	1985	1986	1987	1988	1989
1985	Pearson correlation	1	0.943**	0.933**	0.273**	0.470**
	Significance(double-tailed)		0.000	0.000	0.000	0.000
1986	Pearson correlation	0.943**	1	0.928**	0.310**	0.305**
	Significance(double-tailed)	0.000		0.000	0.000	0.000
1987	Pearson correlation	0.933**	0.928**	1	0.544**	0.512**
	Significance(double-tailed)	0.000	0.000		0.000	0.000
1988	Pearson correlation	0.273**	0.310**	0.544**	1	0.660**
	Significance(double-tailed)	0.000	0.000	0.000		0.000
1989	Pearson correlation	0.470**	0.305**	0.512**	0.660**	1
	Significance(double-tailed)	0.000	0.000	0.000	0.000	
**Indicates a significant correlation at level 0.01(double-tailed)						

Tab.1   Correlation analysis of horizontal displacement observations of Fengman dam from 1985 to 1989

According to the above table, the correlation coefficient of the observation data of Fengman Dam between 1985, 1986, 1987 and over the year is more than 0.9, and the significance is less than 0.05; and the correlation coefficient of the observation data of Fengman Dam between 1988,1989 and over the year is more than 0.3, and the significance is less than 0.05. As the number of samples selected in this paper is large and the average number is more than 300, and the increase in the number of samples increases the difference among samples, which leads to the reduction of correlation coefficient. However, through the significance test, it is proved that there is significant correlation among the observation data over the years, that is to say there is significant periodicity and seasonality in the observation data. In summary, the settlement observation data of Fengman dam is applicable for this experiment(CAO .et al.2018).

Correlation analysis of the observation data of Xiaolangdi dam shows that the correlation coefficient of 2007 and 2008 is 0.922, and the significance is 0 and less than 0.05. Therefore, the observation data of Xiaolangdi dam is also applicable for this experiment.

The observation data of the two dams over the years and the upstream water level and temperature are shown in the figure below:
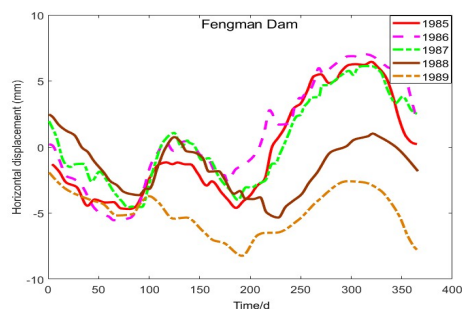

Figure.2   Observed horizontal displacement of the dam crest of Fengman dam from 1985 to 1989
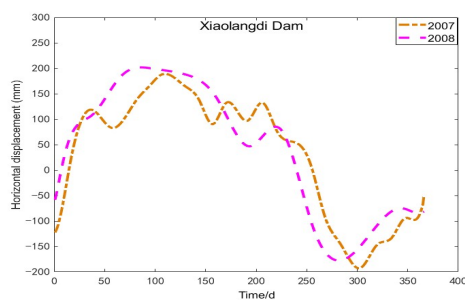

Figure.3   Observed horizontal displacement of the dam crest of Xiaolangdi dam from 2007 to 2008
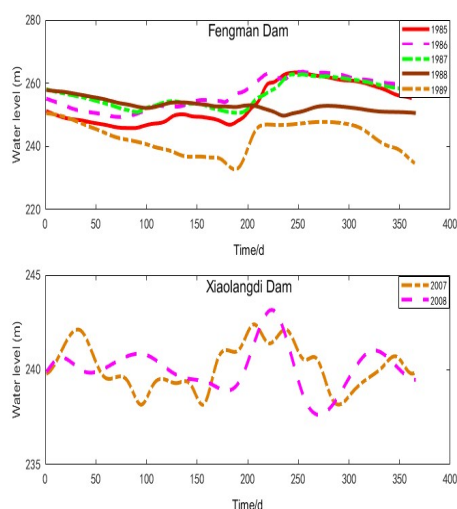

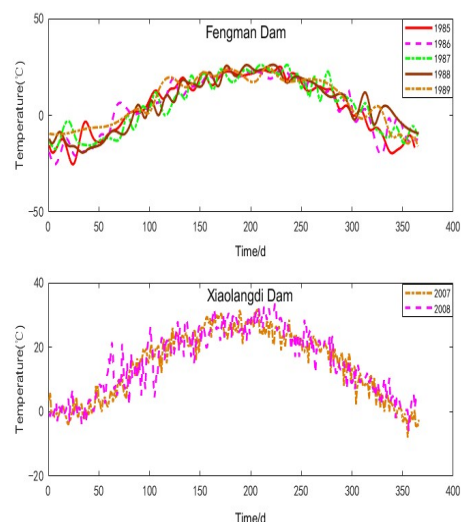Figure.4   Water level change upstream of the two dams

Figure.5   Temperature change at the dam site of two the dams

It can be seen from figure 4 and figure 5 that the two dams due to the different geographical position (the Fengman Dam is located in Jilin City, Jilin Province, and the Xiaolangdi Dam is located in Luoyang City, Henan Province), resulting in different climate differences between the two dam sites. The temperature of the dam site and the difference of the upstream water level cause different directions and sizes of the displacement components generated by the influencing factors, and the weights of the influencing factors also change over time,but for a certain period of the weight can be considered to be relatively stable(HE .et al.2003), and under the joint action of various factors, it leads to the change of the displacement of dam crest. Due to the differences in the influencing factors of the two dams, it is considered that there are seasonal differences in the horizontal displacement of the dam crests. Therefore, the author selects these two group of data for experiments,and believes that the two group of data are representative.
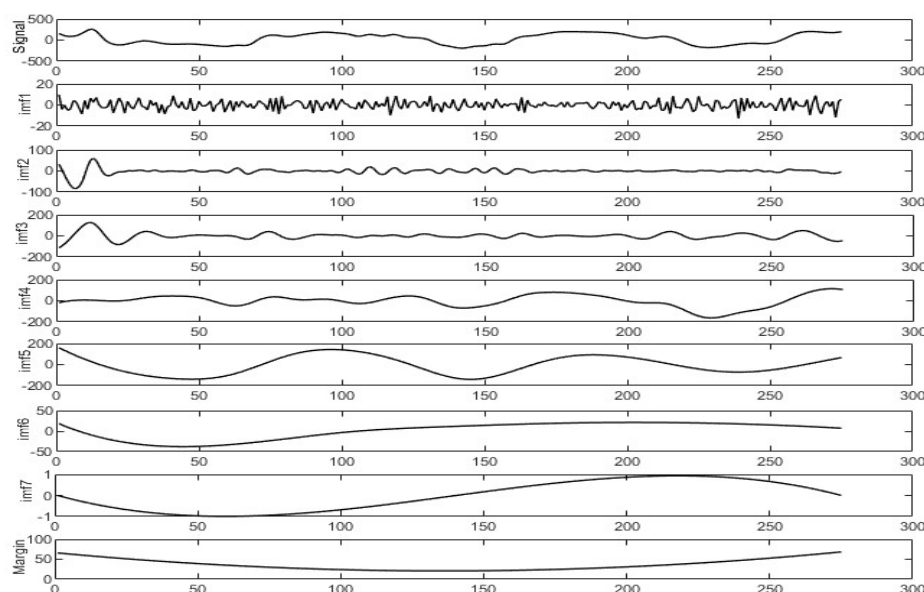
Figure.6   IMF diagram of the horizontal displacement observation of the Xiaolangdi dam crest

2.2   Prediction Experiment

Since the deformation value of the dam is small in one day, the observation data of the two dams are resampled to obtain the observation data with a time resolution of one week, and then EEMD decomposition is carried out respectively (The IMF of the observed displacement values of Xiaolangdi dam crest is shown in figure 6). The prediction experiments of IMF training SARIMA model are also carried out. The previous 215 data were used as training samples. The 10th, 15th, and 20th phases were selected as the prediction step size, and the precipitation data of the 60th phase of the 216th to 275th phases of the IMF were predicted by rolling (Taking 10 phases as the prediction step, and the data of 216-225 phases are forecasted in the first 215 phases, and the data of 226-235 phases are forecasted by 11-225 phases, and so on). After that, the prediction results of the IMF's of each order are superimposed to obtain The final forecast.

By introducing precision indexes, such as Mean Absolute Error (MAE), Mean Relative Error (MRE), Root Mean Square Error (RMSE), and by comparing with the prediction results using only the SARIMA model, the prediction results are analyzed and the model is evaluated. The mathematical expressions of the three evaluation indexes are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right| \qquad (5)$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_i - \hat{X}_i}{X_i}\right| \qquad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2} \qquad (7)$$

Among them, n is the number of prediction phase, X_i is the predicted value of phase i, and $\hat{X}_i$ is the observed value of phase i.

2.3   Analysis of Prediction Results

The prediction experiment was conducted on the observation data of the two dams, and the accuracy comparison of the prediction results was shown in the following table:

Step-size	MAE（mm）		RMSE（mm）		MRE	
	EEMD-SARIMA	SARIMA	EEMD-SARIMA	SARIMA	EEMD-SARIMA	SARIMA
10	4.77	6.65	6.25	14.77	8.24%	18.03%
15	6.88	12.94	8.43	15.73	12.76%	19.71%
20	8.18	19.88	9.57	21.80	16.03%	29.49%
AVG	6.61	13.15	8.08	17.43	12.34%	22.41%

Tab.2   Xiaolangdi dam precision comparison

Step-size	MAE（mm）		RMSE（mm）		MRE	
	EEMD-SARIMA	SARIMA	EEMD-SARIMA	SARIMA	EEMD-SARIMA	SARIMA
10	0.34	0.59	0.49	0.61	7.10%	9.49%

15	0.42	0.91	0.56	0.91	9.22%	15.03%
20	0.48	2.12	0.63	1.34	10.86%	17.73%
AVG	0.41	1.20	0.56	0.95	9.07%	14.08%

Tab.3   Fengman dam precision comparison

It can be seen from table 2 and table 3 that the prediction accuracy of EEMD-SARIMA model is higher than that of SARIMA model under three prediction steps. This is because after the decomposition of EEMD, the factors affecting the dam are decomposed to a certain extent, which weakens the volatility of the data. After the EEMD decomposition of the time series of observation data, each IMF component has a certain physical meaning, rather than the superposition of all influencing factors, the influential factors of each IMF component is reduced, making the periodicity and seasonality of each IMF easier to grasp. Through the establishment of more specific to each IMF component model, can better reflect the deformation trend and deformation law, so as to better fit the settlement data to obtain higher precision prediction values, and the model is more in line with practical significance.

According to the above table, for the EEMD-SARIAM model, when the predicted step size is 10 phases, the predicted MRE values of the two dams are 8.24% and 7.10% respectively, and

the mean value is 7.67%. When the predicted step size is 15 phases, the predicted MRE values of the two dams 12.76% and 9.23% respectively, and the mean value is 10.99%. When the predicted step size is 20 phases, the predicted MRE values of the two dams 16.03% and 10.86% respectively, and the mean value is 13.45%. With the increase of prediction step size, the prediction accuracy gradually decreases, indicating that the EEMD-SARIMA model is better for short-term prediction and less effective for medium- and long-term prediction. The MRE value of all predicted step sizes is 11.6%, indicating that the prediction effect is good, and the EEMD-SARIMA model can be used as a predictive model for dam deformation. For the SARIMA model, when the predicted step size is 20 phases, the MRE of the predicted values of the two dams are 29.49% and 17.73% respectively, and the mean value is 23.61%, and from the prediction process, the reason the prediction accuracy is low is that the trend of deformation can not be well grasped, resulting in a large deviation between the predicted data and the actual observed data, as shown in the following figure:
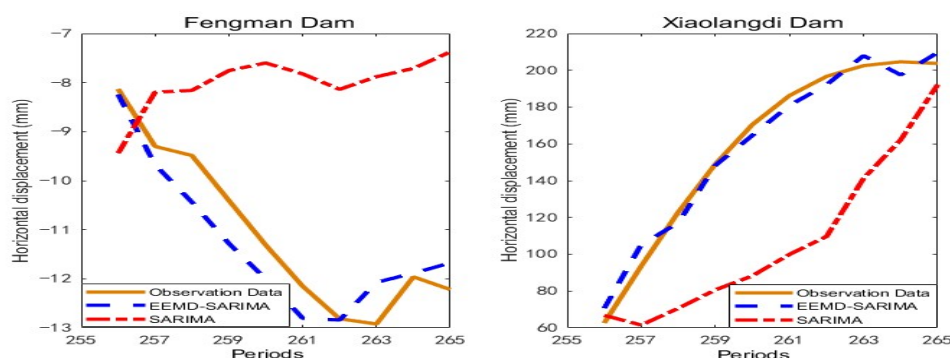

Figure.7   Forecast of the two dams 256-265

In the above two cases, the trend of the original observation data of the prediction results of the SARIMA model is obviously different. Further correlation analysis of the observation data shows that the correlation coefficients are 0.7015 and 0.6882 respectively, which are basically predictive failures. The prediction results of the EEMD-SARIMA model are basically consistent with the trend of the original observation data, and the correlation analysis results show that the correlation coefficients are 0.9726 and 0.9919 respectively. Therefore, compared with the SARIMA model,

the EEMD-SARIMA model can better fit the deformation trend of the horizontal displacement of the dam crest, which has strong applicability and stability. Further comparison of the prediction effects of the EEMD-SARIMA model. Kalman Filter Model and Grey Model are used to predict the data adopted in this paper (Observation data in the first 215 phases are used as training samples, and rolling prediction of data in the 216-275 phases is made with 10 phases as prediction step size). The prediction results are shown in figure 8:
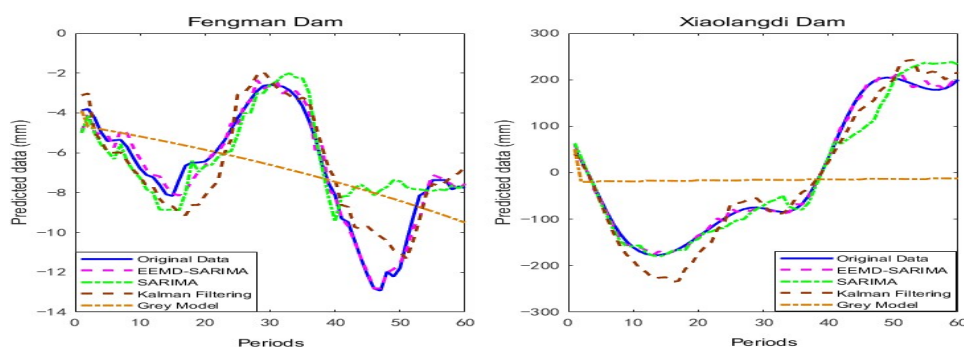

Figure.8   Comparison of prediction results

As can be seen from the figure, compared with other models, EEMD-SARIMA model can better fit the change trend of data. Further analysis of the prediction accuracy shows that, for Fengman Dam, the RMSE, MAE, and MRE of Kalman Filter are 1.12mm, 0.91mm, and 13.97% respectively; while the accuracy indexes of EEMD-SARIMA model are 0.34mm, 0.49mm and 7.1% respectively. In contrast, the prediction accuracy of the model proposed in this paper is better; and the trend of the prediction results of the Gray Model is obviously different from the original trend. For Xiaolangdi Dam, the RMSE, MAE, and MRE of Kalman Filter are 30.12mm, 25.28mm, and 20.64% respectively; while the accuracy indexes of EEMD-SARIMA model are 4.77mm, 6.25mm and 8.24% respectively; and the trend of the prediction results of the Gray Model is obviously different from the original trend. In conclusion, compared with SARIMA model, Kalman Filter and Gray Model, EEMD-SARIMA model proposed in this paper has higher prediction accuracy and better fitting for deformation trend.

CONCLUSIONS

In this paper, the observation data of the horizontal displacement of the dam crest of Xiaolangdi Dam and Fengman Dam are analyzed and the prediction experiment is carried out. From the above analysis process, the author draws the following conclusions:

1) The time series of horizontal displacement of the crest has strong seasonal characteristics. Although the prediction value can be obtained when only using the SARIMA model for prediction, there are many factors affecting the deformation of the dam, and the seasonal characteristics exist difference of various factors, the prediction accuracy is not good. The EEMD-SARIMA model is used to decompose the observations first, so that the various influencing factors are separated, and the influence of the differences between the factors is reduced, thereby improving the prediction accuracy.

2) The EEMD-SARIMA model weakens the volatility of the data to a certain extent in the prediction of dam deformation, it has strong stability, good applicability and certain physical significance. Compared with Kalman Filter and Grey Model, the prediction accuracy of EEMD-SARIMA model is better, and the results show that this model is an effective method for dam deformation prediction.

REFERENCES

CAO Wenhan, ZHANG Qiang, LUO Xiaoqin,et,al. Estimation nitrate of concentration in drinking water besed on cooperative Kriging interpolation[J]. China Rural Water and Hydropower, 2018(06):84-87.

CHEN Yutian, LIU Lilong, LIU Zhongliu, et al. Analysis of the influence of latitude factors on seasonal time series TEC short-term forecast model[J], Journal of Geodesy and Geodynamics, 2018,38(06):581-586.

CHEN Zhong, Fu Hechao. Electromagnetic noise diagnosis for a permanent magnet synchronous motor based on EEMD[J], Journal of VIBRATION and Shock, 2013,32(20):124-128.

FAN Jie, YAN Guangle. Analysis of typical time series model of Shanghai's economic development[J], Statistics and Decision, 2009(21):71-73.

HE Yongbing. Application of multiple grey predicting model for monitoring dam safety[D],Jiangsu:Zhejiang University, 2003:19-25.

JIANG Chengke. Application of neutral network based on genetic algorithm in dam displacement forecast[D]. Liaoning: Dalian University of Technology,2009:68-74.

JIANG Tingchen, ZHANG Qin, ZHOU Li, et al. Research on nonlinear regression model based on wavelet method[J]. Acta Geodaetica et Cartographica Sinica, 2006(04):337-341.

LU Kaiyun. Reserch and analysis about the deformation forecasting methods of Yellow River Xiaolangdi water hydropower dam[D]. Beijing: China Mining University(Beijing),2012:49-75.

LIU Suzhong,ZHAO Guangchao. Review of dam safety monitoring research[J].China Water Transport, 2009,9(11):147-148.

PAN Heping, ZHANG Chengzhao. FEPA:An adaptive integrated prediction model of financial time series[J], Chinese Journal of Management Science, 2018(06):26-38.

RONG Jing, LIU Lilong, GAN Xiangqian, et al. High speed rail deformation prediction based on wavelet analysis and grey-support vector machine[J]. Journal of Geodesy and Geodynamics, 2018,38(05):473-476+545.

WEN Hongyan, Application of wavelet multisolution analysis to deformation analysis[J].Crustal Deformation and Earthquake, 2000(03):27-32.

WEN Hongyan, Reserch on deformation analysis model based upon wavelet transform theory[J]. Acta Geodaetica et Cartographica Sinica, 2005(02):186-187.

WANG Ting. Research on EMD algorithm and its application in signal denoising[D]. HElongjiang: Harbin Engineering University, 2010: 12-26.

Wu Z H, Huang N E. Ensemble empirical mode decomposition: a noise assisted data analysis method [J]. *Advances in Adaptive Data Analysis*, 2009, 1(1): 1-41.

ZHAO Xiaoyu. Reserch and application on adaptive spectra preprocessing methods based on EMD and EEMD[D]. Hebei: YanShan University, 2015: 15-30.