Strain Prediction of Bridge SHM Based on CEEMDAN-ARIMA Model

Siyu Bian*, Jingchao Zhuo^b and Liming Zhu^*

School of Transportation Science & Engineering, Nanjing Tech University, Nanjing, Jiangsu, China

*Corresponding author e-mail: 188383679@qq.com, ^601940038@qq.com, ^b617142003@qq.com

Abstract. In this paper, a model based on CEEMDAN-ARIMA is proposed to predict the strain monitoring data for bridge SHM. In view of the problem that the classical time series theory cannot predict the modal overlapping data effectively, the CEEMDAN method was used to decompose the strain monitoring data for the bridge SHM. To deal with the large number of components after using CEEMDAN, the PE method (permutation entropy) was used to generate a series of new data sequences according to the degree of randomness. Finally, each new data sequence was predicted and the final prediction is obtained by ARIMA model. The method was used to predict the SHM strain data of a cable-stayed bridge in Shanghai. The results show that the proposed combination method is more accurate than the classical time series theory and is promising for engineering applications.

1. Introduction

Bridge structural health monitoring (SHM) system has become an indispensable part of bridge asset management in recent years, not only for super-long bridges, but also for small and medium-span bridges. SHM system collects a large number of monitoring data, and the analysis and prediction of these data is an important step for the follow-up structural assessment and for providing early warning of bridge failure.

The monitoring data of a bridge SHM system are often in the form of time series, therefore the analysis and prediction of monitoring data is mainly performed through analyzing the time series. Tang Hao et al [1] established an ARMA (autoregressive moving average) model for the strain monitoring data of the SHM system of Xi'an Baisheyu Bridge. The results show that the prediction error is essentially less than 10%, which is proven to be a good prediction model. Zeng et al [2] used the Multiple Season ARIMA (autoregressive integrated moving average) model to accurately predict the bridge arch displacements.

In the field of bridge SHM, time series analysis and prediction of monitoring data mostly remain in the application of classical time series analysis theory, and mostly use a single prediction model [3]. Due to the environmental conditions or sensor itself, it is inevitable that there will be strong non-stationary random fluctuations and modal overlapping [4]. As such, using classical time series analysis theory for strain data analysis and prediction may produce relatively large errors.
CEEMDAN (Complete ensemble empirical mode decomposition with adaptive noise) is the latest empirical mode decomposition method [5]. It can decompose signals adaptively according to different frequency bands and is very suitable for the analysis and processing of non-stationary signals. The CEEMDAN method is developed by EMD (empirical mode decomposition) [6], EEMD (ensemble empirical mode decomposition) [7] and CEEMD (complementary ensemble empirical mode decomposition) [8], which effectively solves the problem of modal overlapping and reduces the reconstruction error, saving computation time.

In view of the shortcomings of the classical time series analysis theory in bridge SHM monitoring data analysis and prediction, a prediction model based on CEEMDAN-ARIMA is proposed herein and is verified by SHM measured strain monitoring data of a cable-stayed bridge in Shanghai.

2. Analysis of prediction theories

2.1. CEEMDAN signal decomposition method

CEEMDAN is the latest research progress of EMD method, and its specific algorithm is as follows:

1. Add white noise $\beta_0 w(t)^{(i)} (i = 1, 2, \ldots, I)$ obeying the standard normal distribution to the original signal $x(t)$, $I$ is the number of trials. Five steps are involved in this method.

2. EMD decomposition is performed for each $x(t)^{(i)} = x(t) + \beta_0 w(t)^{(i)} (i = 1, 2, \ldots, I)$ to obtain the first modal component function $imf_1$ and the margin $r_1$:

   $imf_1 = \frac{1}{I} \sum_{i=1}^{I} E_1[x(t)^{(i)}] \tag{1}$

   $r_1 = x(t) - imf_1 \tag{2}$

   where $E$ is the EMD decomposition operator.

3. Adding white noise $\beta_1 E_1[w(t)^{(i)}]$ to the margin $r_1$ to form a new signal, and then performing EMD decomposition to obtain the first modal component function of the new signal as the second modal component function $imf_2$ of the original signal:

   $imf_2 = \frac{1}{I} \sum_{i=1}^{I} E_1(r_1 + \beta_1 E_1[w(t)^{(i)}]) \ (i = 1, 2, \ldots, I) \tag{3}$

4. For $k = 2, \ldots, K$, calculate the $k$th margin:

   $r_k = r_{k-1} - imf_k \tag{4}$

5. At each stage, white noise is added to form a new signal, and the first modal component of the signal is calculated as a new modal component of the original signal. Then the $k$th modal component function is:

   $imf_{k+1} = \frac{1}{I} \sum_{i=1}^{I} E_1(r_k + \beta_k E_1[w(t)^{(i)}]) \ (i = 1, 2, \ldots, I) \tag{5}$

6. Repeat Steps 3 and 4 until the margin cannot be further decomposed by EMD, or meet the IMF condition or less than three local maximums, then all the components $imf$ are found.

   All components are reconstructed to obtain the original signal:

   $x(t) = \sum_{k=1}^{K} imf_k + r_k \tag{6}$

CEEMDAN allows the selection of an appropriate SNR (signal-noise ratio) in the white noise added at each decomposition stage, so the calculation process is self-adaptive.
2.2. **Permutation entropy algorithms (PEA)**

It is possible to obtain a large number of signal time series components by signal decomposition, and the practical significance of these components is usually difficult to be identified. In order to solve this problem, this thesis introduces an algorithm proposed by Bandt et al. [9]: PE (Permutation Entropy) algorithm. The general idea of the PE algorithm is to calculate the average entropy parameter of the time series. The larger the entropy value, the stronger the randomness of the time series; otherwise, the more regular the time series. The calculation steps of the algorithm are:

1. Perform phase space reconstruction on a one-dimensional time series \( \{x(t), t = 1,2,\ldots,n\} \) to obtain a reconstruction matrix:

\[
\begin{bmatrix}
  x(1) & x(1 + r) & \cdots & x(1 + (m - 1)r) \\
  x(2) & x(2 + r) & \cdots & x(2 + (m - 1)r) \\
  \vdots & \vdots & \ddots & \vdots \\
  x(j) & x(j + r) & \cdots & x(j + (m - 1)r)
\end{bmatrix}
\]  

where, \( j = 1,2,\ldots,n - (m - 1)r; \) \( r \) is the delay time; \( m \) is the reconstruction dimension.

2. The reconstructed row vector \( Y_r = [x(j), x(j + r), \ldots, x(j + (m - 1)r)] \), is the index of the column in which each element is located, and the elements are reordered in an ascending order according to the size of the value. If the values of the adjacent elements are equal, they are arranged according to the order of the index, and each set of row vectors of the reconstructed matrix can lead a set of symbol sequences \( S(r) = \{j_1, j_2, \ldots, j_m\} \), where \( r = 1,2,\ldots,l \). It is obvious that the number \( l \) of arrangement of the symbol sequence \( S(r) \) is at most \( m! \).

3. Calculate the probability of occurrence \( P_1, P_2, \ldots, P_l \) of each symbol sequence \( S(r) \), and define the permutation entropy of the time series \( H_P(m) \) in the form of Shannon entropy:

\[
H_P(m) = -\sum_{i=1}^{l} P_i \ln P_i
\]

(8)

4. For convenience, the permutation entropy \( H_P(m) \) is usually normalized. It can be concluded from Equation 8 that there is a maximum value \( \ln(m!) \) when \( P_i = 1/m! \), so the normalization \( H_P(m) \) is performed:

\[
H_p = \frac{H_P(m)}{\ln(m!)}
\]

(9)

The normalized permutation entropy \( H_p \) is obviously in the range of [0, 1], and the \( H_p \) value reflects the complexity of the time series. The more the value is, the more regular the time series is; otherwise, the more random it is. Therefore, the degree of randomness of each component can be quantitatively determined, and the components are classified according to the basis, and the same type of components are combined to achieve the purpose of reducing the signal component and improving the calculation efficiency.

2.3. **ARIMA prediction model**

The ARIMA predictive model is the most classical theory of time series analysis and can predict non-stationary time series. ARIMA \( (p, d, q) \) is expressed as:

\[
\Phi(B)Y^d = \Theta(B)e_t
\]

(10)

where, \( x_t(t = 1, 2, 3\ldots, n) \) is the time series of monitoring data; \( e_t(t = 1, 2, 3\ldots, n) \) for the residual; \( B \) for the delay operator and \( B^n x_t = x_{t-n} \); \( \Phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i \); where \( \phi_i \) is the autoregressive coefficient; \( \theta^d \) for the difference operation and \( \theta^d = (1-B)^d \); \( \theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i \), where \( \theta_i \) is the moving average coefficient.
The general steps of the ARIMA prediction model are: stationary and pure randomness test of time series, model identification, parameter estimation, model checking and data prediction. The specific process is detailed in the literature [10].

3. Model construction
Based on the above-mentioned theories, the SHM monitoring data is decomposed and processed by the signal analysis method (CEEMDAN and PEA), and subsequently the decomposed signals are predicted by the ARIMA model. The process of predicting and analyzing the strain monitoring data based on the CEEMDAN-ARIMA combination model proposed in this paper is shown in Fig. 1:

![Prediction flow based on CEEMDAN-ARIMA combination model](image)

**Figure 1.** Prediction flow based on CEEMDAN-ARIMA combination model.

4. Case study
In this case, a sample of the mid-span web strain health monitoring data of a cable-stayed bridge in Shanghai collected from July 1st to 5th, 2017 was selected, with a total of 720 data, at an interval of 10 minutes. These data samples were used to predict the strain monitoring data and verify the applicability of the proposed model. Fig. 2 is a timing diagram of the mid-span web strain monitoring data of a cable-stayed bridge.

![Strain sequence diagram](image)

**Figure 2.** Strain sequence diagram.
4.1. **CEEMADN signal decomposition**

Taking the strain monitoring data as the signal data, the CEEMDAN method is used to decompose the variable time series. Fig. 3 shows the result of the decomposition. Seven $imf$ components and one margin $r$ are obtained, which are arranged in a descending order of the frequency. It can also be found that the degree of randomness is also arranged from high to low.

4.2. **PE restructuring**

After CEEMDAN decomposition, more strain components are obtained, and the properties are not obvious enough. Only the level of randomness can be determined qualitatively. Direct data prediction will increase the difficulty of calculation and reduce the accuracy of prediction. Therefore, the PE method is used to calculate the entropy value of each component, and then these components are reclassified and recombined [11].

The entropy value of each component is calculated separately. According to our experience, the reconstruction dimension $m$ of PE calculation parameters is usually 3~7, and the delay time $\tau$ is usually 1 [12]. In this example, $m=3$, $\tau=1$, and the PE calculation result is as shown in Fig. 4, and the PE values of the respective components are 0.995, 0.863, 0.692, 0.535, 0.463, 0.430, 0.408, and 0, respectively.

![Figure 3. CEEMDAN decomposition of strain data.](image3)

![Figure 4. PE values of each component.](image4)
Figure 5. New sequence after PE recombination.

It can be seen from Fig. 4 that as the frequency decreases, the PE value of each component also gradually decreases, that is, the degree of randomness becomes smaller and smaller. The \textit{imf}1 component has the highest degree of randomness. Together with \textit{imf}2, \textit{imf}3, and \textit{imf}4, the PE values show a linear declination. Therefore, these components can be grouped together and merged into \textit{c}1. It can be seen from the figure that the PE values of \textit{imf}5, \textit{imf}6, and \textit{imf}7 are a relatively constant, and the degree of randomness is similar, therefore are merged and recombined into \textit{c}2. The PE value of the margin \textit{r} is 0. It can also be seen from Fig. 4 that the margin \textit{r} is a linear sequence and is a very stable sequence, so the margin \textit{r} can be separately listed as a group \textit{c}3. The outcome of the PE reorganization is shown in Fig. 5. As can be seen from Fig. 5, the recombined component has been reduced from the original 8 components to 3 components, greatly reducing the number of computational efforts.

4.3. Model prediction

As can be seen from Fig. 5, Component \textit{c}1 has the strongest degree of randomness, Component \textit{c}2 exhibits a relatively pronounced periodicity, and Component \textit{c}3 is approximately linear. Therefore, the 50th period lengths of Components \textit{c}1, \textit{c}2, and \textit{c}3 are predicted by the ARIMA model, and the linear regression model, respectively.

Firstly, ARMA (3, 1) model is adopted for signal component \textit{c}1, ARIMA (1, 1, 0) model is adopted for signal component \textit{c}2, and linear regression model is adopted for signal component \textit{c}3. After that, the future 50 data of the signal components \textit{c}1, \textit{c}2, \textit{c}3 are predicted respectively. Finally, the prediction data of the signal components \textit{c}1, \textit{c}2, \textit{c}3 are superimposed to obtain the prediction results of the original strain time series \textit{s} in the next 50 periods. In order to verify the effectiveness of the CEEMDAN-ARIMA model, a single ARIMA model was used to predict the strain health monitoring data. The prediction results of the other three models are shown in Figs. 6-7.
4.4. Analysis of strain forecast results

It can be seen clearly from the comparison of Fig. 6 and Fig. 7 that the predictions of the single ARIMA model is poor. Meanwhile the prediction results of the prediction models after CEEMDAN processing is better. In order to evaluate the prediction accuracy of each model quantitatively, the performance of each model was evaluated using three statistical indicators: MSE (mean square error), MAPE (mean absolute percent error) and $R^2$ (coefficient of determination). MSE represents the error between the measured value and the predicted value. It can compare the reliability of the different prediction models. MAPE not only considers the error between the predicted value and the real measurement, but also considers the proportion between the error and the real value, which is a measure of the accuracy of prediction. $R^2$ represents the proportion of the total dispersion square sum that can be explained by the regression square sum, representing the quality of the regression effect. The smaller the MSE and MAPE are, the larger $R^2$ is, indicating the better prediction performance of the model.

| Statistical indicators | MSE  | MAPE   | $R^2$ |
|------------------------|------|--------|-------|
| ARIMA                  | 28.06| $4.94 \times 10^{-2}$ | 0.120 |
| CEEMDAN-ARIMA          | 7.51 | $2.61 \times 10^{-2}$ | 0.765 |

Table 1 shows a comparison of the statistical indicators of the four models. It can be seen from the table that the ARIMA model based on CEEMDAN has better prediction outcome, while the performance of the ARIMA model is poor, with the $R^2$ statistical index of about 0.12, indicating that the prediction of 50 period length is meaningless. The comparison results show that the classical time series prediction model cannot meet the requirements of prediction accuracy in predicting the strain monitoring data of bridge SHM, and the CEEMDAN-ARIMA model proposed in this study can predict the non-stationary strain monitoring data more accurately, which demonstrates obvious advantages.

5. Conclusion

In this study, based on the latest achievement of EMD: CEEMDAN method, combined with the ARIMA model of classical time series analysis theory, a new combination model is proposed, and is used to analyze and predict the strain monitoring data for bridge SHM activities. After verification against the measured strain data of the SHM system of a cable-stayed bridge in Shanghai, the results show that:
(1) The CEEMDAN method can effectively decompose the random load responses, periodic temperature responses and overall temperature rise and fall responses of bridges.

(2) The PE algorithm can effectively reduce the imf component decomposed by the CEEMDAN method, which greatly reduces the calculation time and cost.

(3) Combining the CEEMDAN method with the classical time series ARIMA prediction model, and compared with the single model, the accuracy of the prediction model due to the proposed combination method is effectively improved.

References
[1] Tang Hao, Meng Libo, and Song Gang, et al. (2015). "Study of Predicting Techniques Based on ARMA Used to Process Bridge Monitoring Information", World Bridges, Vol. 43: pp 44-48.
[2] Zeng Faming, and Wang Fuyuan. (2016). "Application of Multiplicative Seasonal ARIMA Model in Bridge Abutment Displacement Monitoring", Jiangxi Kexue, Vol. 34, No. 6: pp 805-808.
[3] Farahani, R. V., and Penumadu, D. (2016). "Full-scale bridge damage identification using time series analysis of a dense array of geophones excited by drop weight", Structural Control and Health Monitoring, Vol. 23, No. 7: pp 982-997.
[4] Zhang, W., Qu, Z., Zhang, K., Mao, W., Ma, Y., and Fan, X. (2017). "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting", Energy conversion and management, Vol. 136: pp 439-451.
[5] Torres M E, Colominas M A, and Schlotthauer G, et al. (2011). "A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise", IEEE International Conference on Acoustics, Speech and Signal Processing.
[6] Huang N E, and ZhengLong S. (1998). "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis", Proceedings Mathematical Physical & Engineering Sciences, Vol. 454: pp 903-995.
[7] Zhaohua W, and Norden E H. (2009). "Ensemble empirical mode decomposition: A noise-assisted data analysis method", Advances in Adaptive Data Analysis, Vol. 1: pp 1-41.
[8] Jia-Rong Y., Jiann-Shing S, and Norden E H. (2010). "Complementary Ensemble Empirical Mode Decomposition: A Novel Noise Enhanced Data Analysis Method", Advances in Adaptive Data Analysis, Vol. 2, No. 2: pp 135-156.
[9] Bandt C., and Pompe B. (2002). "Permutation Entropy: A Natural Complexity Measure for Time Series", Physical Review Letters, Vol. 88: pp 21-24.
[10] Wang Yan. (2015). "Applied Time Series Analysis", 4th ed. Beijing: China Renmin University Press.
[11] Huang, S., Wang, X., Li, C., and Kang, C. (2019). "Da-ta decomposition method combining permutation en-tropy and spectral substitution with ensemble empir-ical mode decomposition", Measurement, Vol. 139: pp 438-453.
[12] Rao Guoqiang, Feng Fuzhou, and Si Aiwei, et al. (2014). "Method for Optimal Determination of Parameters in Permutation Entropy Algorithm", Journal of Vibration and Shock, Vol. 33: pp 188-193.