



**ARTICLE**

# Outlier Detection and Forecasting for Bridge Health Monitoring Based on Time Series Intervention Analysis

Bing Qu\*, Ping Liao and Yaolong Huang

College of Civil Engineering, Putian University, Putian, 351100, China

\*Corresponding Author: Bing Qu. Email: bingqu@ptu.edu.cn

Received: 14 January 2022 Accepted: 18 July 2022

## ABSTRACT

The method of time series analysis, applied by establishing appropriate mathematical models for bridge health monitoring data and making forecasts of structural future behavior, stands out as a novel and viable research direction for bridge state assessment. However, outliers inevitably exist in the monitoring data due to various interventions, which reduce the precision of model fitting and affect the forecasting results. Therefore, the identification of outliers is crucial for the accurate interpretation of the monitoring data. In this study, a time series model combined with outlier information for bridge health monitoring is established using intervention analysis theory, and the forecasting of the structural responses is carried out. There are three techniques that we focus on: (1) the modeling of seasonal autoregressive integrated moving average (SARIMA) model; (2) the methodology for outlier identification and amendment under the circumstances that the occurrence time and type of outliers are known and unknown; (3) forecasting of the model with outlier effects. The method was tested with a case study using monitoring data on a real bridge. The establishment of the original SARIMA model without considering outliers is first discussed, including the stationarity, order determination, parameter estimation and diagnostic checking of the model. Then the time-by-time iterative procedure for outlier detection, which is implemented by appropriate test statistics of the residuals, is performed. The SARIMA-outlier model is subsequently built. Finally, a comparative analysis of the forecasting performance between the original model and SARIMA-outlier model is carried out. The results demonstrate that proper time series models are effective in mining the characteristic law of bridge monitoring data. When the influence of outliers is taken into account, the fitted precision of the model is significantly improved and the accuracy and the reliability of the forecast are strengthened.

## KEYWORDS

Structural health monitoring; time series analysis; outlier detection; bridge state assessment; bridge sensor data; stress forecasting

## 1 Introduction

In recent years, bridge health monitoring (BHM) system has become an inseparable part in not only super-major and major bridges but also small and medium-sized bridges. Vast amounts of monitoring data, which contain a variety of characteristic information of the structure under the operation phase, flow into BHM system every day. How to make effective use of the monitoring data for in-depth mining is of vital importance for bridge early warning and assessment [1,2]. However, in the process of feature



extraction, the complexity of the identification problem has far exceeded the traditional data processing capability and understanding mode in the domain of structural engineering, due to its nonlinearity, incompleteness, and noise interference, which brings huge challenges to bridge monitoring. The phenomenon of “Rich data but poor knowledge” is widely observed in the existing BHM systems [3,4]. Big data theory, which has rapidly risen and improved of late years, provides a possible breakthrough for the processing of massive monitoring data [5–8]. Sun et al. [9,10] investigated the research orientation and application status of big data in bridge health monitoring, established the big data analysis framework and pointed out that current research on big data in the field of bridge health monitoring is still at the initial stage, but it has got a lot of potential.

Time series techniques, originally developed for analyzing long sequences of regularly sampled data, are inherently suitable for BHM. Time series forecasting methods can be broadly classified into two main categories, namely statistical methods and deep learning methods. The performance of each method depends on multiple factors such as trend, seasonality and noise in the data, as well as external conditions and internal damages [11,12].

Deep learning (DL) techniques, which can automatically learn the temporal dependencies present in time series and effectively reduce the complexity of the forecasting pipeline, have proved to be an effective solution in time series forecasting, and have gained considerable prominence [13,14]. Oh et al. [15] proposed a CNN-based architecture to predict strain levels of tall buildings under wind loadings. The training dataset containing displacements and wind speeds was collected from a wind tunnel test of a model of a steel structure. Peng et al. [16] adopted piecewise linear least squares (PLLS) method, fully connected neural network (FCNN) method, and long short-term memory neural network (LSTMNN) method to predict the structural dynamic response of a six-story steel frame under periodic, impact, and seismic loads. Results showed that the LSTMNN method performed better than the other two methods. In addition, the PLLS method was sensitive to noise, while the FCNN and LSTMNN methods based on deep learning were of highly robust and anti-noise performance. Zheng et al. [17] established an LSTM neural network for modeling multiple temperature-displacement correlations. The results revealed that compared with the BP neural network model, the proposed LSTM neural network model could dramatically reduce the reproduction error and prediction error of the thermal displacement.

Time series analysis is one of the statistical procedures applied to simulate the degradation mechanism of bridge structures and make predictions by establishing various time series data mining models and algorithms that can reflect the variation of variables in the time domain [18–20]. van Le et al. [21] used ARIMA model for the analysis of the GPS time series data acquired in a cable-stayed bridge in Vietnam, and data were used to predict the static responses and global deformation. Conclusions were drawn that the AR-MA coefficients plots could be used as the base distributions for the statistical structural condition assessment. Zhu et al. [22] proposed a combination forecast model based on CEEMDAN-NAR-ARIMA, and used it to predict the SHM strain data of a cable-stayed bridge in Shanghai. Results showed that the proposed method was more accurate than classical time series theory. Ahmadivala et al. [23] applied seasonal ARIMA on the mean values of strain monitoring data on Chillon viaducts for every 12 h, so as to explore more details of the loading scenario regarding the seasonal effects of traffic loading. Xin et al. [24] used the Kalman filter to reduce the noise of the bridge’s raw deformation data and developed ARIMA-GARCH model to analyze and predict the structure’s deformation. Shi et al. [25] decomposed the observed SHM time series into three components, namely, level, seasonal and residual, and purged the influence of seasonality, in order to obtain a more agreeable series that could reflect the characteristics of structural damage. Jiang et al. [26] adopted ARIMA model integrated with singular spectrum analysis (SSA) for a more precise prediction of stress monitoring data of Sutong bridge.

It can be seen that an appropriate time series model or combination model does better explain the characteristics of nonlinearity, non-stationarity, high dimensionality and heteroscedasticity of monitoring data. However, these models were built on the assumption that the monitoring data were accurate and reliable [27,28]. In a long-term continuous monitoring system, the regular data collection is inevitably disturbed by accidental extreme loads, external forces on the sensors themselves, abnormal current or voltage, power failure, sensor damage, etc. These external interruptive events, referred to as interventions, may create spurious observations that are inconsistent with the rest of the series and interfere with the identification of the time series model [29]. Therefore, identification and characterization of outlier interventions are important from the point of view of BHM management as these events could probably reduce the precision of the forecast results of structural behavior.

However, from the existing literature, most of the studies focus on the selection of the most convenient type of time series model and its parameter recognition and estimation [30,31]. Research that discusses the intervention effect of outliers on fitted precision of the models and on assessment of structural behavior is rarely seen. Therefore, in this study, the possible abnormal conditions in bridge monitoring data are investigated from the perspective of intervention analysis. First, a brief description of the SARIMA model with outliers is outlined. The methodology of outlier detection and amendment then follows, and the identification and prediction of time series model containing outliers are discussed. The proposed procedure is then used to identify the outliers from strain data recorded by a BHM system and to predict the structural future condition.

The main purpose of outlier correction is to optimize the data in such a way that the normality hypothesis of the SARIMA model can be better accepted. Moreover, by containing outlier intervention effect in the SARIMA model, the residual variance of the model is reduced, the fitting precision of the model is significantly improved, and the accuracy and reliability of the forecast are strengthened.

## 2 Nonstationary SARIMA Models

Under the influence of external loads and structural performance degradation, the monitoring responses of bridge structures exhibit randomness in a short period of time. However, from a longer time perspective, the monitoring time sequences will present certain regularity, such as long-term trend, periodicity, random fluctuation, and mutation. As can be seen from the sequence diagram of observed data, there exists significant non-stationarity in some of the monitoring series, while others have conspicuous seasonal characteristics due to the effect of seasonal temperature difference. Therefore, the seasonal ARIMA model is introduced to fit and analyze the nonstationary bridge monitoring data [32].

Assume  $\{Y_t\}$  is a time series observed from a certain sensor on BHM system, and  $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t\}$  is a zero-mean multivariate Gaussian white noise series. If a stationary series can be obtained by taking a  $d$ th-order nonseasonal difference  $\nabla^d$  and a  $D$ th-order seasonal difference with period  $s$   $\nabla_s^D$ , then the general SARIMA model can be represented analytically as

$$\begin{cases} \Phi(B)\Psi(B^s)\nabla^d\nabla_s^D y_t = \Theta(B)H(B^s)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \text{cov}(\varepsilon_h, \varepsilon_t) = 0, h \neq t \\ \text{cov}(y_h, \varepsilon_t) = 0, \forall h < t \end{cases} \quad (1)$$

The model in Eq. (1) is normally denoted as ARIMA  $(p, d, q) \times (P, D, Q)_s$ , where  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  is the nonseasonal autoregressive (AR) operator of order  $p$ ,  $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  is the nonseasonal moving average (MA) operator of order  $q$ , and polynomials  $\Psi(B^s) = 1 - \psi_1 B^s - \psi_2 B^{2s} - \dots - \psi_p B^{Ps}$  and  $H(B^s) = 1 - \eta_1 B^s - \eta_2 B^{2s} - \dots - \eta_q B^{Qs}$  respectively describe the seasonal AR and MA operators of orders  $P$  and  $Q$  with seasonal period  $s$ ;  $\phi_i, \theta_i, \psi_i, \eta_i$  ( $i = 1, \dots, n$ ) are the coefficients of  $\Phi(B)$ ,  $\Theta(B)$ ,  $\Psi(B^s)$  and  $H(B^s)$ , respectively; and  $B$  is the backward shift operator such that

$B^k y_t = y_{t-k}$ .  $E(\varepsilon_t)$  and  $\text{Var}(\varepsilon_t)$  are the mean function and the variance function of the Gaussian white noise series  $\{\varepsilon_t\}$ , respectively, and  $\text{cov}(\varepsilon_h, \varepsilon_t)$  is the correlation function between  $\varepsilon_h$  and  $\varepsilon_t$ .

In particular, when there is an additive relationship between the seasonal effect and other effects in the series, the seasonal information can be fully extracted by taking a first difference with periodic step length. At this point, the model can be simplified into Eq. (2), which is denoted as ARIMA  $(p, d, q) \times (0, D, 0)_s$ .

$$N_t = \frac{\Theta(B)}{\Phi(B)\nabla^d\nabla_s^D} \varepsilon_t \tag{2}$$

Generally speaking, the measured responses of the structure can be considered as the linear superposition of various load effects when the structure is in the normal operation state. Therefore, for the observed time series with seasonal effect, the seasonal additive model shown in Eq. (2) is adequate to describe the actual state of the structure.

Fig. 1 displays the procedure for the establishment of SARIMA models.

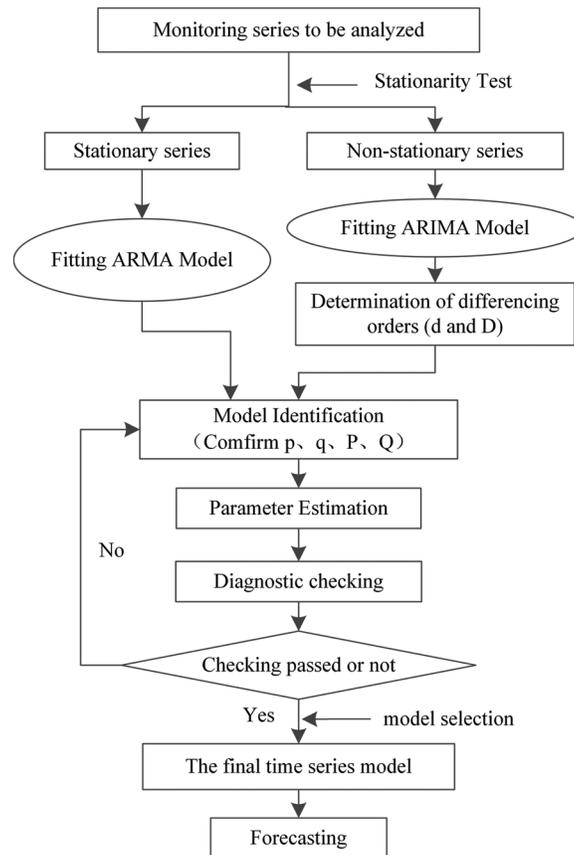


Figure 1: Flow chart of SARIMA modeling process

### 3 Time Series Analysis with Outlier Intervention

#### 3.1 SARIMA Model Containing Outlier Effects

To consider the influence of different types of intervention outliers, the indicator function is introduced. Thus, the general model of time series  $\{Y_t\}$ , which contains outlier information, can be written as the combination of various indicator functions shown in Eq. (3) [33].

$$Y_t = G_t + N_t = c + \sum_{i=1}^k \frac{\omega_i(B)B^{b_i}}{\delta_i(B)} I_{ti} + N_t \quad (3)$$

where  $G_t$  represents the influence of interventions expressed in the form of indicator function using deterministic intervention variables  $I_{ti}$ ;  $k$  is the total number of outliers;  $B$  is the delay operator;  $b$  denotes the time delay for the intervention effects, and for bridge structure,  $b=0$ , on account of the instantaneous response to the external action;  $\omega(B)$  reflects the intensity of intervention and  $\delta(B)$  measures the behavior of the permanent effect of the intervention. Additionally, the time series free of intervention is called the noise series or the undisturbed series and is denoted by  $N_t$ .  $N_t$  can be various kinds of stationary or non-stationary series. For a nonstationary process, the model in Eq. (3) normally does not contain a constant term  $c$ .

When there are outliers detected in monitoring data, the abnormal behavior can be explained by intervention analysis techniques if the timing and causes of interruptions are knowable. However, the timing of interventions is usually unknown. So, it is necessary to detect and estimate the possible effects. Here, two common outlier models, innovational outlier (IO) and additive outlier (AO), are introduced [34].

Let  $N_t$  be an undisturbed process free of interventions and follow an additive SARIMA process ARIMA  $(p, d, q) \times (0, D, 0)_s$  defined in Eq. (2).

(1) If the error  $\varepsilon_t$  of  $N_t$  at time  $T$  is disturbed and turned into  $\varepsilon'_t = \varepsilon_t + \omega P_t^{(T)}$ , the post-disturbed series can be described using an IO model

$$Y_t = \frac{\Theta(B)}{\Phi(B)\nabla^d\nabla_s^D} (\varepsilon_t + \omega P_t^{(T)}) = N_t + \omega \left( \frac{\Theta(B)}{\Phi(B)\nabla^d\nabla_s^D} \right) P_t^{(T)} \quad (4)$$

where  $P_t^{(T)}$  is a pulse function taking place at  $T$  time period which can be written as

$$P_t^{(T)} = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases} \quad (5)$$

(2) If  $N_t$  is subject to additive disturbance at time  $T$ , the AO model is introduced as

$$Y_t = N_t + \omega P_t^{(T)} \quad (6)$$

Hence, an AO affects only the  $T$ th observation by the intensity  $\omega$  when the interruption takes place. While an IO describes the systematic dynamic behavior by  $\Theta(B)/\Phi(B)\nabla^d\nabla_s^D$ , and affects all observations  $Y_T, Y_{T+1}, \dots$  beyond time  $T$ .

(3) If a time series is influenced by  $k$  outliers of different types at different time periods  $T_1, T_2, \dots, T_k$ , a more general model containing various types of outliers is presented as follows:

$$Y_t = N_t + \sum_{i=1}^k \omega_i L_i(B) P_t^{(T_i)} \quad (7)$$

where, for SARIMA model

$$L_i(B) = \begin{cases} \frac{\Theta(B)}{\Phi(B)\nabla^d\nabla_s^D}, & \text{IO} \\ 1, & \text{AO} \\ \dots & \dots \end{cases} \quad (8)$$

### 3.2 Outlier Mining

When outliers exist, the estimated parameters are biased. In this case, appropriate test statistics are constructed from the residuals, which are the discrepancies between the observed and the estimated values, for the detection and correction of outliers. Outlier mining based on time series analysis is to identify the time, size and types of outliers, estimate their impacts, and modify the original time series model affected by outliers so as to improve the accuracy of the model.

Let  $\{Y_t\}$  be the original observed monitoring series, which can be fitted by an additive SARIMA model as Eq. (2). The hypothesis test is performed as follows:

$H_0$ :  $y_T$  is neither an IO nor an AO

$H_1$ :  $y_T$  is an IO

$H_2$ :  $y_T$  is an AO

The process of outliers detection is mainly discussed in the following two cases [35].

(1) Only one type of outlier is included, and the timing of the outlier occurrence is known.

The residual series of IO and AO models can be written respectively as

$$e_t = \begin{cases} \omega_I P_t^{(T)} + \varepsilon_t, & \text{IO} \\ \omega_A \pi(B) P_t^{(T)} + \varepsilon_t, & \text{AO} \end{cases} \quad (9)$$

where  $\pi(B) = \frac{\Phi(B) \nabla^d \nabla_s^D}{\Theta(B)} = 1 - \pi_1 B - \dots - \pi_n B^n - \dots$

According to the least-squares principle, when  $t = T$ , the least-square estimator of the disturbance  $\omega_I$  or  $\omega_A$  (shown in Eq. (10)) is the residual at time  $T$  (IO) or the linear combination of the residuals (AO).

$$\begin{cases} \text{IO: } \hat{\omega}_{I,T} = e_T \\ \text{AO: } \hat{\omega}_{A,T} = -\rho^2 \sum_{t=1}^n \pi_{t-T} e_t \end{cases} \quad (10)$$

where  $\rho^2 = (1 + \pi_1^2 + \pi_2^2 + \dots + \pi_{n-T}^2)^{-1}$ . Thus, the variance of the estimator is

$$\begin{cases} \text{Var}(\hat{\omega}_{I,T}) = \text{Var}(e_T) = \text{Var}(\omega_I P_t^{(T)} + \varepsilon_t) = \sigma_a^2 \\ \text{Var}(\hat{\omega}_{A,T}) = \rho^2 \sigma_a^2 \end{cases} \quad (11)$$

The test statistics for IO and AO at time  $T$  are

$$\begin{cases} \lambda_{I,T} = \frac{\hat{\omega}_{I,T}}{\sigma_a} \\ \lambda_{A,T} = \frac{\hat{\omega}_{A,T}}{\rho \sigma_a} \end{cases} \quad (12)$$

Under the null hypothesis  $H_0$ ,  $\omega_I = 0$  and  $\omega_A = 0$ , and both  $\lambda_{I,T}$  and  $\lambda_{A,T}$  are distributed as  $N(0, 1)$ . Let the significance level be  $\alpha = 0.05$ , then the necessary condition for accepting hypothesis  $H_1$  or  $H_2$  is respectively  $\lambda_{I,T} > 1.96$  or  $\lambda_{A,T} > 1.96$ .

(2) The timing and type of the outliers are unknown.

Case (1) is applicable to situations where the timing and type of the outliers are known. However, more often in practice, the type, timing and number of outliers are unknown and have to be estimated. An iterative

detection procedure is proposed to handle the situation when an unknown number of AO or IO may exist [36,37]. Specifically, the following four steps are included:

**Step 1.** Model the original monitoring series  $\{N_t\}$  with SARIMA supposing that there are no outliers. The residual series can be derived as

$$\varepsilon_t = \pi(B)N_t = \frac{\Phi(B)\nabla^d\nabla_s^D}{\Theta(B)}N_t \quad (13)$$

The initial estimate of the standard deviation of the residuals is

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2} \quad (14)$$

**Step 2.** Perform statistical tests of outliers using the initial residuals estimated at  $t = 1, 2, 3, \dots, n$ . The test statistics for IO and AO refer to Eq. (12). Thus,  $\lambda_{I,T}$  and  $\lambda_{A,T}$  for each time period can be calculated. Define

$$\hat{\lambda}_t = \max_t [\max(|\lambda_{I,t}|, |\lambda_{A,t}|)], \quad t = 1, 2, \dots, n \quad (15)$$

where  $\hat{\lambda}_t$  denotes the maximum outlier at different times.

Bonferroni correction is used to control the overall error rate of multiple tests. Based on 0.05 significance level, if the  $p$ -value of  $\hat{\lambda}_t$  at time  $T$  is greater than the upper percentile of the standard normal distribution, or  $\hat{\lambda}_t > C$ , where  $C$  is a predetermined positive constant usually taken to be some value between 3 and 4, then there is an outlier at time  $T$  with its type determined by  $\lambda_{I,T}$  or  $\lambda_{A,T}$ .

It should be noted that the maximum likelihood estimation of  $\sigma_\varepsilon$  is on the high side in the presence of outliers. Especially when the sample size is small, such deviation may become very great, resulting in a reduction of  $\hat{\lambda}_t$ . To eliminate the influence of outliers on  $\sigma_\varepsilon$ , the robust estimation is adopted to replace the maximum likelihood estimation. In the latter example, the absolute mean of the residual is multiplied by  $\sqrt{2/\pi}$  to get a more robust estimate.

According to different types of outliers, the effect of IO/AO at time  $T$  can be removed by modifying the data using Eq. (16) as

$$\tilde{Y}_t = \begin{cases} Y_t - \omega_{I,T}\pi(B)P_t^{(T)}, & \text{IO} \\ Y_t - \omega_{A,T}\pi(B)P_t^{(T)}, & \text{AO} \end{cases} \quad (16)$$

where  $\pi(B) = \frac{\Phi(B)\nabla^d\nabla_s^D}{\Theta(B)} = 1 - \pi_1B - \dots - \pi_nB^n - \dots$ , and the new residual sequence is defined as

$$\tilde{\varepsilon}_t = \begin{cases} e_t - \hat{\omega}_{I,T}, & \text{IO} \\ e_t - \omega_{A,T}\pi(B)P_t^{(T)}, & \text{AO} \end{cases} \quad (17)$$

The new estimate of the standard deviation of the residuals  $\tilde{\sigma}_\varepsilon$  can be recalculated from the modified residuals.

**Step 3.** Re-compute  $\lambda_{I,T}$  or  $\lambda_{A,T}$  at every time point based on the modified  $\tilde{\varepsilon}_t$  and  $\tilde{\sigma}_\varepsilon$ , and repeat Step 2 to continue detection and modification of outliers until all outliers are identified. In this process, the initial parameter estimates in SARIMA remain unchanged.

**Step 4.** Suppose that  $k$  outliers have been identified at times  $T_1, T_2, \dots, T_k$  through the first three steps. Treat these times as if they have already been known, then the  $k$  outlier parameters  $\omega_1, \omega_2, \dots, \omega_k$  as well as  $\Theta(B)$  and  $\Phi(B)$  can be re-estimated. And a new model containing  $k$  outliers information can be written as

$$Y_t = \frac{\Theta(B)}{\Phi(B)\nabla^d\nabla_s^D} \varepsilon_t + \sum_{j=1}^k \omega_j L_j(B) P_t^{(T_j)} \quad (18)$$

where  $L_j(B)$  is determined by Eq. (8). The new residual series can then be derived from the new fitted model

$$\hat{\varepsilon}_t = \hat{\pi}(B) \left[ Y_t - \sum_{j=1}^k \hat{\omega}_j \hat{L}_j(B) P_t^{(T_j)} \right] \quad (19)$$

and a revised estimate of  $\tilde{\sigma}_\varepsilon$  can further be calculated.

Repeat Step 2 to Step 4 until no new outlier is identified.

### 3.3 Model Forecasting

One of the most important aims of time series analysis is to predict or forecast future development trends of observed data using the fitted models. This is also an important purpose for bridge health monitoring, to figure out the current state of the bridge structure and predict its future long-term development trends through the observed monitoring data.

Consider the general additive seasonal ARIMA  $(p, d, q) \times (0, D, 0)_s$  model shown in Eq. (2). Let  $\Phi^*(B) = \Phi(B)(1-B)^d(1-B^s)^D = 1 - \tilde{\phi}_1 B - \tilde{\phi}_2 B^2 - \dots - \tilde{\phi}_{p+d+sD} B^{p+d+sD}$  be the generalized autocorrelation function. And the model can be rewritten in the form of a linear function expressed by the disturbance term

$$y_t = \frac{\Theta(B)}{\Phi^*(B)} \varepsilon_t = \Gamma(B) \varepsilon_t = (1 + \gamma_1 B + \gamma_2 B^2 + \dots) \varepsilon_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots \quad (20)$$

Let time  $t$  be the forecast origin,  $l$  be the lead time for the forecast, and the forecast that will occur  $l$  time units into the future be  $\hat{y}_t(l)$ . Then, the mean square error (MSE) of the forecast is

$$E[y_{t+l} - \hat{y}_t(l)]^2 = (1 + \gamma_1^2 + \dots + \gamma_{l-1}^2) \sigma_\varepsilon^2 + \sum_{j=0}^{\infty} (\gamma_{l+j} - \gamma_j^*)^2 \sigma_\varepsilon^2 \quad (21)$$

According to the minimum mean square error principle, the MSE shown in Eq. (22) is minimized if and only if  $\gamma_j^* = \gamma_{l+j}$ . Hence, we have the  $l$ -step ahead forecast

$$\hat{y}_t(l) = E(y_{t+l}|y_t, y_{t-1}, \dots) = \gamma_l \varepsilon_t + \gamma_{l+1} \varepsilon_{t-1} + \gamma_{l+2} \varepsilon_{t-2} + \dots \quad (22)$$

The forecast error is

$$e_t(l) = \varepsilon_{t+l} + \gamma_1 \varepsilon_{t+l-1} + \dots + \gamma_{l-1} \varepsilon_{t+1} \quad (23)$$

with the forecast error variance

$$Var[e_t(l)] = Var(y_{t+l}|y_t, y_{t-1}, \dots) = (1 + \gamma_1^2 + \dots + \gamma_{l-1}^2) \sigma_\varepsilon^2 \quad (24)$$

It can be seen that the variance of the forecast is only related to step size  $l$ , and the forecast error becomes larger and larger as the forecast lead time  $l$  increases. Assuming that the value of  $y_{t+l}$  obeys the normal distribution under the condition that  $y_t, y_{t-1}, \dots$ , are known, the confidence interval of the predicted value  $y_{t+l}$  at the confidence level of  $1 - \alpha$  can be obtained as

$$\left( \hat{y}_t(l) - z_{1-\alpha/2} \cdot \sqrt{\text{Var}[e_t(l)]}, \quad \hat{y}_t(l) + z_{1-\alpha/2} \cdot \sqrt{\text{Var}[e_t(l)]} \right) \quad (25)$$

where  $z_{1-\alpha/2}$  is the standard normal percentile, such that  $P(-z_{1-\alpha/2} < Y < z_{1-\alpha/2}) = 1 - \alpha$ .

## 4 Application Analysis and Results

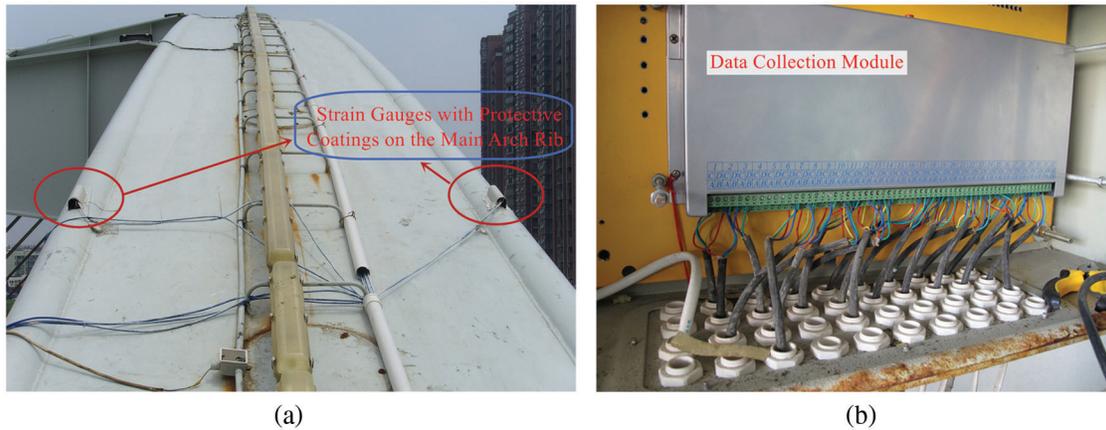
### 4.1 Selection of Monitoring Data Samples

Kunshan Yufeng bridge, located in Kunshan city, Jiangsu Province, is a non-thrusting leaning-type arch bridge with a main span of 110 m (see Fig. 2). In this structural system, the two vertical main arch ribs in the middle and the rigid tie beams (main beams) under the main arch ribs are the major longitudinal load-bearing members, while the inclined arch ribs outward mainly bear part of the dead and live loads and improve the stability of the structure. According to the vulnerability analysis results of the bridge, a total of 43 monitoring points are arranged in 13 sections or positions within the half span of the bridge, aiming at performing a real time monitoring of strain, temperature and support displacement of the critical stressed members. Fig. 3 shows some of the sensors on the bridge and the data collection module.

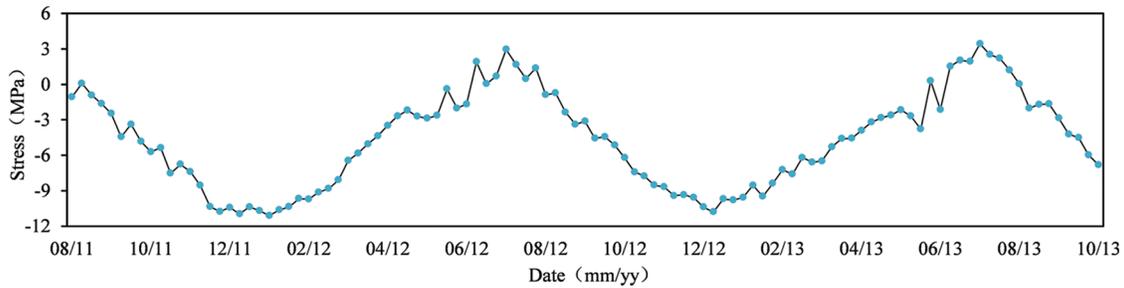


**Figure 2:** Kunshan Yufeng bridge

The monitoring data, which are chosen as the modeling basis, are picked from the stress measuring point at the bottom of the vault of the southwest main arch rib of Kunshan Yufeng bridge (Gauge S1-2), with the date ranging from August 23rd, 2011 to February 28th, 2014. These data are presented in the form of weekly-cycle mean value (M value). The weekly cycle here is not strictly measured by the traditional week. We uniformly divide each month into four weeks, namely, 1st~7th, 8th~15th, 16th~23rd and 24th~30th (31st) (For February, 7th, 14th and 21st are taken as the split nodes). In this way, a year is fixedly divided into 48 weeks, giving a total of 117 sample data. The first 105 monitoring data are used for the model calibration, while the remaining for verification. The weekly-cycle stress M-value of the first 105 data of the sample series is drawn in Fig. 4.



**Figure 3:** Sensors on the bridge. (a) Strain gauges with protective coatings on the main arch rib; (b) Data collection module



**Figure 4:** Weekly-cycle M-value of stresses at gauge S1-2

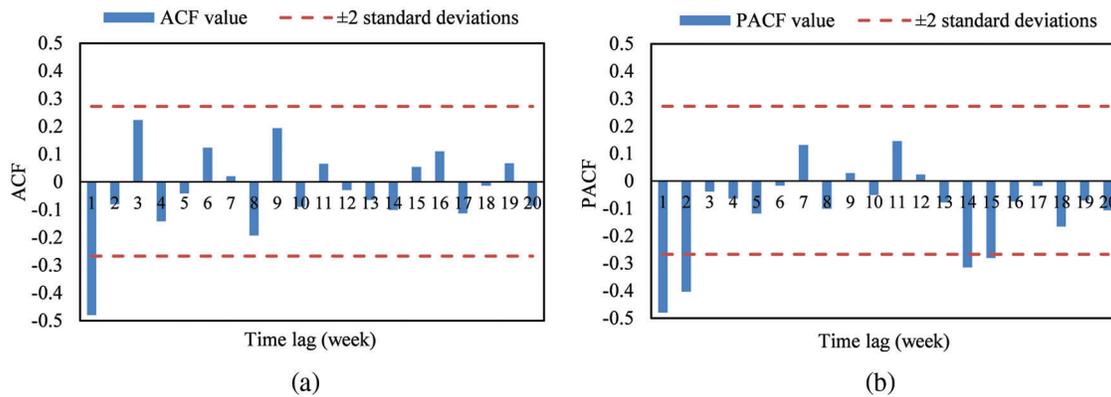
#### 4.2 Establishment of SARIMA Model

The series shown in Fig. 4 implies a clear yearly (48-week) seasonal component, which indicates a typical non-stationary series. Thus, a seasonal ARIMA model is adopted to fit the experimental data. The original data are differenced using operators  $(1 - B)$  and  $(1 - B^{48})$  once each. Then, PP test and KPSS test are implemented to the differenced series. The test results are shown in Table 1 at the 0.05 significance level.

**Table 1:** Unit root stationarity tests for differenced series

Type of test	PP test			KPSS test			Conclusion
	Lag	Stat	<i>p</i> -value	Lag	Stat	<i>p</i> -value	
No drift no trend	3	-68.4	≤0.01	1	0.0493	≥0.1	Stationary
With drift no trend	3	-68.4	≤0.01	1	0.0522	≥0.1	Stationary
With drift and trend	3	-68.2	≤0.01	1	0.0282	≥0.1	Stationary

As illustrated in Table 1, the PP test *p*-values of differenced series under three test types are all ≤0.01, and *p*-values of KPSS test are all ≥0.1, which indicates that the differenced series is significantly stationary. This conclusion is further confirmed by the rapid decay to within ±2 standard deviations of the sample ACF shown in Fig. 5.



**Figure 5:** Sample ACF and PCF of differenced series: (a) Sample ACF; (b) Sample PACF

The values of orders  $p$  and  $q$  can be preliminarily identified by the characteristics of sample ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) [38]. Table 2 summarizes the behavior of ACF and PACF for selecting  $p$  and  $q$ .

**Table 2:** Characteristics of the ACF and PACF for stationary process

	AR ( $p$ )	MA ( $q$ )	ARMA ( $p, q$ )
ACF	Tails off as exponential decay of damped sine wave	Cuts off after lag $q$	Tails off after lag $(q - p)$
PACF	Cuts off after lag $p$	Tails off as exponential decay of damped sine wave	Tails off after lag $(p - q)$

From Table 2, it seems clear that when the ACF or PACF possesses the cutting-off behavior, the identification of the order  $p$  of an AR model and the order  $q$  of an MA model is relatively simple. For a mixed AR-MA model, however, the ACF and PACF all exhibit tapering off property, which makes the identification of the orders  $p$  and  $q$  much more difficult. For more complicated models, Tsay et al. [39] introduced the ESACF (Extended Sample Autocorrelation Function) to estimate the orders  $p$  and  $q$ . In the ESACF table, an ARMA ( $p, q$ ) will have a pattern of a triangle of zeros, with the upper left-hand vertex at the  $(p, q)$  position. However, in the actual identification process, the characteristics of ACF, PACF or ESACF are not always so clear-cut. Therefore, multiple possible models are normally picked according to pertinent information, and the model, of which the statistical indicator is in line with the evaluation criteria, is finally selected. The sample ACF and PCF are plotted in Fig. 5. Table 3 shows the ESACF with indicator symbols.

Note from Fig. 5 that the sample ACF exhibits an alternating decreasing pattern and the sample PACF cuts off after lag 2. In the ESACF, the vertex of the zero triangles is located at the  $(2, 0)$  position. All these above give a preliminary indication of an ARIMA  $(2, 1, 0) \times (0, 1, 0)_{48}$  model. In addition, several adjacent models are also considered. After eliminating the models with insignificant parameters, four adequate models are finally retained for further optimization, which are ARIMA  $(2, 1, 0) \times (0, 1, 0)_{48}$ , ARIMA  $(0, 1, 1) \times (0, 1, 0)_{48}$  and sparse parametric models ARIMA  $(1, 1, (2)) \times (0, 1, 0)_{48}$  and ARIMA  $(1, 1, (2, 3)) \times (0, 1, 0)_{48}$ .

**Table 3:** The ESACF with indicator symbols

AR	MA							
	0	1	2	3	4	5	6	7
0	X	0	X	0	0	0	0	0
1	X	X	X	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0
5	X	X	X	0	0	0	0	0
6	X	0	0	0	0	0	0	0
7	X	0	0	0	0	0	0	0

To quantitatively evaluate the accuracy and stability of the preliminary proposed models, the residual sum of squares ( $RSS$ ),  $\sigma^2$  ( $\hat{\sigma}_\varepsilon^2$ ), log likelihood ( $\ln \hat{L}$ ), Akaike's Corrected Information Criterion (AICc), Bayesian Information Criterion (BIC), adjusted R-squared ( $\bar{R}^2$ ) are utilized for model selection [33]. The calculation formulas are shown as follows:

$$RSS = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (26)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n - p - (p + q + 1)} \quad (27)$$

$$\ln \hat{L} = -\frac{n}{2} \ln \hat{\sigma}_\varepsilon^2 - \frac{n}{2} (1 + \ln 2\pi) \quad (28)$$

$$AICc = AIC + \frac{2(m+1)(m+2)}{n-m-2}, \quad AIC = -2 \ln \hat{L} + 2m \quad (29)$$

$$BIC = -2 \ln \hat{L} + m \ln n \quad (30)$$

$$\bar{R}^2 = 1 - \frac{RSS/(n-m-1)}{TSS/(n-1)}, \quad TSS = \sum_{t=1}^n (y_t - \bar{y})^2 \quad (31)$$

where  $y_t$  is the observed series;  $\hat{y}_t$  is the predicted value of  $y_t$ ;  $\varepsilon_t$  is the residual series;  $n$  is the effective number of observations;  $m$  is the number of parameters in the model; TSS represents the total sum of squares and  $\bar{y}$  denotes the sample mean.

These statistics are normally based on summary statistics from residuals computed from the fitted model. For  $RSS$ ,  $\sigma^2$ , log likelihood, AICc and BIC, lower values specify a better model. Adjusted R-squared values range from 0 to 1. A higher adjusted R-squared value closer to 1 indicates a superior. The statistics of the four models are estimated in Table 4.

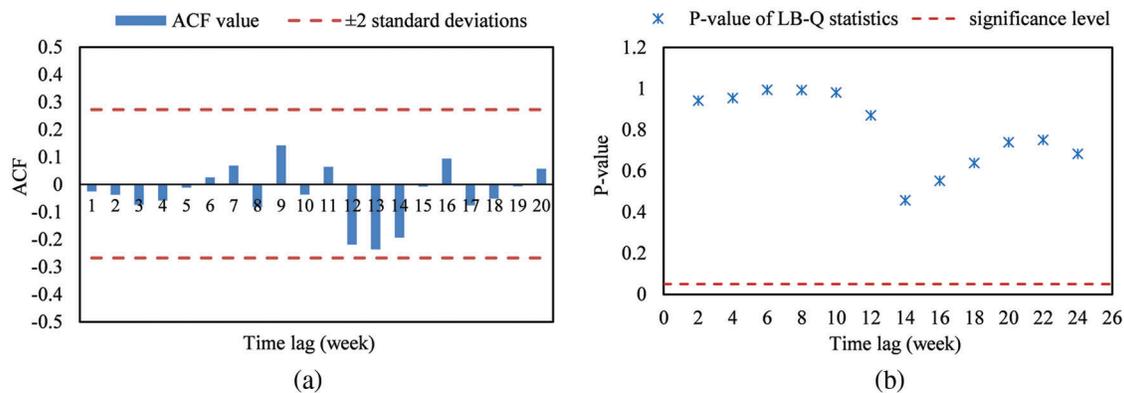
**Table 4:** The statistics of fitted models

	AICc	BIC	sigma <sup>2</sup>	Log likelihood	Adjusted R-squared	RSS
ARIMA (2, 1, 0) × (0, 1, 0) <sub>48</sub>	178.25	184.20	1.294	-86.06	0.9626	69.898
ARIMA (0, 1, 1) × (0, 1, 0) <sub>48</sub>	180.28	184.29	1.370	-88.12	0.9612	75.368
ARIMA (1, 1,(2)) × (0, 1, 0) <sub>48</sub>	181.47	187.31	1.319	-87.61	0.9614	73.883
ARIMA (1, 1,(2, 3)) × (0, 1, 0) <sub>48</sub>	179.30	187.00	1.193	-85.45	0.9636	66.816

According to the statistical information illustrated in Table 4, the ARIMA (2, 1, 0) × (0, 1, 0)<sub>48</sub> model gives the smallest values of AICc and BIC among the four models chosen, while the optimal sigma<sup>2</sup>, Log likelihood, Adjusted R-squared and RSS indices are coming from the ARIMA (1, 1, (2, 3)) × (0, 1, 0)<sub>48</sub> model. After comprehensive comparisons, the relatively concise model ARIMA (2, 1, 0) × (0, 1, 0)<sub>48</sub> is selected for final data fitting. The estimation of this model gives

$$(1 - B)(1 - B^{48})y_t = \frac{\varepsilon_t}{1 + 0.6896B + 0.4919B^2}, \text{Var}(\varepsilon_t) = 1.294 \quad (32)$$

Fig. 6 gives the residual ACF and the Ljung-Box Q (LB-Q) statistics. Results show that the residual ACFs are all within ±2 standard deviations, and the *p*-values of LB-Q statistics are all >0.05, which indicates no relevant information is contained in the residuals any longer and the residual series is a white noise process. Thus, the fitted model ARIMA (2, 1, 0) × (0, 1, 0)<sub>48</sub> is judged adequate for the data.

**Figure 6:** Model adequacy checking results: (a) Residual ACF; (b) *p*-value of LB-Q statistics

#### 4.3 Detection of Outliers

The residual series of the fitted model (see Eq. (26)) is plotted in the legend “Residual of the original model” of Fig. 7. The outlier detecting procedure is then carried out on this residual series. The robust estimator is adopted for the standard deviation of the residuals with the critical value  $c = 3.5$ . As a result, a total of 2 AOs and 4 IOs are identified at the significance level 5%, as shown in Fig. 7. The iterative procedure for outlier detection and the details of outliers are listed in Table 5.

The fitted SARIMA-outlier model can be presented as follows in the form of the main SARIMA model plus outlier interventions:

$$y_t = \frac{1}{(1 - B)(1 - B^{48})(1 + 0.5478B + 0.4078B^2)} \times [2.3443P_t^{(63)} - 3.5260P_t^{(87)} + 3.5541P_t^{(88)} - 2.3106P_t^{(98)} + \varepsilon_t] - 0.7850P_t^{(77)} + 1.2286P_t^{(91)} \quad (33)$$

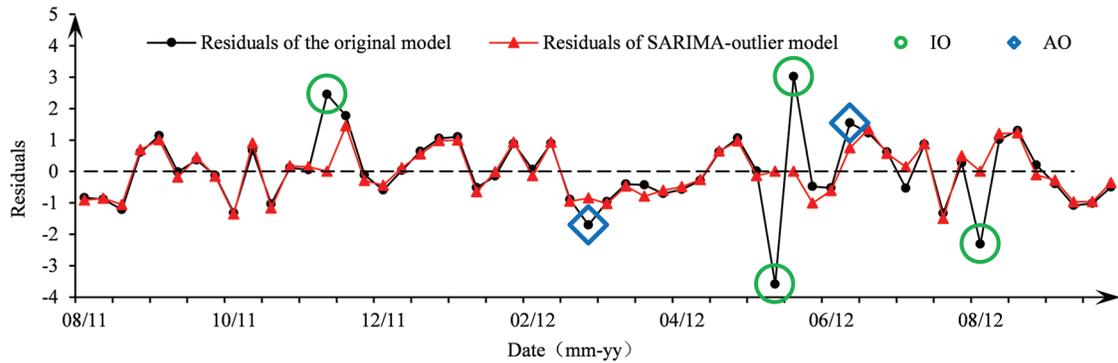


Figure 7: Residual series before and after the identification of outliers

Table 5: Iterative procedure for outlier detection

The first iteration	Model parameters	$\phi_1$	$\phi_2$	—	—	$\hat{\sigma}_\varepsilon^2$	
		-0.6896	-0.4919			1.2481	
	Detected outliers	Time	Type	$\hat{\lambda}_t$	Time	Type	$\hat{\lambda}_t$
		91	AO	3.564134	88	IO	5.226359
63		IO	4.256304	9	IO	-4.007655	
	87	IO	-6.214973		—		
The second iteration	Model parameters	$\phi_1$	$\phi_2$	$\omega_{A,91}$	$\omega_{I,63}$	$\hat{\sigma}_\varepsilon^2$	
		-0.5461	-0.4204	1.2106	2.3466		
		$\omega_{I,87}$	$\omega_{I,88}$	$\omega_{I,98}$		0.6093	
		-3.5170	3.5502	-2.3406			
	Detected outliers	Time	Type	$\hat{\lambda}_t$	—	—	
	77	AO	-4.088277				
The final results	Model parameters	$\phi_1$	$\phi_2$	$\omega_{A,91}$	$\omega_{I,63}$	$\hat{\sigma}_\varepsilon^2$	
		-0.5478	-0.4078	1.2286	2.3443		
		$\omega_{I,87}$	$\omega_{I,88}$	$\omega_{I,98}$	$\omega_{A,77}$	0.5945	
		-3.5260	3.5541	-2.3106	-0.7850		
	Detected outliers	None					
Model statistics with outlier interventions		AICc	BIC	Log likelihood	Adjusted R-squared	RSS	
		149.83	166.56	-65.16	0.9793	33.30	

After two rounds of iteration, as shown in Table 5, no new outliers are identified. The recognized 2 AOs are from time 77 and 91, and the 4 IOs come from times 63, 87, 88, and 98, respectively. The model parameters  $\phi$  change a lot after the adjustment of outliers on the premise of not changing the structure and the order of the fitted model. The value of  $\hat{\sigma}_\varepsilon^2$  is reduced by about 50%, and other statistics, such as AICc, BIC, RSS, etc., are all optimized greatly. The comparison of the model residuals before and after outlier elimination is illustrated in Fig. 7. It can be seen that the residuals of the two models are basically the same at most of the time points. In the region close to the outliers, the residuals of the SARIMA-outlier model are obviously smaller than those of the original model, and thus  $\hat{\sigma}_\varepsilon^2$  gets a significant decrease, indicating that the adjustment of outliers has a significant impact on the model fitted results. The precision of model fitting improves after including the intervention influence of outliers.

The fitted results of the original model and SARIMA-outlier model are illustrated in Fig. 8 (Note: the first 51 data of the fitted curve are lost due to the differences). Compared with the original model, the SARIMA-outlier model shows a better fitting performance.

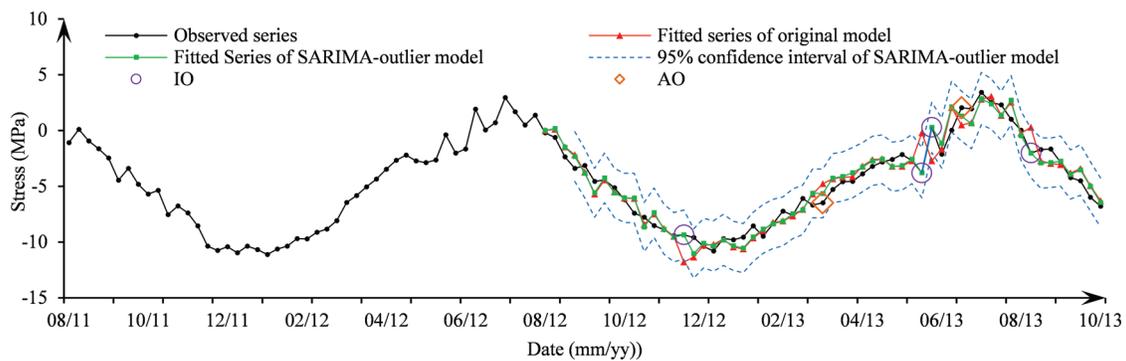


Figure 8: Comparison of the fitted models

#### 4.4 Forecasting Results and Discussion

Use the original model (Eq. (26)) and the SARIMA-outlier model (Eq. (27)) respectively to forecast the weekly mean stress values of measuring point S1.2 in the next 8 weeks. The prediction results are shown in Fig. 9, and the comparison of the accuracy measurement indices of the prediction errors of the two models is presented in Table 6.

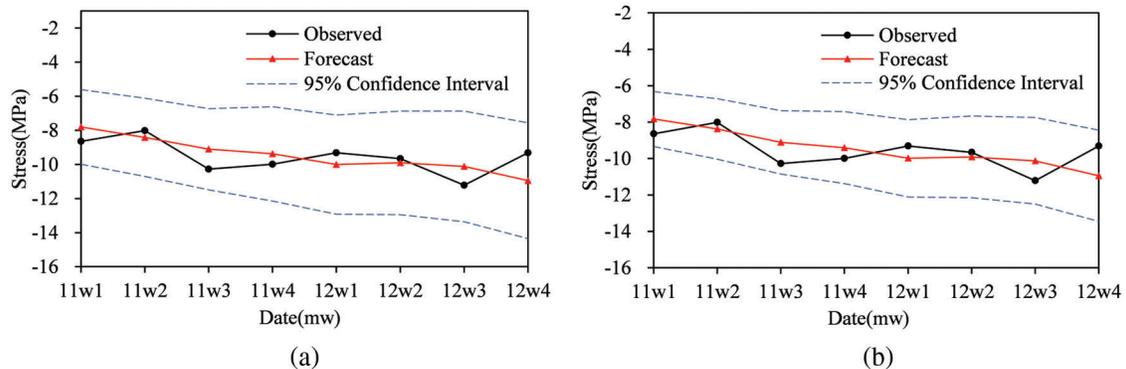


Figure 9: Results of the forecast: (a) Forecasts of the original model (Eq. (26)); (b) Forecasts of the SARIMA-outlier model (Eq. (27))

**Table 6:** Iterative procedure for outlier detection

Indices	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Mean Absolute Percentage Error (MAPE)	Theil Inequality Coefficient (TIC)
Original model	0.5594	0.7678	5.8111%	0.0492
SARIMA-outlier model	0.5481	0.7566	5.6864%	0.0485

As can be seen from Fig. 9, the width of the 95% confidence interval for the prediction values gradually grows in a trumpet shape, indicating a gradual decrease in the reliability of the prediction. Compared with the original model, the 95% prediction confidence interval of the SARIMA-outlier model is narrower, that is, the prediction reliability remains at a higher level. In terms of the accuracy measurement indices of the prediction errors (see Table 6), the indices of the SARIMA-outlier model are smaller, which also shows a better prediction accuracy. In view of the absolute error of the forecasts, errors of the short-term forecast results fluctuate within a reasonable range, which can meet the needs of forecasts and assessment of bridge structures.

On the other hand, the difference between the predicted values of the two models at each time point is not significant, and the maximum absolute value of the difference is only 0.036. The overall prediction results of the two models are quite close. Therefore, although the existence of outliers has a significant impact on the parameter identification and fitted precision of the time series models, it is insensitive to the forecasting results when there are only a small number of outliers and the value of outlier effect  $\lambda$  is not big. The original model is of strong robustness.

## 5 Conclusions

In order to ensure the reliability of monitoring data and improve the accuracy of forecasts, the time series model with outliers for bridge monitoring is established using the intervention analysis theory. IOs and AOs are diagnosed and extracted from observed data. Through comparative analysis with the original model without considering outliers, some conclusions are drawn as follows:

- (1) The additive seasonal ARIMA model is suitable for the fitting of bridge monitoring data with obvious seasonal effects. The residuals of the fitted model are white noise, which indicates that the fitted model is of significant effectiveness. Use this model for forecasts, the errors of which can meet the demand for prediction and assessment of bridge structures.
- (2) The outlier detection algorithm presented can rapidly and efficiently identify the outliers existing in BHM data. The detected IOs and AOs are sensitive to the model parameter estimation. After considering the influence of IOs and AOs, the model parameters vary greatly and the fitting accuracy improves significantly, which also verifies the effectiveness and accuracy of the outlier identification.
- (3) In comparison with the original model, the prediction confidence interval of the SARIMA-outlier model is narrower, indicating a more reliable forecasting result. In the meantime, the accuracy measurement indices are smaller and the prediction accuracy is higher.
- (4) The existence of outliers is insensitive to the forecasting results under the condition that the number of outliers is small and the test statistics for outlier effects are not big. The original time series model has strong prediction robustness.

**Funding Statement:** This work was funded by the Natural Science Foundation of Fujian Province (Grant No. 2020J05207), Fujian University Engineering Research Center for Disaster Prevention and Mitigation of Engineering Structures along the Southeast Coast (Grant No. JDGC03), Major Scientific Research Platform Project of Putian City (Grant No. 2021ZP03), and Talent Introduction Project of Putian University (Grant No. 2018074).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Flah, M., Nunez, I., Ben Chaabene, W., Nehdi, M. L. (2021). Machine learning algorithms in civil structural health monitoring: A systematic review. *Archives of Computational Methods in Engineering*, 28(4), 2621–2643. DOI 10.1007/s11831-020-09471-9.
2. Sony, S., Laventure, S., Sadhu, A. (2019). A literature review of next-generation smart sensing technology in structural health monitoring. *Structural Control and Health Monitoring*, 26(3), e2321. DOI 10.1002/stc.2321.
3. Entezami, A., Sarmadi, H., Behkamal, B., Mariani, S. (2020). Big data analytics and structural health monitoring: A statistical pattern recognition-based approach. *Sensors*, 20(8), 2328. DOI 10.3390/s20082328.
4. Bao, Y., Chen, Z., Wei, S., Xu, Y., Tang, Z. et al. (2019). The state of the art of data science and engineering in structural health monitoring. *Engineering*, 5(2), 234–242. DOI 10.1016/j.eng.2018.11.027.
5. Cai, G., Mahadevan, S. (2018). Big data analytics in uncertainty quantification: Application to structural diagnosis and prognosis. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(1), 04018003. DOI 10.1061/AJRUA6.0000949.
6. Tu, C. F., Liu, Z. J., Zhang, G., Zhou, L. C., Chen, Y. T. et al. (2017). Processing technique and application of big data oriented to long-term bridge health monitoring. *Journal of Experimental Mechanics*, 32, 652–663. DOI 10.7520/1001-4888-17-309.
7. Wang, Y., Wang, P., Tang, H., Liu, X., Xu, J. et al. (2021). Assessment and prediction of high speed railway bridge long-term deformation based on track geometry inspection big data. *Mechanical Systems and Signal Processing*, 158, 107749. DOI 10.1016/j.ymssp.2021.107749.
8. Xu, S., Ma, R., Wang, D., Chen, A., Tian, H. (2019). Prediction analysis of vortex-induced vibration of long-span suspension bridge based on monitoring data. *Journal of Wind Engineering and Industrial Aerodynamics*, 191, 312–324. DOI 10.1016/j.jweia.2019.06.016.
9. Sun, L., Shang, Z., Xia, Y. (2019). Development and prospect of bridge structural health monitoring in the context of big data. *China Journal of Highway and Transportation*, 32(11), 1–20. DOI 10.19721/j.cnki.1001-7372.2019.11.001.
10. Sun, L., Shang, Z., Xia, Y., Bhowmick, S., Nagarajaiah, S. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Journal of Structural Engineering*, 146(5), 04020073. DOI 10.1061/(ASCE)ST.1943-541X.0002535.
11. Ye, X. W., Jin, T., Yun, C. B. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems*, 24(5), 567–585. DOI 10.12989/sss.2019.24.5.567.
12. Dang, H. V., Raza, M., Nguyen, T. V., Bui-Tien, T., Nguyen, H. X. (2021). Deep learning-based detection of structural damage using time-series data. *Structure and Infrastructure Engineering*, 17(11), 1474–1493. DOI 10.1080/15732479.2020.1815225.
13. Lara-Benítez, P., Carranza-García, M., Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(3), 2130001. DOI 10.1142/S0129065721300011.
14. Faloutsos, C., Gasthaus, J., Januschowski, T., Wang, Y. (2018). Forecasting big time series: Old and new. *Proceedings of the VLDB Endowment*, 11(12), 2102–2105. DOI 10.14778/3229863.3229878.

15. Oh, B. K., Glisic, B., Kim, Y., Park, H. S. (2019). Convolutional neural network-based wind-induced response estimation model for tall buildings. *Computer-Aided Civil and Infrastructure Engineering*, 34(10), 843–858. DOI 10.1111/mice.12476.
16. Peng, H., Yan, J., Yu, Y., Luo, Y. (2021). Time series estimation based on deep learning for structural dynamic nonlinear prediction. *Structures*, 29, 1016–1031. DOI 10.1016/j.istruc.2020.11.049.
17. Zheng, Q. Y., Zhou, G. D., Liu, D. K. (2021). Method of modeling temperature-displacement correlation for long-span arch bridges based on long short-term memory neural networks. *Engineering Mechanics*, 38(4), 68–79. DOI 10.6052/j.issn.1000-4750.2020.05.0323.
18. Gul, M., Catbas, F. N. (2011). Damage assessment with ambient vibration data using a novel time series analysis methodology. *Journal of Structural Engineering*, 137(12), 1518–1526. DOI 10.1061/(ASCE)ST.1943-541X.0000366.
19. Pan, H., Lin, Z., Gui, G. (2019). Enabling damage identification of structures using time series-based feature extraction algorithms. *Journal of Aerospace Engineering*, 32(3), 04019014. DOI 10.1061/(ASCE)AS.1943-5525.0000978.
20. Buckley, T., Pakrashi, V., Ghosh, B. (2021). A dynamic harmonic regression approach for bridge structural health monitoring. *Structural Health Monitoring*, 20(6), 3150–3181. DOI 10.1177/1475921720981735.
21. van Le, H., Nishio, M. (2015). Time-series analysis of GPS monitoring data from a long-span bridge considering the global deformation due to air temperature changes. *Journal of Civil Structural Health Monitoring*, 5(4), 415–425. DOI 10.1007/s13349-015-0124-9.
22. Zhu, L., Zhuo, J., Xing, S. (2020). Strain prediction of bridge structural health monitoring based on CEEMDAN-NAR-ARIMA combination model. *Science Technology and Engineering*, 20(4), 1639–1644.
23. Ahmadivala, M., Sawicki, B., Brühwiler, E., Yamalas, T., Gayton, N. et al. (2019). Application of time series methods on long-term structural monitoring data for fatigue analysis. *SMAR 2019-5th International Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures*, pp. 1–8. Postdam, Germany.
24. Xin, J., Zhou, J., Yang, S. X., Li, X., Wang, Y. (2018). Bridge structure deformation prediction based on GNSS data using kalman-ARIMA-GARCH model. *Sensors*, 18(1), 298. DOI 10.3390/s18010298.
25. Shi, H., Worden, K., Cross, E. J. (2019). A cointegration approach for heteroscedastic data based on a time series decomposition: An application to structural health monitoring. *Mechanical Systems and Signal Processing*, 120, 16–31. DOI 10.1016/j.ymsp.2018.09.036.
26. Jiang, L., Chen, Z. (2016). Stress analysis for a bridge cable-tower anchorage zone based on the singular spectrum analysis. *Henan Science*, 34(7), 1107–1113.
27. Amezcua-Sanchez, J. P., Adeli, H. (2016). Signal processing techniques for vibration-based health monitoring of smart structures. *Archives of Computational Methods in Engineering*, 23(1), 1–15. DOI 10.1007/s11831-014-9135-7.
28. Seo, J., Hu, J. W., Lee, J. (2016). Summary review of structural health monitoring applications for highway bridges. *Journal of Performance of Constructed Facilities*, 30(4), 04015072. DOI 10.1061/(ASCE)CF.1943-5509.0000824.
29. Aljoumani, B., Sánchez-Espigares, J. A., Canameras, N., Josa, R., Monserrat, J. (2012). Time series outlier and intervention analysis: Irrigation management influences on soil water content in silty loam soil. *Agricultural Water Management*, 111, 105–114. DOI 10.1016/j.agwat.2012.05.008.
30. Kaloop, M. R., Elbeltagi, E., Hu, J. W., Elrefai, A. (2017). Recent advances of structures monitoring and evaluation using GPS-time series monitoring systems: A review. *ISPRS International Journal of Geo-Information*, 6(12), 382. DOI 10.3390/ijgi6120382.
31. Shan, D., Luo, L., Li, Q. (2020). State-of-the-art review of the bridge health monitoring in 2019. *Journal of Civil and Environmental Engineering*, 42(5), 115–125. DOI 10.11835/j.issn.2096-6717.2020.109.
32. Musa, Y. (2014). Modeling an average monthly temperature of sokoto metropolis using short term memory models. *International Journal of Academic Research in Business and Social Sciences*, 4(7), 382–397. DOI 10.6007/IJARBS/v4-i7/1030.
33. Cryer, J. D., Chan, K. S. (2008). *Time series analysis: With applications in R*, 2nd ed. New York: Springer.

34. Moyo, P., Brownjohn, J. M. (2002). Application of Box-Jenkins models for assessing the effect of unusual events recorded by structural health monitoring systems. *Structural Health Monitoring*, 1(2), 149–160. DOI 10.1177/1475921702001002003.
35. Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1), 1–20. DOI 10.1002/for.3980070102.
36. Wei, W. W. S. (2006). *Time series analysis univariate and multivariate methods*, 2nd ed. Boston, MA: Pearson Addison Wesley.
37. Chang, I., Tiao, G. C., Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2), 193–204. DOI 10.1080/00401706.1988.10488367.
38. Shumway, R. H., Stoffer, D. S., Stoffer, D. S. (2016). *Time series analysis and its applications: With R Examples*, 4th ed. New York: Springer.
39. Tsay, R. S., Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, 79(385), 84–96. DOI 10.1080/01621459.1984.10477068.