Engineering Structures 192 (2019) 220–232

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Modeling and forecasting of temperature-induced strain of a long-span bridge using an improved Bayesian dynamic linear model

Hao Wang[a,*], Yi-Ming Zhang[a], Jian-Xiao Mao[a], Hua-Ping Wan[b], Tian-You Tao[a], Qing-Xin Zhu[a]

[a] Key Laboratory of C&PC Structures of Ministry of Education, Southeast University, Nanjing 211189, China
[b] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Keywords:
Bayesian dynamic linear model
Temperature-induced strain
Strain forecast
Long-span bridge
Structural health monitoring

ABSTRACT

Temperature-driven baseline is highly responsive to anomalous structural behavior of long-span bridges, which means that the discrepancy between the measured and forecasting temperature-induced strain (TIS) can be examined for anomalies. In this regard, it is important to guarantee the accuracy of the forecasting TIS responses for reliable assessment of structural performance. Bayesian dynamic linear model (BDLM) has shown a promising application in the field of structural health monitoring. Traditionally, BDLM is used to forecast structural responses by utilizing its trend form, seasonal form, regression form, or combination of the three forms. However, different features of time series cannot be totally captured by these forms, which would undermine the accuracy of BDLM. To improve the computational accuracy, an improved BDLM, which considers an autoregressive (AR) component in addition to the trend, seasonal and regression components, is presented in this paper. Specifically, the AR component is able to model the component which cannot be captured by other three components. The real-time monitoring data collected from a long-span cable-stayed bridge is utilized to demonstrate the feasibility of the improved BDLM-based method. In particular, the present BDLM-based method allows for probabilistic forecasts, offering substantial information about the target TIS response, such as mean and confidence interval. Results show that the improved BDLM is capable of capturing the relationship between temperature and TIS. Compared to the AR model, multiple linear regression (MLR) model and BDLM without the AR component, the improved BDLM shows better forecasting performance in modeling and forecasting the TIS of a long-span bridge.

1. Introduction

In the past several decades, a large number of long-span bridges which support vital arteries for national transportation systems have been constructed in the world [1–4]. Age-related deterioration is a critical issue for in-service long-span bridges, thus leading to increasing concerns regarding structural safety and service performance [5,6]. Recently, structural health monitoring (SHM) systems are gradually equipped with long-span bridges to trace their health status and assess their performance using various sensors [7–10]. Massive measurement data related to the structural response and environmental variations is continuously collected for evaluation of the structural condition. However, rational interpretation of the massive amount of data to obtain useful information on structural performance remains a challenge in civil engineering [11–14].

The existing data interpretation approaches in SHM can be generally classified into two categories: model-based and data-driven approach [15]. Basically, model-based approaches rely on the finite element (FE) models to reflect the inherent structural characteristics. Usually, the initial FE models should be calibrated to have a better prediction of the structural responses [16–18]. However, it is a big challenge to create a reliable FE model that is able to well characterize the structural behavior of the in-service structure [19]. Alternatively, data-driven approaches show promise to mitigate shortcomings of the model-based method since they do not rely on an FE model and perform analyses directly on the measured time series data. Therefore, the data-driven approaches are more suitable for SHM to track and identify variation in the measurement [20,21].

Long-span bridges are highly sensitive to temperature actions, and the temperature-induced strain (TIS) response may even exceed the strain caused by traffic [22–25]. In recent years, temperature-based measurement interpretation approaches have been gradually applied to assess the structural condition of long-span bridges, leading to an easier structural identification [26]. Through a temperature-based structural

* Corresponding author.
E-mail address: wanghao1980@seu.edu.cn (H. Wang).

https://doi.org/10.1016/j.engstruct.2019.05.006
Received 20 January 2019; Received in revised form 2 May 2019; Accepted 2 May 2019
Available online 10 May 2019

identification (TBSI) method, Yarnold et al. [27] established the relationship between structural displacement and restrained member forces from the measured TIS. Xia et al. [28] presented a TBSI method using the structural transfer function from the input data (temperature) and output data (TIS). Results show that the input-output methods based on thermal response were considered more reliable than traditional methods using output-only data. Kromanis and Kripakaran [29] developed a temperature-based anomaly detection approach by analyzing the differences between measurements and predictions of temperature-induced responses. Apparently, the accuracy of the TBSI method relies largely on the reliability of forecasted TIS. Thus, it is important to develop appropriate approaches that are capable of well predicting TIS response [30].

In the SHM field, time-series analysis based on the autoregressive (AR) model has been extensively applied due to its computational simplicity and efficiency. Sohn and Farrar [31] constructed a two-stage model which combined AR and AR with exogenous inputs (ARX) to locate damage sources. They pointed out that the appropriate model order was important for AR model, since it would affect prediction performance of the AR model. Figueiredo et al. [32] investigated the performance of four methods, namely, Akaike information criterion, partial autocorrelation function, root mean squared error, and singular value decomposition, in obtaining an optimal AR order. The results showed that the AR model order defined by these four methods could provide robust damage detection. The AR model only describes the relationship between the current output and previous output of one variable, which is unable to model the relationship between different variables. Besides, the study of AR model-based method to forecast measurements in SHM field is limited. Bayesian dynamic linear model (BDLM) is considered as a special state-space model in machine learning field, which has shown a promising application in the field of SHM. BDLM has various forms, including polynomial trend models, seasonal models, regression models, and other models. Specifically, the regression model is employed to capture the dependency of state variables related to different observations. Fan [33] predicted the non-uniform extreme stress of I-39 Northbound Bridge using the second-order polynomial BDLM and showed that the predicted extreme stress was basically in line with the measurements. Wang [34] applied BDLM which consisted of a locally linear trend, a seasonal and a regression component to forecast the strain response and demonstrated that the selected optimal model performed well. Goulet and his colleges [35,36] proposed a BDLM framework for estimating the hidden states variables of structural response and the external effects (e.g., wind and temperature). This framework was first validated by the simulated data and later applied to Tamar Bridge in the UK. Nguyen and Goulet [37,38] presented BDLM-based approaches to conduct anomaly detection, and addressed situations where hidden covariates influenced the observed responses of structures. It was found that their approach had a better performance than the method using a superposition of harmonic hidden components.

Most measurements can be well modeled by BDLM with the superposition of trend, seasonality and regression components. Since the characteristics of the forecasting residuals cannot be well captured by the trend, seasonal or regression terms, the forecasting efficiency and accuracy can be influenced to some extent. In the BDLM context, the forecasting residuals can be described by the AR model which reflects the relationship between previous and current values. However, in the SHM field, existing approaches mainly focus on the analysis of measurements and neglect the temperature actions [30]. Besides, the existing BDLM is used to forecast observations by utilizing its trend form, seasonal form, regression form, or combination of the three forms. However, different features of time series cannot be totally captured by these forms, which would undermine the accuracy of BDLM. Furthermore, few studies use BDLM to model the relationship between temperature and TIS, as well as forecast TIS. To improve the computational accuracy, in addition to the trend, seasonal, and regression components, an autoregressive (AR) component is introduced to formulate the BDLM. The feasibility of the improved BDLM-based method is illustrated by using the real-time monitoring data collected from a long-span cable-stayed bridge.

2. BDLM for modeling TIS

BDLM is a system determined by observation and system equations. The observation equation describes how the observations rely on the current state parameters, while the system equation indicates the dynamic and stochastic variation of the interior system. The two equations are as follows:

Observation equation:

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim N(0, V_t) \tag{1}$$

System equation:

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \tag{2}$$

where Y_t is the observation vector; F_t is the observation matrix; θ_t is the state parameter vector; G_t is the evolution matrix; v_t is the observational error; and w_t is the evolution error which represents the stochastic changes in the state parameters from time $t-1$ to t. v_t and w_t are internally and mutually independent zero-mean Gaussian distributions with variance V_t and W_t.

In practice, specifying a model which can capture the different features of time series is a difficult task. To improve model accuracy, BDLM can be formulated by superposition of various elementary components, including trend, seasonality and regression components. As a result, lots of time series problems can be better handled with this model [35,38]. The noise model is employed to reflect the residual that cannot be captured in trend, seasonal or regression terms [39]. Among the various noise models, the AR model is commonly used in the literature [40].

2.1. Description of BDLM

One promising capability of the present BDLM is that the measurement is modeled by using the combination of elementary components: (a) a trend component, (b) a seasonal component, (c) a regression component and (d) an AR component. Furthermore, the measurement can be described by trend, seasonal, regression components as well as the AR component which accounts for the variation of residuals.

(a) Trend component

In time series, polynomial models are widely used for response forecasting, since they can efficiently describe trends which are considered as steady variation over time. First-order polynomial components of BDLM can reflect many significant concepts and features of time series [39]. Therefore, the first-order polynomial is suitable for short-term forecasting either on its own or in combination with other constituents. Specifically, short-term forecasting depends on the time-scales regarding the forecast horizon and the time-step, e.g., a few minutes, hours, and days ahead [41]. In this work, the first-order polynomial is employed as the trend component of BDLM to forecast TIS. Besides, the forecast horizon varies from one step to five steps and the time-step is hours.

The BDLM representation is

$$Y_t^{Trend} = \theta_t^{Trend} + v_t^{Trend}, \quad v_t^{Trend} \sim N(0, V_t^{Trend}) \tag{3}$$

$$\theta_t^{Trend} = \theta_{t-1}^{Trend} + w_t^{Trend}, \quad w_t^{Trend} \sim N(0, W_t^{Trend}) \tag{4}$$

For this case, the block component matrices are

$$\theta_t^{Trend} = \theta_t^{Trend}, \quad F_t^{Trend} = 1, \quad G_t^{Trend} = 1, \quad V_t^{Trend} = (v_t^{Trend})^2, \quad W_t^{Trend} = (w_t^{Trend})^2 \tag{5}$$

(b) Seasonal component

Periodic behavior is evident in TIS. In BDLM, the term seasonality is used to denote any periodic behavior. The seasonality in TIS is considered as an important factor for BDLM. Seasonal components with the different periods are described by their Fourier form as follows

$$\begin{aligned}
&\theta_t^{Season} = [\theta_t^{Season(1)}, \quad \theta_t^{Season(2)}, \cdots, \quad \theta_t^{Season(i)}, \cdots, \quad \theta_t^{Season(n)}]^T \\
&F_t^{Season} = [1 \quad 0 \quad 1 \quad 0 \quad \cdots \quad 1 \quad 0], \quad V_t^{Season} = (v_t^{Season})^2 \\
&G_t^{Season} = block \quad diag(J_1, J_2, \cdots, J_i, \cdots, J_n) \\
&W_t^{Season} = block \quad diag(w_t^{Season(1)}, w_t^{Season(2)}, \cdots, w_t^{Season(i)}, \cdots, w_t^{Season(n)})
\end{aligned} \tag{6}$$

In this form, the i th harmonic component is given by

$$J_i = \begin{bmatrix} \cos p_i & \sin p_i \\ -\sin p_i & \cos p_i \end{bmatrix} \tag{7}$$

where $p_i = 2\pi/C$, in which C is the period of the time series.

The i th evolution error is given by

$$w_t^{Season(i)} = \begin{bmatrix} w_{11}^{Season(i)} & 0 \\ 0 & w_{22}^{Season(i)} \end{bmatrix} \tag{8}$$

(c) Regression component

Regression component is capable of describing the dependency between temperature and TIS. Consider modeling TIS Y_t^{Re} by regressing on temperature x. For each t, the regression model is defined as

$$Y_t^{\mathrm{Re}} = \beta_t \cdot x_t + v_t^{\mathrm{Re}}, \quad v_t^{\mathrm{Re}} \sim N(0, V_t^{\mathrm{Re}}) \tag{9}$$

with temporal evolution β_t given by

$$\beta_t = \beta_{t-1} + w_t^{\mathrm{Re}}, \qquad w_t^{\mathrm{Re}} \sim N(0, W_t^{\mathrm{Re}}) \tag{10}$$

In the BDLM, the regression model can be expressed by

$$\theta_t^{\mathrm{Re}} = \beta_t, \quad F_t^{\mathrm{Re}} = x_t, \quad G_t^{\mathrm{Re}} = 1, \quad V_t^{\mathrm{Re}} = (v_t^{\mathrm{Re}})^2, \quad W_t^{\mathrm{Re}} = (w_t^{\mathrm{Re}})^2 \tag{11}$$

(d) AR component

The AR (p) model, where p is the order of AR model, is described as follows:

$$Y_t^{AR} = \sum_{j=1}^{p} \phi_j Y_{t-j} + v_t^{AR} \tag{12}$$

where $Y_t^{AR}$is the measurement at time t; ϕ_j is the autoregressive coefficient; and v_t^{AR} is the observational error.

In the context of BDLM, the AR noise model is often used to represent the forecasting errors which cannot be well characterized by the trend, seasonal or regression components. The AR model is expressed by

$$\begin{aligned}
&\theta_t^{AR} = [\theta_1^{AR}, \theta_2^{AR}, \cdots, \theta_n^{AR}]^T \quad F_t^{AR} = [1 \quad 0 \quad \cdots \quad 0], \quad V_t^{\mathrm{AR}} = (v_t^{\mathrm{AR}})^2 \\
&W_t^{AR} = [(w_t^{AR})^2, 0, 0, \cdots, 0]^T \qquad G_t^{AR} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_3 & \phi_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}
\end{aligned} \tag{13}$$

Therefore, the measurement Y_t can be defined as

$$Y_t = Y_t^{Trend} + Y_t^{Season} + Y_t^{\mathrm{Re}} + Y_t^{AR} + v_t \tag{14}$$

2.2. Model inference and forecast

BDLM for TIS forecast is detailed as follows [34].

(1). Initialization: The initial prior state parameter is specified as

$$P(\theta_0 \mid D_0) \sim N(m_0, S_0) \tag{15}$$

where m_0 and S_0 are the initial mean and variance of the state parameter, respectively; and D is the state of information.

(2). Prior distribution for θ_t: assuming the posterior distribution of state parameter at time t is defined as

$$P(\theta_t \mid D_t) \sim N(m_t, S_t) \tag{16}$$

The prior distribution of $P(\theta_{t+1} \mid D_t)$state parameter at time $t+1$ can be obtained by

$$\begin{aligned}
&E[\theta_{t+1} \mid D_t] = G_{t+1} E[\theta_t \mid D_t] + E[w_{t+1}] = \mathrm{G}_{t+1} m_t = \mu_{t+1} \\
&Var[\theta_{t+1} \mid D_t] = G_{t+1} Var[\theta_t \mid D_t] G'_{t+1} + Var[w_{t+1}] = G_{t+1} S_t G'_{t+1} + W_{t+1} \\
&\qquad\qquad = R_{t+1}
\end{aligned} \tag{17}$$

(3). One-step-ahead forecast: the expectation is used to forecast TIS, and the associated variance represents the confidence intervals. The one-step-ahead forecast distribution $P(Y_{t+1} \mid D_t)$ for observation at time $t+1$ can be calculated by

$$\begin{aligned}
&E[Y_{t+1} \mid D_t] = G_{t+1} E[\theta_{t+1} \mid D_t] + E[v_{t+1}] = F_{t+1} \mu_{t+1} = f_{t+1} \\
&Var[Y_{t+1} \mid D_t] = F_{t+1} Var[\theta_{t+1} \mid D_t] F'_{t+1} + Var[v_{t+1}] = F_{t+1} R_{t+1} F'_{t+1} + V_{t+1} \\
&\qquad\qquad = Q_{t+1}
\end{aligned} \tag{18}$$

(4). Posterior distribution for θ_{t+1}: After the observation Y_{t+1} is provided, the distribution of state parameter is updated to$P(\theta_{t+1} \mid D_{t+1}) \sim N(m_{t+1}, S_{t+1})$,

$$\begin{aligned}
&m_{t+1} = \mu_{t+1} + K_{t+1} e_{t+1} \\
&S_{t+1} = R_{t+1} - K_{t+1} K_{t+1}^T Q_{t+1}
\end{aligned} \tag{19}$$

where $e_{t+1} = Y_{t+1} - f_{t+1}$ and $K_{t+1} = R_{t+1} F_{t+1} / Q_{t+1}$.

(5). k-step-ahead forecast: the prior distribution of state parameter at time $t+k$ can be expressed as

$$\begin{aligned}
&E[\theta_{t+k} \mid D_t] = G_{t+k} E[\theta_{t+k-1} \mid D_t] + E[w_{t+k}] = \mu_t(k) \\
&Var[\theta_{t+k} \mid D_t] = G_{t+k} Var[\theta_{t+k-1} \mid D_t] G'_{t+k} + Var[w_{t+k}] = R_t(k)
\end{aligned} \tag{20}$$

The k-step-ahead forecast distribution $P(Y_{t+k} \mid D_t)$ for observation series at time $t+k$ can be calculated by

$$\begin{aligned}
&E[Y_{t+k} \mid D_t] = G_{t+k} E[\theta_{t+k} \mid D_t] + E[v_{t+k}] = F_{t+k} \mu_t(k) = f_t(k) \\
&Var[Y_{t+k} \mid D_t] = F_{t+k} Var[\theta_{t+k} \mid D_t] F'_{t+k} + Var[v_{t+k}] \\
&\qquad\qquad = F_{t+k} R_t(k) F'_{t+k} + V_{t+k} = Q_t(k)
\end{aligned} \tag{21}$$

2.3. Model parameter determination

In this paper, the expectation maximization (EM) and Kalman smoother (KS) algorithm are employed to estimate model parameters in the BDLM [35]. Maximum likelihood estimation is widely used to estimate parameters of models, e.g. linear regression model, AR model and autoregressive moving average model [42,43]. However, the likelihood function is difficult to solve when latent variables of the model cannot be directly observed. The EM algorithm provides an efficient solution to identify systems with hidden variables which consists of an estimation step and a maximization step [44]. Therefore, EM algorithm is used here for parameter estimation. To be specific, in the E step, the expectation of the log-likelihood function is derived by Kalman smoother, whereas in the M step, estimation of the parameters is renewed by maximizing the expected log-likelihood function.

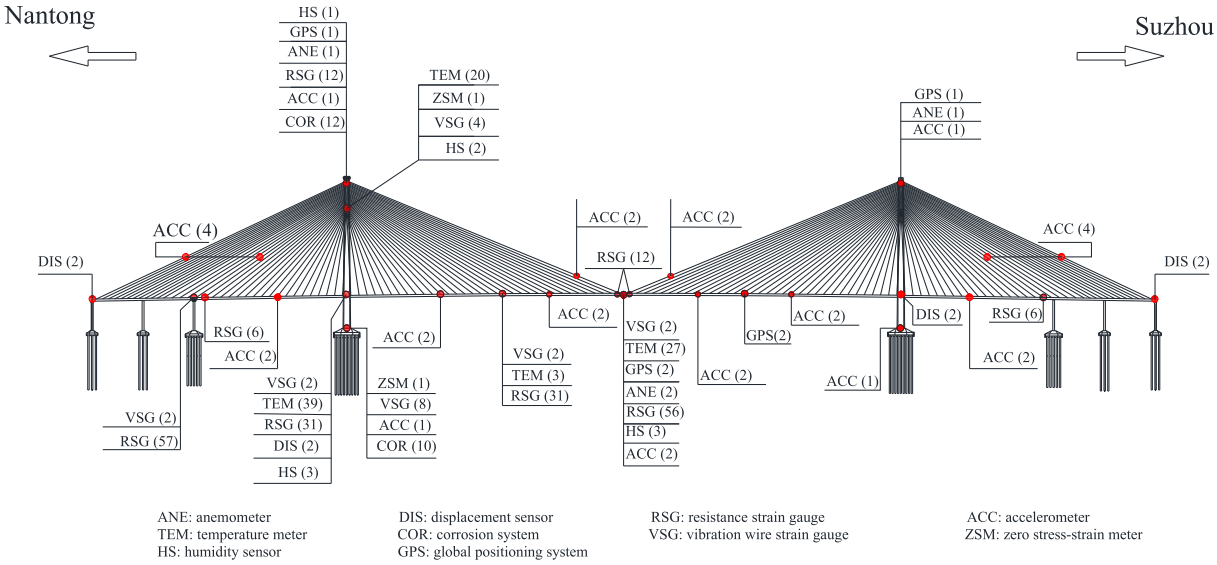

Fig. 1. Sensors layout in the structural health monitoring system of SCB.

E-step: Let $\Gamma = \sum_{t=1}^{T} m_{t|T} m'_{t|T} + S_{t|T}$, $\Upsilon = \sum_{t=1}^{T} m_{t-1|T} m'_{t-1|T} + S_{t-1|T}$, $\Lambda = \sum_{t=1}^{T} m_{t|T} m'_{t-1|T} + S_{t,t-1|T}$. And use the property of the matrix trice, $M \in R^n$, $N \in R^n$, then

$$M^T N = tr(NM^T) \tag{22}$$

Based on Eq. (22) and the results ($m_{t-1|T}$, $S_{t-1|T}$, $S_{t,t-1|T}$) calculated by the Kalman smoother, the expectation of the log-likelihood function can be derived by

$$\begin{aligned} &E[\ln p(\Theta_{0:T}, Y_{1:T} | \Theta)] \\ &= -\frac{1}{2}\ln|S_0| - \frac{T}{2}\ln|W| - \frac{T}{2}\ln|V| - \frac{(a+b)T+a}{2}\ln(2\pi) \\ &\quad -\frac{1}{2}tr\{S_0^{-1}[(m_{0|T} - m_0)(m_{0|T} - m_0)'] + S_{0|T}\} \\ &\quad -\frac{1}{2}\sum_{t=1}^{T} tr\{W^{-1}[\Gamma - 2G_t\Upsilon' + G_t\Lambda G_t']\} \\ &\quad -\frac{1}{2}\sum_{t=1}^{T} tr\{V^{-1}[(Y_t - F_t m_{t|T})(Y_t - F_t m_{t|T})' + F_t S_{t|T} F_t']\} \end{aligned} \tag{23}$$

where T is the sample size; a is the dimension of state vector; and b is the dimension of observation vector.

M-step: Maximize $E[\ln p(\Theta_{0:T}, Y_{1:T} | \Theta)]$ by letting its partial derivative in terms of F, G, V, W, m_0, S_0 be zero, respectively. The parameter estimation can be expressed as

$$\begin{aligned} F_t &= (\sum_{t=1}^{T} Y_t \theta'_{tT})\Gamma^{-1}, \\ G_t &= \Upsilon\Lambda^{-1}, \\ V_t = \frac{1}{T}\sum_{t=1}^{T}(\theta_t\theta'_t - F_t\theta_{tT}Y'_t),\ W_t &= \frac{1}{T}(\Gamma - \Upsilon\Lambda^{-1}\Upsilon') \\ m_0 = m_{0T},\ S_0 &= S_{0T} \end{aligned} \tag{24}$$

The E and M steps are repeated recursively until the variation between iterations is below a small threshold.

2.4. Model assessment

The Akaike information criterion (AIC) is an efficient tool for variable selection in modeling [45,46]. To select a suitable AR model order and measure the effectiveness of BDLM, AIC and root mean squared error (RMSE) are presented as

$$AIC = -2\ln(L) + 2p \tag{25}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(Y_n - \hat{Y}_n)^2}{N}} \tag{26}$$

where L is the likelihood function; p is the number of parameters of BDLM; N is the number of samples; and Y_n is the TIS, $\hat{Y}_n$ is the predicted value. According to the AIC and RMSE criterion, the smaller the AIC and RMSE value the better the model.

3. Sutong Cable-Stayed Bridge and its SHM system

The Sutong Cable-Stayed Bridge (SCB), completed in June 2008, with a main span of 1088 m over the Yangtze River in Jiangsu province of China (Fig. 1). It was the longest cable-stayed bridge in the world when it opened to public traffic. SCB has two inverted Y-shaped reinforced concrete towers with heights of 300.4 m, and 272 symmetrically distributed cables connect the bridge tower with girder. The main girder is a streamlined flat steel box girder with a height of 4 m and a width of 41 m. SCB forms a key component of the highway

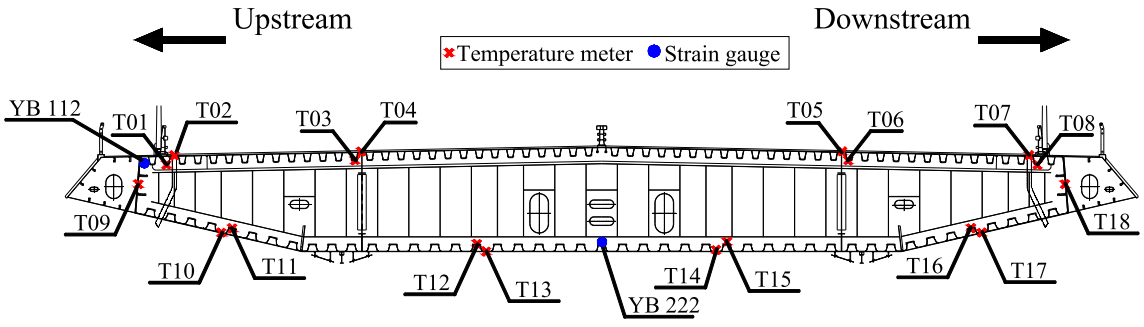

Fig. 2. Arrangement of strain gauges and temperature meters at the mid-span of SCB.

network of Jiangsu province linking Nantong city to Suzhou city. By the end of last year, the average traffic flow over the bridge has increased to 96,900 daily units which are much higher than 27,200 daily units during its first year of operation.

After completing the construction of SCB, a sophisticated long-term SHM system was installed on the bridge to monitor and evaluate its performance in real time. This system can be divided into 4 integrated modules, namely sensory subsystem, data acquisition and transmission subsystem, data management and control subsystem, and structural health evaluation subsystem. The system consists of 10 types of sensors, including anemometer, temperature meter, humidity sensor, displacement sensor, corrosion system, global positioning system, resistance strain gauge, vibration wire strain gauge, accelerometer, and zero stress-strain meter. More details about the SHM system can be found in Wang et al. [47]. The sensor layout of strain gauges and temperature meters at the mid-span of SCB is shown in Fig. 2. 18 temperature meters and 56 strain gauges installed at the mid-span. The strain data are acquired at a sampling frequency of 10 HZ. In this paper, two strain gauges located at the top and bottom flange are selected to extract TIS.

4. Application of BDLM to TIS forecasting

4.1. Temperature-induced strain separation

Without loss of generality, TIS extracted from two strain gauges (YB 112 and YB 222) during two periods (January 1 to February 15, 2009 and June 16 to July 31, 2009) with relatively high and low temperature are selected for illustration. Since the temperature meters and strain gauges are not at the same location, the average of the temperature meters close to the target strain gauge is used to approximate the true temperature. Specifically, the average of two temperature meters (T01 and T02) adjacent to YB 112 is used to forecast TIS extracted from this strain gauge. Similarly, the average temperature measurements (T12, T13, T14, and T15) are selected to forecast TIS separated from strain gauge YB 222. Hourly temperature time histories (a total of 1104 observations during each period) are exhibited in Fig. 3. As shown in Fig. 3, the temperature fluctuates within −10 °C and 30 °C from January 1 to February 15, 2009. In addition, during the period of June 16 to July 31, 2009, the temperature varies from 20 °C to 50 °C, and

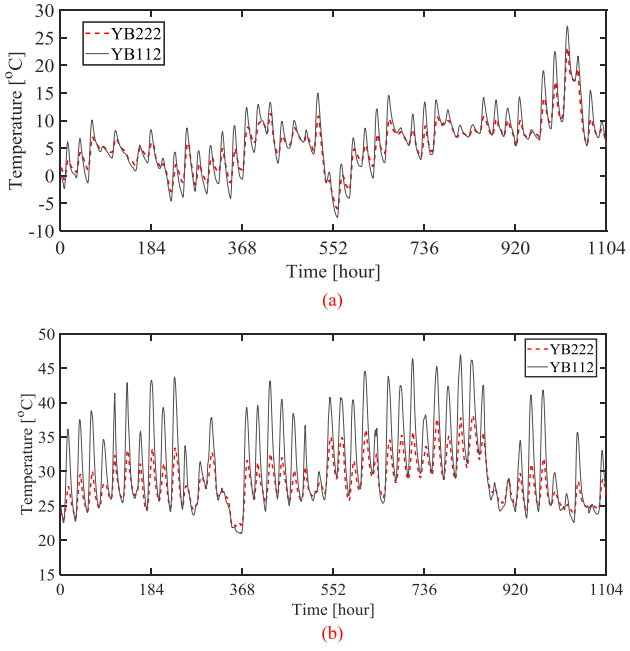

Fig. 3. Temperature measurement of SCB: (a) From January 1 to February 15, 2009; (b) From June 16 to July 31, 2009.

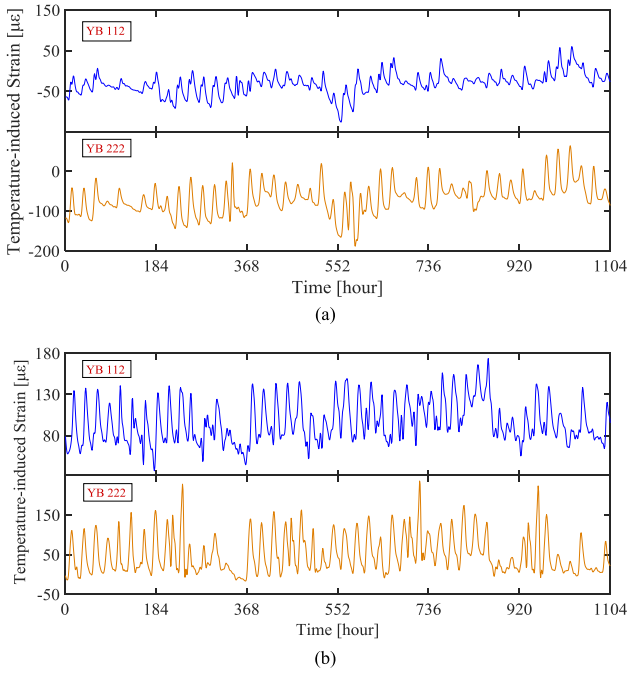

Fig. 5. TIS of SCB: (a) From January 1 to February 15, 2009; (b) From June 16 to July 31, 2009.

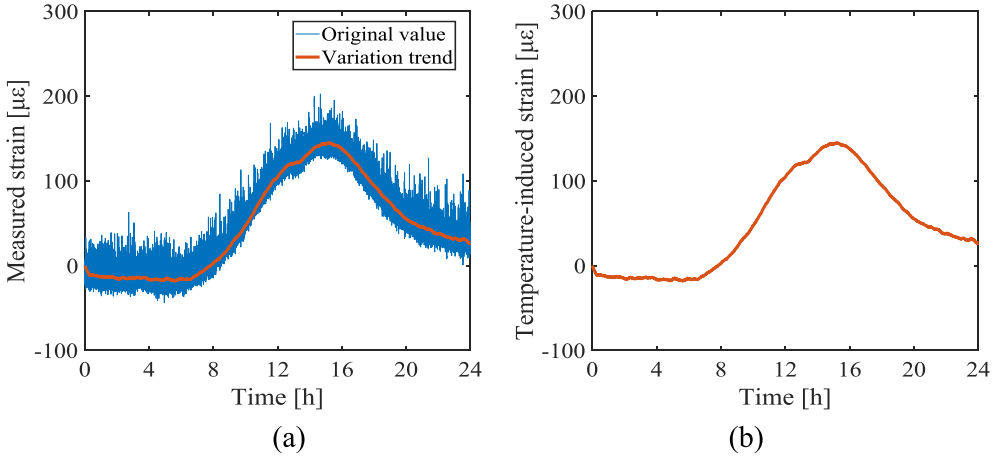

Fig. 4. TIS separation: (a) measured strain; (b) TIS.

obvious temperature differences can be observed in the two periods.

The monitored strain response of SCB mainly induced by three sources which compromise traffic, wind, and temperature. Fig. 4 illustrates the measured strain time history in a one-day period (July 1, 2009). As shown in Fig. 4(a), the strain response has a range of [−50, 200] με. Specifically, the largest strain response acquired at about 4p.m. It can be also observed that the strain response consists of two ingredients: slow-varying trend and dynamic ingredient. The slow-varying trend in strain time history represents the daily cycle effect of temperature variation, which can be obtained through trend extraction methods [48,49]. The moving average method (MAM) is used to extract TIS from the measured strain response [50–52]. Fig. 4(b) shows the time history of separated TIS. The TIS varies between −20 με and 150 με. The extracted TIS over the two periods is presented in Fig. 5.

4.2. Model construction for TIS

The temperature and TIS are modeled and forecasted by utilizing the preceding procedure with BDLM. Four components are involved in the BDLM for TIS: a trend component, a seasonal component, a regression component, and an AR component.

For the seasonal component, accurate identification of the period of time series plays a significant role in improving the forecasting performance. As shown in Fig. 5, seasonal variation of TIS is evident. Spectral analysis method is used to identify the main period of time series by breaking it into multi-periods with different amplitude, phases, and frequencies. Because of its efficiency and simplicity, the spectral analysis method has been widely applied to period identification of time series [53,54]. The power spectral density of TIS is calculated and shown in Fig. 6.

According to the power spectral density in Fig. 6, two obvious peaks occur at a period of 12 h and 24 h. Therefore, evolution matrix in Eq. (6) can be expressed by

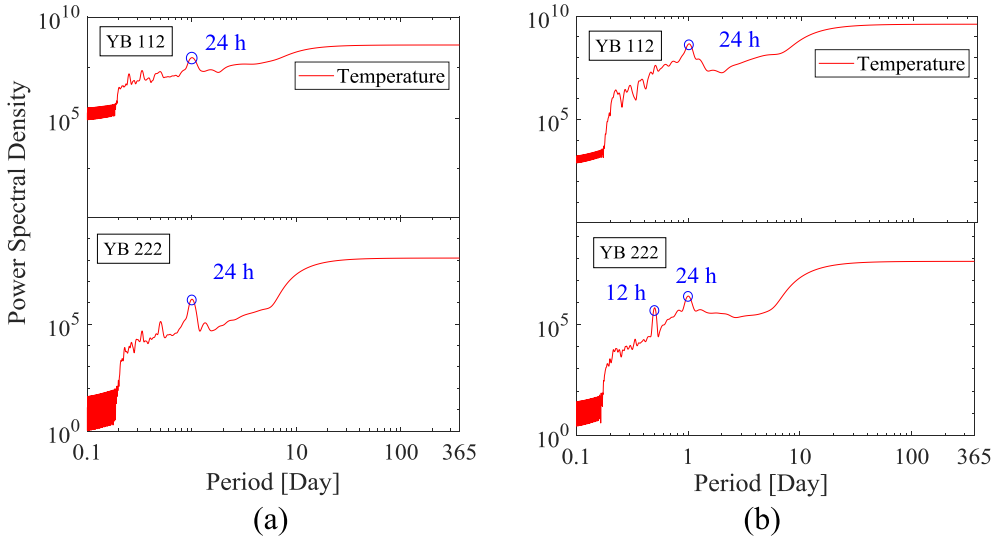

Fig. 6. Power spectral density of TIS: (a) From January 1 to February 15, 2009; (b) From June 16 to July 31, 2009.

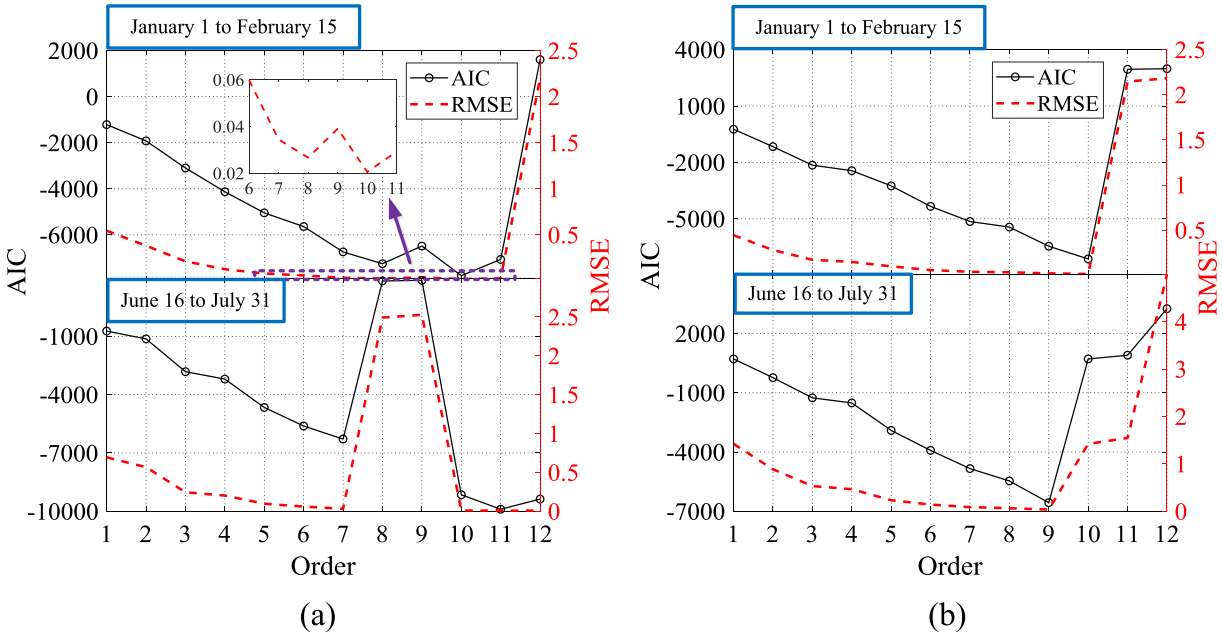

Fig. 7. Results of AIC and RMSE: (a) YB 112 (b) YB 222.

$$G^{Season1} = \begin{bmatrix} \cos(\pi/12) & \sin(\pi/12) \\ -\sin(\pi/12) & \cos(\pi/12) \end{bmatrix}$$

$$G^{Season2} = \begin{bmatrix} \cos(\pi/6) & \sin(\pi/6) & 0 & 0 \\ -\sin(\pi/6) & \cos(\pi/6) & 0 & 0 \\ 0 & 0 & \cos(\pi/12) & \sin(\pi/12) \\ 0 & 0 & -\sin(\pi/12) & \cos(\pi/12) \end{bmatrix} \quad (27)$$

After the period of TIS is confirmed, the overall matrixes defining BDLM are

$$\begin{aligned} Y_t &= [Y_t^{Trend}, Y_t^{Season}, Y_t^{\mathrm{Re}}, Y_t^{AR}]^T \\ \theta_t &= [\theta_t^{Trend}, \theta_t^{Season}, \beta_t, \theta_t^{AR}]^T \\ F_t &= [1, 1, 0, 1, 0, 1, 0, \cdots, 0, Temperature(t)] \\ G_t &= block \quad diag[G_t^{Trend}, G_t^{Season}, G_t^{\mathrm{Re}}, G_t^{AR}] \\ V_t &= block \quad diag[V_t^{Trend}, V_t^{Season}, V_t^{\mathrm{Re}}, V_t^{AR}] \\ W_t &= block \quad diag[W_t^{Trend}, W_t^{Season}, W_t^{\mathrm{Re}}, W_t^{AR}] \end{aligned} \quad (28)$$

A training period of 100 h is used to learn the model parameters. The EM algorithm is employed recursively until the variation of log-likelihood between iterations is below 10^{-5}, or the number of iterations reaches 1000. The set of unknown parameters is defined as follows:

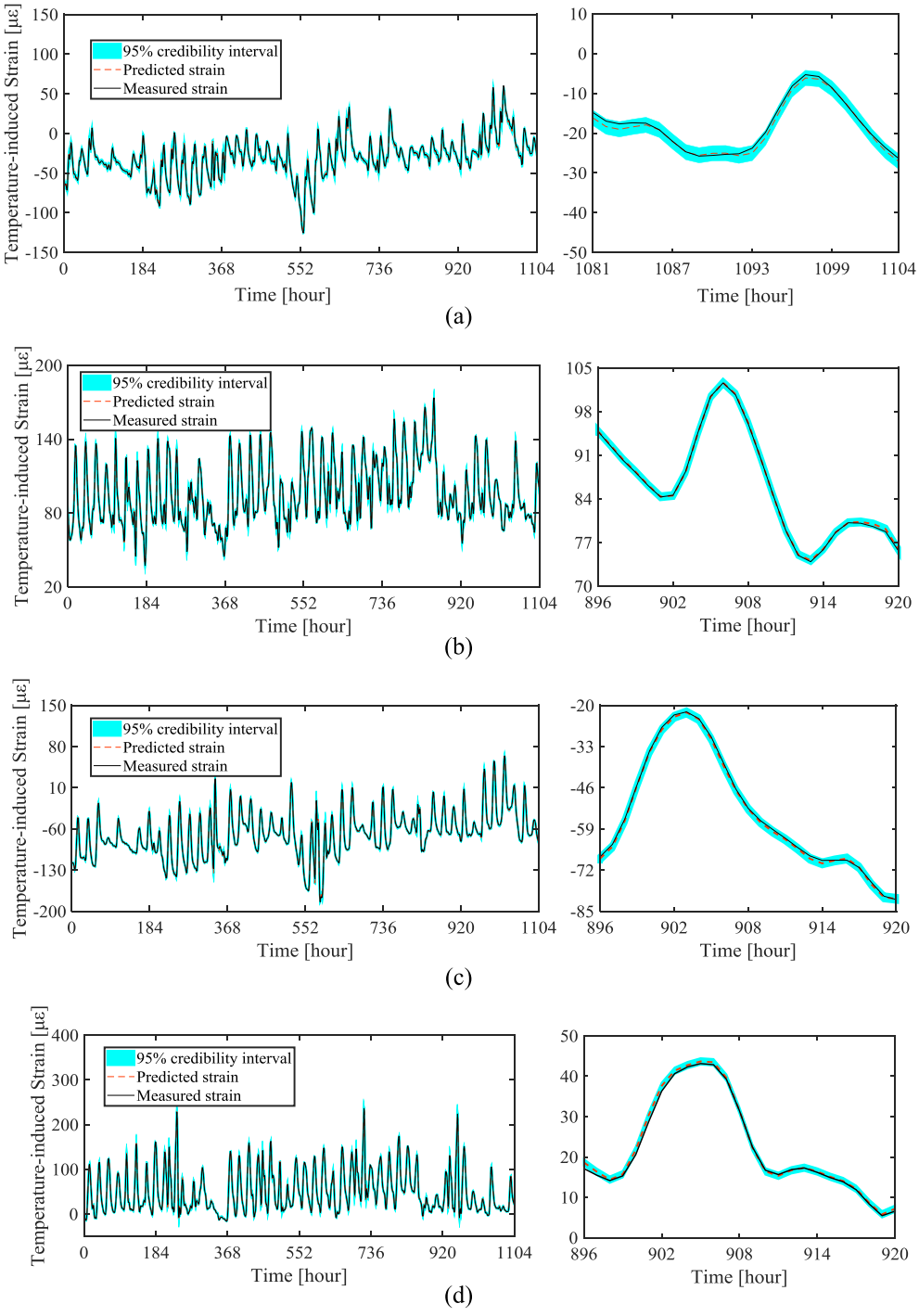

Fig. 8. One-step-ahead forecasting results using BDLM: (a) From January 1 to February 15 for YB 112 with AR (10); (b) From June 16 to July 31 for YB 112 with AR (11); (c) From January 1 to February 15 for YB 222 with AR (10); (d) From June 16 to July 31 for YB 222 with AR (9).

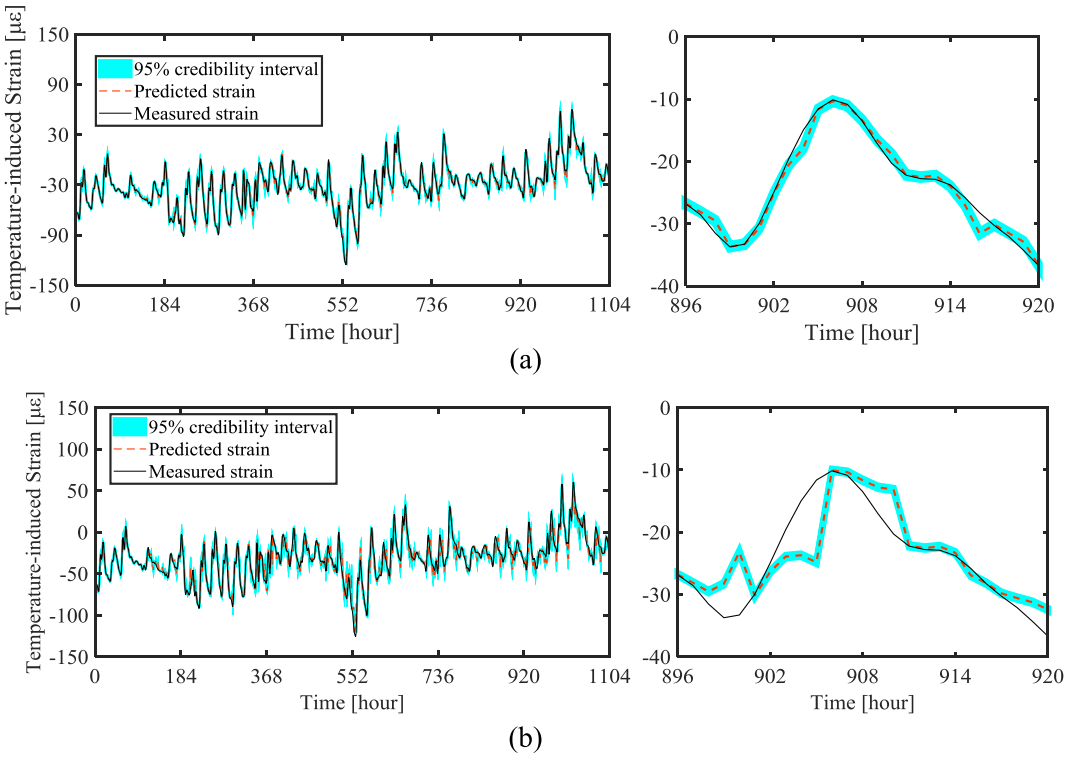

Fig. 9. Multi-step-ahead forecasting results of YB 112 from January 1 to February 15: (a) three-step-ahead forecast; (b) five-step-ahead forecast.

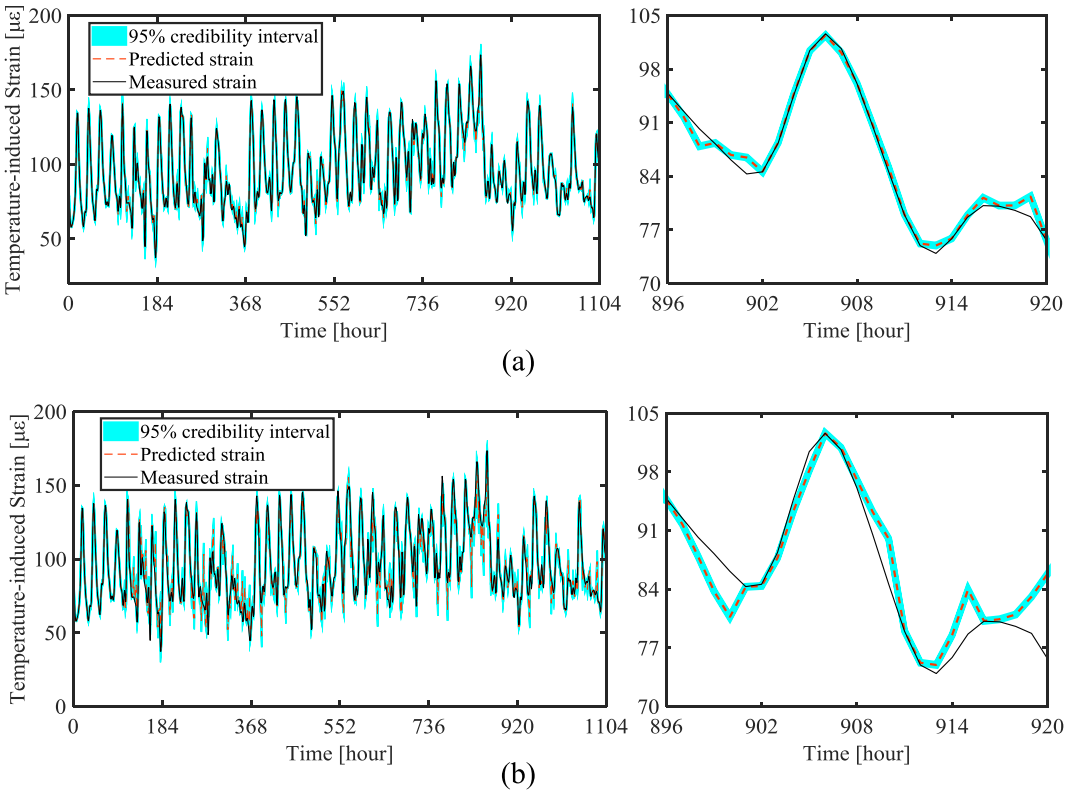

Fig. 10. Multi-step-ahead forecasting results of YB 112 from June 16 to July 31: (a) three-step-ahead forecast; (b) five-step-ahead forecast.

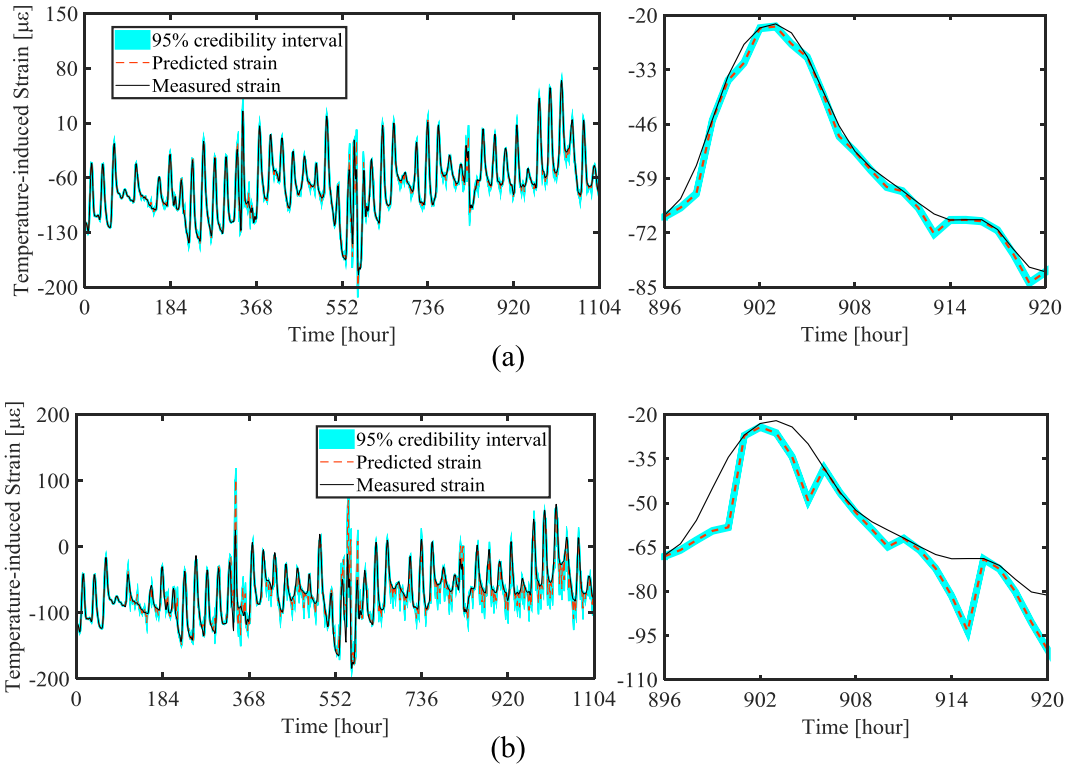

Fig. 11. Multi-step-ahead forecasting results of YB 222 from January 1 to February 15: (a) three-step-ahead forecast; (b) five-step-ahead forecast.

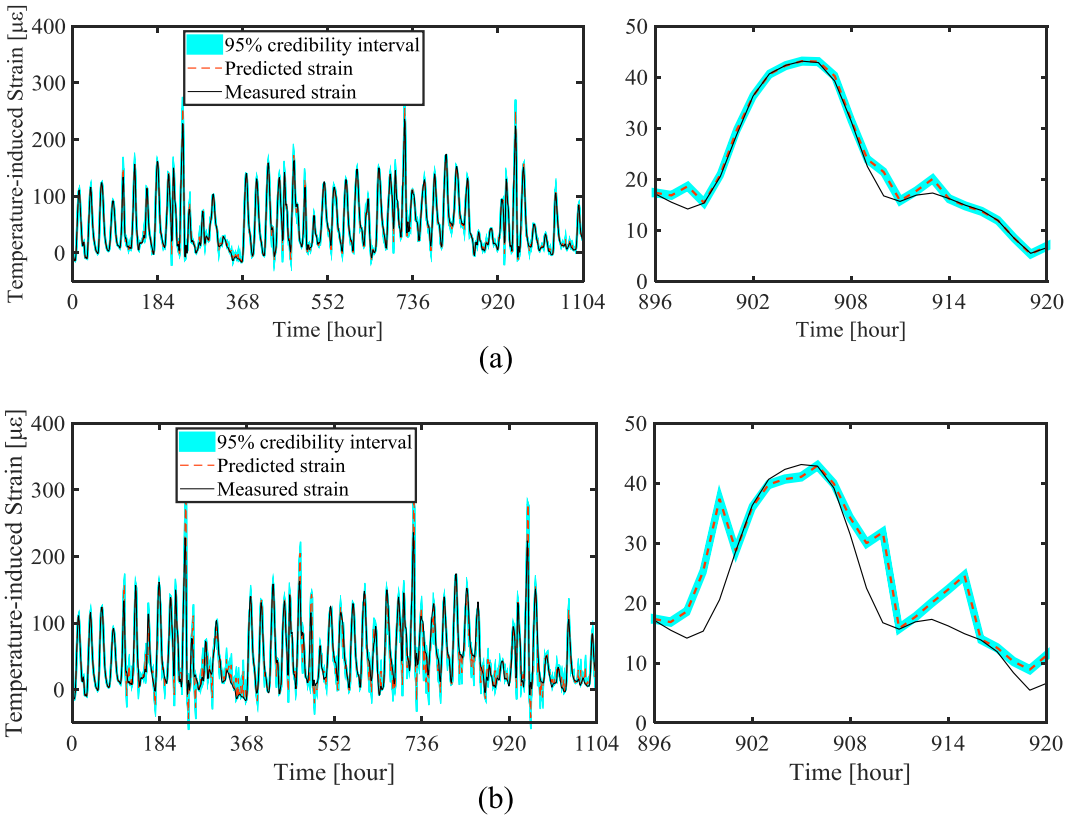

Fig. 12. Multi-step-ahead forecasting results of YB 222 from June 16 to July 31: (a) three-step-ahead forecast; (b) five-step-ahead forecast.

Table 1
AIC and RMSE results of multiple-step-ahead forecasting.

Position	Period	One-step		Three-step		Five-step	
		RMSE	AIC	RMSE	AIC	RMSE	AIC
YB 112	January 1 to February 15	0.021	−7770.5	0.225	−2970.2	0.706	−676.7
	June 16 to July 31	0.007	−9894.3	0.173	−3500.3	0.700	−693.9
YB 222	January 1 to February 15	0.028	−7134.2	0.367	−1.991.7	1.463	786.6
	June 16 to July 31	0.038	−6564.3	0.447	−1.594.4	1.442	756.7

Table 2
RMSE results of multiple-step-ahead forecasting using AR model.

Position	Period	AR order	One-step	Three-step	Five-step
			RMSE	RMSE	RMSE
YB 112	January 1 to February 15	47	0.49	0.95	1.34
	June 16 to July 31	39	0.61	1.25	1.67
YB 222	January 1 to February 15	41	0.73	1.54	2.27
	June 16 to July 31	30	0.53	2.24	3.24

$$\Theta = [\beta_t, v_t, w_t^{Trend}, w_{11}^{Season(1)}, w_{22}^{Season(1)}, w_{11}^{Season(2)}, w_{22}^{Season(2)}, w_t^{Re}, w_t^{AR}, \phi_1, \phi_2, \cdots, \phi_p] \quad (29)$$

4.3. Determination of the optimal BDLM

To choose a suitable order of the AR component in BDLM, two measure criteria of RMSE and AIC are presented. One advantage of the BDLM method is that it allows for probabilistic forecasts, providing massive information about the forecasts, such as mean and confidence interval. More specifically, the confidence interval represents the uncertainty of the target TIS response. A 95% confidence interval is presented in this study. The AIC and RMSE values are presented in Fig. 7.

As shown in Fig. 7, it can be found that the variation of AIC and RMSE values show a similar trend with the increase of AR model order. Specifically, the variation of AIC is more evident than RMSE. Furthermore, AIC and RMSE for YB 112 and YB 222, during the selected two periods, reach the minimum value when the AR model orders are 10, 11, 10 and 9. Therefore, the appropriate orders of BDLMs are determined as 10, 11, 10 and 9. The forecasting results of BDLM with suitable orders are shown in Fig. 8, which demonstrates that most observations lie within the 95% credibility interval and the predictions closely match the observations.

4.4. Multi-step-ahead forecast based on the optimal BDLM

As mentioned above, the suitable orders of AR component in BDLMs are 10, 11, 10 and 9. The optimal BDLMs with these suitable AR orders are utilized for multi-step-ahead forecast. Figs. 9–12 show the multi-step-ahead forecasting results. AIC and RMSE results of TIS of different periods are listed in Table 1. As shown in Figs. 9–12, it can be readily seen that the forecasting performance degrades when the forecast horizon increases. It indicates that the longer forecast horizon implies more stochastic uncertainty for the forecasting TIS. The detailed analysis of forecasting results is discussed below.

From Table 1, it can be concluded that the one-step-ahead forecasting results in the lowest AIC and RMSE values. Both values become larger with the increase of forecast steps. For example, for YB 112 during the period from January 1 to February 15, the one-step-ahead forecast has the minimum AIC value of −7770.5 as opposed to the higher AIC values of −2970.2 and −676.7 for three-step-ahead and five-step-ahead forecast. The one-step-ahead forecast achieves the lowest RMSE value of 0.021while the higher values of 0.225 and 0.706 for three-step-ahead to five-step-ahead forecast. Besides, it is obvious that the fluctuation of the forecasts becomes more intensive as the forecasting horizon increases. The reason is that the forecasting uncertainty increases as it forecasts further ahead.

4.5. Comparison of forecasting performance

In order to evaluate the forecasting performance, the present BDLM approach is compared with the AR model, BDLM without the AR component, and multiple linear regression (MLR) model. Similarly, 100 data are used for training and the rest data for testing. RMSE is performed to compare the forecasting accuracy of the models. According to RMSE

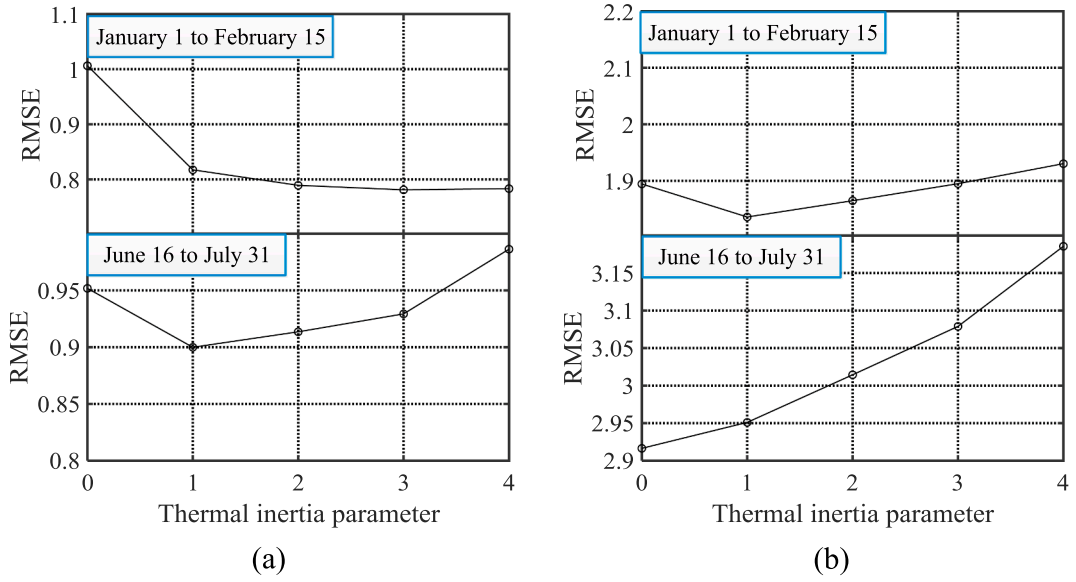

Fig. 13. Thermal inertia parameter: (a) YB 112 (b) YB 222.

criterion, the smaller the value of RMSE is, the more accurate the model.

In this work, for the AR model, AIC is employed to select an appropriate order and the Burg algorithm is chosen to determine the model coefficients [55,56]. The minimum AIC value corresponds to the optimal order of AR model. Table 2 shows RMSE results of AR model with optimal order which employed to forecast TIS for YB 112 and YB 222.

Details about forecasting TIS using MLR model can be found in Ref. [30]. For the MLR model which used to forecast TIS in this paper, explanatory variables derived from the temperature measurements recorded by the temperature meters. Besides, the time-varying regression weights are computed based on least-squares estimator. Thermal inertia that internal temperatures within the structure may lag behind the ambient temperature is also considered. The optimal thermal inertia parameter is determined based on the criterion of minimum forecasting errors. Thermal inertia parameters vary from 0 to 4 to ensure that the regression weights have the unique value. Fig. 13 shows the variation of RMSE values with respect to thermal inertia parameters.

Fig. 14 shows the results of RMSE for multi-step-ahead forecasts by using the AR model, MLR model, and BDLMs with and without the AR component. It should be noted that the MLR only employs temperature variables to forecast TIS at the same time and thus is suitable for one-step-ahead forecast. Therefore, only one-step-ahead forecasting results of MLR model are presented.

As shown in Fig. 14, it can be concluded that the accuracy decreases dramatically with the increased forecast horizon for the AR model, and BDLMs with and without the AR component. Besides, it can be easily observed that BDLM with the AR component has the lowest RMSE value among all different approaches. The results indicate that the present approach outperforms the AR model, MLR model, and BDLM without the AR component. This advantage can be explained by the fact that BDLM with the AR component considers the relationship between the current and previous output of TIS as well as the relationship with temperature.

MLR model leads to the poorest forecasting performance, mainly because the model only considers the relationship between TIS and temperature. As reported in [30], more than six-month measurements are required to capture the full range of temperature and TIS, and in this circumstance of sufficient data information, MLR may have better forecasting performance. In this study, the training data is relatively limited, which may lead to the poorest forecasting performance.

It can also be seen from Fig. 14 that the forecasting errors of AR model are larger than BDLM with and without AR component. This may be attributed that AR coefficients rely on the historical TIS response without the feature of dynamic state update. BDLM without the AR component has better forecasting performance than the AR model, since BDLM can account for stochastic uncertainty to some extent during its dynamic state estimate. Thus the BDLM has a strong capacity of handling stochastic uncertainty. Furthermore, the BDLM with the AR component combines the merits of the AR model and BDLM approach to well account for system uncertainty, thus mitigating forecasting errors. In addition, the BDLM is able to capture the trend and periodic component for TIS. Therefore, the BDLM with the AR component leads to better forecasting performance than other three approaches.

5. Conclusions

In this work, an improved BDLM, which consists of the trend, seasonal, regression, and AR components, is employed to model and forecast TIS. The model parameters are estimated by combing the EM with KS algorithm. The TIS collected from a long-span cable-stayed bridge is applied to illustrate the effectiveness of the presented method. The performance of the BDLMs with different orders of AR component is compared in terms of measure criteria of AIC and RMSE, and then the appropriate order of BDLM is determined for multi-step-ahead forecast. The forecasting performance of this improved BDLM, AR model, MLR model, and BDLM without AR component is fully investigated. The following conclusions can be made:

(1) The improved BDLM is able to capture the hidden state and model TIS efficiently, since it characterizes the behavior of TIS by superposing the trend, seasonal, regression, and AR components. Besides, the improved BDLM approach has the advantage that the model inference involves the dynamic state update, which allows for accounting for stochastic uncertainty through considering the forecasting error.
(2) The period of the TIS plays an important role in the seasonal component, which can directly affect the forecasting accuracy. In this study, the periodic pattern of the TIS can be identified by spectral analysis method.
(3) Compared to AR model, MLR model and BDLM without the AR component, the improved BDLM shows better forecasting performance in modeling and forecasting the TIS. This can be attributed

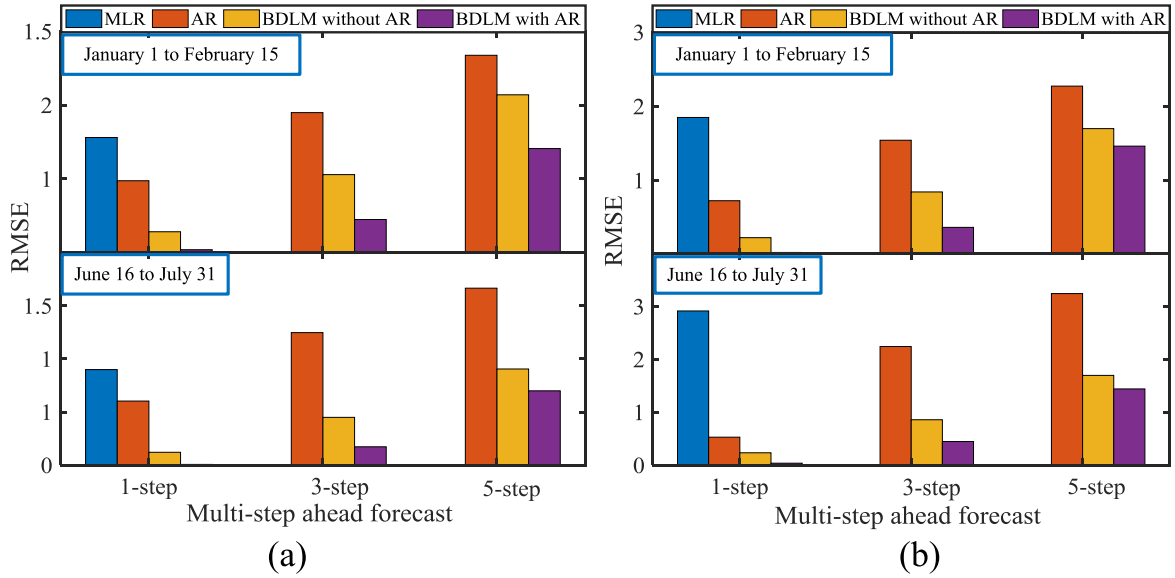

Fig. 14. Comparison of TIS by the AR model, MLR model, and BDLMs with and without the AR component: (a) YB 112 (2) YB 222. (Note that MLR model is only suitable for one-step-ahead forecast).

to the fact that the BDLM takes into account the relationship between the current and previous state of TIS as well as the correlation between TIS and temperature.

(4) The EM algorithm is able to estimate the parameters of BDLM simultaneously. The appropriate order of AR component in BDLM can be determined effectively using AIC and RMSE criterion. The forecasting performance of the BDLM with the suitable AR model order is satisfactory. Therefore, the improved method is suitable for temperature-based anomaly detection.

In the future, some work on using this BDLM-based approach combined with anomaly detection technique for either anomaly detection or damage detection will be delivered.

Acknowledgments

The authors gratefully acknowledge the support from the National Basic Research Program of China (973 Program) (Grant No. 2015CB060000), the National Natural Science Foundation of China (Grant Nos. 51722804 and 51878235), the National Ten Thousand Talent Program for Young Top-Notch Talents (Grant No. W03070080), the Jiangsu Key R & D Plan (Grant No. BE2018120), and the Jiangsu Transportation Scientific Research Project (Grant No. 8505001498). The support from the Jiangsu Health Monitoring Data Center for Long-Span Bridges is also acknowledged.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engstruct.2019.05.006.

References

[1] Ni YQ, Hua XG, Fan KQ, Ko JM. Correlating modal properties with temperature using long-term monitoring data and support vector machine technique. Eng Struct 2005;27:1762–73.
[2] Ou JP, Li H. Structural health monitoring in mainland China: review and future trends. Struct Health Monitor 2010;9:219–31.
[3] Xu YL. Making good use of structural health monitoring systems of long-span cable-supported bridges. J Civ Struct Health Monit 2018;8:477–97.
[4] Wang H, Tao TY, Gao YQ, Xu FY. Measurement of wind effects on a kilometer-level cable-stayed bridge during Typhoon Haikui. J Struct Eng 2018;144:04018142.
[5] Xu ZD, Zeng X, Li S. Damage detection strategy using strain-mode residual trends for long-span bridges. J Comput Civil Eng 2013;29:04014064.
[6] Ni YQ, Wang YW. Bayesian approach for characterizing wind-induced displacement responses of bridge using structural health monitoring data. Mechanics of structures and materials, XXIV. CRC Press; 2016. p. 1425–30.
[7] Spencer BF, Ruiz-Sandoval ME, Kurata N. Smart sensing technology: opportunities and challenges. Struct Control Health Monit 2004;11:349–68.
[8] Spencer BF, Jo H, Mechitov KA, Li J, Sim S-H, Kim RE, et al. Recent advances in wireless smart sensors for multi-scale monitoring and control of civil infrastructure. J Civ Struct Health Monit 2016;6:17–41.
[9] Mao JX, Wang H, Feng DM, Tao TY, Zheng WZ. Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition. Structural Control and Health Monitoring. 2018;25:e2146.
[10] Li J, Mechitov KA, Kim RE, Spencer Jr BF. Efficient time synchronization for structural health monitoring using wireless smart sensor networks. Struct Control Health Monit 2016;23:470–86.
[11] Posenato D, Kripakaran P, Inaudi D, Smith IF. Methodologies for model-free data interpretation of civil engineering structures. Comput Struct 2010;88:467–82.
[12] Brownjohn JM, De Stefano A, Xu Y-L, Wenzel H, Aktan AE. Vibration-based monitoring of civil infrastructure: challenges and successes. J Civ Struct Health Monit 2011;1:79–95.
[13] Alavi AH, Hasni H, Lajnef N, Chatti K, Faridazar F. Damage detection using self-powered wireless sensor data: an evolutionary approach. Measurement 2016;82:254–83.
[14] Wan HP, Ni YQ. Bayesian modeling approach for forecast of structural stress response using structural health monitoring data. J Struct Eng 2018;144:04018130.
[15] Catbas F, Kijewski-Correa T, Aktan AE. Structural identification (st-id) of constructed facilities: approaches, methods and technologies for effective practice of st-id. Am Soc Civ Eng 2011.
[16] Zhou C, Chase JG, Rodgers GW, Xu C. Comparing model-based adaptive LMS filters and a model-free hysteresis loop analysis method for structural health monitoring. Mech Syst Sig Process 2017;84:384–98.
[17] Moriot J, Quaegebeur N, Le Duff A, Masson P. A model-based approach for statistical assessment of detection and localization performance of guided wave–based imaging techniques. Struct Health Monit 2017. 1475921717744679.
[18] Wang H, Li AQ, Li J. Progressive finite element model calibration of a long-span suspension bridge based on ambient vibration and static measurements. Eng Struct 2010;32:2546–56.
[19] Wan HP, Ni YQ. Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. Struct Health Monit 2018. 1475921718794953.
[20] Xu B, He J, Masri SF. Data-based model-free hysteretic restoring force and mass identification for dynamic systems. Comput-Aided Civ Infrastruct Eng 2015;30:2–18.
[21] Neves A, González I, Leander J, Karoumi R. Structural health monitoring of bridges: a model-free ANN-based approach to damage detection. J Civ Struct Health Monit 2017;7:689–702.
[22] Catbas FN, Susoy M, Frangopol DM. Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data. Eng Struct 2008;30:2347–59.
[23] Xu Y, Chen B, Ng C, Wong K, Chan W. Monitoring temperature effect on a long suspension bridge. Struct Control Health Monit 2010;17:632–53.
[24] Brownjohn JM, Koo K-Y, Scullion A, List D. Operational deformations in long-span bridges. Struct Infrastruct Eng 2015;11:556–74.
[25] Zhou L, Xia Y, Brownjohn JM, Koo KY. Temperature analysis of a long-span suspension bridge based on field monitoring and numerical simulation. J Bridge Eng 2015;21:04015027.
[26] Murphy B, Yarnold M. Temperature-driven assessment of a cantilever truss bridge. Structures congress 2017. 2017. p. 461–73.
[27] Yarnold M, Moon F. Temperature-based structural health monitoring baseline for long-span bridges. Eng Struct 2015;86:157–67.
[28] Xia Q, Cheng Y, Zhang J, Zhu F. In-service condition assessment of a long-span suspension bridge using temperature-induced strain data. J Bridge Eng 2016;22. 04016124.
[29] Kromanis R, Kripakaran P. SHM of bridges: characterising thermal response and detecting anomaly events using a temperature-based measurement interpretation approach. J Civ Struct Health Monit 2016;6:237–54.
[30] Kromanis R, Kripakaran P. Predicting thermal response of bridges using regression models derived from measurement histories. Comput Struct 2014;136:64–77.
[31] Sohn H, Farrar CR. Damage diagnosis using time series analysis of vibration signals. Smart Mater Struct 2001;10:446.
[32] Figueiredo E, Figueiras J, Park G, Farrar CR, Worden K. Influence of the autoregressive model order on damage detection. Comput-Aided Civ Infrastruct Eng 2011;26:225–38.
[33] Fan XP. Bridge extreme stress prediction based on Bayesian dynamic linear models and non-uniform sampling. Struct Health Monit 2017;16:253–61.
[34] Wang YW. Bayesian-based methodology for progressive structural health evaluation and prediction by use of monitoring data. The Hong Kong Polytechnic University; 2017.
[35] Goulet JA. Bayesian dynamic linear models for structural health monitoring. Struct Control Health Monit 2017;24:e2035.
[36] Goulet JA, Koo K. Empirical validation of bayesian dynamic linear models in the context of structural health monitoring. J Bridge Eng 2017;23:05017017.
[37] Nguyen LH, Goulet JA. Anomaly detection with the Switching Kalman Filter for structural health monitoring. Struct Control Health Monit 2018;25:e2136.
[38] Nguyen LH, Goulet JA. Structural Health Monitoring with dependence on non-harmonic periodic hidden covariates. Eng Struct 2018;166:187–94.
[39] West M, Harrison J. Bayesian forecasting and dynamic models. Springer Science & Business Media; 2006.
[40] Petris G, Petrone S, Campagnoli P. Dynamic linear models. Dynamic linear models with R. Springer; 2009. p. 31–84.
[41] Lange M, Focken U. Physical approach to short-term wind power prediction. Springer; 2006.
[42] Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 1973;60:255–65.
[43] Harrell Jr. FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.
[44] Guo F, Wu O, Ding Y, Huang B. A data-based augmented model identification method for linear errors-in-variables systems based on EM algorithm. IEEE Trans Ind Electron 2017;64:8657–65.
[45] Akaike H. Factor analysis and AIC. Selected papers of Hirotugu Akaike. Springer; 1987. p. 371–86.
[46] VanBuren J, Cavanaugh J, Marshall T, Warren J, Levy SM. AIC identifies optimal representation of longitudinal dietary variables. J Public Health Dent 2017;77:360–71.
[47] Wang H, Tao TY, Li AQ, Zhang YF. Structural health monitoring system for Sutong cable-stayed bridge. Smart Struct Syst 2016;18:317–34.
[48] Ni YQ, Xia HW, Wong KY, Ko JM. In-service condition assessment of bridge deck using long-term monitoring data of strain response. J Bridge Eng 2011;17:876–85.
[49] Zhu Y, Ni YQ, Jesus A, Liu J, Laory I. Thermal strain extraction methodologies for bridge structural condition assessment. Smart Mater Struct 2018;27:105051.
[50] Mills TC. Signal extraction and two illustrations of the quantity theory. Am Econ Rev 1982;72:1162–8.

[51] Follen CW, Sanayei M, Brenner BR, Vogel RM. Statistical bridge signatures. J Bridge Eng 2014;19:04014022.
[52] Mao JX, Wang H, Li J. Fatigue reliability assessment of a long-span cable-stayed bridge based on one-year monitoring strain data. J Bridge Eng 2018;24:05018015.
[53] Moineddin R, Nie JX, Domb G, Leong AM, Upshur RE. Seasonality of primary care utilization for respiratory diseases in Ontario: a time-series analysis. BMC Health Services Res 2008;8:160.
[54] Marr LC, Harley RA. Spectral analysis of weekday–weekend differences in ambient ozone, nitrogen oxide, and non-methane hydrocarbon time series in California. Atmos Environ 2002;36:2327–35.
[55] Maitre H, Wu Y. A dynamic programming algorithm for elastic registration of distorted pictures based on autoregressive model. IEEE Trans Acoust Speech Signal Process 1989;37:288–97.
[56] Bos R, De Waele S, Broersen PM. Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. IEEE Trans Instrum Meas 2002;51:1289–94.