

SHM-based probabilistic representation of wind properties: statistical analysis and bivariate modeling

X.W. Ye^{*1}, L. Yuan^{1a}, P.S. Xi^{1b} and H. Liu^{2c}

¹Department of Civil Engineering, Zhejiang University, Hangzhou 310058, China

²China Railway Major Bridge (Nanjing) Bridge and Tunnel Inspect & Retrofit Co., Ltd., Nanjing 210061, China

(Received December 6, 2017, Revised March 10, 2018, Accepted March 25, 2018)

Abstract. The probabilistic characterization of wind field characteristics is a significant task for fatigue reliability assessment of long-span railway bridges in wind-prone regions. In consideration of the effect of wind direction, the stochastic properties of wind field should be represented by a bivariate statistical model of wind speed and direction. This paper presents the construction of the bivariate model of wind speed and direction at the site of a railway arch bridge by use of the long-term structural health monitoring (SHM) data. The wind characteristics are derived by analyzing the real-time wind monitoring data, such as the mean wind speed and direction, turbulence intensity, turbulence integral scale, and power spectral density. A sequential quadratic programming (SQP) algorithm-based finite mixture modeling method is proposed to formulate the joint distribution model of wind speed and direction. For the probability density function (PDF) of wind speed, a double-parameter Weibull distribution function is utilized, and a von Mises distribution function is applied to represent the PDF of wind direction. The SQP algorithm with multi-start points is used to estimate the parameters in the bivariate model, namely Weibull-von Mises mixture model. One-year wind monitoring data are selected to validate the effectiveness of the proposed modeling method. The optimal model is jointly evaluated by the Bayesian information criterion (BIC) and coefficient of determination, R^2 . The obtained results indicate that the proposed SQP algorithm-based finite mixture modeling method can effectively establish the bivariate model of wind speed and direction. The established bivariate model of wind speed and direction will facilitate the wind-induced fatigue reliability assessment of long-span bridges.

Keywords: structural health monitoring; wind monitoring data; joint probability distribution; probabilistic modeling; finite mixture distribution; mixture parameter estimation

1. Introduction

With the rapid progress of high-speed railways for stimulating economic growth, amounts of long-span railway bridges have been built in China (Wang *et al.* 2013). As typical flexible structures, the long-span bridges are characterized by low natural frequency and high flexibility and may experience large-amplitude structural vibrations under the action of wind loading. Also, a significant number of stress cycles will occur at the critical structural components under the normal wind conditions, which may lead to accumulated fatigue damage or even structural failure. Up to now, there are several accidents attributed to wind-induced fatigue (Robertson *et al.* 2001, Peil and Behrens 2002). To understand the mechanism of wind-induced fatigue problem, it is necessary to analyze the stochastic properties of wind field nearby the bridge during the periods of construction and operation.

In the past two decades, structural health monitoring (SHM) systems have been installed on more and more

large-scale bridge structures. They can provide an extremely large number of field monitoring data from different types of sensors for grasping the long-term service performance of the instrumented bridges (Ni *et al.* 2010, Ni *et al.* 2012, Ye *et al.* 2016, Ye *et al.* 2017, Ye *et al.* 2018a,b). Koo *et al.* (2013) described the multi-component instrumentation of Tamar Bridge and presented the corresponding performance observations based on the SHM system. Comanducci *et al.* (2015) analyzed the existence of damage in the bridge by using structural vibration and environmental condition data. Ding *et al.* (2013, 2015) presented the wind characteristic parameters based on the monitoring data and used a statistical method to evaluate the coherence function of these parameters.

In order to investigate the stochastic characteristics of wind field around the bridge site, the statistical methods are usually developed for the analysis of the wind monitoring data. In recent years, many researchers have employed the probability density function (PDF) models to describe the distribution of wind characteristics. Ye and Xiang (2011) analyzed the weather station observations for the development of the joint distribution of wind speed and direction near a bridge. Feng *et al.* (2015) developed a modeling method for establishment of joint distribution of wind speed and direction with the aid of wind monitoring data. Gu *et al.* (1999) established the joint PDF of wind speed and direction to estimate the fatigue life of the bridge girder. Many research efforts have been devoted to the

*Corresponding author, Associate Professor
E-mail: cexwe@zju.edu.cn

^a MSc Student

^b MSc Student

^c Chief Engineer

construction of continuous joint PDF models of wind speed and direction. For example, Weber (1991) used an isotropic Gaussian model to describe the joint PDF of wind speed and direction. Qu and Shi (2010) applied the Farlie-Gumbel-Morgenstern (FGM) approach to construct the bivariate joint distribution of wind speed and direction. Johnson and Wehrly (1978) proposed an angular-linear (AL)-based method to establish the joint distribution of bivariate random variables. Erdem and Shi (2011) modeled different bivariate joint distributions by the use of the AL, anisotropic lognormal, and FGM methods.

This paper presents the statistical analysis and bivariate modeling of wind characteristics based on the long-term wind monitoring data acquired by the anemometers installed on an arch railway bridge. One-year wind monitoring data measured by the ultrasonic anemometer are employed to analyze the wind characteristics, such as the mean wind speed and direction, turbulence intensity, turbulence integral scale, and power spectral density (PSD). A sequential quadratic programming (SQP) algorithm-based finite mixture modeling method is proposed to estimate the parameters in the joint PDF of wind speed and direction. For the PDF of wind speed, a double-parameter Weibull distribution function is used, and a von Mises distribution function is applied to represent the PDF of wind direction. The SQP algorithm with multi-start points is employed to estimate the parameters in the bivariate model, namely Weibull-von Mises mixture model. The optimal model is jointly judged by the Bayesian information criterion (BIC) and coefficient of determination, R^2 .

2. The investigated bridge and SHM system

2.1 Description of the investigated bridge

The investigated bridge, as illustrated in Fig. 1, is a long-span continuous steel truss arch bridge. The bridge has an overall length of 9,723 m with the main bridge length of 1,608 m. The main span of the bridge is a double continuous steel truss girder with the span arrangement of 108 m + 192 m + 336 m + 336 m + 192 m + 108 m. The location of the investigated bridge is shown in Fig. 2.



Fig. 1 The investigated bridge

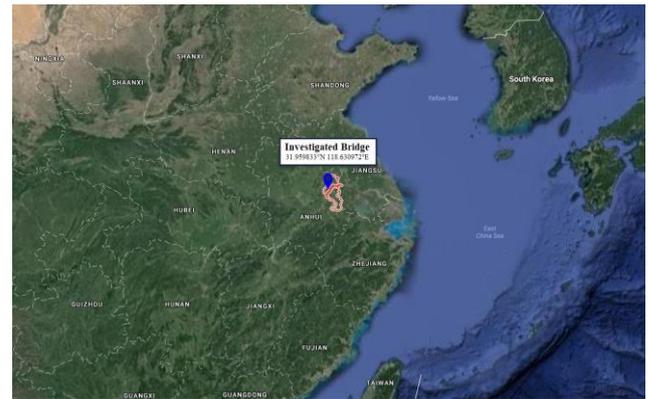


Fig. 2 Location of the investigated bridge

The bridge spans the Yangtze River and has six tracks including two lines for Beijing-Shanghai High-Speed Railway, two lines for Shanghai-Wuhan-Chengdu Railway and two lines for Nanjing Metro. The design speed of railway is set as 350 km/h. The construction of the bridge began in 2006 and completed in 2010 with the official operation in 2011. The bridge is recognized as the longest bridge with a continuous arch in the world.

2.2 SHM system and layout of anemometers

In recognition of the important role of the bridge, a long-term SHM system has been installed to monitor the integrity, durability and reliability of the bridge during the operation period. The SHM system includes more than 100 sensors to monitor the environmental condition, structural temperature, dynamic strain, structural vibration, structural displacement, and traffic condition. The sensors were permanently deployed on the bridge to continuously acquire the monitoring data for structural condition assessment and safety evaluation of the bridge.

For the environmental monitoring at the bridge site, two anemometers were installed at different locations on the bridge to continuously measure the wind speed and direction data. These two anemometers including one mechanical anemometer and one ultrasonic anemometer monitored wind data at different sections and altitudes of the bridge. The detailed installation locations of the anemometers are shown in Fig. 3(a). The section 1-1 represents the mid-span section of the northern main span and the section 2-2 is the southern arch foot of the southern main span. As shown in Figs. 3(b) and 3(c), the mechanical anemometer named as FS-11-1 and the corresponding anemoscope named as FX-11-1 were installed on the downstream vault of the north arch. The ultrasonic anemometer and the corresponding anemoscope, named as FS-17-1 and FX-17-1 were fixed on the steel truss between two high-speed railways. The sampling frequency of the mechanical anemometer is set as 1 Hz and the sampling frequency of the ultrasonic anemometer is set as 10 Hz.

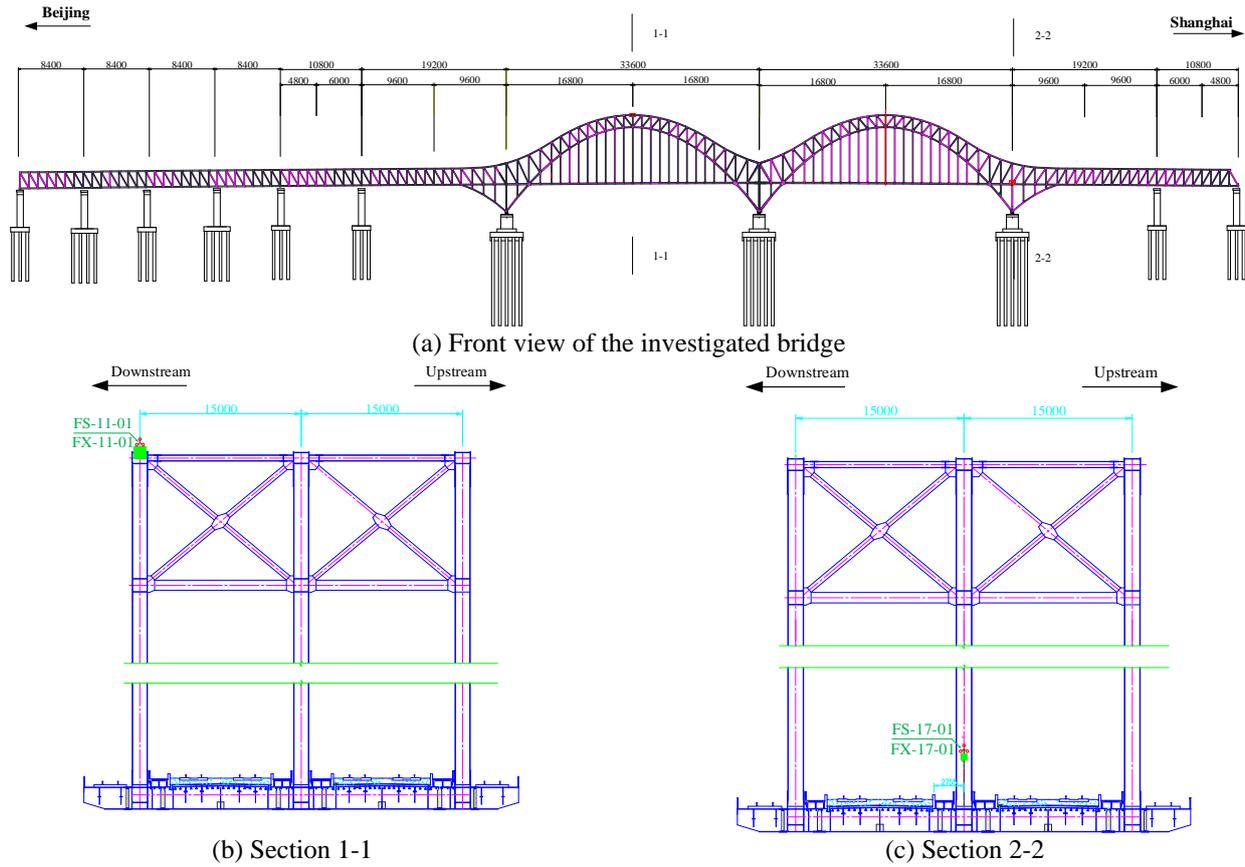


Fig. 3 layout of anemometers on the investigated bridge

3. Statistical characteristics of wind monitoring data

In this section, the wind monitoring data during 2015 are extracted for further analysis. Because there are several typhoons in 2015, the wind monitoring data measured by the high-frequency ultrasonic anemometer are selected to analyze the statistical characteristics of wind field. The average wind characteristics, i.e., mean wind speed and direction, and turbulence wind characteristics, i.e., turbulence intensity, turbulence integral scale and PSD are calculated based on the selected measurement data.

3.1 Mean wind speed and direction

For the ultrasonic anemometer, the measured wind data include the wind speed $U(t)$ and wind direction $\theta(t)$. The wind direction angle 0° denotes north and 90° denotes east, rotating in a clockwise direction. Thus, in the coordinate system, the y -axis is defined as the direction of north and x -axis is set to be the direction of east. The basic time interval T is 10 minutes according to the Chinese standard (Xiang *et al.* 2004). The wind monitoring data are decomposed into two components, i.e., U_x and U_y in the coordinate system within 10 minutes. After the orthogonal decomposition of recorded wind data, the wind speed in the two orthogonal directions for each set of wind data can be calculated. The mean values in the two directions, i.e., \bar{U}_x and \bar{U}_y , can

be computed by the sum of two axis components of wind data within 10 minutes, which are expressed as

$$\bar{U}_x = \frac{1}{T} \int_0^T U(t) \sin \theta(t) dt = \frac{1}{N} \sum_{i=1}^N U_x(i) \quad (1)$$

$$\bar{U}_y = \frac{1}{T} \int_0^T U(t) \cos \theta(t) dt = \frac{1}{N} \sum_{i=1}^N U_y(i) \quad (2)$$

where T is the time interval, and N is the number of wind data within the time interval.

The mean wind speed \bar{U} and mean wind direction $\bar{\theta}$ are defined as an average over the time interval T and can be calculated by

$$\bar{U} = \sqrt{\bar{u}_x^2 + \bar{u}_y^2} \quad (3)$$

$$\bar{\theta} = \arcsin \frac{\bar{u}_x}{\bar{U}} \quad (4)$$

According to the calculated mean wind direction $\bar{\theta}$, the wind monitoring data can be decomposed into the along-wind component $u(t)$ and across-wind component $v(t)$, which are parallel and perpendicular to the mean wind direction, as expressed by

$$u(t) = u_x(t)\cos\phi + u_y(t)\sin\phi - \bar{U} \quad (5)$$

$$v(t) = -u_x(t)\sin\phi + u_y(t)\cos\phi \quad (6)$$

The wind rose diagram of mean wind data at the bridge site during the period of one year is presented in Fig. 4(a). It can be found in Fig. 4(a) that the distribution of wind direction is uneven and dominated by northeastern direction and southwestern direction, which satisfies with the monsoon climate characteristics of the bridge site. In each wind direction, the low wind speed components hold a large proportion. Fig. 4(b) shows the extreme mean wind speed in each direction and indicates that the extreme wind speed varies with different directions. The maximum 10-min mean wind speed is 14.1 m/s in the direction of SWW. This direction is approximately perpendicular to the bridge and the wind load in this direction will cause the greatest impact on the bridge.

3.2 Turbulence intensity

The turbulence intensity reflects the turbulence characteristics of the wind monitoring data. The turbulence intensities in the along-wind and across-wind directions expressed as I_u and I_v are defined as

$$I_u = \frac{\sigma_u}{U} \quad (7)$$

$$I_v = \frac{\sigma_v}{U} \quad (8)$$

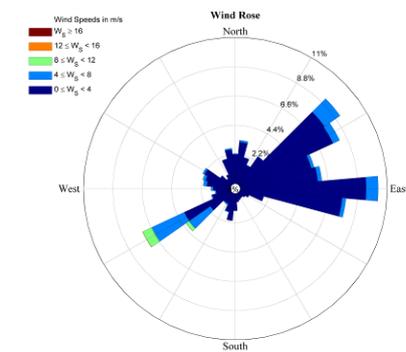
where σ_u and σ_v are the root mean square (RMS) values of the turbulence components in the along-wind and across-wind directions within 10 minutes, respectively, and U is the corresponding 10-min mean wind speed in the longitudinal direction.

Based on the long-term wind monitoring data, the turbulence intensities of 10-min mean wind speed are calculated by Eqs. (7) and (8). The relationships between the mean wind speeds and turbulence intensities in the along-wind and across-wind directions are shown in Fig. 5. It is obvious that the turbulence intensity decreases gradually with the increase of mean wind speed. The value of turbulence intensity ranges from 0 to 0.5 with the corresponding mean wind speed over 4 m/s. The mean value of turbulence intensities in the along-wind and across-wind directions are 0.5202 and 0.4238, respectively, and the ratio of mean turbulence intensities in two directions $I_u : I_v$ is 0.814 which is approximate to the mean turbulence intensities ratio 0.81 calculated by Chen *et al.* (2013).

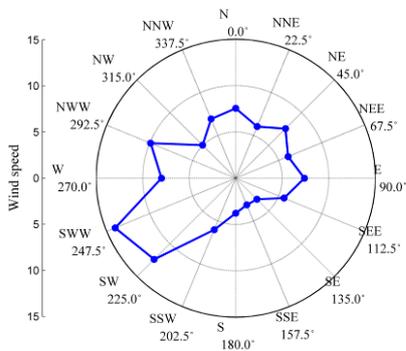
3.3 Power spectral density

For the continuous signals, the PSD describes the distribution of a signal or time series over frequency based on measured data. The PSD can be defined as

$$S_{xx}(\omega) = \lim_{T \rightarrow \infty} \mathbf{E}[|x_T(\omega)|^2] \quad (9)$$

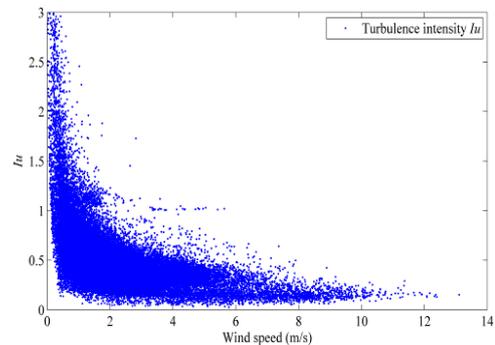


(a) Windrose diagram

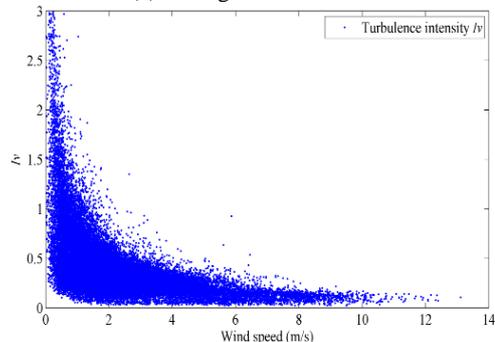


(b) Extreme wind speed diagram

Fig. 4 One-year mean wind speed and direction



(a) Along-wind direction



(b) Across-wind direction

Fig. 5 Relationship between turbulence intensity and mean wind speed

where $\mathbf{E}[\cdot]$ is the expectation function, and $x_T(\omega)$ is the truncated Fourier transform function as expressed by

$$x_T(\omega) = \frac{1}{\sqrt{T}} \int_0^T x(t) e^{-i\omega t} dt \quad (10)$$

The energy distribution of the fluctuating wind can be expressed in the formulation of PSD. In this section, several power spectra including Karman spectrum (Von Karman 1948), Kaimal spectrum (Kaimal *et al.* 1972), and Teunissen spectrum (Teunissen 1980) are adopted to compare the PSD of the wind monitoring data in the along-wind direction and across-wind direction. These spectra are defined as:

For Karman spectrum (along-wind direction)

$$\frac{nS_u(Z,n)}{(u^*)^2} = \frac{4\beta f}{(1+70.8f^2)^{5/6}} \quad (11)$$

$$f = nL_u^x/U \quad (12)$$

where S_u is the auto-PSD of along-wind turbulence, n is the natural frequency of the fluctuating wind, Z is the corresponding height of the fluctuating wind speed, U is the mean wind speed at the standard height, β is the coefficient of the friction wind speed, σ_u is the standard deviation of the fluctuating wind speed, L_u is the longitudinal turbulence integral length scale, and u^* is the friction wind speed which can be calculated by the energy unitary method (Simiu and Scanlan 1996), as expressed by

$$(u^*)^2 = \sigma_u^2 / \beta \quad (13)$$

For Kaimal spectrum (along-wind direction)

$$\frac{nS_u(Z,n)}{(u^*)^2} = \frac{200f}{(1+50f)^{5/3}} \quad (14)$$

$$f = nZ/U \quad (15)$$

For Teunissen spectrum (along-wind direction)

$$\frac{nS_u(Z,n)}{(u^*)^2} = \frac{105f}{(0.44+33f)^{5/3}} \quad (16)$$

$$f = nZ/U \quad (17)$$

For Karman spectrum (across-wind direction)

$$\frac{nS_v(Z,n)}{(u^*)^2} = \frac{4\beta f [1+755.2f^2]}{(1+283.2f^2)^{1/6}} \quad (18)$$

$$f = nL_u^x/U \quad (19)$$

where S_v is the auto-PSD of cross-wind turbulence.

For Kaimal spectrum (across-wind direction)

$$\frac{nS_v(Z,n)}{(u^*)^2} = \frac{3.36f}{1+10f^{5/3}} \quad (20)$$

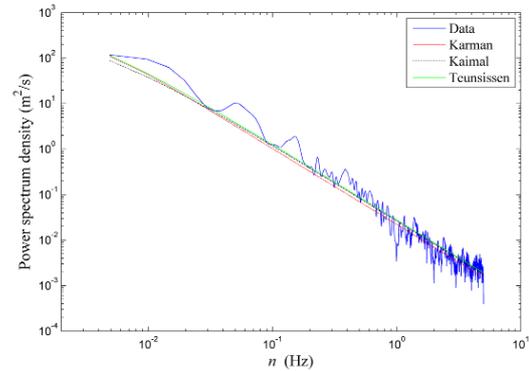
$$f = nZ/U \quad (21)$$

For Teunissen spectrum (across-wind direction)

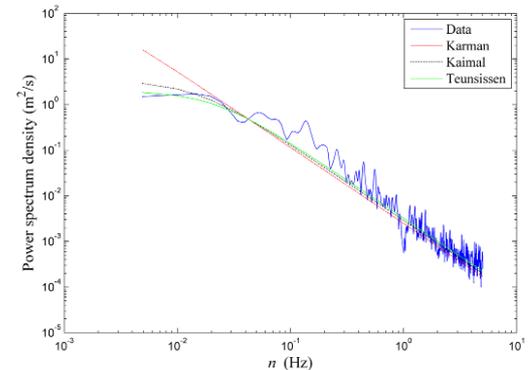
$$\frac{nS_v(Z,n)}{(u^*)^2} = \frac{2f}{1+5.3f^{5/3}} \quad (22)$$

$$f = nZ/U \quad (23)$$

The wind monitoring data measured during monsoon climate and typhoons are used to analyze the turbulence PSD in the along-wind and cross-wind directions. The number of discrete Fourier transform (DFT) points to be used in the PSD calculation is 2,048, and the Hamming window with 50 percent overlap is adopted to reduce the signal leakage in the frequency domain. The calculated PSD of the wind monitoring data is compared with three typical spectra. The comparative study for the monsoon monitoring data is shown in Fig. 6 and that for the typhoon monitoring data is shown in Fig. 7. Fig. 6 reveals that the measured PSD for the along-wind direction fits relatively well with three typical spectra.



(a) Along-wind direction



(b) Across-wind direction

Fig. 6 Wind power spectra of monsoon monitoring data

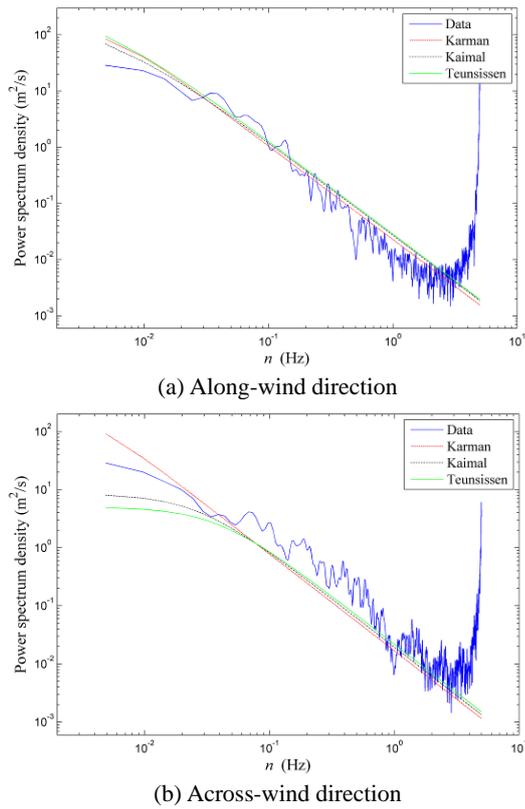


Fig. 7 Wind power spectra of typhoon data

For the PSD of the cross-wind direction, the measured SPD matches well with Kaimal spectrum and Teunissen spectrum, while Karman spectrum overestimates the turbulence PSD in the low-frequency range (0.005 Hz ~ 0.03 Hz). As shown in Fig. 7, the measured PSD of the typhoon monitoring data are totally different with that of the monsoon monitoring data. For both the along-wind and cross-wind directions, the measured PSD increases rapidly in the high-frequency region and the apparent gaps present in the high-frequency region (3 Hz ~ 5 Hz), although three typical spectra match well in the low-frequency and middle-frequency regions. It means that the turbulence wind has a great energy in the high-frequency region during the period of the typhoon.

4. Probabilistic modeling of wind speed and direction

4.1 Joint probability distribution model

The basic finite mixture model of independent scalar or vector observations \mathbf{y} can be expressed as (Richardson and Greem 1997)

$$f(\mathbf{y}; N, \Theta) = \sum_{l=1}^N w_l f(\mathbf{y}; \theta_l) \quad (24)$$

where $f(\mathbf{y}; N, \Theta)$ is the predictive mixture density function with estimated parameters Θ , and $f(\mathbf{y}; \theta_l)$ is the subfunction

family of all component densities which indexed by component parameters θ_l . The number of components in the family is N and w_l is the corresponding weights.

For the wind monitoring data, the observations include the wind speed variable and wind direction variable. It is assumed that the wind speed and direction follow the finite mixture density with conditionally independent component densities and the joint PDF of the bivariate finite mixture distribution of the wind speed and direction is expressed as

$$f(v, \theta; N, \Theta) = \sum_{i=1}^N w_i f_v(v; \alpha_i, \beta_i) f_\theta(\theta; \mu_i, \kappa_i) \quad (25)$$

where $f(v, \theta)$ is the joint PDF of the wind speed and direction, $f_v(v)$ is the PDF of the wind speed, and $f_\theta(\theta)$ is PDF of the wind direction. w_j is the weight of each mixture component and satisfies that the summation equals one as expressed as

$$\sum_{j=1}^N w_j = 1 \quad (26)$$

In the finite mixture distribution function, the two-parameter Weibull distribution is chosen as the wind speed distribution function $f_v(v)$, which is expressed by

$$f_v(v; \alpha_i, \beta_i) = \frac{\beta_i}{\alpha_i} \left(\frac{v}{\alpha_i}\right)^{\beta_i-1} \exp\left[-\left(\frac{v}{\alpha_i}\right)^{\beta_i}\right] \quad (27)$$

where α_i ($i=1, \dots, N$) represents the scale parameter, β_i ($i=1, \dots, N$) is the shape parameter. For the PDF of the wind direction $f_\theta(\theta)$, the von Mises distribution is employed which can be expressed by

$$f_\theta(\theta; \mu_i, \kappa_i) = \frac{1}{2\pi I_0(\kappa_i)} \exp[\kappa_i \cos(\theta - \mu_i)] \quad (28)$$

where κ_i ($i=1, \dots, N$) is the measure of the location, and μ_i ($i=1, \dots, N$) represents the concentration parameter. In Eq. (5), $I_0(\kappa_i)$ is the modified Bessel function of first kind and order zero and can be calculated by

$$I_0(\kappa_i) = \frac{1}{\pi} \int_0^\pi e^{\kappa_i \cos(x)} dx \quad (29)$$

4.2 Mixture parameter estimation

To estimate the parameters in the finite mixture model, the SQP method is used to estimate the parameters in the finite mixture model by minimizing the value of the objective function based on the function `fmincon` with the linear inequality constraint in Matlab. The process of mixture parameter estimation is the minimization of real-valued objective function f_{obj} in the parameter-space D defined by the constraints of each parameter, as expressed by

$$\begin{aligned} \Theta &= \min_{\Theta} f_{obj}(\Theta) \\ \text{s.t. } b(\Theta) &\geq 0 \quad (\text{i.e. } \mathbf{lb} \leq \Theta \leq \mathbf{ub}) \end{aligned} \quad (30)$$

where Θ is the estimated parameters, \mathbf{lb} is the vector of lower bounds, and \mathbf{ub} is the vector of upper bounds.

The SQP method is an iterative method for nonlinear optimization which transforms the complex optimization

problem into a sequence of optimization subproblems, and each of them optimizes a quadratic model of the objective subjected to the corresponding constraints. At the i th iteration, the SQP method defines an appropriate search direction d_i as the resolution to the quadratic programming subproblem, as expressed by

$$\begin{aligned} \min_{d_i} f(\Theta_i) + \nabla f(\Theta_i)^T d_i + \frac{1}{2} d_i^T \nabla^2 L(\Theta_i, \lambda_i) d_i \\ \text{s.t. } b(\Theta_i) + \nabla b(\Theta_i)^T d_i \geq 0 \end{aligned} \quad (31)$$

where λ is Lagrange multiplier, and L is the Lagrange function of the objective function

$$L(\Theta, \lambda, \sigma) = f(\Theta) - \lambda^T b(\Theta) \quad (32)$$

The joint distribution of the wind speed and direction exhibits the multimodal and two-dimensional characteristic. In this case, the objective function f_{obj} has a lot of local minima, and finding the global minimum is different. There are two main factors to determine the calculation speed and precision: the initial parameter values and the objective function. On one hand, the MultiStart algorithm is selected to eliminate the effect of the initial parameter values by using the multiple start point from the problem structure. On the other hand, the establishment of the fitness function is also crucial.

Assuming that we have wind monitoring data $\mathbf{y}=[\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_n]^T$ which is a two-dimensional and n size dataset, $\mathbf{y}_i=[y_{i1}, y_{i2}]$ as a two-dimensional vector including the wind speed variable y_{i1} and the wind direction variable y_{i2} , we select sections V_d ($d=1,2$) which just contain all data for each dimension and then divide the section into Π_{vd} equal intervals with the width s_{jd} , and v_d depicts the number of intervals for each dimension. For the two-dimensional wind monitoring data, the size of the j th area ($j_d=1,2,3, \dots, v_d$) is $S_j=[s_{j1}, s_{j2}]$ and the proportion of the j th area ξ_j is a rectangle with sides s_{j1} and s_{j2} , as $\xi_j=s_{j1} \times s_{j2}$. Using a histogram, the data are counted into the nonoverlapping and equally sized areas, and the quantity of the observations N_j falling into the j th area V_j is counted out. When $f(\mathbf{y}; \Theta)$ is the continuous PDF, the probability that the observation \mathbf{y} will fall inside the j th area is given by

$$\iint_{\xi_j} f(\mathbf{y}; \Theta) dx_1 dx_2 \approx f(\mathbf{y}_j; \Theta) \xi_j \quad (33)$$

where ξ_j represents the j th area, and \mathbf{y}_j is the center point of the j th area. The frequency of the data falling in the j th area p_j is given by

$$p_j = \frac{N_j}{N} \approx f(\mathbf{y}_j; \Theta) \xi_j \quad (34)$$

where N represents the total number of the data, and N_j is the number of the observation data which fall into the j th area.

The objective function measures the quality of the produced parameters and judges whether these parameters are optimal solutions. The closer the model and the measured data distribution, the smaller the gap between the frequency p_j and $f(\mathbf{y}; \Theta)$. In this study, we use the

following function as the objective function

$$f_{obj}(\Theta) = \frac{1}{\sum_{j=1}^v \left(\frac{p_j - f(\mathbf{y}_j; \Theta)}{f(\mathbf{y}_j; \Theta)} \xi_j \right)^2} \quad (35)$$

4.3 Optimal model selection

In this study, the BIC is employed to evaluate the quality of the predicted model of the wind monitoring data Schwarz (1978). When the predicted model fits well with the wind monitoring data, it can find that the likelihood function value increases with the number of parameters, while the unlimited increase may result in overfitting. This problem can be avoided by the penalty term in BIC. For the wind monitoring data, the data size n is much larger than the number of parameters k in the model. Thus, BIC is more appropriate for selecting an optimal model from the set of candidate models. The BIC value of the model is defined as

$$\text{BIC} = \ln(n)k - 2\ln(L) \quad (36)$$

where k is the number of unknown parameters to be estimated in the finite mixture distribution, L is the maximized value of the likelihood function of the model, and n is the number of data point. When determining the optimal model, the one with the lowest BIC is preferred.

In this study, the R^2 criterion is also used to determine the optimal model based on the proportion of the total variation between the expected and observed frequencies of the bins. It is obvious that the R^2 value is between zero and one, and the higher R^2 value is, the more likely the measured data match the distribution model. The R^2 is defined as

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (37)$$

where the term SS_{tot} denotes the total sum of squares which shows the proportional to the variance of the data. It is defined as the sum of the squared differences between the observed and average frequency of all bins. The term SS_{res} denotes the sum of squares of residuals which reflects the total discrepancy between the observed data and the estimation model, as expressed by

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (38)$$

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (39)$$

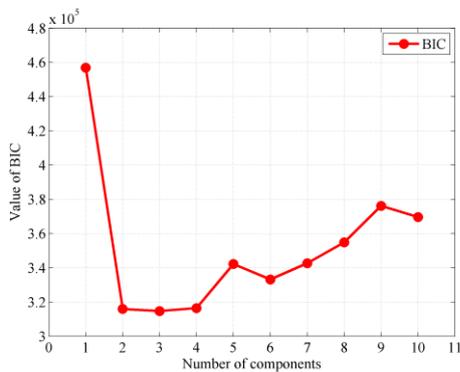
4.4 Application to the investigated bridge

The one-year wind monitoring data measured by SHM system of the investigated bridge are applied for the construction of the joint distribution of wind speed and direction based on the proposed SQP algorithm-based finite

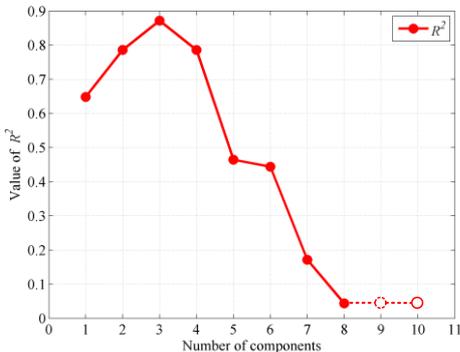
mixture modeling method. In consideration of several typhoons in the summer of 2015, the monitoring data measured by a high-frequency ultrasonic anemometer are selected for further analysis.

As presented above, the joint probability distribution model adopts the Weibull-von Mises finite mixture distribution model. Because the increase of the component number will lead to the radical increase of the number of parameters and the dimension of parameter domain, the maximum component number of the finite mixture distribution model is set to ten. The number of multiple start points within the parameter also needs to increase to eliminate the effect of the initial value of parameters. The number of multiple start points is set to change with the number of parameters linearly. After estimating the parameters of all models, the optimal joint probability distribution model is selected from the ten predicted models with the component number from one to ten. For each bivariate model, the BIC and R^2 are used to evaluate the fitness performance of each model and choose the optimal one.

In this study, the SQP method is chosen to estimate the parameters for the Weibull-von Mises finite mixture distribution model, i.e., $w_i, \alpha_i, \beta_i, \mu_i,$ and κ_i . According to the characteristics of Weibull distribution and von Mises distribution, the low bound vector of the corresponding estimated parameters is $[0,0,0,0,0]$ and the upper bound vector is $[1,10,20,2\pi,100]$. In consideration of the increase of parameter number and the constraint of computation time, the number of multiple start points is set as ten thousand times the number of parameters.

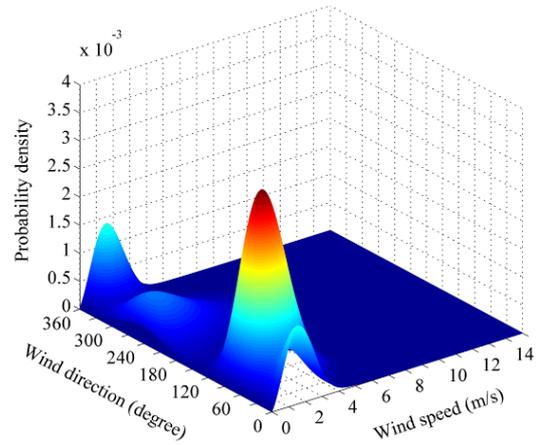


(a) BIC values of different component numbers

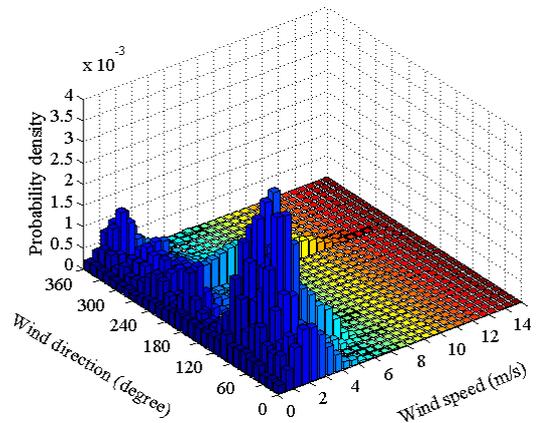


(b) R^2 values of different component numbers

Fig. 8 Wind power spectra of typhoon data



(a) Predicted joint distribution model



(b) Histogram of wind speed and direction

Fig. 9 Joint distribution model of wind speed and direction

To maximize the likelihood function value of the wind monitoring data, the wind speed and direction ranging from their minimum value and maximum value are divided into 30 bins with an equal width. Based on the probability of the wind monitoring data, the parameters of the proposed models with different components are estimated. Fig. 8 shows the values of BIC and R^2 with different numbers of components. It can be seen from Fig. 8(a) that the BIC value of one component model is large and gradually decreased when the number of components is less than four. The BIC values reach the minimum value at three components and begin to gradually increase when the number of components is more than five. Fig. 8(b) indicates that R^2 values have the same tendency. The R^2 values reach the maximum value at three components and then go down. What's more, the R^2 values of nine components and ten components are less than zero which means that the sum of squares of residuals is greater than the total sum of squares. This is because there are not enough initial start points to cover the parameter domain when the number of components is more than five. Limited start points converge to a local optimum without the global optimum. According to the values of BIC and R^2 , the proposed bivariate model with three components is regarded as the optimal joint

Table 1 Estimated parameters of the optimal model

Distribution	Parameter value		
Component (N)	1	2	3
Weight (w)	0.5966	0.2181	0.1853
Weibull distribution			
Scale parameter (α)	2.1519	2.5870	2.5485
Shape parameter (β)	1.6979	2.4204	2.4893
von Mises distribution			
Location parameter (μ)	5.8358	1.5896	0.9893
Concentration parameter (κ)	0.2910	33.9939	32.6738

probabilistic model of wind speed and direction, and the parameters of the selected optimal bivariate model are listed in Table 1. The joint probabilistic model of the wind speed and direction established based on the proposed modeling method is illustrated in Figs. 9(a), and 9(b) shows the histogram of the wind speed and direction. As shown in Fig. 9, the histogram of the wind speed and direction exhibits a multi-peak characteristic which has three main peaks. The predicted finite mixture model with three components including three peaks can fit well with the monitoring data of wind speed and direction.

5. Conclusions

This paper addressed the statistical analysis of wind characteristics nearby a bridge site by using the long-term wind monitoring data collected by the SHM system, and the bivariate modeling of the wind speed and direction based on the proposed SQP algorithm-based finite mixture modeling method. The wind characteristics were analyzed and presented based on one-year wind monitoring data. The calculated mean wind speed and direction were adopted to conduct the bivariate modeling of the wind speed and direction by using the proposed SQP algorithm-based finite mixture modeling method. The joint distribution model was formulated by the finite mixture model, in which a double-parameter Weibull distribution was selected to represent the wind speed distribution and a von Mises distribution is applied to represent the wind direction distribution. The SQP algorithm with multi-start points was employed to estimate the parameters in the proposed finite mixture model, namely Weibull-von Mises mixture model and the optimal model is jointly determined by the BIC and R^2 .

By statistical analysis and bivariate modeling of the wind monitoring data, the results show: (i) the probabilistic properties of the wind field at the bridge site can be effectively characterized by the long-term monitoring data measured by the SHM system; (ii) the proposed SQP algorithm-based finite mixture modeling method has good performance in bivariate modeling of the wind speed and direction; (iii) the predicted joint distribution model of the wind speed and direction can well reflect the multi-peak characteristic of the wind monitoring data. The achieved results from this study will facilitate the structural condition assessment and wind-induced fatigue damage estimation of

long-span railway bridges.

Acknowledgments

The work described in this paper was jointly supported by the National Key R&D Program of China (Grant No. 2017YFC0806100), the National Science Foundation of China (Grant No. 51778574), and the Fundamental Research Funds for the Central Universities of China (Grant No. 2017QNA4024).

References

- Chen, B., Yang, Q.S., Wang, K. and Wang, L.N. (2013), "Full-scale measurements of wind effects and modal parameter identification of Yingxian wooden tower", *Wind Struct.*, **17**(6), 609-627.
- Comanducci, G., Ubertini, F. and Materazzi, A.L. (2015), "Structural health monitoring of suspension bridges with features affected by changing wind speed", *J. Wind Eng. Ind. Aerod.*, **141**, 12-26.
- Ding, Y.L., Wang, G.X., Sun, P., Wu, L.Y. and Yue, Q. (2015), "Long-term structural health monitoring system for a high-speed railway bridge structure", *Sci. World J.*, 1-17.
- Ding, Y.L., Zhou, G.D., Li, A.Q. and Deng, Y. (2013), "Statistical characteristics of sustained wind environment for a long-span bridge based on long-term field measurement data", *Wind Struct.*, **17**(1), 43-68.
- Erdem, E. and Shi, J. (2011), "Comparison of bivariate distribution construction approaches for analysing wind speed and direction data", *Wind Energy*, **14**(1), 27-41.
- Feng, J., Shen, W.Z. and Sciubba, E. (2015), "Modelling wind for wind farm layout optimization using joint distribution of wind speed and wind direction", *Energies*, **8**(4), 3075-3092.
- Gu, M., Xu, Y.L., Chen, L.Z. and Xiang, H.F. (1999), "Fatigue life estimation of steel girder of Yangpu cable-stayed bridge due to buffeting", *J. Wind Eng. Ind. Aerod.*, **80**(3), 383-400.
- Johnson, R.A. and Wehrlyb, T.E. (1978), "Some angular-linear distributions and related regression models", *J. Am. Stat. Assoc.*, **73**(363), 602-606.
- Kaimal, J.C., Wyngaard, J., Izumi, Y. and Cote, O.R. (1972), "Spectral characteristics of surface - layer turbulence", *Q. J. Roy. Meteor. Soc.*, **98**(417), 563-589.
- Koo, K.Y., Brownjohn, J.M.W., List, D.I. and Cole, R. (2013), "Structural health monitoring of the Tamar suspension bridge", *Struct. Control Health.*, **20**(4), 609-625.
- Ni, Y.Q., Ye, X.W. and Ko, J.M. (2010), "Monitoring-based fatigue reliability assessment of steel bridges: analytical model and application", *J. Struct. Eng. - ASCE*, **136**(12), 1563-1573.
- Ni, Y.Q., Ye, X.W. and Ko, J.M. (2012), "Modeling of stress spectrum using long-term monitoring data and finite mixture distributions", *J. Eng. Mech. - ASCE*, **138**(2), 175-183.
- Peil, U. and Behrens, M. (2002), "Fatigue of tubular steel lighting columns under wind load", *Wind Struct.*, **5**(5), 463-478.
- Qu, X. and Shi, J. (2010), "Bivariate modeling of wind speed and air density distribution for long-term wind energy estimation", *Int. J. Green Energy*, **7**, 21-37.
- Richardson, S. and Green, P.J. (1997), "On Bayesian analysis of mixtures with an unknown number of components", *J. R. Stat. Soc. B.*, **59**(4), 731-792.
- Robertson, A.P., Hoxey, R.P., Short, J.L., Burgess, L.R., Smith, B.W. and Ko, R.H.Y. (2001), "Wind-induced fatigue loading of tubular steel lighting columns", *Wind Struct.*, **4**(2), 163-176.

- Schwarz, G. (1978), "Estimating the dimension of a model", *Ann. Stat.*, **6**(2), 461-464.
- Simiu, E. and Scanlan, R.H. (1996), *Wind Effects on Structures*, John Wiley and Sons, New York, USA.
- Teunissen, H.W. (1980), "Structure of mean winds and turbulence in the planetary boundary layer over rural terrain", *Bound.-Lay. Meteorol.*, **19**(2), 187-221.
- Von Karman, T. (1948), "Progress in the statistical theory of turbulence", *P. Natl. Acad. Sci.*, **34**(11), 530-539.
- Wang, J.J., Xu, J. and He, J. (2013), "Spatial impacts of high-speed railways in China: A total-travel-time approach", *Environ. Plann. A*, **45**(9), 2261-2280.
- Weber, R.O. (1991), "Estimator for the standard deviation of wind direction based on moments of the Cartesian components", *J. Appl. Meteorol.*, **30**(9), 1341-1353.
- Xiang, H.F., Bao, W.G., Chen, A.R., Lin, Z.X., and Liu, J.X. (2004), *Wind-Resistant Design Specification for Highway Bridges*, Ministry of Communications of the People's Republic of China, Beijing, China.
- Ye, X.W., Su, Y.H. and Xi, P.S. (2018a), "Statistical analysis of stress signals from bridge monitoring by FBG system", *Sensors*, **18**(2), 1-14.
- Ye, X.W., Su, Y.H., Xi, P.S., Chen, B. and Han, J.P. (2016), "Statistical analysis and probabilistic modeling of WIM monitoring data of an instrumented arch bridge", *Smart Struct. Syst.*, **17**(6), 1087-1105.
- Ye, X.W., Xi, P.S., Su, Y.H. and Chen, B. (2017), "Analysis and probabilistic modeling of wind characteristics of an arch bridge using structural health monitoring data during typhoons", *Struct. Eng. Mech.*, **63**(6), 809-824.
- Ye, X.W., Xi, P.S., Su, Y.H., Chen, B. and Han, J.P. (2018b), "Stochastic characterization of wind field characteristics of an arch bridge instrumented with structural health monitoring system", *Struct. Saf.*, **71**, 47-56.
- Ye, Z.W. and Xiang, Y.Q. (2011), "Bridge design basic wind speed based on the joint distribution of wind speed and direction", *Appl. Mech. Mater.*, **90-93**, 805-812.