



Probabilistic Regression Model for OMA-Based Damping Estimates of a Cable-Stayed Bridge Under Environmental and Operational Variability

Sunjoong Kim¹ (✉) , Doyun Hwang² , and Ho-Kyung Kim² 

¹ University of Seoul, Seoul 02592, South Korea
sunjoong@uos.ac.kr

² Seoul National University, Seoul 08826, South Korea

Abstract. Modal parameters are key factors in vibrational serviceability assessments of long-span cable-supported bridges. The recent development of automated operational modal analysis (OMA) has enabled modal tracking according to environmental and operational conditions (EOCs). However, in contrast to the successful implementations for natural frequencies, there are still challenges in the investigation of damping ratio due to (1) inherent errors in damping estimates and (2) epistemic uncertainties on the effects of EOCs. In this regard, this study proposes the framework to establish a reliable probabilistic regression model for the damping ratio of an actual cable-stayed bridge by incorporating various machine learning (ML) algorithms. First, a conventional automated OMA algorithm is improved by employing a displacement reconstruction algorithm and the optimized unsupervised clustering method. These approaches can reduce the model-order dependencies in OMA-based damping estimates with minimum user intervention. The proposed framework is then employed to estimate the long-term damping ratio from 2.5 years of structural health monitoring data. The monthly fluctuation and amplitude dependency of long-term damping characteristics are discussed. Subsequently, the deep Gaussian Process (DGP) model is applied to damping estimates for developing a probabilistic regression model. A statistic-based and knowledge-based data cleansing strategies are proposed to enhance the model's regression performance. A comparative study with different regression models validates the robustness of DGPs. Finally, the predictability of the trained model is validated using the actual monitoring datasets.

Keywords: Damping ratio · Automated Operational Modal Analysis · Cable-stayed bridge · DBSCAN · Deep Gaussian Process

1 Introduction

Automated operational modal analysis (AMA) has seen significant growth in recent years, with the technology being widely adopted for the condition assessment of large-scale civil infrastructures [1–6]. AMA algorithms aim to reduce user intervention by

automating various processes, including algorithmic parameter selection, spurious pole removal, and grouping physical poles into structural modes. This automated implementation allows for near real-time tracking of dynamic properties and the automatic detection of abnormal structural behaviors or damages [3–5].

Determining the damping ratio of large-scale structures is crucial for assessing their vibrational serviceability; however, many AMA studies have ignored identifying the damping ratio [4–6] due to the high complexity and uncertainty associated with damping estimates. In particular, the accuracy of AMA-based damping estimates is highly dependent on their parameter selection, and using automatically selected parameters can lead to increased uncertainty in the results. Thus, accurate damping estimation remains a significant challenge within the modal analysis community.

The clustering algorithm is also a significant issue in automated damping estimations. Even after applying a stabilization diagram, several spurious poles can remain within the modal estimate in long-term applications, and the accuracy of damping estimates is much more sensitive to the existence of spurious poles compared to the accuracy of natural frequency and/or mode shape estimates [7, 8]. In other words, the clustering process can lead to large variability in damping estimates even if it is successfully carried out based on the similarity of natural frequency and mode shape; thus, an appropriate technique is required to discard the spurious poles properly.

The uncertainty stemming from a lack of understanding of the underlying damping mechanisms is also a noteworthy challenge in modeling damping estimates. Indeed, this uncertainty is further compounded by ambiguous and sometimes conflicting relationships between environmental and operational conditions (EOCs) and damping, as reported in various studies [8–11]. Traditional deterministic regression approaches, which rely on a pre-determined set of parameters, are ill-suited to address the complexities and uncertainties involved in assessing the damping characteristics of large-scale structures.

In this regard, this study aims to enhance the accuracy of automated damping estimates and evaluate the damping characteristics of a cable-stayed bridge from long-term monitoring data. We employed a displacement reconstruction algorithm to minimize the required model order in AMA and DBSCAN for eliminating spurious poles. Based on these high-quality damping estimates, we further developed an enhanced regression model for the damping ratio using ML approaches. The training and testing data were established from 2.5 years of monitoring data. Data cleansing strategies based on statistical techniques and knowledge-based criteria were employed to ensure that only stable damping estimates were retained. Gaussian process (GP) models were used to derive the regression model. The impact of different data cleansing strategies and input features on model performance was examined to gain deeper insight into the damping mechanisms.

2 Methodology

2.1 Data Pre-processing: Displacement Reconstruction

Kim et al. [12] have demonstrated that the use of displacement reconstruction effectively suppresses high-mode components in the measured acceleration data, resulting in a more reliable and robust damping estimate for low-frequency modes. The displacement

reconstruction algorithm is defined by the minimization problem, as shown in Eq. (1).

$$\min_u \Pi(u) = \frac{1}{2} \int_{T_1}^{T_2} \left(\frac{d^2 u}{dt^2} - \bar{a} \right)^2 dt + \frac{(1 - \alpha_t)(2\pi f_t)^4}{2\alpha_t} \int_{T_1}^{T_2} u^2 dt \quad (1)$$

where u , \bar{a} , T_1 , and T_2 represent the displacement to be reconstructed, the measured acceleration, and the initial and final times of a time window, respectively. The regularization term, the second term in Eq. (1), can filter out low-frequency noise in the measured acceleration. For a more detailed description of this algorithm, readers are referred to previous studies [12–14].

2.2 Data Establishment: NExT-ERA-DBSCAN

2.2.1 NExT-ERA

The cross-correlation function between the measurement and reference response vectors can satisfy the homogeneous equation of motions under the assumption that the excitation is a stationary white-noise process. Accordingly, such correlation functions can be regarded as impulse response functions (IRFs) [15]. Then, the system matrices are calculated from the Hankel matrices composed of the IRF matrix. The Hankel matrix is subsequently decomposed using a singular value decomposition, allowing for the extraction of modal parameters from the system matrix \mathbf{A} [16]. Equations (2) and (3) demonstrate this process, where \mathbf{U} and \mathbf{V} are the singular vector matrices and \mathbf{S} is the diagonal matrix composed of the square root of the singular values of either $\mathbf{H}^T(0)\mathbf{H}(0)$ or $\mathbf{H}(0)\mathbf{H}^T(0)$.

$$\mathbf{H}(0) = \mathbf{H}\mathbf{S}\mathbf{V}^T \quad (2)$$

$$\mathbf{A} = \mathbf{S}^{-1/2}\mathbf{U}^T\mathbf{H}(1)\mathbf{U}\mathbf{S}^{1/2} \quad (3)$$

2.2.2 DBSCAN

Once the modal estimates are accumulated with the different system orders, all estimates should be separated appropriately into physical and spurious poles. To this end, we utilized a density-based clustering algorithm (DBSCAN) [17]. It does not require the specification of the number of clusters in advance, as well as the built-in concept of “noise” enables the elimination of outliers in the stabilization diagram as spurious poles within the clustering process. The overall steps of DBSCAN include the following tasks: (1) selecting and initializing an observation or point as the first cluster; (2) obtaining a set of points within a specified neighborhood search radius; (3) iterating through each neighbor until all points in the dataset are labeled; (4) merging clusters with varying densities that have border points with a distance less than the predefined parameter, ϵ . In this study, a nondimensional modal similarity distance measure between poles i and j is defined as Eq. (4).

$$d_{ij} = \alpha \frac{|f_i - f_j|}{\min(f_i, f_j)} + (1 - \alpha)(1 - \text{MAC}_{ij}) \quad (4)$$

In Eq. (4), α is a weighting value determining the importance of each modal parameter, $\min(\bullet)$ represents the minimum operator, f_i is the modal frequency of the i -th pole, and MAC_{ij} is the modal assurance criterion (MAC) value between the i -th and j -th poles.

2.3 Data Cleansing: Knowledge- and Statistics-Based Approaches

Data cleansing is a crucial step in the application of ML techniques as it ensures the integrity and accuracy of the data. In this study, outliers in damping estimates likely to include analytical errors are removed as an initial step. The outliers are detected based on the two statistical criteria: exceedance of 1.5 interquartile range (IQR) or three local scaled median absolute deviations (MAD) from a moving window with five points. This approach is referred to as a statistics-based strategy in this study.

Additionally, the system's input should ideally be a stationary white-noise process to ensure accurate estimation of modal parameters. Therefore, the presence of strong sinusoidal excitation may result in underestimating the structure's damping ratio. Therefore, damping estimates from vortex-induced vibration (VIV) conditions that introduce harmonic excitation must be discarded from the dataset. In that sense, the data point is discarded if any of the ten-minute mean wind speeds from the one-hour wind speed records were within the VIV wind speeds of 10.4 to 16 m/s. This approach is referred to as a VIV-based strategy in this study.

2.4 Probabilistic Regression: Deep GPR

Gaussian Process Regression (GPR) is a non-parametric method for modeling the relationship between inputs and outputs in a dataset [18]. It uses a probability distribution over functions, called a Gaussian process, to model the underlying function that generated the data. GPR makes predictions by assuming that the function's value at any point is a Gaussian random variable and that the function's values at different points are correlated. The correlation between function values at different points is determined by a kernel function, which can be chosen based on prior knowledge about the function or learned from data. The deep Gaussian Process (DGP) is a hierarchical extension of GPR that allows for the modeling of complex and high-dimensional data by stacking multiple GPRs (see Fig. 1) [19]. Each GPR in the stack models a different level of abstraction in the data, with the output of one GPR serving as input for the next one. This configuration allows for the modeling of multi-scale and non-linear relationships in the data. Also, it can overcome the limited expressiveness of kernel functions of a single GPR.

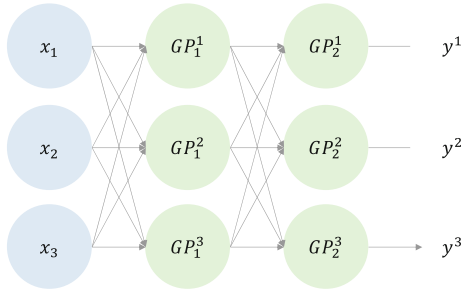


Fig. 1. The illustration of a two-layer DGP

3 Data

The target bridge in this study is a twin cable-stayed bridge located in the southern part of the Korean Peninsula (see Fig. 1). This bridge comprises two adjacent bridge structures, with a mid-span of 344 m and two side spans of 73.79 m. The first of these bridges (referred to as Bridge 1), the target bridge in this study, was built in 1984; the second bridge (referred to as Bridge 2) was built in 2005 to accommodate the increased traffic flow. In 2011, a significant vortex-induced vibration (VIV) occurred at Bridge 2, with a 10-min average wind velocity ranging from 9.8 to 11.5 m/s. The low damping ratio of Bridge 2 was found to be the reason for this unexpected vibration [20, 21].

The bridge is equipped with a wired SHM system consisting of 83 sensors, such as accelerometers, GNSS, strain gauges, ultrasonic anemometers, thermometers, and tiltmeters. Such data is collected at a sampling rate of 100 Hz. From the accumulated dataset, we utilized long-term monitoring data from January 1st, 2016, to August 31st, 2018, in this study. Note that the limited vortex-induced vibrations (VIV) were observed within the wind velocity range of 12 to 16 m/s, consistent with the results from a previous wind tunnel test [21].



Fig. 2. Target bridge

4 Results

4.1 Damping Estimation

The proposed OMA-based automated damping estimation algorithm was employed for the 2.5-year monitoring data of Bridge 1. The fundamental vertical mode was investigated in this study. The optimal parameters derived from previous studies for the NeXT-ERA and DBSCAN were used [20, 22], as listed in Table 1. The percentage of successful modal parameter identification from the total number of datasets was almost perfect, 99.99%, highlighting the robust nature of the automated clustering process with the optimal parameters with minimum user intervention.

Table 1. Algorithmic parameters for NeXT-ERA and DBSCAN

Parameter	Value
Target frequency for displacement reconstruction	0.97 Hz
Filter size for displacement reconstruction	Long filter
Number of FFT for NeXT	2^{15}
Overlapping for NeXT	2^{14} (50%)
Model order range for ERA	1 to 20
Epsilon for DBSCAN	0.1
Minimum points for DBSCAN	6

4.1.1 Monthly Variation

Figure 3 depicts the fluctuations of the modal damping ratio during the monitoring period. It can be characterized by a gradual decrease from winter to summer and an increase from summer to winter. Figure 4(b) confirms this monthly variation in damping estimates. This finding aligns with previous research on cable-stayed bridges, which showed that modal damping ratios typically exhibit a negative correlation with temperature.

4.1.2 Amplitude Dependency

As shown in Fig. 4(a), damping estimates increase during daytime hours and decrease at night. The bivariate histograms of the RMS acceleration and damping estimates (See Fig. 4(b)) indicates a clear correlation between the vibration amplitude and the damping ratio. The damping ratio increased proportionally from 0.5 to 1.2%, with the RMS acceleration from 0.00 to 0.05 m/s².

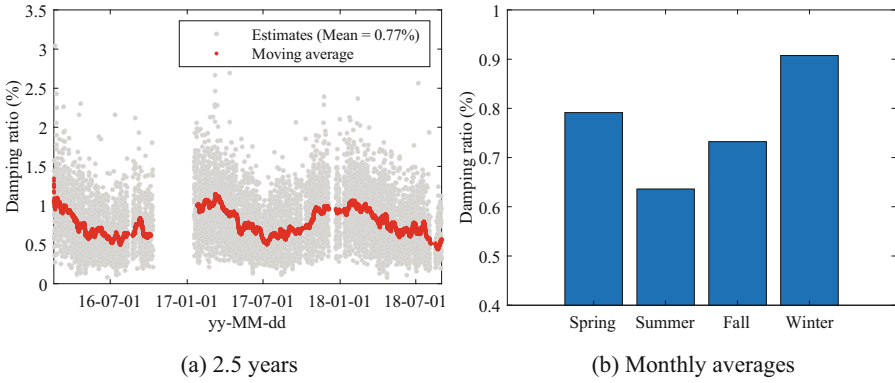


Fig. 3. Fluctuation of damping estimates

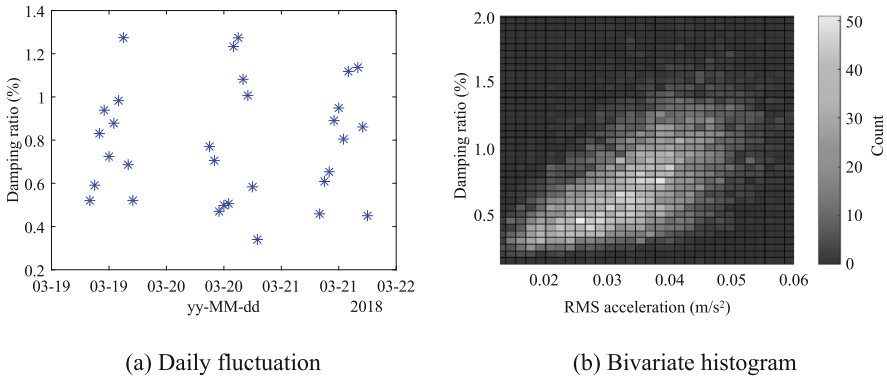


Fig. 4. Daily fluctuation of damping estimates

4.1.3 GPR Modeling

In this study, the input features for estimating the damping ratio were selected as follows. First, the RMS of acceleration and reconstructed displacement, as well as their power spectral density (PSD) amplitudes at the natural frequency were utilized as measures of the excitation level. In addition, one-hour mean wind speed and direction were considered in terms of wind conditions. Temperature data were acquired from the closest automatic weather stations. The established database of damping ratio and corresponding EOC was then processed by the two data-cleansing strategies described in Sect. 2.3. The details of the three datasets used in this study are presented in Table 2.

4.1.4 Model Performance

We conducted a comparative study between several regression models. A total of three DGP models with 2, 3, and 4 layers were constructed, referred to as DGP2, DGP3, and DGP4, respectively. We additionally employed a multivariate Linear Regression

Table 2. Dataset according to the data cleansing strategies

Dataset name	Descriptions
D1	None of data cleansing strategies is applied
D2	Statistics-based cleansing strategy is applied
D3	Statistics-based and VIV-based cleansing strategies are applied

Model (LM) and Generalized Linear Model (GLM) as a comparison group. The model performance is evaluated by 10-fold cross-validation with DS3.

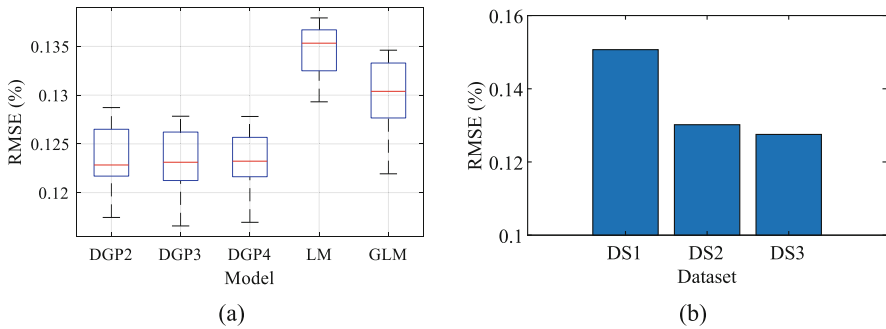
**Fig. 5.** Model performance according to (a) model and (b) data cleansing strategies (TBU)

Figure 5(a) shows the box plots of RMSE from each model, indicating the superior performance of the DGPs compared to the classical regression models. The GP models had not only low RMSE but also a little variance, highlighting their stability and robustness. Among the GP models, DGP2 exhibited slightly better performance compared to others. In addition, Fig. 5(b) illustrates the impact of data cleansing through a comparison of the model performance of the DGP2 using datasets DS1, DS2, and DS3. The 80/20 holdout method was applied, and the RMSEs of each test dataset were calculated. The cleaned datasets, DS2 and DS3, showed improved performance compared to the raw dataset, DS1. The comparison between DS2 and DS3 also confirms that the aerodynamic effects from interference VIVs can lead to uncertainties in the damping estimates and corresponding modeling errors.

4.1.5 Prediction Performance

The final regression model was established using DGP2 on DS3. Figure 6(a) presents the prediction results of the DGP2 model trained on dataset DS3. The model exhibits a high level of accuracy, as demonstrated by the RMSE value of 0.123 and adjusted R2 value of 84.7% (See Fig. 6(b)), indicating a favorable regression interval. The results demonstrate the model's capability to effectively capture the dynamic fluctuations of the damping ratios, with a substantial majority of 94% of the test data estimates falling within the 95% confidence interval.

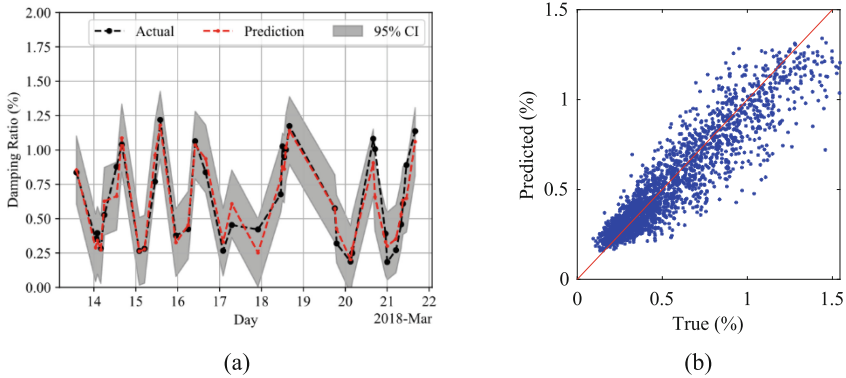


Fig. 6. Predicted damping ratio from DGP2: (a) one-week example and (b) scatter diagram

5 Conclusions

This study proposed a probabilistic regression model for predicting the damping characteristics of a cable-stayed bridge from environmental and operational conditions. A comprehensive evaluation of the long-term damping characteristics of the bridge was conducted through the application of the proposed automated damping estimation framework to the 2.5-year monitoring data. The results exhibited a consistent trend in damping estimates even under varying environmental and operational conditions by minimizing analytical errors and excluding aerodynamic uncertainties. It is highlighted that the amplitude dependency of the estimated damping ratio and the monthly trends in estimated damping ratios having a negative correlation with temperature.

The proposed Deep Gaussian Process (DGP) model demonstrated successful prediction of the variability in damping with good accuracy, as 94% of the damping estimates were located within the 95% confidence interval of the model prediction. A comparative study with other regression techniques confirmed the superiority of the DGP models in terms of accuracy, stability, and robustness across different datasets. Data cleansing was carried out effectively by incorporating the knowledge gained from Operational Modal Analysis (OMA) and interactive vortex-induced vibration (VIV), thereby improving the performance of the regression model.

Acknowledgments. This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) [grant numbers RS-2023-00213436]. The photo in Fig. 2 was provided by the Jindo County Office and used in this manuscript according to Korea Open Government License (KOGL).

References

1. Rainieri, C., Fabbrocino, G.: Automated output-only dynamic identification of civil engineering structures. *Mech. Syst. Signal Process.* **24**(3), 678–695 (2010)
2. Reynders, E., Houbrechts, J., De Roeck, G.: Fully automated (operational) modal analysis. *Mech. Syst. Signal Process.* **29**, 228–250 (2012)

3. Peeters, B., De Roeck, G.: One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Eng. Struct. Dynam.* **30**(2), 149–171 (2001)
4. Magalhães, F., Cunha, Á., Caetano, E.: Vibration based structural health monitoring of an arch bridge: from automated OMA to damage detection. *Mech. Syst. Signal Process.* **28**, 212–228 (2012)
5. Lorenzo, E.D., Petrone, G., Manzato, S., Peeters, B., Desmet, W., Marulo, F.: Damage detection in wind turbine blades by using operational modal analysis. *Struct. Health Monit.* **15**(3), 289–301 (2016)
6. Soyoz, S., Feng, M.Q.: Long-term monitoring and identification of bridge structural parameters. *Comput. Aid. Civil Infrastruct. Eng.* **24**(2), 82–92 (2009)
7. Kim, S., Kim, H.K., Hwang, Y.C.: Enhanced damping estimation for cable-stayed bridges based on operational monitoring data. *Struct. Eng. Int.* **28**(3), 308–317 (2018)
8. Kim, S., Kim, H.-K., Spencer, B.F.: Automated damping identification of long-span bridge using long-term wireless monitoring data with multiple sensor faults. *J. Civ. Struct. Heal. Monit.* **12**(2), 465–479 (2022). <https://doi.org/10.1007/s13349-022-00556-6>
9. Asadollahi, P., Li, J.: Statistical analysis of modal properties of a cable-stayed bridge through long-term wireless structural health monitoring. *J. Bridg. Eng.* **22**(9), 04017051 (2017)
10. Moser, P., Moaveni, B.: Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mech. Syst. Signal Process.* **25**(7), 2336–2357 (2011)
11. Xia, Y., Hao, H., Zanardo, G., Deeks, A.: Long term vibration monitoring of an RC slab: Temperature and humidity effect. *Eng. Struct.* **28**(3), 441–452 (2006). <https://doi.org/10.1016/j.engstruct.2005.09.001>
12. Kim, S., Park K.Y., Kim H.K., Lee H.S.: Damping estimates from reconstructed displacement for low-frequency dominant structures. *Mech. Syst. Signal Proces.* **136**, 106533 (2020)
13. Lee, H.S., Hong, Y.H., Park, H.W.: Design of an FIR filter for the displacement reconstruction using measured acceleration in low-frequency dominant structures. *Int. J. Numer. Meth. Eng.* **82**(4), 403–434 (2010)
14. Hong, Y.H., Kim, H.K., Lee, H.S.: Reconstruction of dynamic displacement and velocity from measured accelerations using the variational statement of an inverse problem. *J. Sound Vib.* **329**(23), 4980–5003 (2010)
15. James, G.H., Carne, T.G., Lauffer, J.P.: The natural excitation technique (NExT) for modal parameter extraction from operating wind turbines. Sandia National Labs., Albuquerque, New Mexico, USA (1993)
16. Juang, J.N., Pappa, R.S.: An eigensystem realization algorithm for modal parameter identification and model reduction. *J. Guid. Control. Dyn.* **8**(5), 620–627 (1985)
17. Ester M., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. pp. 226–231, AIAA, Portland, Oregon, USA (1996)
18. Williams, C.K., Rasmussen, C.E.: *Gaussian Processes for Machine Learning*, vol. 2. MIT Press, Cambridge (2006)
19. Damianou, A., Lawrence, N.D.: Deep Gaussian Processes. In: Carvalho, C.M., Ravikumar, P. (eds.) *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research*, vol. 31, pp. 207–215. PMLR, Scottsdale, Arizona, USA (2013)
20. Kim S.J., Kim H.K., Calmer R., Park J., Kim G.S., Lee D.K.: Operational field monitoring of interactive vortex-induced vibrations between two parallel cable-stayed bridges. *J. Wind Eng. Indust. Aerodyn.* **123** (Part A), 143–154 (2013)
21. Seo, J.W., Kim, H.K., Park, J., Kim, K.T., Kim, G.N.: Interference effect on vortex-induced vibration in a parallel twin cable-stayed bridge. *J. Wind Eng. Ind. Aerodyn.* **116**, 7–20 (2013)
22. Hwang, D., Kim, S., Kim, H.K.: Long-term damping characteristics of twin cable-stayed bridge under environmental and operational variations. *J. Bridg. Eng.* **26**(9), 04021062 (2021)