sustainability

MDPI

Article

Steady-State Data Baseline Model for Nonstationary Monitoring Data of Urban Girder Bridges

Shaoyi Zhang [1,2,*], Yongliang Wang [3] and Kaiping Yu [2]

[1] School of Transportation Science and Engineering, Harbin Institute of Technology, 73 Huanghe Road, Harbin 150090, China
[2] Department of Astronautic Science and Mechanics, Harbin Institute of Technology, No. 92 West Dazhi Street, Harbin 150001, China
[3] Jinan Huanghe Luqiao Construction Group Co., Ltd., 5111 Aoti Middle Road, Jinan 250000, China
* Correspondence: zhangshaoyi@hit.edu.cn

Abstract: In bridge structural health monitoring systems, an accurate baseline model is particularly important for identifying subsequent structural damage. Environmental and operational loads cause nonstationarity in the strain monitoring data of urban girder bridges. Such nonstationary monitoring data can mask damage and reduce the accuracy of the established baseline model. To address this problem, a steady-state data baseline model for bridges is proposed. First, for observable effects such as ambient temperature, a directional projection decoupling method for strain monitoring data is proposed, which can reduce the nonstationary effect of ambient temperature, and the effectiveness of this method is proven using equations. Second, for unobservable effects such as traffic load, a k-means clustering method for steady state of traffic loads is proposed; using this method, which can divide the steady and nonsteady states of traffic loads and reduce the nonstationary effect of traffic loads on strain monitoring data, a steady-state baseline model is established. Finally, the effectiveness of the steady-state baseline model is verified using an actual bridge. The results show that the proposed baseline model can reduce the error caused by nonstationary effects, improve the modelling accuracy, and provide useful information for subsequent damage identification.

Keywords: structural health monitoring; baseline model; nonstationary environmental effects; principal component analysis; cluster analysis

check for updates

Citation: Zhang, S.; Wang, Y.; Yu, K. Steady-State Data Baseline Model for Nonstationary Monitoring Data of Urban Girder Bridges. *Sustainability* **2022**, *14*, 12134. https://doi.org/10.3390/su141912134

Academic Editors: Yang Liu, Hongye Gou and Wanshui Han

Received: 28 August 2022
Accepted: 21 September 2022
Published: 25 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Structural health monitoring (SHM) systems [1–5] can collect structural response monitoring data from bridges in real time online and accurately diagnose the operation status of a bridge structure through in-depth analysis of the monitoring data; this approach has become an effective means to ensure the safe operation of bridge structures [6,7]. In damage diagnosis, a baseline model is usually compared with a test model, so an accurate baseline model is particularly important [8,9]. According to the model category, baseline models can be divided into finite element (FE) or data baseline models. FE baseline models [10–12] require a complex finite element model (FEM) of a bridge and then perform model updating through many iterative operations to obtain the real modal parameters of the bridge. This process is very time-consuming. Due to the high computational cost of FE baseline models, they cannot meet the requirements of online real-time monitoring systems [13]. Different from FE baseline models, data baseline models [14,15] directly use the structural response data or extracted damage features and statistical methods to identify damage. Data baseline models do not require a complex FEM, so they have a high computational efficiency and are widely used in online real-time bridge monitoring systems [16,17].

However, operating bridge structures are inevitably affected by environmental factors, and the monitoring data obtained are nonstationary [18]. Mosavi et al. [19] found that when the top flange temperature increased by 8.1 °C in one day, the frequency of the first

five orders of magnitude of a two-span steel—concrete composite bridge in North Carolina changed by 0.9% to 2.18%. Catbas et al. [20] found that when the ambient temperature increased by 41.6 °C, the bending strains of the hanger bar of the Commodore Barry Bridge in the United States changed by approximately 400 microstrain, and the maximum traffic load in the year caused the bending strain of the hanger bar to change by 40 microstrain. Peeters et al. [21] found that the frequency of the first four orders of the Z24 bridge changed by approximately 14% to 16% in a 10-month period. When the Z24 bridge suffered major damage, the change in the first four orders of magnitude was still less than 10%. The influence of environmental factors on the monitoring response data of bridges cannot be ignored, and the environmental effects usually mask changes in monitoring data caused by damage [22,23], which brings great challenges to the diagnosis and evaluation of the service performance of bridges [24,25]. Therefore, if the environmental effects cannot be reduced, accurately diagnosing the state of a bridge based on a large amount of monitoring data is difficult.

Two main data-driven solutions, supervised and unsupervised solutions, are used to assess the influence of environmental factors. Supervised methods [26] develop a relationship model between the environmental factors and the damage features of a bridge to reduce the environmental effects. However, accurately establishing a relationship model is often difficult. Unsupervised methods [27–30] consider environmental factors as latent variables and use projection or variance analysis to determine damage features that are not sensitive to environmental factors to reduce the environmental effects. However, these algorithms lose some information in the process of projective transformation of damage features. If damage information is lost, the effectiveness of the algorithm decreases. Therefore, the effect of these two methods alone is not ideal.

Compared with the traffic load of existing bridges, the traffic load of new bridges is not stable at the beginning of operation, but it gradually becomes stable as the operation time increases. Therefore, the monitoring data collected by SMH systems contain both nonsteady-state data and steady-state data. To address this problem, a steady-state data baseline model is proposed. For observable nonstationary effects such as the ambient temperature, a directional projection decoupling method for strain monitoring data with ambient temperature effects is developed. For unobservable nonstationary effects such as traffic loads, a clustering method of nonsteady- and steady-state traffic loads is studied. The establishment of a high-accuracy data baseline model for urban girder bridges is realized.

2. Theory of Steady-State Data Baseline Model for Nonstationary Monitoring Data

2.1. The Directional Projection Decoupling Method for Strain Monitoring Data under Ambient Temperature

Assuming that a bridge structure is not damaged in the reference state, the nonstationary change in the bridge strain is mainly influenced by coupling effects such as the ambient temperature and traffic load. According to whether the nonstationary effect can be obtained through sensor observation, these nonstationary effects can be divided into observable nonstationary effects (such as the ambient temperature) and unobservable nonstationary effects (such as the traffic load). The strain ε can be obtained using the following equation:

$$\varepsilon = \varepsilon^{O} + \varepsilon^{N} \quad (1)$$

where ε^{O} is the strain caused by observable nonstationary effects, and ε^{N} is the strain caused by unobservable nonstationary effects.

The observable nonstationary effects can be obtained through sensor observation, such as the ambient temperature load, which can be obtained using temperature sensors. Through principal component analysis (PCA) directional projection, the observable and unobservable nonstationary effects in the strain data of a bridge can be separated. The strain data processed by the directional projection can eliminate the influence of the observable nonstationary effects, which is the strain caused by the unobservable nonstationary effects. The proposed PCA directional projection decoupling method is different from PCA

projection. The first difference is that PCA directional projection uses different types of data, such as strain data and ambient temperature data, to construct a covariance matrix. The second difference is that PCA directional projection can artificially control the projection direction by giving weights, and the obtained projection data are the direction of the strain data. For observable nonstationary effects, taking the ambient temperature as an example, a directional projection method is proposed to weaken the influence of the observable nonstationary effects. The specific description of the proposed method is as follows.

The strain monitoring dataset collected by the bridge SHM system is assumed to be a matrix $\mathbf{X}$. For n strain measurement points in the monitoring system and m sampling points over time for each measurement point, the strain monitoring dataset $\mathbf{X} \in \mathbb{R}^{m\times n}$ is defined as follows:

$$\mathbf{X} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_i, \cdots, \boldsymbol{\omega}_n] \tag{2}$$

where i is an arbitrary strain measurement point, $i \in (1,\ 2,\ \cdots,\ n)$, and $\boldsymbol{\omega}_i$ is the strain monitoring dataset at the ith measurement point, $\boldsymbol{\omega}_i \in \mathbb{R}^{m\times 1}$, which is defined as follows:

$$\boldsymbol{\omega}_i = \{X_{1i}, X_{2i}, \cdots, X_{ki}, \cdots, X_{mi}\}^{\mathrm{T}} \tag{3}$$

In the equation, $(\cdot)^{\mathrm{T}}$ represents the transpose of the matrix; k is an arbitrary sampling point at time $k \in (1,\ 2,\ \cdots,\ m)$; and X_{ki} is the strain monitoring data at the kth sampling point of the ith measurement point.

Using the calculation steps from Equation (1) to Equation (11) in reference [31], the denoised strain monitoring dataset at the ith measurement point is obtained as $\boldsymbol{\omega}'_i$ ($\boldsymbol{\omega}'_i \in \mathbb{R}^{m\times 1}$). The set $\boldsymbol{\varepsilon}_i$ is obtained by centralization of $\boldsymbol{\omega}'_i$, ($\boldsymbol{\varepsilon}_i \in \mathbb{R}^{m\times 1}$), which is defined as follows:

$$\boldsymbol{\varepsilon}_i = \boldsymbol{\omega}'_i - \overline{\boldsymbol{\omega}}'_i \tag{4}$$

where $\boldsymbol{\varepsilon}_i$ is the strain at the ith measurement point after centralization; $\boldsymbol{\omega}'_i$ is the reconstruction matrix of $\boldsymbol{\omega}_i$ after denoising; $\overline{\boldsymbol{\omega}}'_i$ is the mean vector of the reconstruction matrix $\boldsymbol{\omega}'_i$, $\overline{\boldsymbol{\omega}}'_i \in \mathbb{R}^{m\times 1}$, and the elements of $\overline{\boldsymbol{\omega}}'_i$ are $\overline{\omega}'_i$, the expression of which is as follows:

$$\overline{\omega}'_i = \frac{1}{m}\sum_{k=1}^{m} X'_{ki} \tag{5}$$

where $\overline{\omega}'_i$ is the mean value of the strain monitoring data from the ith sensor after denoising, and X'_{ki} is the reconstruction matrix of X_{ki} after denoising.

Similarly, the monitoring dataset at the ith measurement point is denoted as $\mathbf{T}_i$ ($\mathbf{T}_i \in \mathbb{R}^{m\times 1}$), and the dataset $\mathbf{T}_i$ after centralization is denoted as $\boldsymbol{\xi}_i$ ($\boldsymbol{\xi}_i \in \mathbb{R}^{m\times 1}$), which is defined as follows:

$$\boldsymbol{\xi}_i = \mathbf{T}_i - \overline{\mathbf{T}}_i \tag{6}$$

where $\boldsymbol{\xi}_i$ is the ambient temperature dataset at the ith measurement point after centralization, and $\overline{\mathbf{T}}_i$ is the mean vector of the dataset $\mathbf{T}_i$, $\overline{\mathbf{T}}_i \in \mathbb{R}^{m\times 1}$.

Equations (4) and (6) are used to centralize the monitoring data and construct a matrix $\mathbf{B}$ ($\mathbf{B} \in \mathbb{R}^{m\times 2}$), which is defined as follows:

$$\mathbf{B} = \begin{bmatrix} \alpha\boldsymbol{\xi}_i & \boldsymbol{\varepsilon}_i \end{bmatrix} \tag{7}$$

where α is the weight value of $\boldsymbol{\xi}_i$. The role of α is to make the standard deviation of $\boldsymbol{\xi}_i$ magnified by α times much larger than the standard deviation of $\boldsymbol{\varepsilon}_i$. After the PCA decomposition, the main axis direction represents the ambient temperature, and the minor axis direction represents the strain. Therefore, the value of α usually needs to take a relatively large value to satisfy the following equation:

$$\alpha\sigma_{\xi} \gg \sigma_{\varepsilon} \tag{8}$$

where σ_{ξ} is the standard deviation of $\boldsymbol{\xi}_i$, and σ_{ε} is the standard deviation of $\boldsymbol{\varepsilon}_i$.

According to PCA theory, the covariance matrix of the construction matrix $\mathbf{B}$ is calculated as follows:

$$\mathbf{\Gamma} = \frac{1}{m-1}\mathbf{B}^{\mathrm{T}}\mathbf{B} \tag{9}$$

where $\mathbf{\Gamma}$ is the covariance matrix of the construction matrix $\mathbf{B}$, $\mathbf{\Gamma} \in \mathbb{R}^{2\times 2}$. After the eigenvalue decomposition of the covariance matrix $\mathbf{\Gamma}$, the following equation can be obtained:

$$\mathbf{\Gamma} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^{\mathrm{T}} \tag{10}$$

where $\mathbf{\Lambda}$ is the eigenvalue matrix of $\mathbf{\Gamma}$, which is a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{2\times 2}$, and $\mathbf{\Psi}$ is the eigenvector matrix of $\mathbf{\Gamma}$.

The eigenvector matrix $\mathbf{\Psi}$ is also the projection vector matrix of the construction matrix $\mathbf{B}$. Using the PCA principle, the pth principal direction $\mathbf{Y}_p$ ($\mathbf{Y}_p \in \mathbb{R}^{m\times 1}$) of the construction matrix $\mathbf{B}$ can be obtained. The calculation equation is as follows:

$$\mathbf{Y}_p = \mathbf{B}\mathbf{\Psi}_p \tag{11}$$

where $\mathbf{\Psi}_p$ is the pth column vector of the eigenvector matrix $\mathbf{\Psi}$, $\mathbf{\Psi}_p \in \mathbb{R}^{m\times 1}$, and $p \in (1, 2)$.

The column vector $\mathbf{Y}_1$ is the first principal component of the construction matrix $\mathbf{B}$, which represents the ambient temperature monitoring data expanded by α times. The column vector $\mathbf{Y}_2$ is the second principal component of the construction matrix $\mathbf{B}$, which represents the strain after removing the ambient temperature trend, which is also the strain $\boldsymbol{\varepsilon}^{\mathrm{N}}$ caused by the unobservable nonstationary effects.

2.2. Theoretical Proof of the Effectiveness of the Directional Projection Decoupling Method for Strain Monitoring Data

In the proposed PCA directional projection decoupling method, the theoretical basis for constructing the second principal component $\mathbf{Y}_2$ of the construction matrix $\mathbf{B}$ to represent the strain $\boldsymbol{\varepsilon}^{\mathrm{N}}$ caused by unobservable nonstationary effects is as follows.

Equation (7) shows that the covariance matrix $\mathbf{\Gamma}$ of the constructed matrix $\mathbf{B}$ can be expressed by the following equation:

$$\mathbf{\Gamma} = \begin{bmatrix} \alpha^2\sigma_{\xi}^2 & \alpha\sigma_{\xi,\varepsilon} \\ \alpha\sigma_{\xi,\varepsilon} & \sigma_{\varepsilon}^2 \end{bmatrix} \tag{12}$$

where σ_{ξ}^2 is the variance of $\boldsymbol{\xi}_i$; σ_{ε}^2 is the variance of $\boldsymbol{\varepsilon}_i$; and $\sigma_{\xi,\varepsilon}$ is the covariance of $\boldsymbol{\xi}_i$ and ε_i. The expressions of σ_{ξ}^2, σ_{ε}^2 and $\sigma_{\xi,\varepsilon}$ are as follows:

$$\sigma_{\xi}^2 = \frac{1}{m-1}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\xi}_i \tag{13}$$

$$\sigma_{\varepsilon}^2 = \frac{1}{m-1}\boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i \tag{14}$$

$$\sigma_{\xi,\varepsilon} = \frac{1}{m-1}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i \tag{15}$$

The covariance $\sigma_{\xi,\varepsilon}$ has the following relationship with the variance σ_{ξ}^2 and the variance σ_{ε}^2:

$$\sigma_{\xi,\varepsilon}^2 = \rho^2\sigma_{\xi}^2\sigma_{\varepsilon}^2 \tag{16}$$

where ρ is the correlation coefficient between $\boldsymbol{\xi}_i$ and $\boldsymbol{\varepsilon}_i$.

Suppose the expression of the eigenvalue matrix $\mathbf{\Lambda}$ of the covariance matrix $\mathbf{\Gamma}$ is

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{17}$$

In the equation, λ_1 is the first eigenvalue of $\mathbf{\Gamma}$, and λ_2 is the second eigenvalue of $\mathbf{\Gamma}$.

From the covariance matrix $\mathbf{\Gamma}$ in Equation (12) and the eigenvalue matrix $\mathbf{\Lambda}$ in Equation (17), the following relationship can be obtained, and the eigenvalue matrix $\mathbf{\Lambda}$ is solved:

$$|\mathbf{\Gamma} - \mathbf{\Lambda I}| = 0 \tag{18}$$

In the equation, $|\cdot|$ represents the determinant; $\mathbf{I}$ is a unit matrix, $\mathbf{I} \in \mathbb{R}^{2\times 2}$; and 0 is a zero matrix.

By solving the problem in Equation (18) and using the relationship in Equation (16), the expression for the eigenvalue λ_1 can be obtained as

$$\begin{aligned} \lambda_1 &= \frac{\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2+\sqrt{\left(\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2\right)^2-4\left(\alpha^2\sigma_\xi^2\sigma_\varepsilon^2-\alpha^2\sigma_{\xi,\varepsilon}^2\right)}}{2} \\ &= \frac{\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2+\sqrt{\left(\alpha^2\sigma_\xi^2\right)^2+2\alpha^2(2\rho^2-1)\sigma_\xi^2\sigma_\varepsilon^2+\left(\sigma_\varepsilon^2\right)^2}}{2} \end{aligned} \tag{19}$$

The relationship in Equation (8) shows that in the calculation of the square root of Equation (19), σ_ε^2 relative to $\alpha^2\sigma_\xi^2$ is a negligible minimum term, and adjusting the coefficient of the minimal term σ_ε^2 by $\left(2\rho^2-1\right)^2 \in (-1,1)$ times can also be ignored. Finally, the following approximate expression can be obtained:

$$\begin{aligned} \lambda_1 &\approx \frac{\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2+\sqrt{\left(\alpha^2\sigma_\xi^2\right)^2+2\alpha^2(2\rho^2-1)\sigma_\xi^2\sigma_\varepsilon^2+(2\rho^2-1)^2\left(\sigma_\varepsilon^2\right)^2}}{2} \\ &= \alpha^2\sigma_\xi^2 + \rho^2\sigma_\varepsilon^2 \end{aligned} \tag{20}$$

Similarly, the eigenvalue λ_2 is obtained as

$$\begin{aligned} \lambda_2 &= \frac{\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2-\sqrt{\left(\alpha^2\sigma_\xi^2+\sigma_\varepsilon^2\right)^2-4\left(\alpha^2\sigma_\xi^2\sigma_\varepsilon^2-\alpha^2\sigma_{\xi,\varepsilon}^2\right)}}{2} \\ &= \left(1-\rho^2\right)\sigma_\varepsilon^2 \end{aligned} \tag{21}$$

Substituting Equations (20) and (21) into Equation (17), the expression of the eigenvalue matrix $\mathbf{\Lambda}$ is as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} \alpha^2\sigma_\xi^2 + \rho^2\sigma_\varepsilon^2 & 0 \\ 0 & \left(1-\rho^2\right)\sigma_\varepsilon^2 \end{bmatrix} \tag{22}$$

Furthermore, the eigenvector matrix $\mathbf{\Psi}$ of the covariance matrix $\mathbf{\Gamma}$ is assumed to have the following expression:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \tag{23}$$

In the equation, $\{\psi_{11} \quad \psi_{21}\}^{\mathrm{T}}$ is an eigenvector of $\mathbf{\Gamma}$, and $\{\psi_{12} \quad \psi_{22}\}^{\mathrm{T}}$ is another eigenvector of $\mathbf{\Gamma}$.

From Equation (10), the following relationship can be deduced:

$$\mathbf{\Gamma\Psi} = \mathbf{\Psi\Lambda} \tag{24}$$

Substituting the covariance matrix $\mathbf{\Gamma}$ in Equation (12), the eigenvalue matrix $\mathbf{\Lambda}$ in Equation (22) and the eigenvector matrix $\mathbf{\Psi}$ in Equation (23) into Equation (24), the eigenvector $\{\psi_{11} \quad \psi_{21}\}^{\mathrm{T}}$ can be solved, and the elements ψ_{11} and ψ_{21} satisfy the following expressions:

$$\alpha^2\sigma_\xi^2\psi_{11} + \alpha\sigma_{\xi,\varepsilon}\psi_{21} = \left(\alpha^2\sigma_\xi^2 + \rho^2\sigma_\varepsilon^2\right)\psi_{11} \tag{25}$$

$$\frac{\psi_{11}}{\psi_{21}} = \frac{\alpha\sigma_{\xi,\varepsilon}}{\rho^2\sigma_\varepsilon^2} = \frac{\alpha\sigma_{\xi,\varepsilon}\sigma_\xi^2}{\rho^2\sigma_\varepsilon^2\sigma_\xi^2} = \frac{\alpha\sigma_\xi^2}{\sigma_{\xi,\varepsilon}} \tag{26}$$

In the same way, using Equation (24) to solve the eigenvector $\{\psi_{12} \quad \psi_{22}\}^{\mathrm{T}}$, the following relationship between ψ_{12} and ψ_{22} can be obtained:

$$\alpha\sigma_{\xi,\varepsilon}\psi_{12} + \sigma_\varepsilon^2\psi_{22} = \left(\left(1-\rho^2\right)\sigma_\varepsilon^2\right)\psi_{22} \tag{27}$$

$$\frac{\psi_{12}}{\psi_{22}} = \frac{-\rho^2\sigma_\varepsilon^2}{\alpha\sigma_{\xi,\varepsilon}} = \frac{-\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2} \tag{28}$$

Because the covariance matrix $\mathbf{\Gamma}$ is a symmetric matrix, the eigenvector matrix $\mathbf{\Psi}$ is also a symmetric matrix, and the following relationship exists:

$$\psi_{12} = \psi_{21} \tag{29}$$

Substituting Equations (26), (28) and (29) into Equation (23), the expression of the eigenvector matrix $\mathbf{\Psi}$ is obtained as follows:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2}\psi_{11} \\ \frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2}\psi_{11} & -\psi_{11} \end{bmatrix} \tag{30}$$

Substituting covariance matrix $\mathbf{\Gamma}$ from Equation (12), eigenvalue matrix $\mathbf{\Lambda}$ from Equation (22) and eigenvector matrix $\mathbf{\Psi}$ from Equation (30) into Equation (10), the coefficient ψ_{11} can be obtained to satisfy the following relationship:

$$\frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2}\psi_{11}^2\left(\alpha^2\sigma_\xi^2 + \rho^2\sigma_\varepsilon^2\right) - \frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2}\psi_{11}^2\left(1-\rho^2\right)\sigma_\varepsilon^2 = \alpha\sigma_{\xi,\varepsilon} \tag{31}$$

From the above equation,

$$\psi_{11}^2 = \frac{\alpha^2\sigma_\xi^2}{\alpha^2\sigma_\xi^2 + (2\rho^2-1)\sigma_\varepsilon^2} \approx 1 \tag{32}$$

Substituting the coefficient ψ_{11} obtained in Equation (32) into Equation (30), the expression of the eigenvector matrix $\mathbf{\Psi}$ can be obtained as follows:

$$\mathbf{\Psi} = \begin{bmatrix} -1 & -\frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2} \\ -\frac{\sigma_{\xi,\varepsilon}}{\alpha\sigma_\xi^2} & 1 \end{bmatrix} \tag{33}$$

Substituting Equations (1) and (33) into Equation (11), the second component $\mathbf{Y}_2$ of the constructed matrix $\mathbf{B}$ is expressed as follows:

$$\begin{aligned} \mathbf{Y}_2 = \mathbf{B}\mathbf{\Psi}_2 &= -\boldsymbol{\xi}_i\frac{\sigma_{\xi,\varepsilon}}{\sigma_\xi^2} + \boldsymbol{\varepsilon}_i = -\boldsymbol{\xi}_i\frac{\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i}{\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\xi}_i} + \boldsymbol{\varepsilon}_i \\ &= -\boldsymbol{\xi}_i\frac{\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i\sqrt{\boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i}}{\sqrt{\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\xi}_i}\sqrt{\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\xi}_i}\sqrt{\boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\varepsilon}_i}} + \boldsymbol{\varepsilon}_i \\ &= -\boldsymbol{\xi}_i\cos(\beta)\frac{\|\boldsymbol{\varepsilon}_i\|}{\|\boldsymbol{\xi}_i\|} + \boldsymbol{\varepsilon}_i = -\boldsymbol{\varepsilon}_i^{\mathrm{O}} + \boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_i^{\mathrm{N}} \end{aligned} \tag{34}$$

where $\cos(\beta)$ is the cosine of the angle between $\boldsymbol{\xi}_i$ and $\boldsymbol{\varepsilon}_i$.

Equation (34) shows that the second principal component column vector $\mathbf{Y}_2$ is also the strain $\boldsymbol{\varepsilon}^{\mathrm{N}}$ caused by the unobservable nonstationary effects. Additionally, the weight α used in the constructed matrix $\mathbf{B}$ is eliminated in the solution calculation, and the weight α does not affect the calculation results of the second principal component $\mathbf{Y}_2$.

2.3. Classification of the Strain Monitoring Steady-State Data under Traffic Load

The unobservable nonstationary effects cannot be obtained through sensor observations. For example, the traffic load is not easily obtained using a sensor. A special matrix $\mathbf{B}$ must be constructed, and the previously proposed PCA directional projection decoupling method must be used to obtain the second principal component $\mathbf{Y}_2$, that is, the strain $\boldsymbol{\varepsilon}^{\mathrm{N}}$ caused by the unobservable nonstationary effects. The state of the strain $\boldsymbol{\varepsilon}^{\mathrm{N}}$ includes the nonsteady state and the steady state. Taking the traffic load as an example, a classification method is proposed to weaken the influence of the unobservable nonstationary effects. The specific description of the proposed method is as follows.

Equation (11) shows that the first principal component $\mathbf{Y}_1$ and the second principal component $\mathbf{Y}_2$ are obtained by performing PCA directional projection on the construction

matrix **B**. The two principal components dataset **Y** ($\mathbf{Y} \in \mathbb{R}^{m\times 2}$) contains m samples, which are defined as follows:

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k, \cdots, \mathbf{y}_m]^{\mathrm{T}} \tag{35}$$

where $\mathbf{y}_k$ is the kth sample of the dataset **Y**, $\mathbf{y}_k \in \mathbb{R}^{1\times 2}$, which is defined as follows:

$$\mathbf{y}_k = \{Y_{1,k}, Y_{2,k}\} \tag{36}$$

where Y_{k1} is the kth sample of the vector $\mathbf{Y}_1$, and Y_{k2} is the kth sample of the vector $\mathbf{Y}_2$.

Using the K-means cluster analysis algorithm, the m samples of dataset **Y** are randomly divided into two categories, the first cluster group $\mathbf{\Omega}_1$ and the second cluster group $\mathbf{\Omega}_2$. Taking the first cluster group $\mathbf{\Omega}_1$ as an example, its definition is as follows:

$$\mathbf{\Omega}_1 = \left[\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_j, \cdots, \mathbf{y}_v\right]^{\mathrm{T}} \tag{37}$$

where $\mathbf{\Omega}_1$ is the first clustering group in the K-means cluster analysis algorithm; v is the total number of samples in the first clustering group $\mathbf{\Omega}_1$; j is an arbitrary sample in the first clustering group $\mathbf{\Omega}_1$, $j \in (1,\ 2,\ \cdots,\ v)$; and $\mathbf{y}_j$ is the jth sample in the first clustering group $\mathbf{\Omega}_1$, $\mathbf{y}_j \in \mathbb{R}^{1\times 2}$.

Then, the centre point dataset $\mathbf{c}_1$ of the first cluster group $\mathbf{\Omega}_1$ can be solved as follows:

$$\mathbf{c}_1 = \frac{1}{v} \sum_{\mathbf{y}_j \in \mathbf{\Omega}_1} \mathbf{y}_j \tag{38}$$

Similarly, the centre point dataset $\mathbf{c}_2$ of the second cluster group $\mathbf{\Omega}_2$ can be obtained. After determining $\mathbf{c}_1$ and $\mathbf{c}_2$, the m samples in dataset **Y** are reclustered according to the distance from the centre point, and the centre point datasets $\mathbf{c}_1$ and $\mathbf{c}_2$ are updated according to Equation (38). The above update process adopts the method of optimization and solution, and the constructed optimization objectives are as follows:

$$\text{minimize} \left\{ \sum_{\mathbf{y}_k \in \mathbf{\Omega}_q} dist(\mathbf{y}_k, \mathbf{c}_q) \right\} \tag{39}$$

where q is an arbitrary category in the K-means cluster analysis algorithm, $q \in (1,2)$; $\mathbf{c}_q$ is the centre point dataset of the qth clustering group; $\mathbf{\Omega}_q$ is the qth clustering group; and $dist(\mathbf{y}_k, \mathbf{c}_q)$ represents the distance from sample $\mathbf{y}_k$ to the centre point dataset $\mathbf{c}_q$, which usually adopts the Euclidean distance, and the expression is as follows:

$$dist(\mathbf{y}_k, \mathbf{c}_q) = \|\mathbf{y}_k - \mathbf{c}_q\| \tag{40}$$

where $\|\cdot\|$ represents the norm.

When solving the objective function of Equation (39), to improve the efficiency of finding the objective function, David Arthur proposed a method for selecting the centre point dataset. In principle, if the centre point dataset $\mathbf{c}_1$ is known, select the sample $\mathbf{y}_k$ with the highest probability of becoming another centre point dataset. The probability $P(\mathbf{y}_k)$ of becoming the centre point is calculated using the following equation:

$$P(\mathbf{y}_k) = \frac{D(\mathbf{y}_k)}{\sum_{\mathbf{y}_k \in \mathbf{Y}} D(\mathbf{y}_k)} \tag{41}$$

where $D(\mathbf{y}_k)$ is the distance from sample $\mathbf{y}_k$ to the nearest centre point, and the expression is as follows:

$$D(\mathbf{y}_k) = \min\{dist(\mathbf{y}_k, \mathbf{c}_q)\} \tag{42}$$

where $\min\{\cdot\}$ represents the minimum value of the set.

The process of solving the above optimization problem is realized through MATLAB programming. By solving the objective function Equation (39), the first cluster group $\mathbf{\Omega}_1$ and the second cluster group $\mathbf{\Omega}_2$ can be obtained. The steady state is discriminated by comparing the variance of cluster group $\mathbf{\Omega}_1$ and cluster group $\mathbf{\Omega}_2$.

2.4. Construction of a Steady-State Data Baseline Model

The strain monitoring data in the traffic load steady state are extracted to establish a steady-state data baseline model. If the cluster group $\mathbf{\Omega}_1$ is the steady state, the expression for the bridge steady-state dataset $\underset{\cdot}{\mathbf{X}}$ is as follows:

$$\underset{\cdot}{\mathbf{X}} = \begin{bmatrix} \underset{\cdot}{Y}_{2,11} & \underset{\cdot}{Y}_{2,12} & \cdots & \underset{\cdot}{Y}_{2,1i} & \cdots & \underset{\cdot}{Y}_{2,1n} \\ \underset{\cdot}{Y}_{2,21} & \underset{\cdot}{Y}_{2,22} & \cdots & \underset{\cdot}{Y}_{2,2i} & \cdots & \underset{\cdot}{Y}_{2,2n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \underset{\cdot}{Y}_{2,j1} & \underset{\cdot}{Y}_{2,j2} & \cdots & \underset{\cdot}{Y}_{2,ji} & \cdots & \underset{\cdot}{Y}_{2,jn} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ \underset{\cdot}{Y}_{2,v1} & \underset{\cdot}{Y}_{2,v2} & \cdots & \underset{\cdot}{Y}_{2,vi} & \cdots & \underset{\cdot}{Y}_{2,vn} \end{bmatrix} \tag{43}$$

where $\underset{\cdot}{\mathbf{X}}$ is the steady-state data, $\underset{\cdot}{\mathbf{X}} \in \mathbb{R}^{v \times n}$; v is the total number of samples in the first cluster group $\mathbf{\Omega}_1$; j is an arbitrary sampling point at time in the first cluster group $\mathbf{\Omega}_1$, $j \in (1, 2, \cdots, v)$; and $\underset{\cdot}{Y}_{2,ji}$ is the second principal component $\mathbf{Y}_2$ of the jth sampling point in the ith measurement point in steady state.

Equation (43) shows that the steady-state dataset of the jth sampling point is recorded as $\underset{\cdot}{\mathbf{X}}_j$ ($\underset{\cdot}{\mathbf{X}}_j \in \mathbb{R}^{1 \times n}$), which is defined as follows:

$$\underset{\cdot}{\mathbf{X}}_j = \left\{ \underset{\cdot}{Y}_{2,j1}, \underset{\cdot}{Y}_{2,j2}, \cdots, \underset{\cdot}{Y}_{2,ji}, \cdots, \underset{\cdot}{Y}_{2,jn} \right\} \tag{44}$$

The baseline model of steady-state data $\underset{\cdot}{\mathbf{X}}$ is constructed using statistical methods, and the expression of its mean value, $\underset{\cdot}{\overline{\mathbf{X}}} \in \mathbb{R}^{1 \times n}$, is as follows:

$$\underset{\cdot}{\overline{\mathbf{X}}} = \frac{1}{v} \sum_{j=1}^{v} \underset{\cdot}{\mathbf{X}}_j \tag{45}$$

Then, the covariance of the steady-state data $\underset{\cdot}{\mathbf{X}}$ is recorded as $\underset{\cdot}{\mathbf{\Gamma}} \in \mathbb{R}^{n \times n}$, and the following equation is used:

$$\underset{\cdot}{\mathbf{\Gamma}} = \frac{1}{v-1} \sum_{j=1}^{v} \left(\underset{\cdot}{\mathbf{X}}_j - \underset{\cdot}{\overline{\mathbf{X}}} \right)^{\mathrm{T}} \left(\underset{\cdot}{\mathbf{X}}_j - \underset{\cdot}{\overline{\mathbf{X}}} \right) \tag{46}$$

The mean $\underset{\cdot}{\overline{\mathbf{X}}}$ and covariance $\underset{\cdot}{\mathbf{\Gamma}}$ of the steady-state data together construct a baseline model for the steady-state data.

The overall calculation steps of the proposed algorithm are as follows:

Step 1: Use Equation (2) to establish the strain monitoring dataset $\mathbf{X}$ for all measurement points of the bridge with monitoring systems.

Step 2: Select the ith measurement point and obtain the strain monitoring dataset $\boldsymbol{\omega}_i$ and the ambient temperature monitoring dataset $\mathbf{T}_i$.

Step 3: Centralize the dataset to obtain the strain dataset $\boldsymbol{\varepsilon}_i$ and the ambient temperature dataset $\boldsymbol{\xi}_i$ based on Equations (4) and (6).

Step 4: Use the weight value α to ensure that the standard deviation of $\boldsymbol{\xi}_i$ is magnified by α times, which is much larger than the standard deviation of $\boldsymbol{\varepsilon}_i$, and establish a construction matrix $\mathbf{B}$ based on Equation (7).

Step 5: Obtain the first principal component $\mathbf{Y}_1$ and the second principal component $\mathbf{Y}_2$ of the construction matrix $\mathbf{B}$ based on Equation (11) and then construct the projection matrix $\mathbf{Y}$ shown in Equation (35).

Step 6: Solve the optimization problem in Equation (39) to perform a cluster analysis of dataset $\mathbf{Y}$ to find the steady-state clustering group $\mathbf{\Omega}_1$.

Step 7: Repeat steps 2 to 5 to calculate $\mathbf{Y}_2$ for each measurement point and then construct the steady-state data $\dot{\mathbf{X}}$ in clustering group $\mathbf{\Omega}_1$ based on Equation (43).

Step 8: Use Equations (45) and (46) to obtain the mean value $\overline{\dot{\mathbf{X}}}$ and covariance $\dot{\mathbf{\Gamma}}$ of the steady-state strain monitoring data; that is, construct a baseline model for the strain monitoring data in a steady state.

3. Example with an Actual Bridge

3.1. Brief Introduction to the Actual Bridge with a Monitoring System

The Yingxiongshan interchange bridge is in Jinan, China, and a photograph of this interchange bridge is shown in Figure 1. Three medium- and small-span beam bridges in this interchange bridge were selected to establish an SHM system. The entire SHM system includes 36 displacement sensors, 18 acceleration sensors, 64 strain and temperature sensors. In this section, one three-span prestressed continuous box girder bridge in the SHM system is selected as an example. The span combination of the selected bridge #1 is 30 m + 30 m + 30 m, the main beam is a variable-width beam, and its width varies from 26.3 m to 28.3 m. The bridge is made of prestressed concrete with a strength of C50, and the cross-section adopts the structure of a single box with six chambers. In this SHM system, there are totally 25 strain and temperature sensors on the main beam of bridge #1, and the layout of the measurement points of bridge #1 is shown in Figure 2. The measurement points #1~#5, #11~#15, and #21~#25 are located on the bottom flange of the main beam of the bridge, and measurement points #6~#10 and #16~#20 are located on the top flange of the main beam of the bridge. The field installation of pre-embedded strain and temperature sensors is shown in Figure 3.

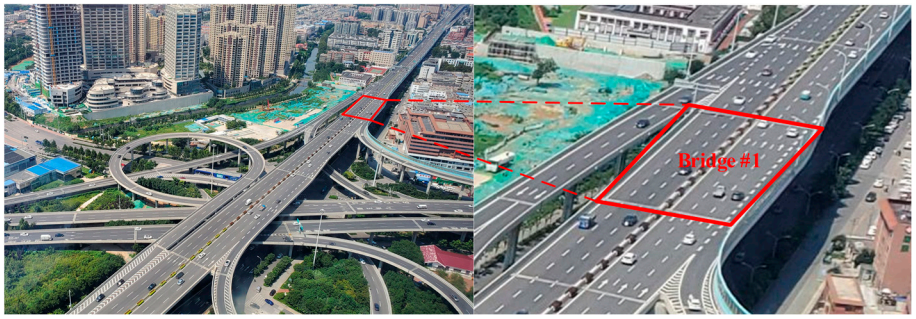


Figure 1. Photo of the Yingxiongshan interchange bridge.

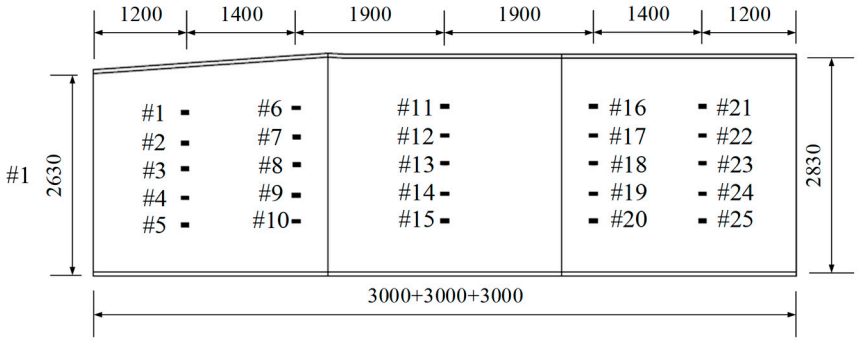


Figure 2. Measurement points of bridge #1 (Unit: cm).

(**a**) (**b**)

Figure 3. Field installation of the SHM system: (**a**) pre-embedded strain and temperature sensors; (**b**) data acquisition instrument.

The SHM system of the Yingxiongshan interchange bridge completed commissioning and verification work on 1 October 2016 and entered the trial operation stage. The strain and temperature were sampled once every 10 min. Due to the limitations of the actual site conditions, the system did not achieve continuous 24 h data collection during the trial operation stage. The SHM system of the Yingxiongshan interchange bridge entered the formal operation stage on 1 April 2017. The interchange bridge was opened to traffic on 27 April 2017. The traffic load flow in the initial operation phase of this interchange bridge was small and unstable, sometimes in an empty state. As time went on, the traffic load began to increase gradually and tended to be saturated and stable. To establish an accurate data baseline model, the traffic load in the early stage of opening to traffic belongs to the nonsteady stage and should be eliminated, and the traffic load after reaching saturation belongs to the steady state and should be used to establish a steady-state data baseline model. To establish and verify the steady-state data baseline model, the strain monitoring data selected in this section start from the trial operation stage, using 1 October 2016, as the starting point for monitoring data analysis, and 19 months of continuous monitoring data are used to establish the steady-state data baseline model. The strain monitoring data of all measurement points of bridge #1 are shown in Figure 4.

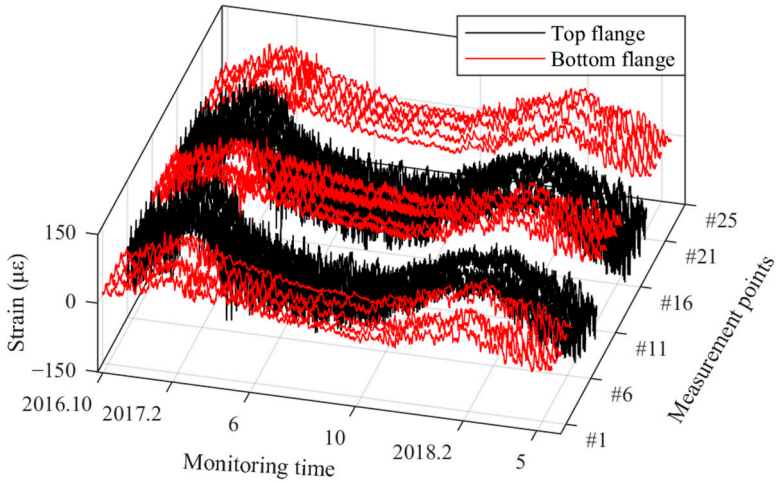


Figure 4. Strain monitoring data of all measurement points of bridge #1.

3.2. Construction of the Steady-State Data Baseline Model

The construction of the steady-state data baseline model of the Yingxiongshan interchange bridge is divided into three steps. First, the observable effects and the strain monitoring data are decoupled to obtain the strain response caused by the unobservable effects. Second, the strain response caused by the unobservable effects is divided into two categories: nonsteady-state data and steady-state data. Finally, the data baseline

model is constructed for the steady-state data. Taking measurement point #8 on the top flange as an example, to correctly evaluate the fluctuation variance of the monitoring data, Equations (4) and (6) are used to centrally process the structural temperature monitoring data and the denoised strain monitoring data, respectively. The result is shown in Figure 5, which shows that the fluctuation range of the centralized temperature monitoring data is ±30 °C, and the fluctuation range of the centralized strain monitoring data is ±100 με. The fluctuation of the strain data is greater than that of the temperature data. Therefore, when building the construction matrix **B** using Equation (7), a weight w must be assigned to the temperature monitoring data so that the fluctuation of the temperature data is greater than that of the strain data. In this way, the first principal component $\mathbf{Y}_1$ of the construction matrix **B** is the temperature data magnified w times. The second principal component $\mathbf{Y}_2$ of the construction matrix **B** is the strain caused by the unobservable effects. According to the actual condition, the weight w is set to 1000. The second principal component $\mathbf{Y}_2$ of the construction matrix **B** calculated by Equation (11) is shown in Figure 6a, and the K-means clustering results of the scatter distribution of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of the construction matrix **B** are shown in Figure 6b.

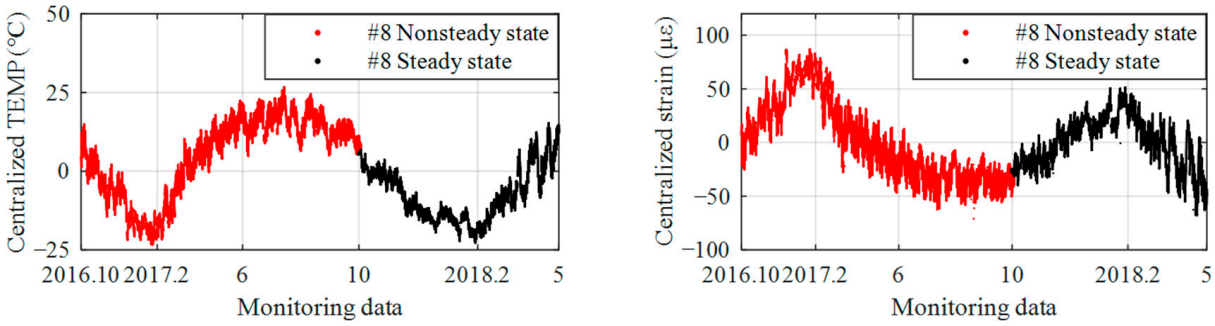


Figure 5. Data of the strain of measurement point #8, for which the mean value is removed: (**a**) centralized temperature; (**b**) centralized strain.

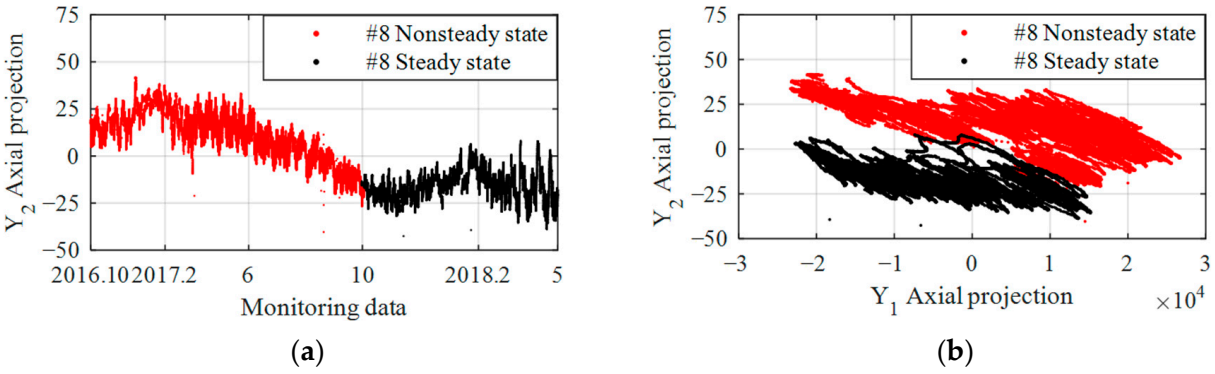


Figure 6. Results of the data cluster analysis of the strain at measurement point #8: (**a**) $\mathbf{Y}_2$ axial projection; (**b**) results of cluster analysis.

Figure 6a shows that the $\mathbf{Y}_2$ axial projection of the construction matrix **B**, that is, the strain caused by unobservable effects, such as traffic loads, presents three stages. The first stage is from 1 October 2016 to 27 April 2017, which is a temporary steady state of approximately no load before opening to traffic. The second stage is from 27 April 2017 to 3 October 2017, which is the gradual increasing trend of the traffic load in the early stage of opening to traffic. The third stage is from 3 October 2017 to 1 May 2018, which is a steady state of traffic saturation after opening to traffic. This shows that the extracted $\mathbf{Y}_2$ axial projection can correctly reflect the change in the traffic load. As shown in Figure 6b, in the K-means clustering results of the scatter distribution of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of the construction

matrix **B**, the red part is the nonsteady-state data (no-load state and traffic flow increase stage), and the black part is the steady-state data (flow saturation state). The red and black parts along the $\mathbf{Y}_2$ axis are clearly distinguished due to the different traffic load effects between the nonsteady-state data and the steady-state data. The data can be easily divided into two cluster groups by solving the objective function in Equation (38), thus obtaining the steady-state data of the black part. The results in Figure 6 show that the proposed method can effectively find the traffic load trend of the data on the top flange and extract the steady-state monitoring data.

To compare the effectiveness of the algorithm, the scatter plot obtained from the original strain monitoring data **X** and the temperature monitoring data is directly used, as shown in Figure 7. For a more obvious comparison, the classification results obtained in Figure 6b are marked in Figure 7.

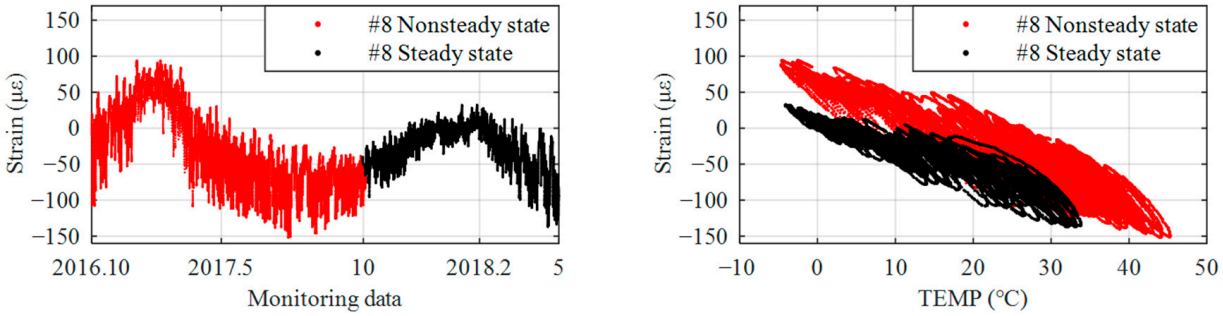


Figure 7. Original data of the strain and temperature of measurement point #8: (**a**) strain monitoring data; (**b**) results of cluster analysis.

Figure 7b shows that separating the steady-state data in the black part from the nonsteady-state data in the red part by directly using the original strain monitoring data is difficult. The black data are offset from the red data but are not completely separated from them. The reason for the offset of the black data is that the effect of the traffic load has changed. In Figure 7b, the direction of action of the traffic load is not a translation along the coordinate axis but an oblique translation, and the translation effect is not obvious. In addition, the influence of other factors in the data also obscures the effect of the traffic load. Performing K-means clustering using the original strain monitoring data is not feasible. Comparing Figures 6b and 7b shows that the proposed method has a good effect in reducing the influence of environmental factors on the strain monitoring data on the top flange. Due to the limited space, only measurement point #8 on the top flange is selected for a detailed description. Figures 6 and 7 show the second principal component $\mathbf{Y}_2$ and the results of the K-means cluster analysis of the other representative measurement points.

Figure 8 shows that the variation law of the $\mathbf{Y}_2$ axis projection obtained using Equation (11) is consistent, indicating that the effect of the traffic load on each measurement point is consistent. This conclusion is also consistent with the actual situation. The selected representative measurement points are from the same bridge, so the temperature and traffic loads are basically the same.

Figure 9 shows that the steady-state data in the black part of each measurement point are completely separated from the nonsteady-state data in the red part along the $\mathbf{Y}_2$ axis direction, indicating that the proposed method can successfully extract the steady-state data of all measurement points. Table 1 shows the results of selecting the starting sampling point k of the steady-state data of all measurement points.

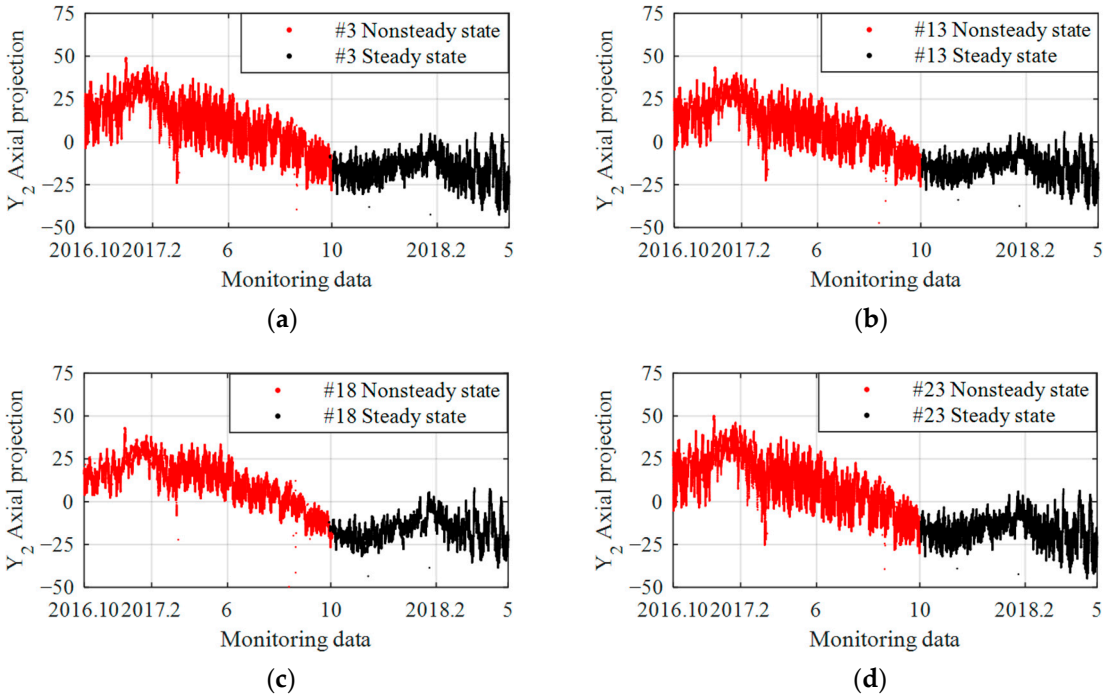


Figure 8. Data of the strain of all measurement points located at the midspan of the bridge, which reduces the effect of environmental temperature: (**a**) measurement point #3; (**b**) measurement point #13; (**c**) measurement point #18; (**d**) measurement point #23.

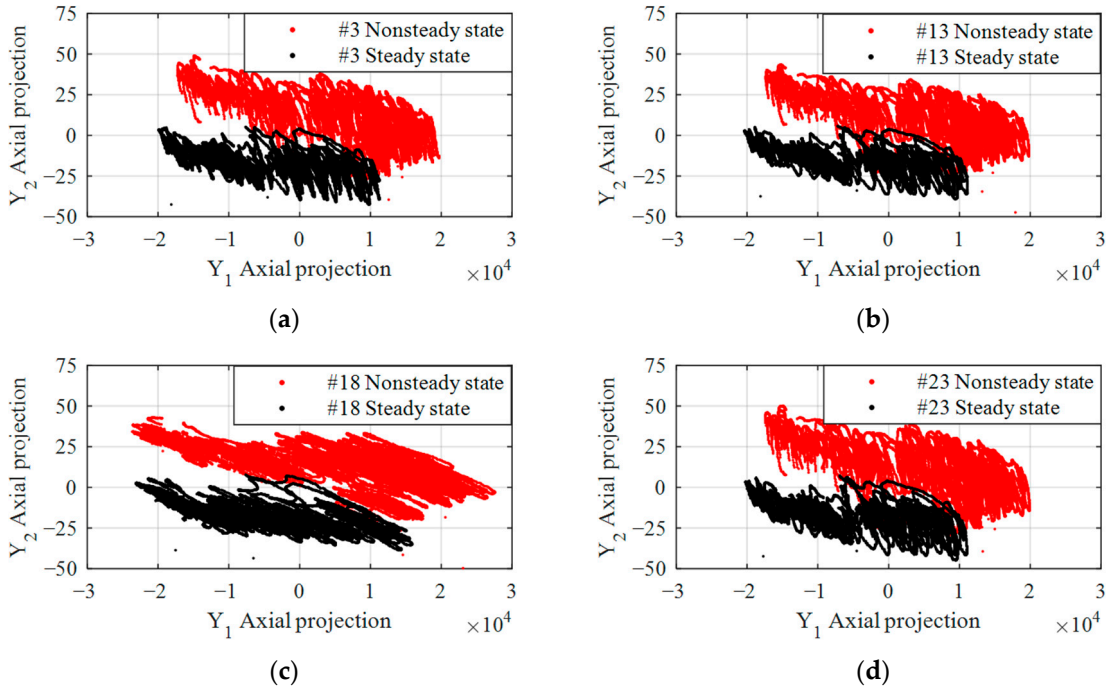


Figure 9. Results of cluster analysis of strain data of all measurement points located at the midspan of the bridge: (**a**) measurement point #3; (**b**) measurement point #13; (**c**) measurement point #18; (**d**) measurement point #23.

Table 1. Results of starting point selection of steady-state data of all measurement points.

Measurement Point	Monitoring Time	Starting Point	Measurement Point	Monitoring Time	Starting Point	Measurement Point	Monitoring Time	Starting Point
#1	2018.10.2	39165	#10	2018.10.2	39298	#19	2018.10.3	39301
#2	2018.10.2	39170	#11	2018.10.2	39172	#20	2018.10.3	39303
#3	2018.10.2	39160	#12	2018.10.2	39163	#21	2018.10.2	39178
#4	2018.10.2	39168	#13	2018.10.2	39175	#22	2018.10.2	39171
#5	2018.10.1	39150	#14	2018.10.2	39165	#23	2018.10.2	39170
#6	2018.10.3	39300	#15	2018.10.2	39174	#24	2018.10.2	39175
#7	2018.10.2	39283	#16	2018.10.3	39301	#25	2018.10.1	39120
#8	2018.10.2	39292	#17	2018.10.2	39284			
#9	2018.10.3	39304	#18	2018.10.2	39291			

Table 1 shows that the results of the steady-state data of all measurement points are extracted independently, but the starting point results of the steady-state data of all measurement points are basically the same. The earliest starting point of steady-state data is located at measurement point #25, which is the 39,120th sampling point. The latest starting point of the steady-state data is located at measurement point #9, which is the 39,304th sampling point. The results of the earliest and latest starting points differ by approximately 31 h. Each measurement point is calculated independently, the maximum error between all measurement points is approximately 31 h, and the total selected research objects are up to 19 months. In contrast, the error of 31 h accounts for 0.23% of the overall sample. The size of the error does not affect the calculation process and accuracy of the algorithm, and all measurement points yield similar results, which also demonstrates the accuracy of the method. Based on the above results, the starting point of the steady-state data is selected as the result of measurement point #9; that is, the period from 3 October 2017, to 1 May 2018, is the steady-state data period of the bridge. The steady-state data baseline model obtained by Equations (45) and (46) and some results are shown in Table 2.

Table 2. Results of the steady-state data baseline model.

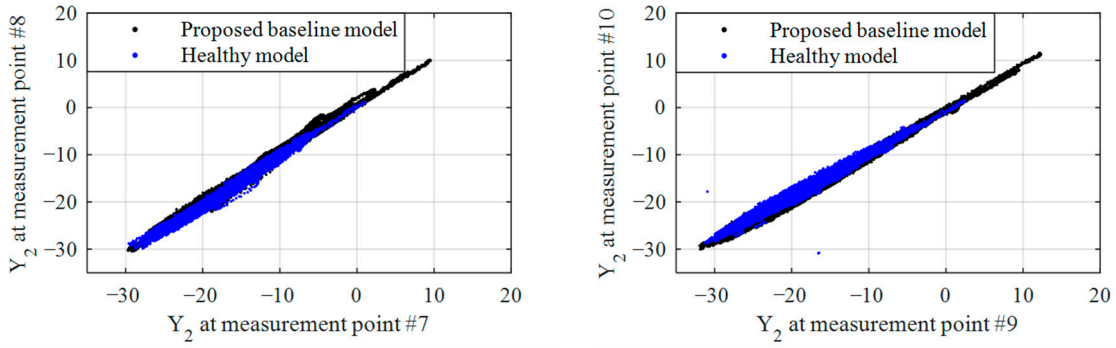
Parameter	Baseline Model	Healthy Model	Error
$\overline{\mathbf{Y}}_{\cdot 2}^{\#7}$	−15.1530	−16.8612	1.7081
$\overline{\mathbf{Y}}_{\cdot 2}^{\#8}$	−15.0544	−17.6427	2.5882
$\overline{\mathbf{Y}}_{\cdot 2}^{\#9}$	−14.6999	−16.6824	1.9825
$\overline{\mathbf{Y}}_{\cdot 2}^{\#10}$	−14.5927	−15.7925	1.1998
$Var(\mathbf{Y}_{\cdot 2}^{\#7})$ *	39.4604	27.0709	
$Var(\mathbf{Y}_{\cdot 2}^{\#8})$	43.4576	29.1966	
$Var(\mathbf{Y}_{\cdot 2}^{\#9})$	49.5208	35.6062	
$Var(\mathbf{Y}_{\cdot 2}^{\#10})$	43.5048	29.2437	
$Cov(\mathbf{Y}_{\cdot 2}^{\#7}, \mathbf{Y}_{\cdot 2}^{\#8})$	41.2107	27.8485	
$Cov(\mathbf{Y}_{\cdot 2}^{\#7}, \mathbf{Y}_{\cdot 2}^{\#9})$	43.6631	30.1595	
$Cov(\mathbf{Y}_{\cdot 2}^{\#7}, \mathbf{Y}_{\cdot 2}^{\#10})$	41.1291	27.4678	
$Cov(\mathbf{Y}_{\cdot 2}^{\#8}, \mathbf{Y}_{\cdot 2}^{\#9})$	46.0370	31.6078	
$Cov(\mathbf{Y}_{\cdot 2}^{\#8}, \mathbf{Y}_{\cdot 2}^{\#10})$	43.1732	28.4656	
$Cov(\mathbf{Y}_{\cdot 2}^{\#9}, \mathbf{Y}_{\cdot 2}^{\#10})$	46.1945	31.7308	
ρ	1.0000	0.9748	2.52%

* $\mathbf{Y}_{\cdot 2}^{\#7}$ represents the second principal component $\mathbf{Y}_2$ at measurement point #7 in steady state.

3.3. Performance Comparison between the Proposed Method and a Conventional Method

The K-means cluster analysis indicates that the period from 3 October 2017 to 1 May 2018 is the steady state of this interchange bridge. The strain monitoring data during the steady state are projected onto the $\mathbf{Y}_2$ axis using Equation (11). The data of measurement points #7 to #10 are arranged in the form of Equation (43). The steady-state data baseline model is obtained using Equations (45) and (46). The period from 1 May 2018, to 1 September 2018, is selected as the healthy state of this interchange bridge. Similarly, the mean and covariance of the healthy model are obtained to test the effectiveness of the proposed method. The results are shown in Table 2.

Table 2 shows that the error of the covariance between the healthy model and the baseline model of the proposed method is 2.52%, the maximum mean error is 2.5882 for measurement point #8, and the minimum is 1.1998 for measurement point #10. To visually represent the difference between the healthy model and the baseline model of the proposed method, scatter plots of measurement points #7 and #8 and scatter plots of measurement points #9 and #10 are drawn in Figure 10. In this figure, the baseline model is represented by the black part, and the healthy model is represented by the blue part. Although the environment of the healthy model using summer data is quite different from that of the baseline model using winter data, the outlines of the scatter plots between the two are in good agreement, indicating that the proposed baseline model can effectively weaken the environmental effects.



(**a**) (**b**)

Figure 10. Comparison results between the baseline model and the healthy model using the proposed method: (**a**) measurement points #7 and #8; (**b**) measurement points #9 and #10.

To compare the modelling accuracy of the proposed method, the conventional method [32] is used to establish a baseline model, and the period from 1 October 2016 to 1 May 2018, after the monitoring system was in operation is used as the reference state. The strain monitoring data from measurement points #7 to #10 are used directly to build the baseline model. Similarly, the period from 1 May to 1 September 2018 is selected as the healthy state of this interchange bridge. The mean and covariance of the healthy model are obtained to test the effectiveness of the conventional method. The results are shown in Table 3.

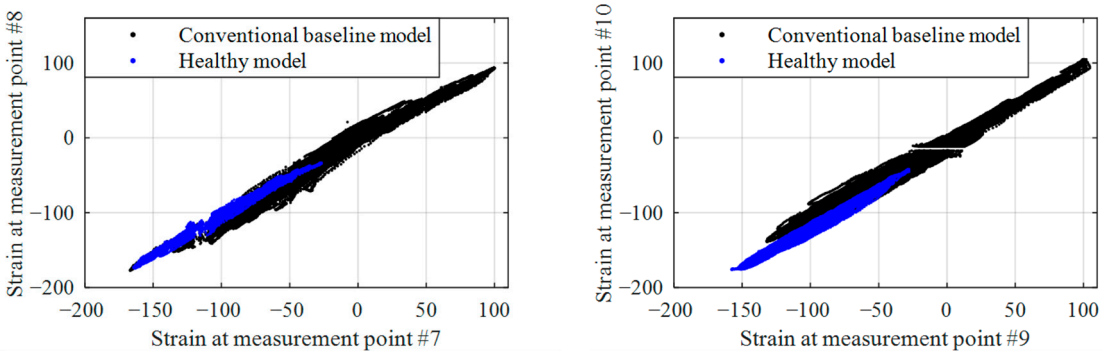
Table 3 shows that the covariance error of the conventional method is 82.65%, the maximum mean error is 83.7621 for measurement point #10, and the minimum is 71.7821 for measurement point #8. To visually represent the difference between the healthy model and the baseline model of the conventional method, scatter plots of measurement points #7 and #8 and scatter plots of measurement points #9 and #10 are drawn in Figure 11. In this figure, the black part is the baseline model, and the blue part is the healthy model. The outline of the scatter plots of the healthy model is much smaller than that of the baseline model due to the different environmental conditions of the two. The healthy model uses summer data, while the baseline model uses the data for the entire year. Conventional methods cannot exclude environmental effects, and the monitoring data are impacted by these effects, which can cover up part of the damage. Therefore, large errors may be introduced when establishing a baseline model by directly using monitoring data, making damage

detection difficult. Comparing Tables 2 and 3 and Figures 8 and 9 shows that compared with the conventional method, the proposed method can greatly improve the modelling accuracy of the baseline model, from the 82.65% error in the conventional method to the 2.52% error in the proposed method, which greatly reduces the environmental effects.

Table 3. Results of the baseline model using the conventional method.

Parameter	Baseline Model	Healthy Model	Error
$\overline{\omega}_7$	−23.8281	−97.7235	73.8953
$\overline{\omega}_8$	−29.0736	−100.8557	71.7821
$\overline{\omega}_9$	−21.1149	−94.4568	73.3419
$\overline{\omega}_{10}$	−30.4165	−114.1785	83.7621
$Var(\omega_7)$ *	1.8985×10^3	796.34	
$Var(\omega_8)$	2.2287×10^3	879.16	
$Var(\omega_9)$	2.1399×10^3	691.16	
$Var(\omega_{10})$	2.2441×10^3	779.63	
$Cov(\omega_7, \omega_8)$	2.0409×10^3	830.61	
$Cov(\omega_7, \omega_9)$	1.9582×10^3	717.96	
$Cov(\omega_7, \omega_{10})$	2.0333×10^3	778.62	
$Cov(\omega_8, \omega_9)$	2.1194×10^3	743.16	
$Cov(\omega_8, \omega_{10})$	2.1940×10^3	808.27	
$Cov(\omega_9, \omega_{10})$	2.1663×10^3	726.59	
ρ	1.0000	0.1735	82.65%

* ω_7 represents the strain monitoring dataset at measurement point #7.



(**a**) (**b**)

Figure 11. Comparison results between the baseline model and the healthy model using the conventional method (Unit: με): (**a**) measurement points #7 and #8; (**b**) measurement points #9 and #10.

4. Conclusions

Aiming at the problem of developing accurate nonstationary monitoring data baseline models for urban girder bridges, a steady-state data baseline model of bridges is proposed. This model is suitable for damage diagnosis of urban bridge structures under the influence of the environment. The main conclusions of this study are summarized as follows.

1. The weight α is proven to only control the projection direction of the construction matrix **B**; it does not participate in the calculation of the second principal component $\mathbf{Y}_2$. The physical meaning of $\mathbf{Y}_2$ is clarified as the strain ε^N caused by unobservable non-stationary effects.
2. The proposed directional projection decoupling method can effectively reduce observable effects such as the temperature from the projection data and weaken the influence of the observable effects on the monitoring data. The results obtained from a case study of an actual bridge show that after reducing the temperature effect, the obtained $\mathbf{Y}_2$ is consistent with the actual traffic conditions of the bridge.

3. K-means cluster analysis can be used to effectively divide unobservable effects such as the traffic load into steady- and nonsteady-state data, thereby reducing the influence of unobservable effects on the monitoring data. The independent operation results of each measurement point show that the error of the steady state starting points of different measurement points is only 0.23%, which shows that the proposed steady-state segmentation method has a good robustness.
4. An actual bridge example is used to compare the proposed baseline model using steady-state data to a conventional baseline model using strain monitoring data. This comparative analysis shows that the proposed steady-state data baseline model can effectively improve the accuracy of the baseline model, and the error is reduced from 82.65% to 2.52%, which greatly reduces the influence of environmental effects on the strain monitoring data.

Given the limitations of the proposed method, the steady-state data baseline model is only applicable to new constructed bridge with SHM systems installed.

Author Contributions: Conceptualization, S.Z.; methodology, S.Z.; validation, K.Y. and Y.W.; investigation, S.Z. and Y.W.; resources, S.Z. and K.Y.; data curation, S.Z. and Y.W.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z., K.Y. and Y.W.; supervision, K.Y.; project administration, S.Z.; funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Heilongjiang Postdoctoral Fund, grant number LBH-Z20011.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ou, J.; Li, H. Structural health monitoring in mainland China: Review and future trends. *Struct. Health Monit.* **2010**, *9*, 219–231.
2. Fan, W.; Qiao, P. Vibration-based damage identification methods: A review and comparative study. *Struct. Health Monit.* **2011**, *10*, 83–111. [CrossRef]
3. Cawley, P. Structural health monitoring: Closing the gap between research and industrial deployment. *Struct. Health Monit.* **2018**, *17*, 1225–1244. [CrossRef]
4. Abdulkarem, M.; Samsudin, K.; Rokhani, F.Z.; Rasid, M.F.A. Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction. *Struct. Health Monit.* **2020**, *19*, 693–735. [CrossRef]
5. Wu, R.T.; Jahanshahi, M.R. Data fusion approaches for structural health monitoring and system identification: Past, present, and future. *Struct. Health Monit.* **2020**, *19*, 552–586. [CrossRef]
6. Li, S.; Ou, J.; Wang, J.; Gao, X.; Yang, C. Level 2 safety evaluation of concrete-filled steel tubular arch bridges incorporating structural health monitoring and inspection information based on China bridge standards. *Struct. Control Health Monit.* **2019**, *26*, e2303. [CrossRef]
7. Lei, X.; Sun, L.; Xia, Y.; He, T. Vibration-based seismic damage states evaluation for regional concrete beam bridges using random forest method. *Sustainability* **2020**, *12*, 5106. [CrossRef]
8. Feng, M.Q.; Kim, D.K.; Yi, J.H.; Chen, Y.B. Baseline Models for Bridge Performance Monitoring. *J. Eng. Mech-ASCE* **2004**, *130*, 562–569. [CrossRef]
9. Liu, C.; DeWolf, J.T.; Kim, J.H. Development of a baseline for structural health monitoring for a curved post-tensioned concrete box–girder bridge. *Eng. Struct.* **2009**, *31*, 3107–3115. [CrossRef]
10. Liu, Y.; Zhang, S. Probabilistic baseline of finite element model of bridges under environmental temperature changes. *Comput.-Aided Civ. Inf.* **2017**, *32*, 581–598. [CrossRef]
11. Kamariotis, A.; Chatzi, E.; Straub, D. Value of information from vibration-based structural health monitoring extracted via Bayesian model updating. *Mech. Syst. Signal Pract.* **2022**, *166*, 108465. [CrossRef]
12. Zhan, J.; Wang, C.; Fang, Z. Condition Assessment of Joints in Steel Truss Bridges Using a Probabilistic Neural Network and Finite Element Model Updating. *Sustainability* **2021**, *13*, 1474. [CrossRef]
13. Zhou, C.; Chase, J.G.; Rodgers, G.W. Support vector machines for automated modelling of nonlinear structures using health monitoring results. *Mech. Syst. Signal Pract.* **2021**, *149*, 107201. [CrossRef]
14. Azim, M.R.; Gül, M. Data-driven damage identification technique for steel truss railroad bridges utilizing principal component analysis of strain response. *Struct. Infrastruct. E* **2021**, *17*, 1019–1035. [CrossRef]

15. Li, X.Y.; Guan, Y.H.; Law, S.S.; Zhao, W. Monitoring abnormal vibration and structural health conditions of an in-service structure from its SHM data. *J. Sound Vib.* **2022**, *537*, 117185. [CrossRef]
16. Barbosh, M.; Sadhu, A.; Vogrig, M. Multisensor-based hybrid empirical mode decomposition method towards system identification of structures. *Struct. Control Health Monit.* **2018**, *25*, e2147. [CrossRef]
17. Li, S.; Xu, H.; Zhang, X.; Gao, M.S.; Sumarac, D.; Novak, D. Automatic uncoupling of massive dynamic strains induced by vehicle-and temperature-loads for monitoring of operating bridges. *Mech. Syst. Signal Pract.* **2022**, *166*, 108332. [CrossRef]
18. Worden, K.; Baldacchino, T.; Rowson, J.; Cross, E.J. Some recent developments in SHM based on nonstationary time series analysis. *Proc. IEEE* **2016**, *104*, 1589–1603. [CrossRef]
19. Mosavi, A.A.; Seracino, R.; Rizkalla, S. Effect of temperature on daily modal variability of a steel-concrete composite bridge. *J. Bridge Eng.* **2012**, *17*, 979–983. [CrossRef]
20. Catbas, F.N.; Susoy, M.; Frangopol, D.M. Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data. *Eng. Struct.* **2008**, *30*, 2347–2359. [CrossRef]
21. Peeters, B.; De Roeck, G. One-year monitoring of the Z 24-Bridge: Environmental effects versus damage events. *Earthq. Eng. Struct. D* **2001**, *30*, 149–171. [CrossRef]
22. Maeck, J.; Peeters, B.; De Roeck, G. Damage identification on the Z24-bridge using vibration monitoring. *Smart Mater. Struct.* **2001**, *10*, 512–517. [CrossRef]
23. Xu, Y.L.; Chen, B.; Ng, C.L.; Wong, K.Y.; Chan, W.Y. Monitoring temperature effect on a long suspension bridge. *Struct. Control Health Monit.* **2010**, *17*, 632–653. [CrossRef]
24. Cao, Y.; Yim, J.; Zhao, Y.; Wang, M. Temperature effects on cable stayed bridge using health monitoring system: A case study. *Struct. Health Monit.* **2011**, *10*, 523–537.
25. Xiao, F.; Hulsey, J.L.; Balasubramanian, R. Fiber optic health monitoring and temperature behavior of bridge in cold region. *Struct. Control Health Monit.* **2017**, *24*, e2020. [CrossRef]
26. Oh, C.K.; Sohn, H. Damage diagnosis under environmental and operational variations using unsupervised support vector machine. *J. Sound Vib.* **2009**, *325*, 224–239. [CrossRef]
27. Yan, A.M.; Kerschen, G.; De Boe, P.; Golinval, J.C. Structural damage diagnosis under varying environmental conditions—part I: A linear analysis. *Mech. Syst. Signal Pract.* **2005**, *19*, 847–864. [CrossRef]
28. Yan, A.M.; Kerschen, G.; De Boe, P.; Golinval, J.C. Structural damage diagnosis under varying environmental conditions—Part II: Local PCA for non-linear cases. *Mech. Syst. Signal Pract.* **2005**, *19*, 865–880. [CrossRef]
29. Deraemaeker, A.; Reynders, E.; De Roeck, G.; Kullaa, J. Vibration-based structural health monitoring using output-only measurements under changing environment. *Mech. Syst. Signal Pract.* **2008**, *22*, 34–56. [CrossRef]
30. Soleimani-Babakamali, M.H.; Sepasdar, R.; Nasrollahzadeh, K.; Sarlo, R. A system reliability approach to real-time unsupervised structural health monitoring without prior information. *Mech. Syst. Signal Pract.* **2022**, *171*, 108913. [CrossRef]
31. Zhang, S.; Liu, Y. Damage detection of bridges monitored within one cluster based on the residual between the cumulative distribution functions of strain monitoring data. *Struct. Health Monit.* **2020**, *19*, 1764–1789. [CrossRef]
32. Deraemaeker, A.; Worden, K. A comparison of linear approaches to filter out environmental effects in structural health monitoring. *Mech. Syst. Signal Pract.* **2018**, *105*, 1–15. [CrossRef]