Damage detection under temperature conditions using PCA - an application to the Z24 Bridge

William Soo Lon Wah Ph.D.
Department of Civil Engineering, Waikato Institute of Technology
Yung-Tsang Chen Ph.D.
Department of Civil Engineering, The University of Nottingham Ningbo

Ahmed Elamin Ph.D.
Department of Civil Engineering, The University of Nottingham Ningbo
Gethin W Roberts Ph.D.
University of the Faroe Islands

Vibration properties of civil structures, which are commonly analysed for damage detection, are affected by the changing environmental and operational conditions, and most notably are subjected to bilinear effects from the changing ambient temperature conditions. Therefore, damage detection of structures during the past decade has focused on eliminating the effects of the changing environments affecting the vibration properties of structures. Several methods have been proposed in the literature to tackle the nonlinear effects from the changing environments. However, these methods can only analyse systems (e.g. natural frequency-environmental and operational system) that have an incremental change in relationship; they cannot model the bilinear effects from the changing temperature conditions, which may lead to false alerts. Hence, a damage detection method is proposed in this paper to tackle the piecewise effects from the changing temperature conditions. The method makes use of Gaussian Mixture Model to separate the different effects affecting the structures and uses Principal Component Analysis for data processing. The method is applied to the Z24 Bridge, in Switzerland which was subjected to bilinear effects from the changing temperature conditions. The results obtained demonstrate that the proposed method successfully takes into account the piecewise effects to indicate the presence of damage.

Notation

$\mathbf{A}$	Matrix of singular values
a	Coefficient defining the model of extreme observations
b	Coefficient defining the model of extreme observations
c	Coefficient defining the model of extreme observations
d	Deviation index
$\mathbf{D}$	Original matrix of natural frequencies
$\mathbf{I}$	Identity matrix
LCL	Lower control limit
m	Number of principal components selected
$\overline{NI}$	Standard deviation of training data set
ob	Monitored observation
p	Number of modes of vibration
$P1$	First principal component score
$P2$	Second principal component score
q	Number of observations
$\mathbf{S}$	Matrix of natural frequencies after mean centering
$\mathbf{T}$	Transformation matrix
$\mathbf{U}$	Matrix of eigenvectors of covariance matrix of $\mathbf{S}$
UCL	Upper control limit
$\mathbf{Y}$	Score matrix
μ	Mean of training data set
σ	Singular value

1. Introduction

Damage detection of civil engineering structures during the past decade has focused on eliminating the effects of the changing environmental and operational conditions affecting the vibration properties of structures. This is because, the changing environmental and operational conditions affect the vibration properties of structures which are commonly analysed for damage detection (Li *et al.* (2011); Soo Lon Wah and Chen (2018)). If these conditions are not taken into account, false alerts may occur.

Several studies have been conducted in the literature to determine which environmental and operational conditions affect the vibration properties of structures. It was found that temperature was the dominant factor affecting the properties (Farrar *et al.* (1996); Alampalli (1998); Rohrmann (2000); Peeters and De Roeck (2001); Ko *et al.* (2003); Cross *et al.* (2013)), and usually a linear or bilinear

relationship exists between the ambient temperature conditions and the natural frequencies of structures (Peeters and De Roeck (2001); Moser and Moaveni (2011); Jin et al. (2015); Salcher et al. (2016)). Two different relationships exist for temperature conditions below and above zero degrees, respectively. For example, the first four natural frequencies of the Z24 Bridge in Switzerland had a bilinear relationship with the ambient temperature conditions (Peeters and De Roeck (2001)). Peeters and De Roeck (2001) suggested that the bilinear relationship was attributed to the asphalt layer on the structure. The asphalt layer at temperatures below zero degrees contributed to the stiffness of the structure, while at warmer temperatures it had less influence. For the Meriden Bridge, a bilinear relationship also existed between the natural frequencies of the structure and the temperature conditions (Jin et al. (2015)). This bilinear relationship was attributed to the freezing of the supports of the bridge and the asphalt layer on the structure (Jin et al. (2015)). Similar bilinear relationships were also found for the Yonghe Bridge in China, a steel truss footbridge (Moser and Moaveni (2011)), and a railway bridge (Salcher et al. (2016)). Therefore, it is important to take into account the bilinear effects from the changing ambient temperature conditions to avoid false alerts.

Besides the effects of temperature, other environmental and operational conditions (e.g. traffic loading) were also found to affect the vibration properties of civil engineering structures. For example, traffic loadings on bridge structures were found to have an effect on the vibration properties (Zhang et al. (2002); Cross et al. (2013); Cabboi (2014); Meng et al. (2018); Mao et al. (2018)). Humidity and wind loading were also found to contribute to the variations of the vibration properties of civil engineering structures (Xia et al. (2006); Siringoringo and Fujino (2008); Cross et al. (2013)). Although several environmental and operational conditions affect the vibration properties of structures, many researchers have concluded that temperature is the dominant effect affecting the properties (Cabboi (2014); Meng et al. (2018); Mao et al. (2018)).

Several approaches have been adopted in the literature for damage detection under changing environmental and operational conditions. For example, one popular approach is to create a natural frequency-temperature model using a regression analysis to be used for future comparison (Worden et al. (2002); Dervilis et al. (2015)). The model is used to represent the normal conditions (undamaged conditions) of the structure and any large deviation from the model is attributed to damage. Several methods have been proposed in the literature using a linear or nonlinear regression analysis to take into account the linear or nonlinear effects from the changing environments (Peeters and De Roeck (2001); Worden et al. (2002); Dervilis et al. (2015); Hu et al. (2017)). However, these methods cannot analysis systems (e.g. temperature-natural frequency system) where piecewise effects from the changing environments are present. The piecewise effect is when an abrupt

change in relationship (e.g. relationship between temperature and natural frequency) is created by the changing environmental and operational conditions which can be represented by several linear relationships. The methods can only analyse systems where the change in relationship is incremental (i.e. the relationship changes gradually and can be represented by curves). Therefore, the models created may not represent the exact situations which may lead to false alerts.

Another popular approach proposed for damage detection is to use linear multivariate statistical tools (e.g. Principal Component Analysis (PCA)) to extract new non-correlated features to separate the effects of damage from the effects of the changing environmental and operational conditions. By analysing the extracted features, damage detection can be achieved (Manson (2002); Yan et al. (2005a); Cross et al. (2011)). This approach does not require the environmental and operational conditions for damage detection, which is an advantage over the regression analysis approach. In this approach, each new feature created will represent an effect (e.g. damage effect or temperature effect); therefore, the performance of the methods depends on the success of separating the different effects using the new features and on choosing the number of features to analyse for damage detection. Although conceptually simple, this approach has the limitation that the number of features to extract and analyse for damage detection is difficult to determine (Santos et al. (2013); Comanducci et al. (2016)). This is because, it is difficult to determine which feature represents which effect. If too many or too few features are extracted and analysed, the effects of damage may be overestimated or underestimated, which may lead to false alerts.

Extension of the linear approach using nonlinear analysis tools has been proposed in the literature to take into account the nonlinear effects from the changing environments (Hsu and Loh (2010); Santos et al. (2015); Shi et al. (2016)). The principle behind these methods is to use nonlinear analysis tools to extract the nonlinear relationships among the features being analysed to be used for damage detection. However, similar to the nonlinear regression analysis approach, these methods do not perform well for cases where piecewise effects from the changing environmental and operational conditions, exist (Fuentes (2017)). To take the piecewise effects into account, researchers have instead proposed to create piecewise-linear models, followed by the application of the multivariate statistical tools for damage detection (Kerschen and Golinval (2002); Yan et al. (2005b); Kullaa (2014)). This approach consists of first clustering the nonlinear data set into several smaller linear data sets, followed by the application of the damage detection methods on each set. Although these methods can take into account the piecewise effects, they face the same problem as the linear methods. It is difficult to determine the number of features to analyse, which may lead to false alerts.

Therefore, a damage detection method is proposed in this paper to take into account the piecewise effects from the changing environmental and operational conditions. The proposed method combines the multivariate approach (the use of PCA) with the regression approach to overcome the problem of the piecewise effects and the problem of choosing the number of features to analyse. In this paper only the first two features are analysed for damage detection. Gaussian Mixture Model (GMM), which is a clustering technique, is used to group the observations into different data sets based on the environmental and operational conditions they face. PCA and an outlier analysis are then adopted to determine whether the structure is damaged or not.

The proposed method is applied to the Z24 Bridge in this paper. The bridge structure was subjected to bilinear effects from the changing temperature conditions. The results obtained demonstrate that the undamaged and damaged cases are well classified even though the structure was subjected to bilinear effects from the changing temperature conditions. PCA was adopted to extract new features and only the first two principal components were analysed for damage detection. This eliminates the problem of choosing the number of features to analyse.

2. Methodology

The damage detection method developed in this paper is introduced in this section. PCA, GMM and outlier analysis, which will be used for data processing are first introduced, followed by a detailed description of the damage detection method.

2.1. Principal Component Analysis

PCA is a multivariate statistical tool used to reduce the dimensions of a data set while still retaining most of the information. It creates new non-correlated variables (latent variables) termed 'principal components' to represent the different factors affecting the data set. These principal components are obtained through a rotational transformation of the original data set as shown in Figure 1. These components highlight the directions of maximum variances in the data set which represent the main factors affecting the observations. Mathematically, these principal components are formed by a linear combinations of the variables in the data set (Jolliffe (2011)). For damage detection, these principal components may represent the different directions of the variations of vibration properties caused by the changing environmental and operational conditions, and damage. Thus, natural frequencies can be processed using PCA to highlight the different directions where the frequencies are affected. A description of the principle behind PCA is given below.

Consider a $p \times q$ data set $\mathbf{D}$ composed of natural frequencies captured from q observations. For each observation, p natural
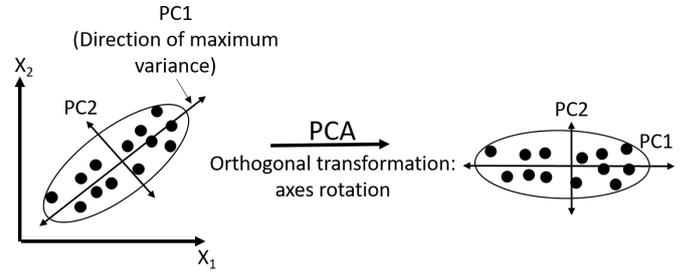

Figure 1. Orthogonal transformation with PCA.

frequencies are obtained, therefore, p represents the number of modes of vibration selected.

$$(1) \quad \mathbf{D} = \begin{bmatrix} D_{1,1} & \dots & D_{1,q} \\ \vdots & \ddots & \vdots \\ D_{p,1} & \dots & D_{p,q} \end{bmatrix}$$

To perform PCA on the data set $\mathbf{D}$, mean centering of the set is first required. This can be achieved by subtracting the mean of each row of the data set from each measurement in that row. Note that mean centering does not change the relative location of each point to each other; it only changes the centroid location to zero coordinates and the mean to zero. The resulting matrix $\mathbf{S}$ after mean centering $\mathbf{D}$ has the same dimensions ($p \times q$) as the original matrix $\mathbf{D}$.

PCA transforms the data set $\mathbf{S}$ into a new data set $\mathbf{Y}$ (dimensions $p \times q$) which contains the principal components highlighting the main factors contributing to the variances in the original data set. The relationship between the newly formed data set $\mathbf{Y}$ and the data set $\mathbf{S}$ can be expressed using a $p \times p$ matrix $\mathbf{T}$.

$$(2) \quad \mathbf{Y} = \mathbf{TS}$$

where,

$\mathbf{Y}$, called the score matrix, combines all the principal components into one single matrix, and

$\mathbf{T}$, termed the loading matrix, contains coefficients used to compute the score matrix $\mathbf{Y}$.

The score matrix $\mathbf{Y}$ contains the principal components of the data set $\mathbf{S}$. These components are arranged in descending order in the matrix according to the amount of variances they contribute in the original data set. The first principal component represents the

factor(s) creating the largest variance in the data set, the second principal component represents the second largest variance, and so on.

In each principal component, every observation is assigned to a score. The score of each observation for each principal component can be thought of as a coordinate representing the location of that observation along that principal component axis. Thus, the first principal component will have the largest span of scores while the last principal component will have the smallest span as shown in Figure 1. PCA makes the first principal component to have the largest span by rotating the cloud of data of the data set in such a way to minimise the distance of each point in the cloud to the first principal component axis, while assuring that the axis goes through the zero centroid. It is for this reason that PCA requires mean centering of the original data set $\mathbf{D}$ prior to application.

The loading matrix $\mathbf{T}$ contains the coefficients which are used to compute the score matrix $\mathbf{Y}$. The loading matrix can be obtained by extracting the eigenvectors of the covariance matrix of $\mathbf{S}$. To obtain $\mathbf{T}$, the matrix $\mathbf{S}$ is first decomposed using singular value decomposition, followed by using that decomposition to construct the covariance matrix of $\mathbf{S}$ as follows.

$$(3) \qquad \frac{1}{q-1}\mathbf{S}\mathbf{S}^T = \mathbf{U}\frac{\mathbf{A}^2}{q-1}\mathbf{U}^T$$

Where, $\mathbf{U}$ is an orthonormal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) whose columns represent the eigenvectors of the covariance matrix of $\mathbf{S}$ (hence $\mathbf{T} = \mathbf{U}^T$), and $\mathbf{A}$ is given as

$$(4) \qquad \mathbf{A} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

with the diagonal terms representing the singular values $\mathbf{A}_1$ and $\mathbf{A}_2$, where $\mathbf{A}_1 = diag(\sigma_1, \sigma_2, ..., \sigma_m)$ and $\mathbf{A}_2 = diag(\sigma_{m+1}, \sigma_{m+2}, ..., \sigma_p)$.

To reduce the dimensions of the data set, only the first m eigenvectors are used to reconstruct the data set, hence, $\mathbf{Y}$ will have dimensions $m \times q$.

2.2. Gaussian Mixture Model

GMM is used in this paper to cluster observations into different data sets based on their environmental and operational conditions, thus, a brief introduction on the application of GMM in the context of this paper is given here. For more detailed information on the principle behind GMM, the readers are referred to Bishop (2006). GMM is a probabilistic clustering method which assumes that a data

set which is not normally distributed can be represented by a set of normally distributed components. Thus, observations captured under piecewise effects of changing environmental and operational conditions, can be clustered into different data sets (e.g. one data set for temperatures below zero degrees and one data set above zero degrees) before the application of damage detection methods.

GMM is used as the clustering technique because it uses the mean and covariance of the clusters as basis of clustering. This has the advantage that the relationship between the variables being clustered are taken into account. Representing the relationship between the different variables is important so that observations subjected to the same environmental and operational effects can be grouped together. For example, Figure 2 shows the clustering of a typical natural frequency-temperature plot. As mentioned previously, natural frequencies of civil engineering structures are usually subjected to bilinear effects from the changing temperature conditions, thus, two clusters surround observations gathered below and above zero degrees.
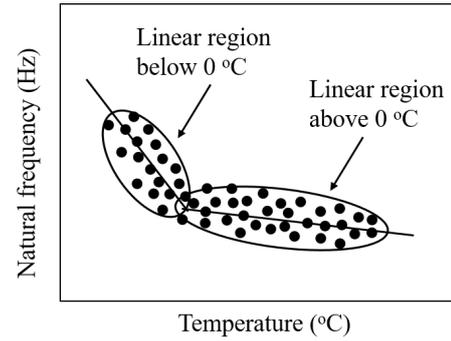

Figure 2. Representation of natural frequency clustering using GMM.

2.3. Outlier analysis

Outlier analysis is commonly performed to classify whether observations gathered are from normal or abnormal conditions. For damage detection, the normal condition represents the undamaged structure subjected to changing environmental and operational conditions, while the abnormal condition is due to damage of structural components. Control charts which are statistical tools are usually adopted for this purpose. Control charts consist of a deviation index (or damage index) plotted against observation number, and horizontal lines (an upper control limit and a lower control limit) which define the range of variations of the deviation index due to normal condition. Any future observation lying outside the control limits can then be attributed to an abnormal condition and in this paper, damage.

The control limits are obtained using a training data set from the normal condition and the limits are calculated using Equation 5 (Yan *et al.* (2005a)). The control limits in Equation 5 represent

the confidence interval that for a normal distribution, 99.7 % of the observations will lie within the control limits.

$$
\begin{aligned}
UCL &= \mu + 3\overline{NI} \\
LCL &= \mu - 3\overline{NI}
\end{aligned}
\tag{5}
$$

where,

UCL is the upper control limit,

LCL is the lower control limit,

μ is the mean of the training data set, and

$\overline{NI}$ is the standard deviation of the training data set.

2.4. Damage detection method

The damage detection method proposed in this paper is described here. The method makes use of PCA to highlight the directions where the variances are the largest. This allows the direction of the variations of the natural frequencies of structures caused by the main environmental and operational conditions, to be highlighted, as well as the direction where damage affects the natural frequencies to be obtained.

Every measurement captured from civil engineering structures can be represented by two components: the model component and the noise component. Figure 3 represents the variations of the first and second natural frequencies of a structure with temperature conditions. In this case, the model component is the line (regression model) across all the undamaged measurements (represented by dots), while the noise component is the direction away from the regression model. The model component represents the main environmental and operational effects affecting the natural frequencies, and in this case, the temperature conditions. The regression model gives the relationship between the natural frequency and temperature conditions, and represents the ideal value of frequency with temperature effect only. From Figure 3, it can be seen that the observations (both the undamaged and damaged) do not lie exactly along the regression model. This is because, other minor environmental and operational conditions (e.g. traffic loading), noise and damage affect the natural frequencies of the structure. The variations in frequencies due to these effects are reflected away from the regression model in the noise space.

Kullaa (2014) mentioned that the effect of damage on the natural frequencies of structures is usually reflected in the noise space as shown in Figure 3. In Figure 3(a), it can be seen that the deviation from the regression model is larger for the damaged observations (represented by circles) than the undamaged

observations. The damaged observations have a larger deviation because, the deviation is due to the changing environmental and operational conditions, which the undamaged observations also face, and damage. Therefore, for damage detection, it is ideal to analyse the noise space to detect the presence of damage. This is the basis of the regression based damaged detection method.

Figure 3(a) is the case where only one mode of vibration (the first natural frequency) is present. However, in reality, more modes of vibration are analysed and usually, each mode of vibration is affected differently by the presence of damage. Some modes of vibration will be affected severely (Figure 3(a)), while other modes of vibration will be less affected (cannot distinguish between damaged and undamaged) similar to Figure 3(b). To detect the presence of damage, several modes of vibration needs to be analysed, which is not practical.

Therefore, to reduce the dimensions of the data set being analysed and to highlight the effects of damage, in this paper, it is proposed to apply PCA on the natural frequencies database. This will highlight the directions of maximum variances (maximum variations of natural frequencies) in the database which is caused by the changing environmental and operational conditions, and damage. It is proposed to use PCA to extract the first two principal components for analysis. The first principal component, which corresponds to the direction with maximum variance, will represent the direction where the changes in natural frequencies is due to the effect of temperature. This is because, as mentioned previously, temperature is the dominant effect creating the variations in natural frequencies of structures. The second principal component, which represents the direction with the second largest variance, will represent the noise direction where other minor environmental and operational conditions (e.g. traffic loading), and damage are usually reflected.

The benefit of applying PCA on the natural frequencies data set is that the dimensions of the set being analysed can be reduced. The plot of the first two principal components can be analysed, instead of analysing several plots of natural frequency-temperature conditions. Another advantage of processing the data set using PCA is that, the direction where damage is most influential can be highlighted. This increases the sensitivity of the method to detect damage. In addition, temperature conditions is not required for analysis which is important for some structures where the environmental and operational measurements are not available during future monitoring.

As mentioned previously, natural frequencies of structures are usually subjected to piecewise effects from the changing environmental and operational conditions, especially the bilinear effects from the changing temperature conditions. To take these effects into account, it is proposed to use different data sets to represent different range of conditions (e.g. one group for temperatures below zero degrees and one group for temperatures
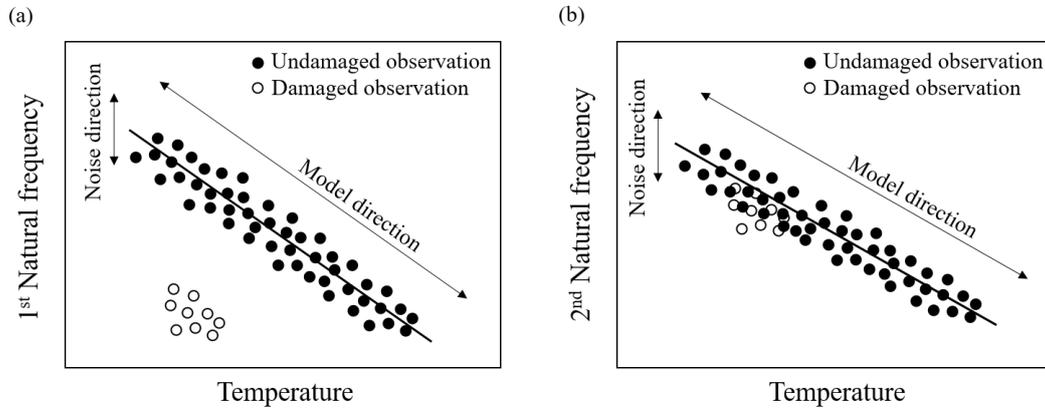
Figure 3. Representation of regression model of (a) first natural frequency and (b) second natural frequency with temperature conditions.

above zero degrees). To separate and create the different data sets, it is proposed to use GMM as the clustering technique. It is proposed to first apply PCA on the observations, followed by the application of GMM on the first principal component and temperature conditions. GMM is applied on the first principal component because the first principal component highlights the main factor affecting the data set of natural frequencies, which is attributed to the main environmental and operational conditions affecting the frequencies.

It should be noted that since different data sets representing different temperature range are used, each new monitored observation needs to be compared to the appropriate data set. To determine the appropriate set, it is proposed to use observations obtained at two extreme and opposite temperature conditions to represent the range of temperature conditions covered by each data set, as shown in Figure 4. Then, the monitored observation can be compared to the appropriate set if it lies within the extreme and opposite temperature conditions as shown in Figure 4. In Figure 4, the blue dots represent observations with temperature conditions below zero degrees, while the red dots represent observations with conditions above zero degrees. The blue and red scales represent the range of observations having the same natural frequency-temperature relationship for conditions below and above zero degrees, respectively. To obtain the observations at the extreme and opposite conditions, a regression analysis can be performed using the observations of each range of conditions.

As mentioned previously, temperature conditions during future monitoring may not be available which may prevent the appropriate range of conditions to be obtained for comparison. To overcome this, it is proposed to add the monitored observations to each extreme and opposite observations data set, followed by the application of PCA on each set. By analysing the first and second
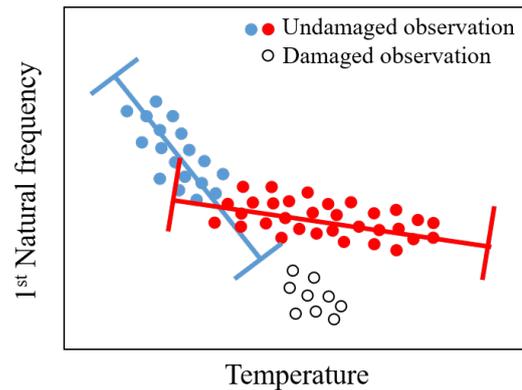
Figure 4. Representation of the range of temperature conditions covered by extreme and opposite temperature conditions.

principal components plot of each data set, the appropriate range of conditions can be obtained as shown in Figure 5.

In Figure 5, two plots are present; one for conditions below zero degrees and one for conditions above zero degrees using the extreme conditions. The crosses in the plots represent the extreme conditions, while the dots and the circles represent the undamaged and damaged cases being analysed, respectively. The blue dots are observations gathered below zero degrees, while the red one are those with conditions above zero degrees. The line joining all the extreme observations together represents the regression model while the dotted lines represent the range covered by that temperature range. Therefore, if the monitored observation lies within the dotted lines, it means that the temperature condition of that observation is in between the range covered by the extreme observations. Thus, this range should be used for analysis. For

(a) Temperature range below zero degrees covered by extreme observations

(b) Temperature range above zero degrees covered by extreme observations
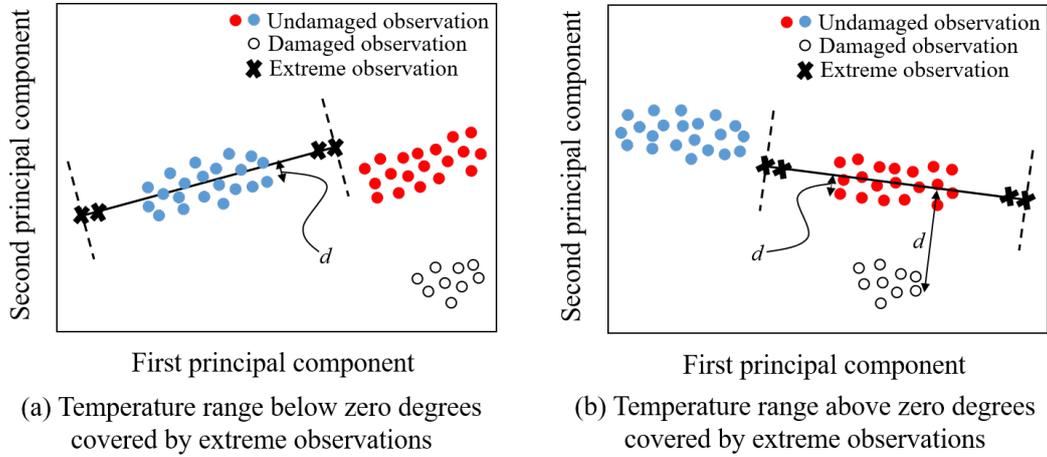
Figure 5. Representation of the first and second principal components plot for temperature range (a) below zero degrees and (b) above zero degrees.

example, in Figure 5, the undamaged cases obtained above zero degrees (red dots) and the damaged cases (circles) lie within the range of the temperature conditions above zero degrees only (Figure 5(b)), therefore, only that range needs to be analysed for damage detection. Similarly, only the range below zero degrees should be analysed for the blue dots. It should be noted that the dotted lines are perpendicular to the regression model.

In Figure 5, it can also be seen that the monitored observations (both damaged and undamaged) do not lie exactly along the regression model. This is because these observations are subjected to the effects of different environmental and operational conditions (e.g. traffic loading), noise and damage. The regression model is an ideal case where temperature is the only effect affecting the observations. The deviation from the model is represented by 'd' and is perpendicular to the model. The deviation 'd' can be obtained as in Equation 6.

$$(6) \qquad d = \frac{|a(P1_{ob}) + b(P2_{ob}) + c|}{\sqrt{a^2 + b^2}}$$

where, a, b and c are the coefficients defining the line $aP1 + bP2 + c = 0$ joining all the extreme observations together,

$P1_{ob}$ is the first principal component score of the monitored observation ob, and

$P2_{ob}$ is the second principal component score of the monitored observation ob.

In Figure 5, the deviation from the regression model is larger for the damaged cases than the undamaged cases. This is because, in addition to the effects of the changing environmental and operational conditions, damage is also affecting the natural frequencies of the structure. To determine whether the deviation 'd' is due to the effects of the changing environmental and operational conditions only or also due to the effect of damage, an outlier analysis can be performed on this deviation. If the deviation is within the range of normal conditions, no damage alert is raised, but if the deviation is outside the limits, the damage alert is raised. The control limits for the outlier analysis can be obtained using Equation 5. It should be noted that Figure 5 contains many observations being analysed, however, for damage detection, it is proposed to analyse one measurement at a time to promote near real-time monitoring. Therefore, the plots will only contain the crosses and a dot or a circle depending on the case being analysed.

To summarise, it is proposed to first obtained a database of natural frequencies and corresponding temperature measurements captured from the undamaged state of the structure under consideration. PCA and GMM can be applied on the database for data clustering to create different data sets representing different range of temperature conditions. A data set containing observations obtained at two extreme and opposite temperature conditions can then be created using a regression analysis for each temperature range (each new data set obtained after clustering). New measurements to be analysed can be added one at a time to each extreme data set, followed by the application of PCA. The first and second principal components can be extracted and analysed to obtain the deviation 'd' and to determine which range of temperature conditions to analyse for damage detection. An outlier analysis can then be performed on the deviation 'd' to classify the cases. The

observations used to create the control limits in the outlier analysis are the measurements of the data sets obtained after clustering.

A flow chart on the procedures to follow for damage detection is also given in Figure 6.

3. Z24 Bridge

The damage detection method is applied to the Z24 Bridge in this section. A brief introduction on the bridge structure is first given, followed by the application of the proposed method and the presentation of the results obtained.

3.1. Description of the Z24 Bridge

The Z24 Bridge (Figure 7) was a three-span bridge in Switzerland connecting Koppigen and Utzenstorf, and overpassing the A1 highway between Bern and Zurich. It was monitored for almost a year to capture several environmental and operational parameters (e.g. temperature) it was faced with, as well as acceleration measurements of the structure. The acceleration measurements were captured for almost every hour, and an automatic system identification system was in place to derive the natural frequencies of the structure.

The bridge was subjected to progressive damage scenarios near the end of the monitoring period to create a case study for the structural health monitoring community. The damaged cases the structure was subjected with are presented in Table 1; a more detailed information on the cases can be found in Krämer *et al.* (1999).

Table 1. Description of the progressive damage cases applied to the Z24 Bridge (Adapted from Reynders *et al.* (2014)).

Test Nº	Case description
1	Reference state
2	System installation for pier settlement
3	20 mm settlement of pier
4	40 mm settlement of pier
5	80 mm settlement of pier
6	95 mm settlement of pier
7	Foundation tilt
8	New reference state
9	12 m^2 chipping of concrete
10	24 m^2 chipping of concrete
11	Landslide
12	Concrete hinges failure
13	2 anchor heads failure
14	4 anchor heads failure
15	Rupture of 2 tendons out of 16
16	Rupture of 4 tendons out of 16
17	Rupture of 6 tendons out of 16

The first four natural frequencies of the structure along with several environmental parameters were made available to researchers. The plot of the first four natural frequencies of the bridge structure with ambient temperature conditions is given in Figure 8. From Figure 8, it can be seen that a bilinear relationship exists between the

frequencies and the temperature conditions. Peeters and De Roeck (2001) suggested that this bilinear relationship was attributed to the asphalt layer on the bridge, which at temperatures below zero degrees contributed to the stiffness of the structure, while at warmer temperatures, it had limited influence. The damage detection method proposed in this paper can be applied to take into account this bilinear effect.

3.2. Damage detection of the Z24 Bridge

The damage detection method developed in this paper is applied to the natural frequencies of the Z24 Bridge for damage detection. The first 2000 undamaged observations are used to create a database to represent the normal undamaged state of the structure, and to create the extreme observations to be used for damage detection. The rest of the undamaged observations and all the damaged cases are analysed using the proposed method. The first 2000 observations cover temperature conditions below and above zero degrees. Including measurements below and above zero degrees in the database allow the bilinear effects from the changing temperature conditions to be taken into account. It should be noted that the natural frequencies of some observations were not extracted, therefore, these observations cannot be analysed.

PCA and GMM are first applied to the 2000 observations database for data clustering. The first principal component with temperature conditions of the observations are analysed using GMM to group the observations under the same temperature effects, together as shown in Figure 9. From Figure 9, it can be seen that two clusters exist, which represent observations gathered below (blue dots) and above zero degrees (red dots). These two groups of observations are formed because two different temperature mechanisms affected the natural frequencies of the bridge structure.

After clustering, the observations of each group are used to create the extreme and opposite limits for damage detection. A linear regression analysis is used to obtain the natural frequencies of the observations at the extreme and opposite temperature conditions. The group of extreme observations which covers a temperature range below zero degrees is termed 'cold baseline', while the group which covers a range of temperature conditions above zero degrees is termed 'hot baseline'. The temperature conditions adopted for the cold baseline are -40 °C to -36 °C and 2 °C to 6 °C, while for the hot one the temperatures are -20 °C to -16 °C and 76 °C to 80 °C with 1 °C interval. Some temperature conditions of each baseline overlap with each other because natural frequencies around 0 °C may fall in any of the two baselines due to the effects of other minor environmental and operational conditions. Thus, to cover all possible effects using the extreme observations, the temperature condition is extended. For the hot baseline, the range of temperature considered covers a large part of temperature conditions below zero degrees. For the cold baseline, it covers a smaller range of conditions above zero degrees. This is because, the linear regression of the cold baseline is steeper than that of the hot one. If a larger
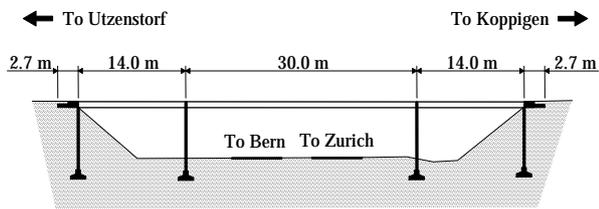
Start

Obtain the structure at its undamaged state.

Collect natural frequencies in the form of data set **D** in Equation 1 and corresponding temperature measurements from the undamaged structure.

Apply PCA on data set **D** to obtain the principal components (score matrix **Y** in Equation 2).

Plot the first principal component against temperature, and apply GMM for data clustering.

Create the extreme observations data set using natural frequencies obtained at two extreme and opposite temperature conditions for each (cold and hot) cluster. A regression analysis can be adopted to obtained the observations at the extreme conditions.

Add the natural frequencies of each cluster to the corresponding extreme data set one at a time to form data set **D** in the form of Equation 1, and apply PCA for data processing.

Plot the first principal component against the second principal component similar to Figure 5, followed by recording the deviation 'd between the monitored case and the extreme observations using Equation 6. Gather the deviations of all the measurements together.

Calculate the mean and standard deviation of the collected deviation values and calculate the UCL using Equation 5 to create a control chart for each group of measurements.

Obtain the structure at a later stage for monitoring.

Collect natural frequencies only for new observations and add them one at a time to each extreme data set created to form data set **D** in the form of Equation 1. Apply PCA on each newly formed data set to obtain the principal components (score matrix **Y**).

Plot the first principal component against the second principal component and retain the plot where the monitored observation falls in the range covered by the extreme observations.

Record the deviation 'd between the monitored observation and the extreme observations.

Compare the deviation to the control limit (UCL) created in the control chart.

Monitored observation outside control limits?

No → No damage is present → End

Yes → Damage is present → End

Figure 6. Flow chart of the procedures to follow for damage detection.

range of temperature was adopted for the cold baseline, most of the observations gathered above zero degrees would have fallen into the cold baseline, as the frequencies of the extreme cases would have dropped consequently due to the steep gradient.

After the extreme observations data sets have been created, the testing cases (all the undamaged (except the first 2000 observations) and damaged cases) can be analysed. One measurement is added at a time to each baseline (cold and hot baselines), and PCA is applied for data processing. The first and second principal components are

Figure 7. Representation of the Z24 Bridge (adapted from Peeters and De Roeck (2001)).

extracted and analysed to obtain the deviation 'd'. The plot of the first principal component against the second principal component of a randomly chosen undamaged case and a randomly chosen damaged case is given in Figure 10. In this figure, the dots represent the extreme and opposite observations, while the cross represents the monitored observation. The line that joins all the extreme observations together is the model, while the dotted lines, which are perpendicular to the model, define the range of conditions covered by the extreme cases. The deviation of the monitored case from the model is given by 'd' and is perpendicular to the model.

For the undamaged case given in Figure 10(a), it can be seen that the monitored case lies in between the range covered by the extreme observations in both baselines. Therefore, both baselines are analysed for damage detection. The monitored observation falls in between both range of conditions because, the temperature condition of that observation is around zero degrees. Since both extreme data sets covered temperature conditions around zero degrees, the monitored observation falls in between both range of conditions. For the damaged case (Figure 10(b)), only the hot baseline should be analysed using the outlier analysis because the monitored observation falls in between the range of the hot baseline only. It can also be seen from Figure 10 that the damaged case has a larger deviation 'd' than the undamaged case. This larger deviation is attributed to the effect of damage which produces an additional effect on the natural frequencies.

After the deviation 'd' has been obtained, an outlier analysis can be performed to determine whether the deviation is due to the effects of environmental and operational conditions only or also due to damage. The results of the outlier analysis for all the undamaged and damaged cases analysed are given in Figure 11. The dots in the figures represent the different cases being analysed while the horizontal dotted line defines the normal range of variations (control limit). To determine the control limits, Equation 5 is used. The training database used to calculate the control limits for the cold baseline and the hot baseline are the observations obtained after clustering was applied to the 2000 undamaged observations. Each observation is added to the extreme observations data set created and PCA is applied to obtained the deviation 'd'. It should be noted that only the UCL is adopted in the outlier analysis. This is because

the deviation 'd' is taken as the magnitude only, and damage is indicated with a large value.

It should be noted that some undamaged observations fell within both baselines, similar to the undamaged case given in Figure 10(a). Therefore, the deviation 'd' in both baselines are analysed using the outlier analysis, for damage detection. These observations are the observations with temperature conditions around zero degrees. For the damaged cases, all the observations fell in the range of the hot baseline only, therefore, only the hot baseline is analysed using the outlier analysis.

From Figure 11, it can be seen that some undamaged cases are well above the UCL. This may be attributed to some errors while collecting the data or while deriving the natural frequencies of the bridge. As mentioned by Soo Lon Wah *et al.* (2019), a database of vibration properties usually contains some observations which are subjected to large errors. These observations are usually not within the normal range of variation.

For the damaged cases, although some observations are also wrongly classified, the performance of the proposed method is deemed satisfactory. The successful rates of the proposed method to indicate healthy condition for the undamaged cases is 99 %, while alerting damage for the damaged cases is 97.2 %. Although the structure was subjected to bilinear effects from the changing temperature conditions, the proposed method successfully takes the piecewise effects into account by using different range of conditions. Moreover, even though the temperature conditions was used to create different data sets to represent the normal conditions of the bridge structure, the proposed method successfully distinguished between the effect of damage, and other environmental and operational conditions (e.g. wind and humidity) acting on the structure.

As mentioned previously, the advantage of using PCA for data processing is that, the dimensions of the data set is reduced while also increasing the sensitivity to detect damage. Moreover, only the first two principal components were analysed which eliminates the risk of overestimating or underestimating the effects of damage by analysing too many or too few components. The temperature conditions of new measurements were also not required during future monitoring.

4. Conclusion

A damage detection method is proposed in this paper to take into account the piecewise effects from the changing environmental and operational conditions, and most notably the bilinear effects from the changing temperature conditions. The method makes use of different data sets to represent different range of temperature conditions affecting the natural frequencies of structures. The method also makes use of PCA for data processing. This has the advantage that the dimensions of the data set to analyse is
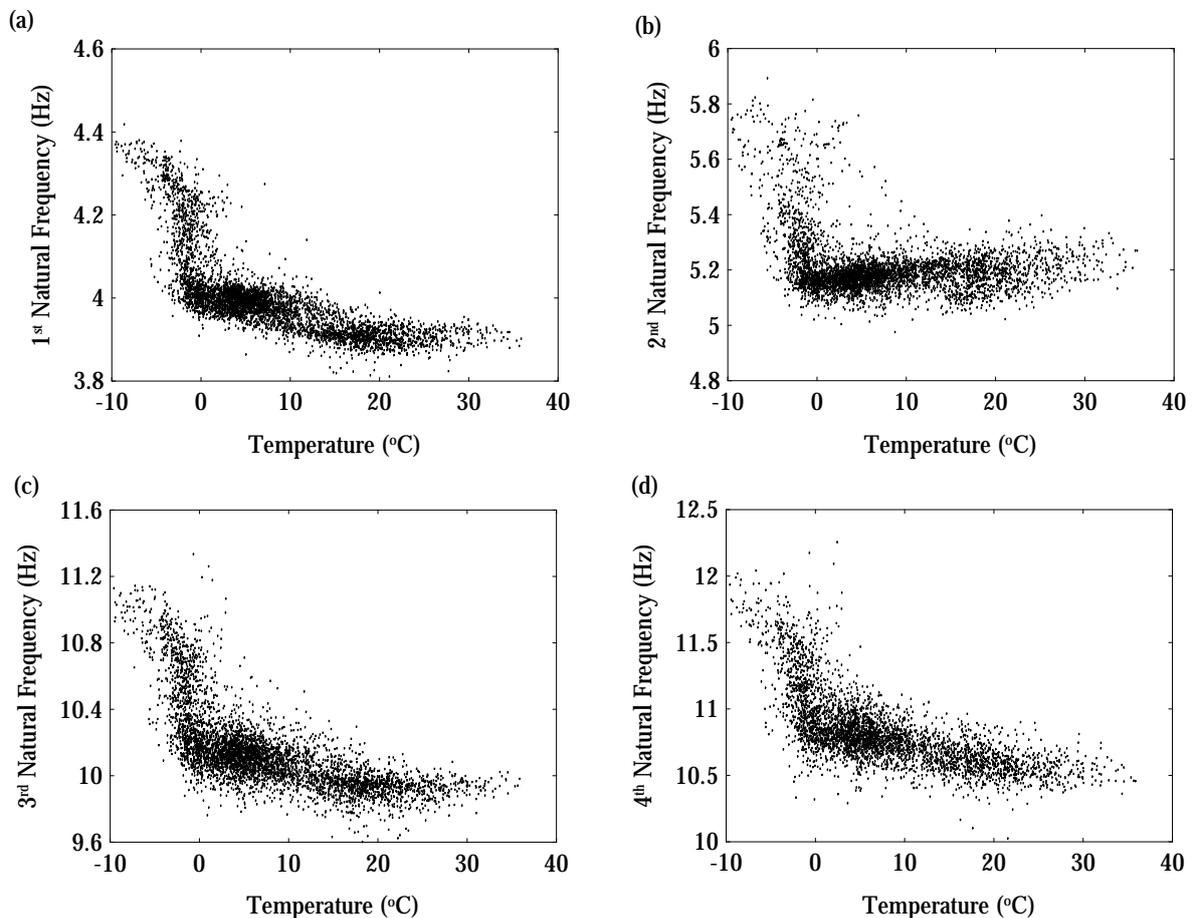
Figure 8. Graph of temperature versus (a) first natural frequency, (b) second natural frequency, (c) third natural frequency and (d) fourth natural frequency of the Z24 Bridge.

reduced while also increasing the sensitivity of the method to detect damage. The proposed method is applied to the Z24 Bridge, in Switzerland which was subjected to bilinear effects from the changing temperature conditions. The results obtained demonstrate that the method successfully classify between the effects of damage and the effects of the changing environmental and operational conditions.

5. Acknowledgment

REFERENCES

Alampalli S (1998) Influence of in-service environment on modal parameters. In *Proceedings-SPIE The International Society For Optical Engineering*, vol. 1, Citeseer, pp. 111–116.

Bishop CM (2006) *Pattern recognition and machine learning*. springer.

Cabboi A (2014) *Automatic operational modal analysis: challenges and applications to historic structures and infrastructures*. PhD thesis, University of Cagliari, Italy.

Comanducci G, Magalhães F, Ubertini F and Cunha Á (2016) On vibration-based damage detection by multivariate statistical techniques: Application to a long-span arch bridge. *Structural Health Monitoring* **15(5)**: 505–524.

Cross E, Koo K, Brownjohn J and Worden K (2013) Long-term monitoring and data analysis of the tamar bridge. *Mechanical Systems and Signal Processing* **35(1-2)**: 16–34.

Cross EJ, Worden K and Chen Q (2011) Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467(2133)**: 2712–2732.

Dervilis N, Worden K and Cross E (2015) On robust regression analysis as a means of exploring environmental and operational
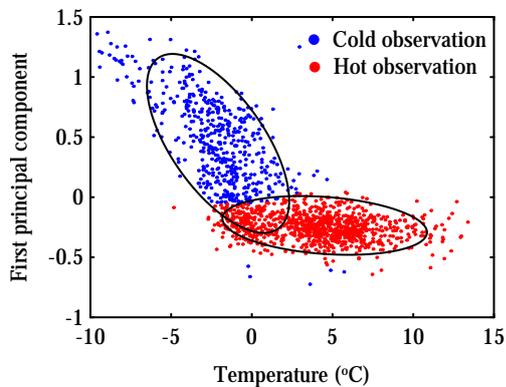
Figure 9. Clustering of first principal component with temperature of the Z24 Bridge.

conditions for shm data. *Journal of Sound and Vibration* **347**: 279–296.

Farrar CR, Doebling SW, Cornwell PJ and Straser EG (1996) *Variability of modal parameters measured on the Alamosa Canyon Bridge. Technical report*, Los Alamos National Lab., NM (United States).

Fuentes R (2017) *On Bayesian Networks for Structural Health and Condition Monitoring*. PhD thesis, University of Sheffield.

Hsu TY and Loh CH (2010) Damage detection accommodating nonlinear environmental effects by nonlinear principal component analysis. *Structural Control and Health Monitoring* **17(3)**: 338–354.

Hu WH, Cunha Á, Caetano E, Rohrmann R, Said S, Teng J *et al.* (2017) Comparison of different statistical approaches for removing environmental/operational effects for massive data continuously collected from footbridges. *Structural Control and Health Monitoring* **24(8)**.

Jin C, Li J, Jang S, Sun X and Christenson R (2015) Structural damage detection for in-service highway bridge under operational and environmental variability. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2015*, vol. 9435, International Society for Optics and Photonics, p. 94353A.

Jolliffe I (2011) *Principal component analysis*. Springer.

Kerschen G and Golinval JC (2002) Non-linear generalization of principal component analysis: from a global to a local approach. *Journal of Sound and Vibration* **254(5)**: 867–876.

Ko JM, Chak K, Wang J, Ni YQ and Chan TH (2003) Formulation of an uncertainty model relating modal parameters and environmental factors by using long-term monitoring data. In *Smart Structures and Materials 2003: Smart Systems and Nondestructive Evaluation for Civil Infrastructures*, vol. 5057, International Society for Optics and Photonics, pp. 298–307.

Krämer C, De Smet C and De Roeck G (1999) Z24 bridge damage detection tests. In *IMAC 17, the International Modal Analysis Conference*, vol. 3727, Society of Photo-optical Instrumentation Engineers, pp. 1023–1029.

Kullaa J (2014) Structural health monitoring under nonlinear environmental or operational influences. *Shock and Vibration* **2014**.

Li H, Huang Y, Ou J and Bao Y (2011) Fractal dimension-based damage detection method for beams with a uniform cross-section. *Computer-Aided Civil and Infrastructure Engineering* **26(3)**: 190–206.

Manson G (2002) Identifying damage sensitive, environment insensitive features for damage detection. In *Proceedings of the third international conference on identification in engineering systems*, pp. 187–197.

Mao JX, Wang H, Feng DM, Tao TY and Zheng WZ (2018) Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition. *Structural Control and Health Monitoring* **25(5)**.

Meng X, Nguyen DT, Xie Y, Owen JS, Psimoulis P, Ince S, Chen Q, Ye J and Bhatia P (2018) Design and implementation of a new system for large bridge monitoring geoshm. *Sensors* **18(3)**: 775–797.

Moser P and Moaveni B (2011) Environmental effects on the identified natural frequencies of the dowling hall footbridge. *Mechanical Systems and Signal Processing* **25(7)**: 2336–2357.

Peeters B and De Roeck G (2001) One-year monitoring of the z24-bridge: environmental effects versus damage events. *Earthquake engineering & structural dynamics* **30(2)**: 149–171.

Reynders E, Wursten G and De Roeck G (2014) Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring* **13(1)**: 82–93.

Rohrmann R (2000) Structural causes of temperature affected modal data of civil structures obtained by long time monitoring .

Salcher P, Pradlwarter H and Adam C (2016) Reliability assessment of railway bridges subjected to high-speed trains considering the effects of seasonal temperature changes. *Engineering Structures* **126**: 712–724.

Santos ADF, Silva MFM, Sales CDS, Costa JCWA and Figueiredo E (2015) Applicability of linear and nonlinear principal component analysis for damage detection. In *Proceedings of The 2015 IEEE International Instrumentation and Measurement Technology Conference*.

Santos JP, Crémona C, Orcesi AD and Silveira P (2013) Multivariate statistical analysis for early damage detection. *Engineering Structures* **56**: 273–285.

Shi H, Worden K and Cross E (2016) A nonlinear cointegration approach with applications to structural health monitoring. In
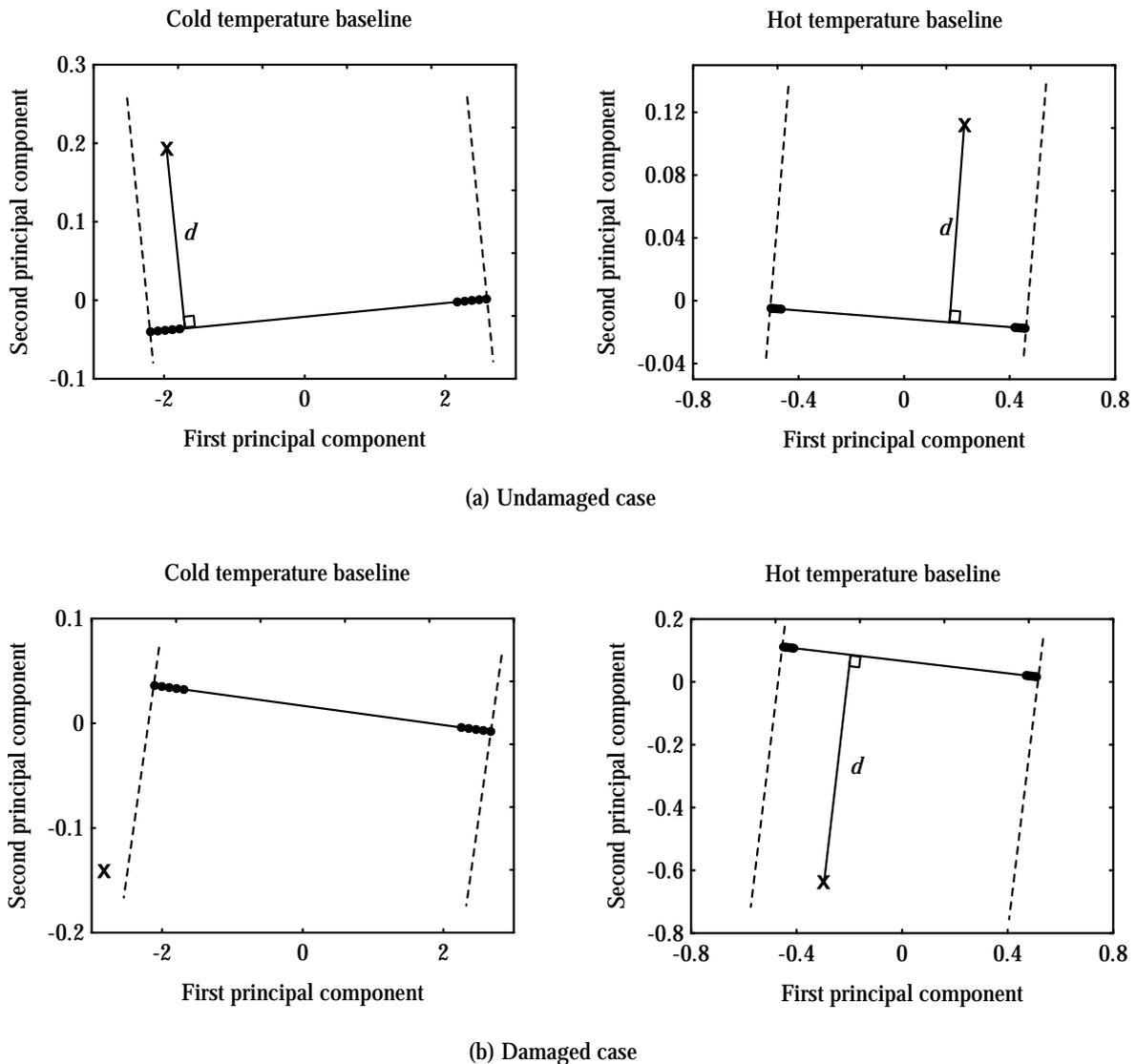
Cold temperature baseline

Hot temperature baseline

(a) Undamaged case

Cold temperature baseline

Hot temperature baseline

(b) Damaged case

Figure 10. Plot of first principal component against second principal component for (a) an undamaged case and (b) a damaged case of the Z24 Bridge.

Journal of Physics: Conference Series, vol. 744, IOP Publishing, p. 012025.

Siringoringo DM and Fujino Y (2008) System identification of suspension bridge from ambient vibration response. *Engineering Structures* **30(2)**: 462–477.

Soo Lon Wah W and Chen YT (2018) A new approach toward damage localization and quantification of structures under changing temperature condition. *Journal of Low Frequency Noise, Vibration and Active Control* : 1–16.

Soo Lon Wah W, Owen JS, Chen YT, Elamin A and Roberts GW (2019) Removal of masking effect for damage detection of

structures. *Engineering Structures* **183**: 646–661.

Worden K, Sohn H and Farrar CR (2002) Novelty detection in a changing environment: regression and interpolation approaches. *Journal of sound and vibration* **258(4)**: 741–761.

Xia Y, Hao H, Zanardo G and Deeks A (2006) Long term vibration monitoring of an rc slab: temperature and humidity effect. *Engineering Structures* **28(3)**: 441–452.

Yan AM, Kerschen G, De Boe P and Golinval JC (2005a) Structural damage diagnosis under varying environmental conditionspart i: a linear analysis. *Mechanical Systems and Signal Processing* **19(4)**: 847–864.
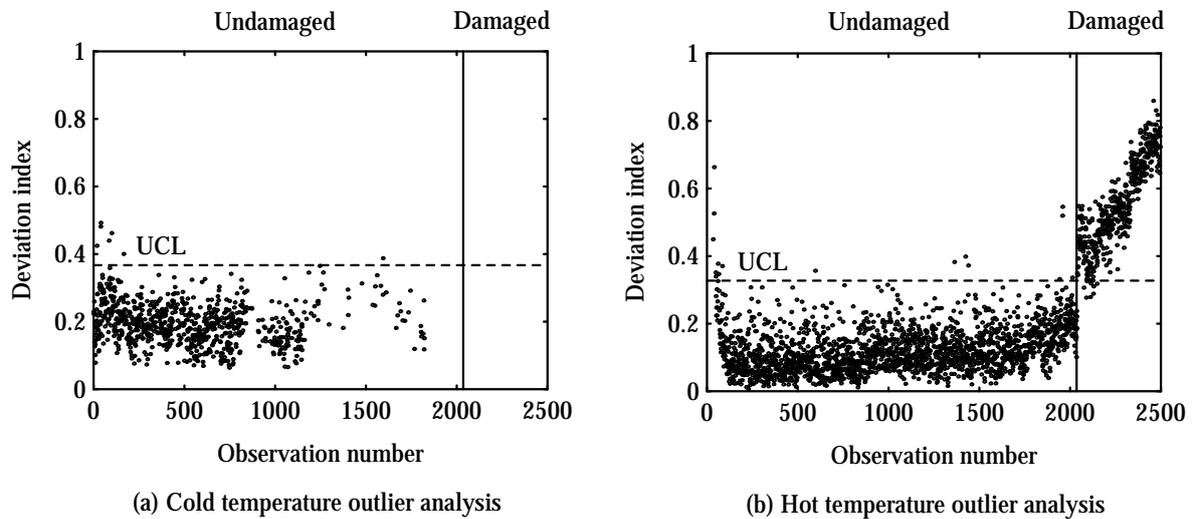
Figure 11. Outlier analysis of undamaged and damaged cases of the Z24 Bridge using (a) cold baseline and (b) hot baseline.

Yan AM, Kerschen G, De Boe P and Golinval JC (2005b) Structural
damage diagnosis under varying environmental conditionspart
ii: local pca for non-linear cases. *Mechanical Systems and
Signal Processing* **19(4)**: 865–880.

Zhang Q, Fan L and Yuan W (2002) Traffic-induced variability in
dynamic properties of cable-stayed bridge. *Earthquake
Engineering and Structural Dynamics* **31(11)**: 2015–2021.