Research Article

Structural Health Monitoring under Nonlinear Environmental or Operational Influences

Jyrki Kullaa[1,2]

[1] *Department of Applied Mechanics, Aalto University, P.O. Box 14300, 00076 Espoo, Finland*
[2] *Helsinki Metropolia University of Applied Sciences, P.O. Box 4021, 00079 Helsinki, Finland*

Correspondence should be addressed to Jyrki Kullaa; jyrki.kullaa@metropolia.fi

Vibration-based structural health monitoring is based on detecting changes in the dynamic characteristics of the structure. It is well known that environmental or operational variations can also have an influence on the vibration properties. If these effects are not taken into account, they can result in false indications of damage. If the environmental or operational variations cause nonlinear effects, they can be compensated using a Gaussian mixture model (GMM) without the measurement of the underlying variables. The number of Gaussian components can also be estimated. For the local linear components, minimum mean square error (MMSE) estimation is applied to eliminate the environmental or operational influences. Damage is detected from the residuals after applying principal component analysis (PCA). Control charts are used for novelty detection. The proposed approach is validated using simulated data and the identified lowest natural frequencies of the Z24 Bridge under temperature variation. Nonlinear models are most effective if the data dimensionality is low. On the other hand, linear models often outperform nonlinear models for high-dimensional data.

1. Introduction

In structural health monitoring (SHM), changes in damage-sensitive features are an indication of damage. Also other sources of deviation are often present, for example, environmental or operational variability. If these effects are not taken into account, they can result in false identifications of damage or a loss of sensitivity to detect minor damage. It is important to distinguish between the two sources of changes in the dynamic characteristics. One option is to make a physical model of different environmental or operational phenomena, but it can be too expensive and inaccurate. An alternative is to include the normal variability in the training data and build a model solely based on the data. Using multivariate statistics, the environmental or operational effects can be eliminated even without measuring the underlying variables (see [1] and the references therein). Also a third source of change in the monitoring data is sensor fault. Kullaa [1] proposed a unified model to distinguish between the three sources of changes in a monitoring system.

Most of the models assume linear correlation between the measured variables or features. However, the environmental or operational variations often cause nonlinear effects. For example, as the temperature falls below zero, its influence on the natural frequencies can change abruptly. This often results also in nonlinear correlation between the features, especially if the data dimensionality is low. On the other hand, a linear model may be sufficient with a large data dimensionality, because the correlation structure may become linear [1]. There are only few studies of nonlinear models. Kullaa [2] used the mixture of factor analyzers [3] model to compensate the nonlinear effects. A similar approach was used by Yan et al. [4] having local PCA models for local regions in the data space. Sohn et al. [5] used an autoassociative neural network that can be thought as a nonlinear PCA [6]. Figueiredo et al. [7] applied the Bayesian approach to a mixture model and the Mahalanobis squared distance for the mixture components.

A nonlinear model is studied in this paper. A Gaussian mixture model (GMM) is proposed in Section 2 to compensate for the nonlinear effects. It is based on the mixture of linear models, each modelling a region in the input space. The approach needs a clustering algorithm to assign each new measurement to the corresponding class. Clustering can be performed independently of the local linear models.

Therefore, clustering is first performed identifying a Gaussian mixture model followed by local linear models to eliminate the underlying effects within each class. The number of classes is often unknown but can also be estimated. Minimum mean square error (MMSE) estimation is applied to the local linear models, which is described in Section 3. Damage can be detected from the residuals between the data and the model. Damage detection is discussed in Section 4.

The first applications in Section 5 are numerical studies, in which the objective is to validate the proposed approach. Section 5.3 shows the experimental results of the Z24 Bridge, in which the natural frequencies varied due to the temperature. Finally, concluding remarks are given in Section 6.

2. Gaussian Mixture Model (GMM)

Let $\mathbf{x}$ be the multivariate measurement data (also subsequently called variables), which can be time series (e.g., acceleration or strains) from a simultaneously sampled sensor network, or a feature vector comprising identified dynamic properties of the structure (e.g., natural frequencies or mode shapes). Nonlinear data are not normally distributed and cannot be modelled as a single Gaussian distribution. One may try a mixture of Gaussian components, in which the distribution can be written as a linear superposition of K Gaussian densities in the form [8]

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \qquad (1)$$

which is called a mixture of Gaussians. Each Gaussian density $N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is called a component of the mixture and has its own mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$. The parameters π_k are called mixing coefficients, which are positive and are summed to one.

The first step is to identify the model parameters. The difficulty lies in the fact that the data points are unlabeled; that is, it is typically not known which component was responsible for generating each data point. The data labels can be considered as latent variables and the expectation-maximization (EM) algorithm can be used to identify the mixture model. It is momentarily assumed that the number of components is known.

The EM algorithm is iterative and consists of two steps: the E step and the M step. In the expectation step, or E step, the model parameters are held fixed and the posterior probability of the component k (latent variable) given the data point $\mathbf{x}$ is evaluated. In the maximization step, or M step, the latent variables are assumed to be known, and the model parameters are obtained by maximizing the log-likelihood function.

A K-dimensional binary random variable $\mathbf{z}$ is introduced having a 1-of-K representation in which a particular element z_k is equal to 1 and all other elements are zero. For an observation $\mathbf{x}_n$, z_{nk} denotes the kth component of $\mathbf{z}_n$.

The algorithm is outlined as follows. In the E step, the expected value of the indicator variable z_{nk} under the posterior distribution is

$$E[z_{nk}] = \frac{\pi_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \qquad (2)$$

In the M step, the model parameters are updated to maximize the log-likelihood function, resulting in [8]

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} E[z_{nk}] \mathbf{x}_n,$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} E[z_{nk}] (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T, \qquad (3)$$

$$\pi_k^{\text{new}} = \frac{N_k}{N},$$

where N is the number of observations and

$$N_k = \sum_{n=1}^{N} E[z_{nk}]. \qquad (4)$$

The log-likelihood is then evaluated:

$$\ln p(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}. \qquad (5)$$

The steps are repeated until the log-likelihood converges. It is not guaranteed that the algorithm converges to the global maximum. Therefore, it is often advised to run the algorithm a couple of times with different initial guesses of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ to find a satisfactory maximum. An example of convergence to a local maximum is given in Section 5.1. Another problem is that the number of components is often unknown. To that end, different models can be identified by varying K, and the model resulting in the highest log-likelihood is chosen. In order to avoid overfitting, a penalty term $-(1/2)M \ln N$ is added to the log-likelihood [8], where N is the number of training samples and M is the number of model parameters:

$$M = K \left[1 + p + \frac{p(p+1)}{2} \right], \qquad (6)$$

where p is the data dimensionality.

Once the model parameters are identified and fixed, the objective is to decide if the new data are generated by the model (undamaged) or by another model (damage). To this end a residual is estimated, which is the difference of the true data point and that estimated by the model:

$$\mathbf{r}_n = \mathbf{x}_n - E[\hat{\mathbf{x}}_n \mid \mathbf{x}_n], \qquad (7)$$

where

$$E[\hat{\mathbf{x}}_n \mid \mathbf{x}_n] = \sum_{k=1}^{K} \hat{\mathbf{x}}_{nk} p(z_{nk} = 1 \mid \mathbf{x}_n). \qquad (8)$$

The last term in the RHS of (8) is obtained using Bayes' theorem:

$$p\left(z_{nk} = 1 \mid \mathbf{x}_n\right) = \frac{p\left(z_{nk} = 1\right) p\left(\mathbf{x}_n \mid z_{nk} = 1\right)}{\sum_{j=1}^{K} p\left(z_{nj} = 1\right) p\left(\mathbf{x}_n \mid z_{nj} = 1\right)}$$

$$= \frac{\pi_k N\left(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{j=1}^{K} \pi_j N\left(\mathbf{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)}, \tag{9}$$

which is the same as (2). The first term in the RHS of (8) is given by the local linear model, which in this paper is the minimum mean square error (MMSE) estimate for each component [9]:

$$\widehat{\mathbf{x}}_{nk} = \boldsymbol{\mu}_k + \mathbf{A}_k\left(\mathbf{x}_n - \boldsymbol{\mu}_k\right), \tag{10}$$

where the coefficient matrix $\mathbf{A}_k$ is composed of rows estimating the variable corresponding to that row using the remaining variables. Therefore, the diagonal components of $\mathbf{A}_k$ are zero. The MMSE estimation is discussed in the next section.

3. Local Linear Models Using MMSE Estimation

With enough redundancy, a subset of observation $\mathbf{x}$ can be estimated using the remaining variables. Each observation is divided into observed variables $\mathbf{v}$ and missing variables $\mathbf{u}$. It is assumed here that $\mathbf{u}$ is the ith variable x_i and the remaining variables are collected in vector $\mathbf{v}$:

$$\mathbf{x} = \begin{Bmatrix} \mathbf{v}_a \\ \mathbf{u} \\ \mathbf{v}_b \end{Bmatrix}, \qquad \mathbf{v} = \begin{Bmatrix} \mathbf{v}_a \\ \mathbf{v}_b \end{Bmatrix}. \tag{11}$$

The partitioned covariance matrix $\boldsymbol{\Sigma}$ of the training data is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uv} \\ \boldsymbol{\Sigma}_{vu} & \boldsymbol{\Sigma}_{vv} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_{uu} & \boldsymbol{\Gamma}_{uv} \\ \boldsymbol{\Gamma}_{vu} & \boldsymbol{\Gamma}_{vv} \end{bmatrix}^{-1}, \tag{12}$$

where the precision matrix $\boldsymbol{\Gamma}$ is defined as the inverse of the covariance matrix $\boldsymbol{\Sigma}$ and is also written in the partitioned form. A linear minimum mean square error (MMSE) estimate for $\mathbf{u} \mid \mathbf{v}$ ($\mathbf{u}$ given $\mathbf{v}$) is obtained by minimizing the mean square error (MSE)

$$\text{MSE} = E\left[\left(\mathbf{u} - \widehat{\mathbf{u}}\right)^T \left(\mathbf{u} - \widehat{\mathbf{u}}\right)\right] \tag{13}$$

resulting in [9]

$$\widehat{\mathbf{u}} = E\left(\mathbf{u} \mid \mathbf{v}\right) = \boldsymbol{\mu}_u - \boldsymbol{\Gamma}_{uu}^{-1} \boldsymbol{\Gamma}_{uv}\left(\mathbf{v} - \boldsymbol{\mu}_v\right)$$

$$= \boldsymbol{\mu}_u + \mathbf{H}\left(\mathbf{v} - \boldsymbol{\mu}_v\right), \tag{14}$$

where $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_v$ are the mean of $\mathbf{u}$ and $\mathbf{v}$, respectively, and $\mathbf{H} = -\boldsymbol{\Gamma}_{uu}^{-1}\boldsymbol{\Gamma}_{uv}$. The error covariance is

$$\text{cov}\left(\mathbf{uv}\right) = \boldsymbol{\Gamma}_{uu}^{-1}. \tag{15}$$

An MMSE model is estimated for each mixture component. For component k, the estimate of variable x_i is given by (14). Then, the ith row of matrix $\mathbf{A}_k$ in (10) is composed of the partitioned row matrix $\mathbf{H}$ and a zero:

$$\left[\mathbf{A}_k\right]_{i,\bullet} = \begin{bmatrix} \mathbf{H}_a & 0 & \mathbf{H}_b \end{bmatrix}. \tag{16}$$

The partitioning should be clear from (11). The zero element hits the diagonal in $\mathbf{A}_k$, originating from the fact that x_i is not used to estimate itself, but all the remaining variables are only used. Therefore, the diagonal elements of matrix $\mathbf{A}_k$ are all zeros. The other rows of $\mathbf{A}_k$ are obtained similarly by estimating all variables in turn using the remaining variables. Matrix $\mathbf{A}_k$ is estimated for each mixture component k.

To show the relation between (10) and (14), compute the estimate of the ith variable x_i for a fixed mixture component k using (10). For clarity, the component index k is omitted. Consider

$$\widehat{\mathbf{x}}_i = \boldsymbol{\mu}_i + \begin{bmatrix} \mathbf{H}_a & 0 & \mathbf{H}_b \end{bmatrix}\left(\mathbf{x} - \boldsymbol{\mu}\right)$$

$$= \boldsymbol{\mu}_i + \begin{bmatrix} \mathbf{H}_a & 0 & \mathbf{H}_b \end{bmatrix} \begin{Bmatrix} \mathbf{v}_a - \boldsymbol{\mu}_a \\ \mathbf{u} - \boldsymbol{\mu}_u \\ \mathbf{v}_b - \boldsymbol{\mu}_b \end{Bmatrix}$$

$$= \boldsymbol{\mu}_i + \begin{bmatrix} \mathbf{H}_a & \mathbf{H}_b \end{bmatrix} \begin{Bmatrix} \mathbf{v}_a - \boldsymbol{\mu}_a \\ \mathbf{v}_b - \boldsymbol{\mu}_b \end{Bmatrix}$$

$$= \boldsymbol{\mu}_i + \mathbf{H}\left(\mathbf{v} - \boldsymbol{\mu}_v\right), \tag{17}$$

which is equal to $\widehat{\mathbf{u}}$ in (14).

4. Damage Detection

Using the mixture model for damage detection introduces an issue of residual scaling, because each class may have a different error variance (15). Therefore, the residual of each variable within each class is divided with the corresponding standard deviation, which is the square root of (15). Also the data dimensionality may be too high for statistical reliability (curse of dimensionality). Therefore, the first principal component scores [10] of residual (7) are used for damage detection. Control charts [11] are used for damage detection. The control chart used in this study is the Shewhart chart [11], and the plotted variable is the subgroup mean of successive observations. It is believed that the robustness of damage detection increases, because (1) additional variability due to environmental or operational influences can be removed, in this paper using a nonlinear model; (2) PCA is applied to the residuals avoiding the curse of dimensionality; and (3) control charts utilize averaging for better statistical reliability.

5. Experimental Results

The proposed nonlinear model and the subsequent SHM functions are applied to two numerical studies and the experimental data of the Z24 Bridge.

5.1. Five Gaussian Components. The first numerical example is a mixture of five Gaussian components in a two-dimensional space. Each component has 10,000 data points. This
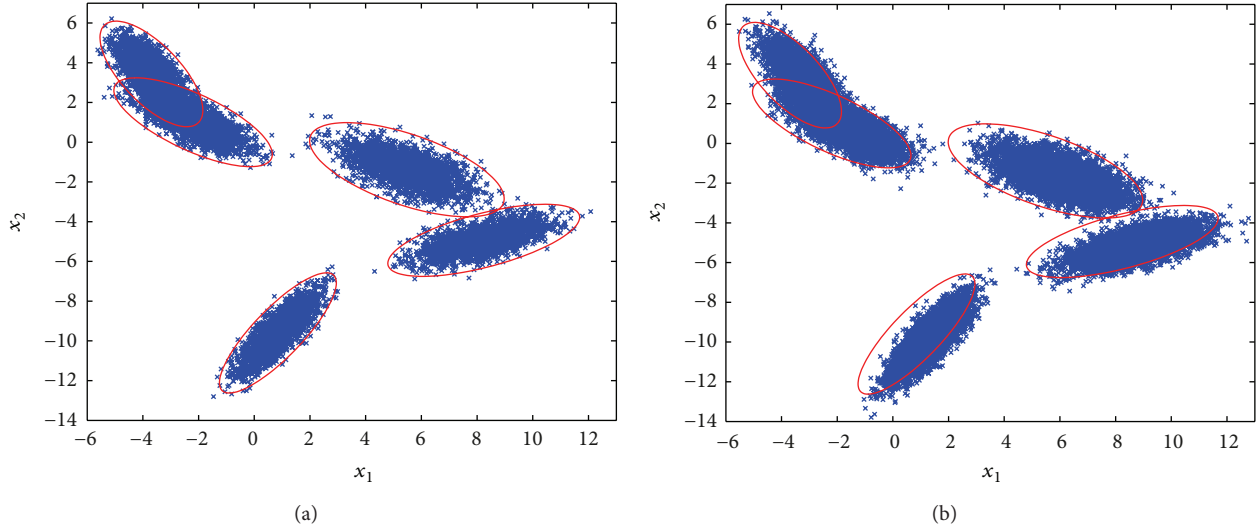
(a)

(b)

FIGURE 1: Training data (a) and test data (b) with a change in mean. The identified GMM model is shown in red.

example was chosen because the model assumptions are satisfied. In addition to damage detection, the objective is to test the model identification performance and the number of components selected by the algorithm.

The data were created as follows. The data dimensionality was two and the number of components was five. For each component k, the components of the mean vector $\boldsymbol{\mu}_k$ were sampled from a uniform random distribution between -10 and 10. The covariance matrix was generated by first generating the variances of the principal directions, resulting in a diagonal covariance matrix:

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \qquad (18)$$

where λ_1 and λ_2 were uniform random variables, λ_1 varying between 1 and 2 and λ_2 varying between 0.01 and 0.5. This diagonal covariance matrix was then rotated in a random orientation φ, resulting in covariance matrix $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\Sigma}_k = \mathbf{U}\mathbf{D}\mathbf{U}^T, \qquad (19)$$

where

$$\mathbf{U} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}. \qquad (20)$$

The data were then generated by sampling from a multivariate Gaussian distribution

$$N\left(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right). \qquad (21)$$

Once data from each component were generated, all data were concatenated and random permutation was applied to randomize the data labels.

The data are plotted in Figure 1. The training data are the first 10,000 data points shown in Figure 1(a) which were used to identify the model. Because the number of components is often unknown, different models were identified varying the

number of components between 1 and 10. The log-likelihoods with the penalty term are plotted in Figure 2(a) for different number of components. The maximum was correctly found with a five-component model. It should be noted that sometimes the components were not correctly identified, but the solution converged to a local maximum (Figure 2(b)). Therefore, it is suggested that the identification is repeated until a satisfying model is obtained, and the model with the highest log-likelihood is selected.

Damage was an equal shift in mean for all components. A bias vector $\mathbf{x}_d = \begin{bmatrix} 0.5 & -0.25 \end{bmatrix}^T$ was added to each data point. The damaged data were the last 25,000 data points plotted in Figure 1(b) together with the identified model for the training data. The shift of mean can be visually observed. The residuals were estimated for all data points and the first principal component scores were used for the control chart. The Shewhart chart was designed with a subgroup size 100 and the in-control samples 1–10,000. The shift in mean was clearly detected (Figure 3).

5.2. Three Piecewise Linear Components. The second example is a more realistic one in which the data are continuous with piecewise linear correlation between the two monitored variables. The piecewise linear regions are not necessarily Gaussian. Also, the variances are different in each region.

The data were created as follows. The data dimensionality was two and the number of piecewise linear components was three. Variable x_1 was uniformly distributed between 0 and 1. Variable x_2 was a piecewise linear function of x_1:

$$x_2 = a_k + b_k x_1, \qquad (22)$$

where the parameters and their validity regions are given in Table 1. Gaussian noise was added to the variables, with the standard deviations within each component shown in Table 1.

The test data came from a limited region of component 2. The first half of the test data was healthy and the second half
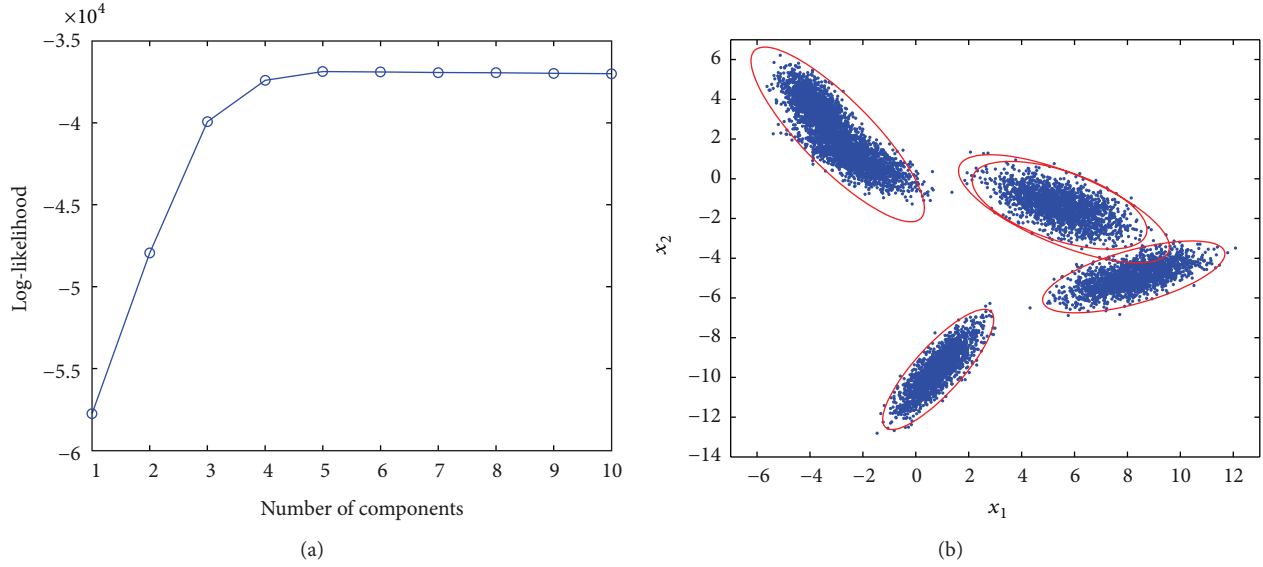
(a)

(b)

FIGURE 2: Log-likelihood with the penalty term (a) and an identified GMM model (in red) converged to a local maximum (b).
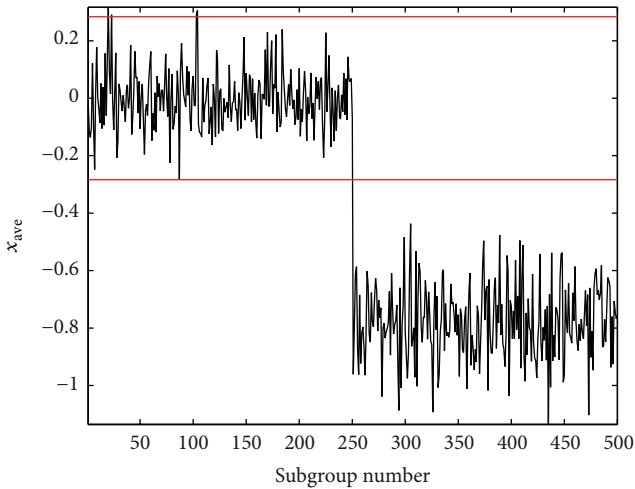

FIGURE 3: Shewhart control chart for damage detection.

TABLE 1: Parameters of the three linear components (22).

Class k	$x_{1,min}$	$x_{1,max}$	a_k	b_k	σ_1	σ_2
1	0	0.4	7	−0.2	0.01	0.02
2	0.4	0.6	6.4	1.3	0.02	0.01
3	0.6	1.0	7.84	−1.1	0.04	0.03

included damage, which was a shift in mean. A bias vector $\mathbf{x}_d = \begin{bmatrix} 0.1 & -0.05 \end{bmatrix}^T$ was added to those data points.

The data are plotted in Figure 4. They consist of 20,000 data points in a two-dimensional space. The first half is randomly distributed in all three regions (Figure 4(a)), but the last 10,000 data points are confined to the middle region (Figure 4(b)).

The training data include a larger variability than the test data. This is also more realistic as the training typically consists of monitoring under a full range of environmental or operational conditions, while the test data often come from a limited number of measurements at more or less constant conditions.

The training data were the first 5,000 data points. The model identification suggested 7 Gaussian components (Figure 5(a)), which are plotted in Figure 4.

Data points 10,001–15,000 are from the undamaged case, and data points 15,001–20,000 are from the damaged case with a shift in mean shown in Figure 4(b) together with the identified model for the training data.

The residuals were estimated for all data points and the first principal component scores were used for damage detection. The Shewhart chart was designed with a subgroup size 100 and the in-control samples 1–5,000. Damage was clearly detected with no false alarms (Figure 5(b)).

5.3. The Z24 Bridge. The data in the last case are the four lowest identified natural frequencies of the Z24 Bridge (see [12] for details) shown in Figure 6. Their pairwise correlation is plotted in Figure 7. It was reported that the frequencies varied considerably due to environmental effects and can be seen to be nonlinearly correlated. The physical reason was the different behaviour of the bridge below and above the freezing point.

Progressive damage test scenarios were introduced: settlement of pier, spalling of concrete, landslide at abutment, failure of a concrete hinge, failure of anchor heads, and rupture of tendons [12]. The first damage was introduced around measurement number 3517, shown with a vertical dashed line in Figure 6.

The training data were the first 3,000 samples shown in blue in Figure 7 and with another vertical dashed line in Figure 6. The training algorithm suggested 6 Gaussian components (Figure 8). After identifying the Gaussian mixture model, the residuals were estimated for all data points and
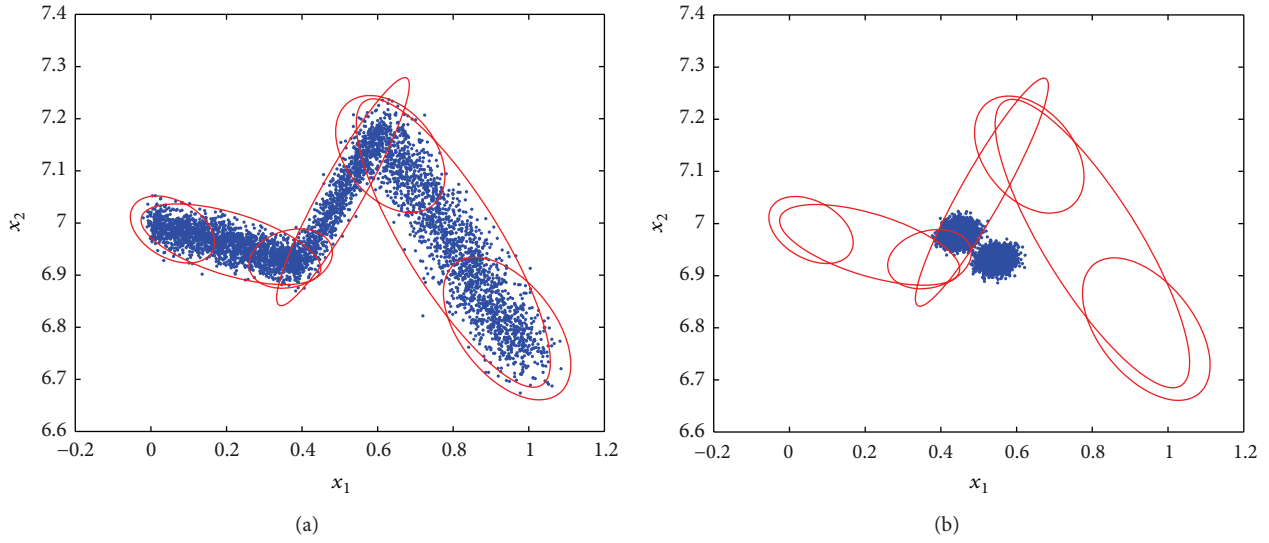
(a)

(b)

FIGURE 4: Training data (a) and test data (b) with a change in mean. The identified GMM model is shown in red.
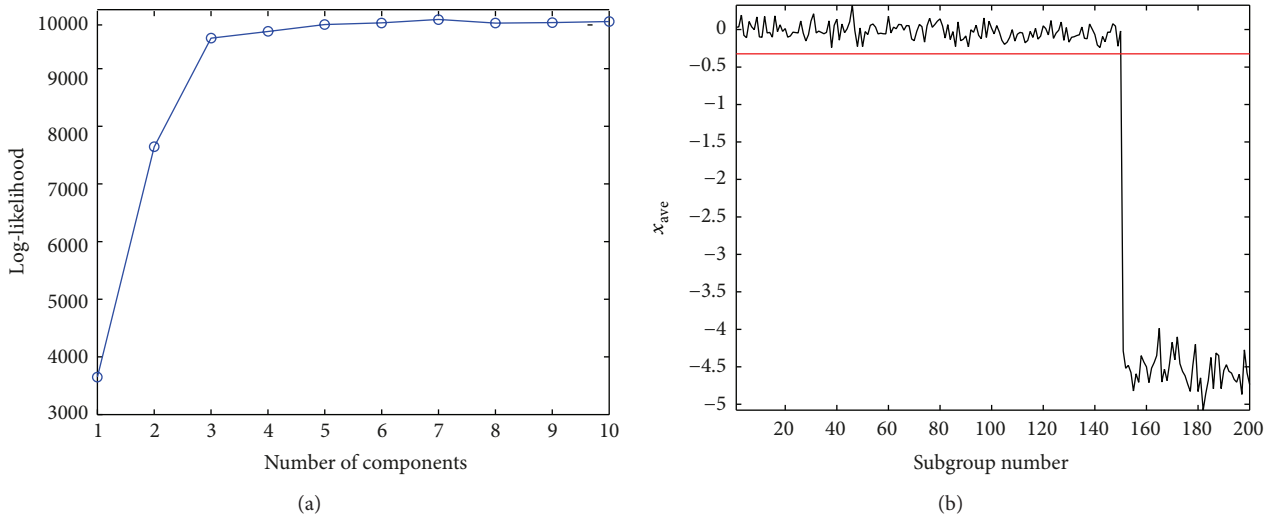

(a)

(b)

FIGURE 5: Log-likelihood with the penalty term (a) and Shewhart control chart for damage detection (b).
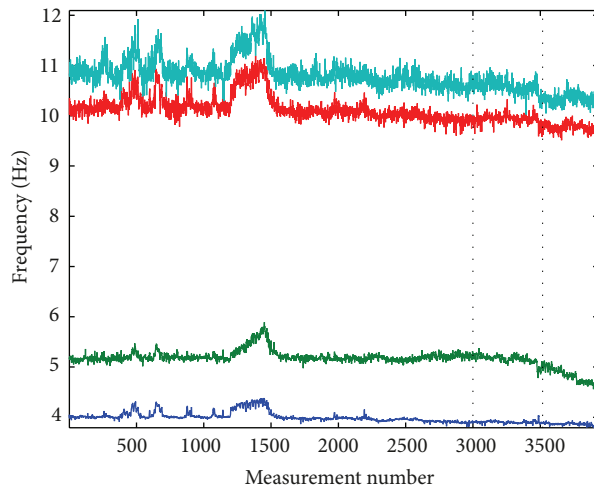

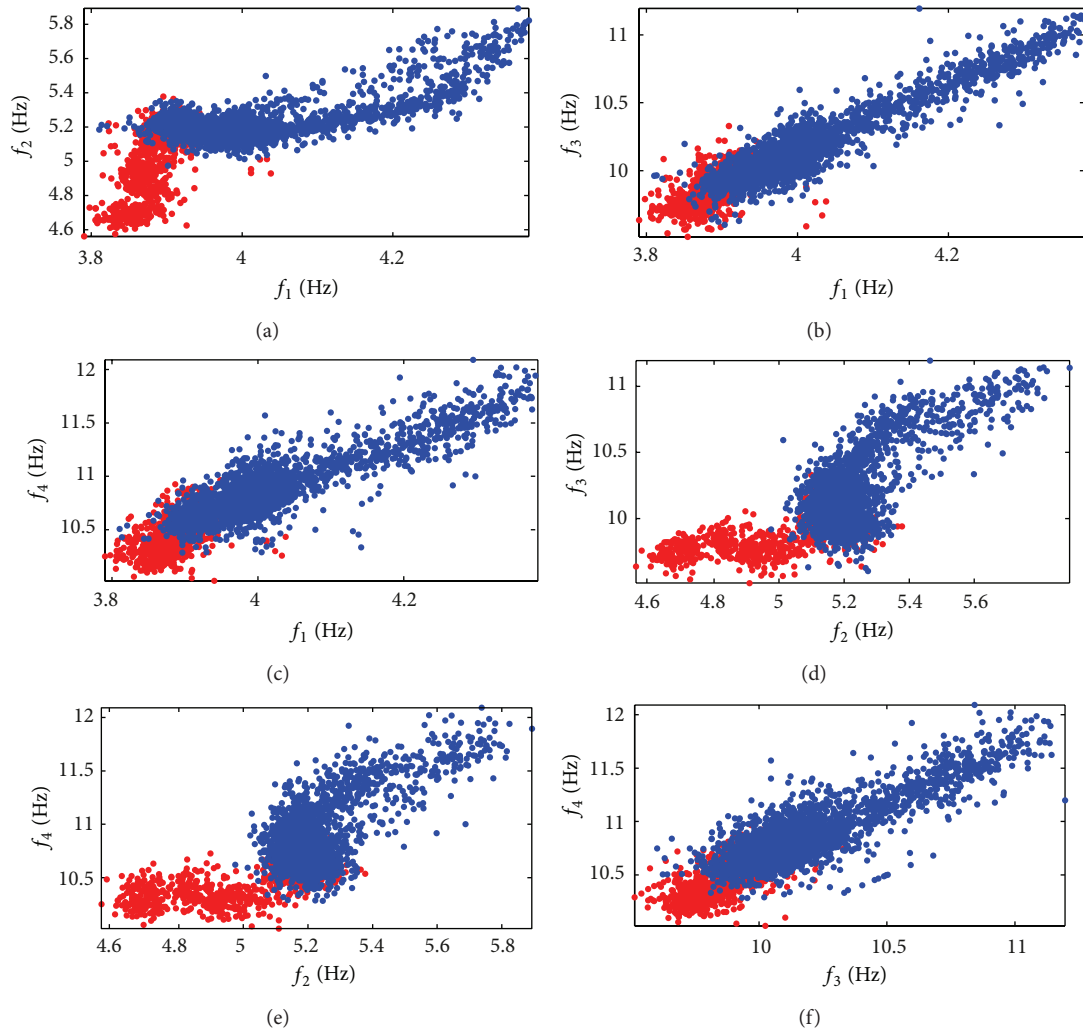FIGURE 6: Identified lowest natural frequencies of the Z24 Bridge.

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 7: Correlation between the four lowest natural frequencies of the Z24 Bridge. Blue symbols indicate the training data.
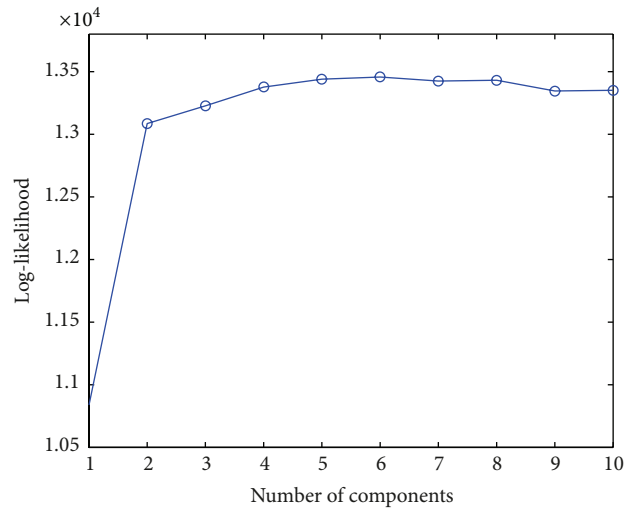

FIGURE 8: Log-likelihood including the penalty term.
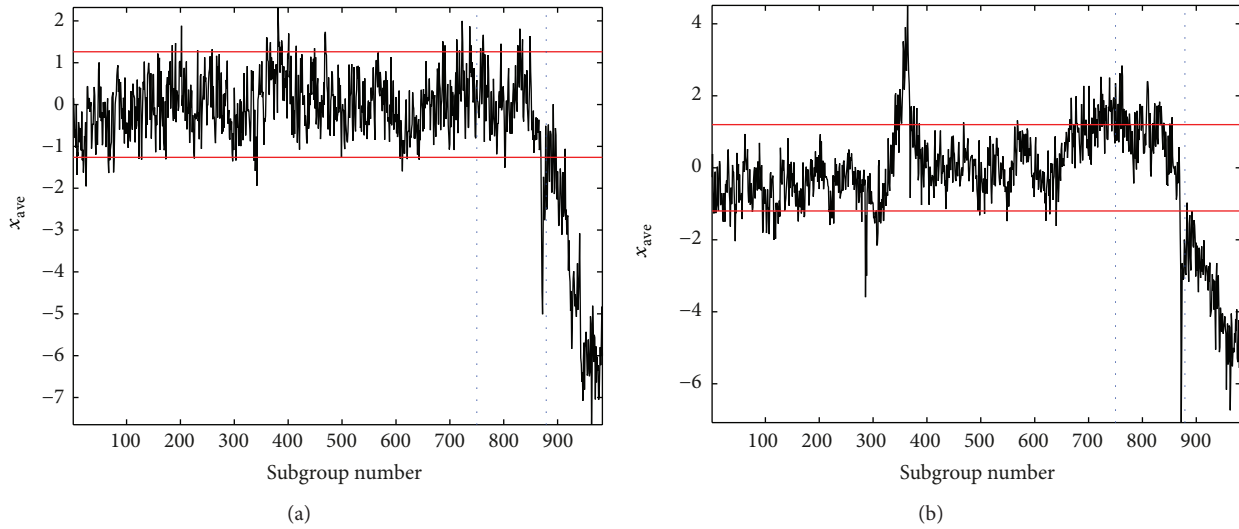
(a)

(b)

FIGURE 9: Shewhart control charts for damage detection using GMM (a) and a corresponding linear model (b).

the first principal component scores were used for damage detection. The Shewhart control chart was designed with a subgroup size 4 and the in-control samples 1–3,000. Damage was clearly detected (Figure 9(a)). The control limits are probably too tight resulting in several false alarms. In particular, a few false indications can be observed just prior to damage. Some activity was reported as the settlement system was installed [12], which may have changed the natural frequencies. Another control chart is shown in Figure 9(b) after using a corresponding linear model (with one component only). Compared to the chart in Figure 9(a), it can be concluded that GMM outperformed the linear model in this case.

6. Conclusion

A Gaussian mixture model was proposed to eliminate nonlinear environmental or operational influences from structural health monitoring data. The main advantages are that (1) the measurement of the underlying variables is not necessary, (2) the number of Gaussian components can be estimated, (3) the GMM model can be identified independently of the local linear models, and (4) it is a data-based method; no finite element model is needed. The main disadvantages are that the EM algorithm is not guaranteed to find the global maximum and that the training may be quite slow. Nonlinear models are most effective if the data dimensionality is low. Linear models often outperform nonlinear models for high-dimensional data [1]. The number of environmental or operational variables is usually relatively small. Therefore, their influences on the data are virtually located in a low-dimensional subspace, and a linear analysis is capable of removing this subspace from the subsequent analysis, thus eliminating the environmental or operational effects from the data.

Once the GMM model was identified, MMSE was applied to each component to take into account the local linear

correlation. The Mahalanobis distance or whitening transformation [13] could also be applied to the linear components.

The question of how small damage can be detected was not addressed in this study. Detection performance depends on the signal-to-noise ratio (SNR), in which signal is the shift or variance change due to damage and noise is often the measurement error or more generally everything that cannot be explained by the model. SNR should be as low as possible. In this study, noise was decreased by building an accurate model for nonlinear data.

Damage detection comprises several functions and models, many of which are classical. Also, many methods have been applied to SHM by the author and other researchers. This paper focused on the residual generation when the data are nonlinearly correlated. The remaining functions were merely referred to by name.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] J. Kullaa, "Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural

health monitoring," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 2976–2989, 2011.

[2] J. Kullaa, "Is temperature measurement essential in structural health monitoring?" in *Proceedings of the 4th International Workshop on Structural Health Monitoring*, pp. 717–724, Stanford University, DEStech, Stanford, Calif, USA, September 2003.

[3] Z. Ghahramani and G. E. Hinton, "The EM-algorithm for mixtures of factor analyzers," Tech. Rep. CRG-TR-96-1, University of Toronto, 1996.

[4] A.-M. Yan, G. Kerschen, P. de Boe, and J.-C. Golinval, "Structural damage diagnosis under varying environmental conditions—part II: local PCA for non-linear cases," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 865–880, 2005.

[5] H. Sohn, K. Worden, and C. R. Farrar, "Statistical damage classification under changing environmental and operational conditions," *Journal of Intelligent Material Systems and Structures*, vol. 13, no. 9, pp. 561–574, 2002.

[6] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal*, vol. 37, no. 2, pp. 233–243, 1991.

[7] E. Figueiredo, L. Radu, R. Westgate et al., "Applicability of a Markov-Chain Monte Carlo method for damage detection on data from the Z-24 and Tamar suspension bridges," in *Proceedings of the 6th European Workshop on Structural Health Monitoring*, C. Boller, Ed., pp. 747–754, Dresden, Germany, July 2012.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[9] J. Kullaa, "Sensor validation using minimum mean square error estimation," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1444–1457, 2010.

[10] S. Sharma, *Applied Multivariate Techniques*, John Wiley & Sons, New York, NY, USA, 1996.

[11] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1997.

[12] B. Peeters, *System identification and damage detection in civil engineering [M.S. thesis]*, K.U.Leuven, 2000.

[13] J. Kullaa, "Whitening transformation in damage detection. Smart structures," in *Proceedings of the 5th European Conference on Structural Control (EACS '12)*, C. D. ROM, A. E. del Grosso, and P. Basso, Eds., Erredi Grafiche Editoriali, Genoa, Italy, June 2012.