NOVELTY DETECTION BY NONLINEAR FACTOR ANALYSIS FOR STRUCTURAL HEALTH MONITORING

V. Lämsä

Aalto University
School of Science and Technology
Dep. of Applied Mechanics
P.O. Box 14300, 00076 Aalto, Finland

T. Raiko

Aalto University
School of Science and Technology
Dep. of Information and Computer Science
P.O. Box 15400, 00076 Aalto, Finland

ABSTRACT

In vibration-based structural health monitoring damage in structure is tried to detect from damage-sensitive features. Because neither prior information nor data about expected damage are normally available, damage detection problem must be solved by using a novelty detection approach. Features, which are sensitive to damage, are often sensitive to environmental and operational variations. Therefore elimination of these variations is essential for reliable damage detection. At present many of the damage detection methods are linear, though it has been shown that many of the vibration changes in structures are bilinear or nonlinear. This paper proposes to use nonlinear factor analysis to detect damage via elimination of external effects from damage features. The effectiveness of the proposed method is demonstrated by analyzing the experimental Z24 Bridge data with a comparison to a linear method. It is shown that elimination of adverse effects and damage detection are feasible.

1. INTRODUCTION

The main premise in vibration-based structural health monitoring (SHM) is that damage in structure can be identified from damage-sensitive features, such as natural frequencies, mode shapes or frequency transmissibilities. Damage detection problem must be solved by using a novelty detection approach applied to damage features, because data from the damaged structure is not usually available. In general, only existence and sometimes location of a damage can be identified in an unsupervised learning mode. On the other hand, identification of the type and the severity of the damage can be done only in a supervised learning mode, where known and identified reference data from a different damage levels and scenarios are available.

This research was partially funded by the Multidisciplinary Institute of Digitalisation and Energy (MIDE) research programme at Aalto University School of Science and Technology. The natural frequencies and the temperature ratings of the Z24 Bridge were identified and provided by Dr. Bart Peeters, whose contribution is greatly acknowledged.

To detect novel events, like damages and deflections, the machine learning system is first trained with the features extracted in normal conditions. Anyhow, it is commonly known that changing environmental and operational conditions can affect vibration measurements, which makes novelty detection approach substantially more difficult. Sohn [1] has listed some of the most common ambient vibration causes. Adverse variations in vibration measurements can be for example due to temperature, humidity, boundary conditions, mass loading or wind induced variation. Because these ambient variations can have similar effects on the vibration response as the damage itself, it is essential to separate structural changes from environmental and operational variation.

In recent years researchers have tried to tackle this problem in many ways. The ideal situation in SHM is to use damage features, which are sensitive to damage and insensitive to environmental and operational variations. For instance, Manson et al. [2] used genetic algorithm to select the best damage features in supervised fashion. Recently also Toivola et al. [3] proposed a method to select the best damage features by estimating probabilistic classifiers and their detection performance in supervised fashion with promising results.

In situations, where some of the underlying quantities can be measured, it is possible to model their relationship to the damage features. Peeters [4] studied autoregressive models to compensate environmental changes by using partially measured environmental variables. Ruotolo et al. [5] and Vanlanduit et al. [6] used singular value decomposition, in which measurements from the different working conditions were known. Basseville et al. [7] have studied parametric temperature-adjusted null space approach for this problem. A different approach was taken by Worden et al. [8]. They used outlier analysis as the novelty detector with Mahalanobis distance as the novelty index.

The natural extension to the preceding approach is to model the relationships of different factors behind the observations without measuring the underlying quantities. The basic idea is that a chosen model can explain the observations in normal conditions, while damage and unseen data from other environmental conditions are seen as outliers or damage. Even simple linear models have shown to be able to reduce the effects of environmental and operational variations. For example Kullaa [9, 10] studied factor analysis and missing data analysis with successful results. Yan et al. [11] used principal component analysis (PCA) for linear or weakly nonlinear cases. However, it has been shown that many of the vibration changes in structures are bilinear or nonlinear, especially in structures with moving parts and joints. To take into account also the nonlinear behaviour of the extracted damage features, nonlinear and piecewise linear latent variable models have been proposed with encouraging results. Sohn et al. [12] have proposed a nonlinear autoassociative neural network based method, which relies on autoregressive models. Kullaa [13] used piecewise linear mixture of factor analyzers model, while Lämsä et al. [14] studied nonlinear factor analysis (NFA) [15, 16] to eliminate environmental and operational variation from vibration measurements. Recently also Eciolaza et al. [17] have proposed Gaussian process latent variable model (GPLVM) for fault detection in a case of predefined steady-state stages over different operating conditions with encouraging results.

In the present work we demonstrate how NFA as a neural network based method can be used for separating structural changes from environmental and operational variation, and thereafter also for damage detection. In Section 2 NFA model is given and explained. The main idea is to learn the latent structure of the given observation vectors by using variational Bayesian (VB) learning methods [18] in unsupervised learning mode. Section 3 introduces the damage detection scheme and details to produce the damage indicator. In Section 4 NFA method and the damage indicator are applied for experimental data set. Feasibility and effectiveness are evaluated by comparing NFA to variational Bayesian principal component analysis (VBPCA), which can be seen as a linear version of NFA. Finally Section 5 concludes the paper.

2. NONLINEAR FACTOR ANALYSIS

In NFA [15, 16] the relationship among the observations (e.g. natural frequencies) is described in terms of a few underlying but unobservable factors. The main objective is to eliminate the adverse effects of these underlying factors from the observations resulting in new variables that can be used in damage detection. Primarily NFA is based on an assumption that the underlying factors are independent and normally distributed. It is easy to see that in SHM the independency assumption is not necessarily true for all factors, e.g. temperature and humidity are correlated. Furthermore, it is assumed that the damage features are generated through nonlinear mapping according to

$$\begin{aligned}\mathbf{x}(t) &= \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) \\ &= \mathbf{B}\tanh(\mathbf{A}\mathbf{s}(t) + \mathbf{a}) + \mathbf{b} + \mathbf{n}(t)\,. \end{aligned} \quad (1)$$

Equation (1) defines the observations $\mathbf{x}(t)$ which are assumed to be generated from hidden factors $\mathbf{s}(t)$ at time t. The nonlinear mapping $\mathbf{f}$ from hidden factors to observations is modelled by a multi-layer perceptron (MLP) network having two layers. The matrices $\mathbf{A}$ and $\mathbf{B}$ are the weights of the first and second layer of the MLP network and, $\mathbf{a}$ and $\mathbf{b}$ are the corresponding bias vectors. The noise $\mathbf{n}(t)$ is assumed to be independent and normally distributed. The nonlinear activation function in $\mathbf{f}$ is chosen to be hyperbolic tangent and it is applied to each component of the input vector separately.

The observation vector $\mathbf{x}(t)$ contains both the damage features $\mathbf{y}(t)$ and the measured environmental and operational variables $\mathbf{z}(t)$, that is, vector $\mathbf{x}(t) = [\mathbf{y}(t)^T \ \mathbf{z}(t)^T]^T$. This can be interpreted as some of the originally explanatory factors have become dependent variables. The main premise is that once all the accessible data is used, the model has eventually ability to explain normal variation of $\mathbf{y}(t)$ in conditions given in vector $\mathbf{z}(t)$.

As a comparison method, VBPCA [19] is used. This can be seen as a linear version of NFA, where Equation (1) is replaced with $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{a} + \mathbf{n}(t)$.

2.1. Variational Bayesian learning

Basically there are several possible factors and infinitely many models which can explain the measurement data in matrix $\mathbf{X} = [\mathbf{x}(1) \ \mathbf{x}(2) \dots \mathbf{x}(T)]$. Moreover, too complex models results in overfitting, while simple models (e.g. a linear model) results in underfitting. In either case some characteristics of data are lost. The Bayesian solution to the problem is that instead of choosing a single model, all models should be used and weighted according to how probable they seem in light of measurement data. This information is coded into posterior probabilities.

Since exact treatment of the posterior probability density function (pdf) of the unknown variables is impossible, approximative methods have to be used. Basic idea in VB learning [18] is that posterior pdf is approximated and thus it is an approximative inference method. The approximation is found by iteratively minimizing the misfit between the true posterior pdf and its approximation. Denote the exact posterior pdf by $p(\mathbf{S}, \boldsymbol{\theta} \mid \mathbf{X})$ and its parametric approximation by $q(\mathbf{S}, \boldsymbol{\theta})$, where $\mathbf{S}$ are the hidden factors, $\mathbf{X}$ are the observations (e.g. natural frequencies and temperature ratings) and $\boldsymbol{\theta}$ includes all the unknown parameters such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{a}$ and $\mathbf{b}$. Then the misfit can be measured by Kullback-Leibler (KL) divergence [20] used as a cost function $\mathcal{C}_{\mathrm{KL}}$.

It can be written

$$\mathcal{C}_{\mathrm{KL}} = \int_{\mathbf{S}} \int_{\boldsymbol{\theta}} q(\mathbf{S}, \boldsymbol{\theta}) \log \frac{q(\mathbf{S}, \boldsymbol{\theta})}{p(\mathbf{S}, \boldsymbol{\theta} \mid \mathbf{X})} d\boldsymbol{\theta} d\mathbf{S}. \tag{2}$$

Because the KL divergence is sensitive to probability mass rather than to probability density, overfitting can be largely avoided. According to the Bayes rule, the posterior pdf of the unknown variables $\mathbf{S}$ and $\boldsymbol{\theta}$ is

$$p(\mathbf{S}, \boldsymbol{\theta} \mid \mathbf{X}) = \frac{p(\mathbf{X} \mid \mathbf{S}, \boldsymbol{\theta}) p(\mathbf{S} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{X})}. \tag{3}$$

The factor $p(\mathbf{X} \mid \mathbf{S}, \boldsymbol{\theta})$ can be obtained directly from Equation (1), namely the distribution of the observed data vector $\mathbf{x}(t)$ is the same as the distribution of the noise vector $\mathbf{n}(t)$, but with the mean vector $\mathbf{f}(\mathbf{s}(t))$. If the variance of the *i*th component of $\mathbf{n}(t)$ is denoted by σ_i^2, the distribution $p(x_i(t) \mid \mathbf{s}(t), \boldsymbol{\theta})$ is normally distributed with mean $\mathbf{b}_i^T \mathbf{f}(\mathbf{A}\mathbf{s}(t) + \mathbf{a}) + b_i$ and variance σ_i^2. Here $x_i(t)$ is the *i*th component of $\mathbf{x}(t)$, $\mathbf{b}_i^T$ denotes the *i*th row vector of $\mathbf{B}$, and b_i is the *i*th component of the bias vector $\mathbf{b}$. Furthermore, the noise components $n_i(t)$ are assumed to be independent, and therefore

$$p(\mathbf{X} \mid \mathbf{S}, \boldsymbol{\theta}) = \prod_{t,i} p(x_i(t) \mid \mathbf{s}(t), \boldsymbol{\theta}). \tag{4}$$

The factors $p(\mathbf{S} \mid \boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ are products of simple normal distributions and they are obtained directly from the prior. The normalizing constant $p(\mathbf{X})$ in Equation (3) does not depend on the unknown parameters $\boldsymbol{\theta}$ or the hidden factors $\mathbf{S}$ and can be neglected.

It is also assumed that the hidden factors $\mathbf{S}$ are independent of the other unknown parameters $\boldsymbol{\theta}$, so that the approximation $q(\mathbf{S}, \boldsymbol{\theta})$ can be decomposed into two parts as

$$q(\mathbf{S}, \boldsymbol{\theta}) = q(\mathbf{S}) q(\boldsymbol{\theta}). \tag{5}$$

A normally distributed density with a diagonal covariance matrix is used for the parameters $\boldsymbol{\theta}$. The term $q(\boldsymbol{\theta})$ is a product of independent distributions:

$$q(\boldsymbol{\theta}) = \prod_i q_i(\theta_i) = \prod_i \mathcal{N}\left(\theta_i; \bar{\theta}_i, \tilde{\theta}_i\right). \tag{6}$$

The parameters of each normally distributed density component $q_i(\boldsymbol{\theta}_i)$ are its mean $\bar{\theta}_i$ and variance $\tilde{\theta}_i$. The definition of the pdf $q(\mathbf{S})$ is similar. Because the approximate posterior pdf $q(\mathbf{S}, \boldsymbol{\theta})$ is a product of simple normally distributed terms, the cost function can be simplified to expectations of many simple terms.

Learning is done by finding $\bar{\theta}_i$, $\tilde{\theta}_i$, $\bar{s}_i(t)$, and $\tilde{s}_i(t)$ that minimize the cost for the training data (see [15], [16], [18] for details). After the model has been learned, it can be applied to new data by using fixed $\bar{\theta}_i$ and $\tilde{\theta}_i$, and infering $\bar{s}_i(t)$ and $\tilde{s}_i(t)$ that minimize the cost for the test data. Because in many SHM applications computations needs to be done preferably online, we try to keep NFA model computationally simple and use early stopping of the learning.

3. DAMAGE DETECTION

Damage detection is based on reconstructing the observations by first infering the latent factors $\mathbf{s}(t)$ from observations $\mathbf{x}(t)$ (as explained above) and then reconstructing the observations from the latent factors. The residual error is the difference between the reconstructed observation vector $\hat{\mathbf{x}}(t) = E_q\{\mathbf{f}(\mathbf{s}(t))\}$ and the original observation $\mathbf{x}(t)$.

Since we are not interested in modelling the already measured environmental and operational variables $\mathbf{z}(t)$, but only in their influence to damage features $\mathbf{y}(t)$ through the latent factors, only the residual errors of the damage feature vectors are used as the novelty indicators (recall that $\mathbf{x}(t) = [\mathbf{y}(t)^T \ \mathbf{z}(t)^T]^T$). The remaining residual error vector is denoted by $\mathbf{u}(t) = \hat{\mathbf{y}}(t) - \mathbf{y}(t)$. The influence of damage or unseen data from other environmental conditions can be seen in a residual term. Anyhow, it is not possible to distinguish if damage has actually occured or if unseen environmental conditions have arisen after the learning phase.

If statistical process control charts are used for damage detection, usually the average residual error from subgroups of size four to six data points are charted [21]. Reason for this is to make observations within each group more similar than observations between groups [22]. For the same purpose we will use moving average. The *n* point moving average for the residual error vectors is defined as $\bar{\mathbf{u}}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{u}(t - j + 1)$. The final damage indicator $d(t)$ to be used for damage detection is the squared norm of the vector $\bar{\mathbf{u}}(t)$, that is, $d(t) = \sum_i \bar{u}_i(t)^2$. A pleasant advantage is that the resulting damage indicator $d(t)$ is already one-dimensional so dimensionality reduction is not needed like in [9, 10, 13, 14].

4. EXPERIMENTS

The main objectives of the experiments were to show the effectiveness of the NFA method in special case of novelty detection, namely damage detection, with a real-world data and compare it to the linear VBPCA method. Additional objective was to show that the NFA method partially eliminates the adverse effects of the environmental variability by learning the latent structure of the damage features.

The Z24 Bridge was demolished at the end of 1998. Before complete demolition the bridge was subjected to monitoring and progressive damage tests between November 11th 1997 and September 11th 1998. The accelerations of the bridge were measured and the environmental monitoring system was used to capture environmental parameters such as deck temperatures in different positions and humidity. In the end of the monitoring period several damage scenarios were applied to the bridge. These original raw data measurements were used in further analyses, where the natural frequencies from the healthy and the damaged bridge

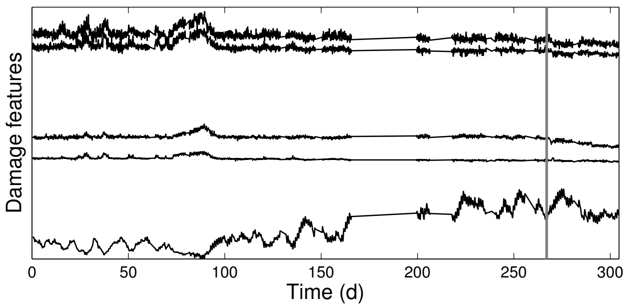

Fig. 1. The identified damage features and mean of the standardized bridge deck temperature measurements (lowest curve) for the Z24 Bridge.

were extracted and identified by using the stochastic subspace identification method [4].

Altogether the feature dataset $\mathbf{X}$ contained the four lowest identified natural frequencies as $\mathbf{y}(t)$ and mean of the standardized bridge deck temperature measurements from 21 different points as a one-dimensional vector $\mathbf{z}(t)$. The dimensions of the feature dataset were 5×3932. The whole feature dataset are presented in Figure 1, where the gray line indicates the time instant after installation of the settlement system used for damaging the bridge. Because the bridge was monitored for quite a long time, some variation of the identified natural frequencies can be seen. For example the bridge deck stiffening resulting the natural frequencies to increase during the cold periods is obvious. Moreover, the temperature effect results in quite high variation to the natural frequencies in comparison to the damage in the final stage of the monitoring period. Because the bridge was damaged gradually, there are no clear boundaries between damaged and undamaged cases.

The NFA and the VBPCA models were trained by using measurements from day 58 to 230 as training data, i.e. 5×2000 data points. Note that all the training data were from the undamaged bridge. To preserve simplicity, one factor models were used. For same reason the number of neurons for the NFA model was assumed to be equal to the dimensionality of the measurement matrix $\mathbf{X}$, i.e. five. For the training phase of the NFA only 500 iterations were taken. Note that the learning did not converged within the taken iterations. Number of iterations for the learning phase of the VBPCA was 100.

The test data were all the available data, i.e. 5×3932 data points, since we want to generate the damage indicator for each data point. An additional 100 iterations were taken for the testing phase for the NFA, whereas for the VBPCA additional iterations are not needed.

Original standardized natural frequencies $\mathbf{y}(t)$ and their reconstructions $\hat{\mathbf{y}}(t)$ are presented in Figure 2. It can be seen that the reconstructions fit quite well with the training data, and even with the unseen data from the undamaged bridge. It is clearly visible that the reconstructions differ from the originals with the data from the damaged bridge. Although accuracy of the VBPCA reconstructions seem to be quite good, some clear reconstruction errors are visible. The VBPCA reconstructions have much more variation than the NFA reconstructions. One can assume that this is because of the linear behaviour of the VBPCA method.

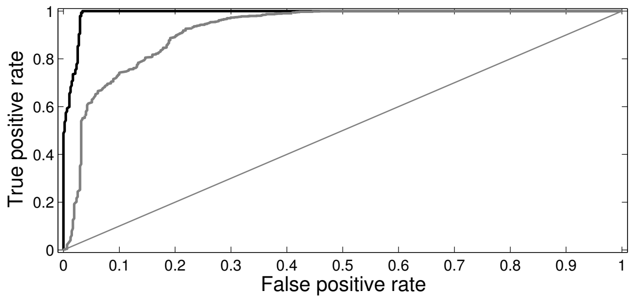

Fig. 4. The ROC curve for the NFA and the VBPCA as a function of the threshold.

Thereafter the damage indicators for all the data points for the NFA and the VBPCA models were formed by first averaging the reconstructions $\hat{\mathbf{y}}(t)$ over five samples and then taking the squared norm. The damage indicators are presented in Figure 3. To illustrate the effect of taking moving average, also the squared norm for the original residual error vectors (without averaging) are presented. Again, gray lines indicate the instant from which onward the bridge is damaged. It is clear that moving avarage tries to smooth the peaks of the damage indicator. It can be seen that with the NFA model damage can be almost correctly detected when the averaged damage indicator is used. There are only few false alerts even in the part of the data that was not used for training. Moreover, the average level of the indicator value for the VBPCA model is lower than with the NFA model, but damage detection is still difficult or impossible. It is very likely that the results for the VBPCA model would be better, if more factors were used.

The results of the damage detection are summarized in Figure 4 as receiver operating characteristic (ROC) curves. The ROC curves were calculated as a function of the threshold limit taken from the training data. Obviously the maximum and the minimum of the threshold interval are simply the maximum and the minimum of the damaged indicator value from the training data. The whole interval was devided into 1000 segments. From the Figure 4 it is clearly visible that the NFA outperforms the VBPCA in damage detection.

5. CONCLUSIONS

The NFA method was first recapitulated and thereafter applied to the experimental vibration data of the Z24 Bridge.

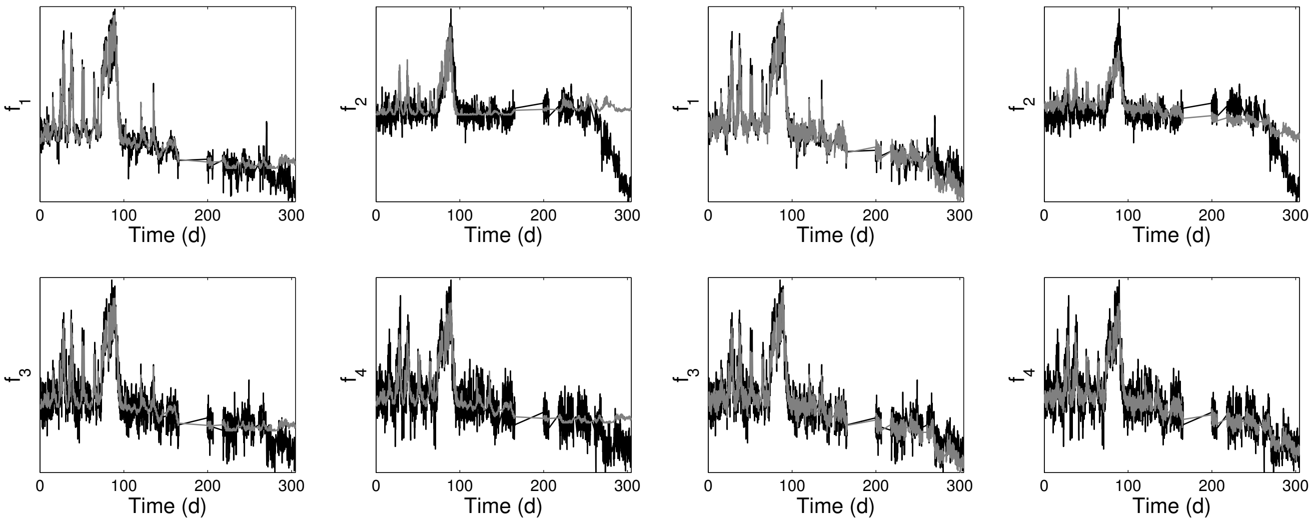

Fig. 2. Original standardized natural frequencies (black) and left: their NFA reconstructions (gray), right: their VBPCA reconstructions (gray).

It was shown that damage detection via elimination of environmental and operational effects from the damage features is feasible. Comparison to the VBPCA method addresses the usability of the NFA method in SHM applications, since it can take nonlinearities into account. The NFA model overcomes many of the problems related to overfitting, and with the early stopping of the learning computational burden is tolerable providing still good results.

Future research encompass utilization of VB learning methods more efficiently in different SHM applications.

6. REFERENCES

[1] Hoon Sohn, "Effects of environmental and operational variability on structural health monitoring," *Phi. Trans. A of the Royal Society on Structural Health Monitoring*, vol. 365, no. 1851, pp. 539–560, 2007.

[2] Graeme Manson and Keith Worden, "Feature selection for a neural network damage diagnostic using a genetic algorithm," in *Proc. of the 3rd European Workshop on Structural Health Monitoring*, Alfredo Güemes, Ed. 2006, pp. 683–690, Onera DEStech Publications.

[3] Janne Toivola and Jaakko Hollmén, "Feature extraction and selection from vibration measurements for structural health monitoring," in *Advances in Intelligent Data Analysis VIII*, Niall Adams, Ed. 2009, pp. 213–224, Springer-Verlag.

[4] Bart Peeters, *System identification and damage detection in civil engineering*, Ph.D. thesis, Department of Civil Engineering, K.U. Leuven, Belgium, 2000.

[5] Romualdo Ruotolo and Cecilia Surace, "Damage detection using singular value decomposition," in *Proc. of DAMAS 97: Structural Damage Assessment Using Advanced Signal Processing*, J.M. Dulieu-Smith, Ed. 1997, pp. 87–96, Sheffield Academic Press.

[6] Steve Vanlanduit, Eli Parloo, Bart Cauberghe, Patrick Guillaume, and Peter Verboven, "A robust singular value decomposition for damage detection under changing operating conditions and structural uncertainties," *Journal of Sound and Vibration*, vol. 284, pp. 1033–1050, 2005.

[7] Michéle Basseville, Frédéric Bourquin, Laurent Mevel, Houssein Nasser, and Fabien Treysséde, "Handling the temperature effect in shm: combining a subspace-based statistical test and a temperature-adjusted null space," in *Proc. of the 3rd European Workshop on Structural Health Monitoring*, Alfredo Güemes, Ed., Granada, Spain, July 2006, pp. 635–642, Onera DEStech Publications.

[8] Keith Worden, Hoon Sohn, and Charles Farrar, "Novelty detection in a changing environment: regression and interpolation approaches," *Journal of Sound and Vibration*, vol. 258, no. 4, pp. 741–761, 2002.

[9] Jyrki Kullaa, "Is temperature measurement essential in structural health monitoring?," in *Proc. of the 4th Int. Workshop on Structural Health Monitoring*, Fu-Kuo Chang, Ed. 2003, pp. 717–724, Onera DEStech Publications.

[10] Jyrki Kullaa, "Damage detection under a varying environment using the missing data concept," in *Proc.*

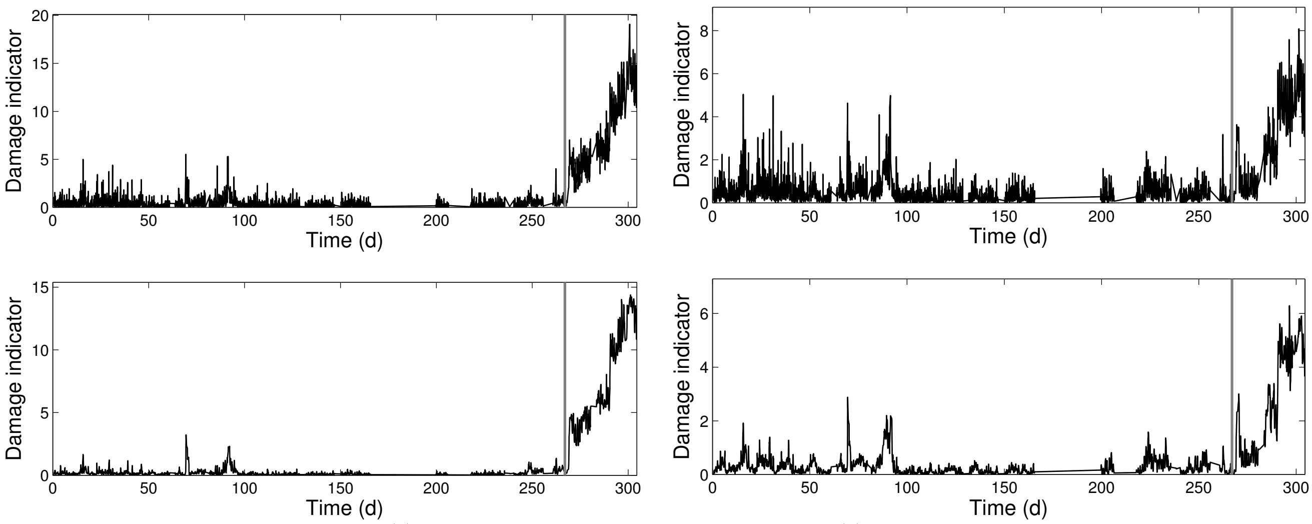

Fig. 3. Left: the damage indicator $d(t)$ by NFA, right: the damage indicator $d(t)$ by VBPCA. Above: the damage indicator without taking average. Below: moving average of the damage indicator over 5 samples.

of the 5th Int. Workshop on Structural Health Monitoring, Fu-Kuo Chang, Ed. 2005, pp. 565–573, Onera DEStech Publications.

[11] Ai-Min Yan, Gaëtan Kerschen, Pascal De Boe, and Jean-Claude Golinval, "Structural damage diagnosis under varying environmental conditions - part i: A linear analysis," *Mechanical Systems and Signal Processing*, vol. 19, pp. 847–864, 2005.

[12] Hoon Sohn, Keith Worden, and Charles Farrar, "Statistical damage classification under changing environmental and operational conditions," *Journal of Intelligent Material Systems and Structures*, vol. 13, no. 9, pp. 561–574, 2001.

[13] Jyrki Kullaa, "Removing nonlinear environmental influences from structural features," in *Proc. of the 3rd European Workshop on Structural Health Monitoring*, Alfredo Güemes, Ed. 2006, pp. 635–642, Onera DEStech Publications.

[14] Ville Lämsä and Jyrki Kullaa, "Nonlinear factor analysis in structural health monitoring to remove environmental effects," in *Proc. of the 6th Int. Workshop on Structural Health Monitoring*, Fu-Kuo Chang, Ed. 2007, pp. 1092–1099, Onera DEStech Publications.

[15] Harri Lappalainen and Antti Honkela, "Bayesian nonlinear independent component analysis by multi-layer perceptrons," in *Advances in Independent Component Analysis*, M. Girolami, Ed., pp. 93–121. Springer-Verlag, Berlin, 2000.

[16] Antti Honkela and Harri Valpola, "Unsupervised variational Bayesian learning of nonlinear models," in *Advances in Neural Information Processing Systems 17*, L. Saul, Y. Weiss, and L. Bottou, Eds., pp. 593–600. MIT Press, Cambridge, MA, USA, 2005.

[17] Luka Eciolaza, Muhammed Alkarouri, Neil Lawrence, Visakan Kadirkamanathan, and Peter Fleming, "Gaussian process latent variable models for fault detection," in *Proc. of the 2007 IEEE Symposium on Computational Intelligence and Data Mining*, 2007, pp. 287–292.

[18] Harri Lappalainen and James Miskin, "Ensemble learning," in *Advances in Independent Component Analysis*, Mark Girolami, Ed. 2000, pp. 75–92, Springer-Verlag.

[19] Alexander Ilin and Tapani Raiko, "Practical approaches to principal component analysis in the presence of missing values," *Journal of Machine Learning Research*, 2010, to be published.

[20] Solomon Kullback and Richard A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951.

[21] Jyrki Kullaa, "Damage detection of the z24 bridge using control charts," *Mechanical Systems and Signal Processing*, vol. 17, no. 1, pp. 163–170, 2003.

[22] Michael Fugate, Hoon Sohn, and Charles Farrar, "Vibration-based damage detection using statistical process control," *Mechanical Systems and Signal Processing*, vol. 15, no. 4, pp. 707–721, 2001.