Effects of environmental and operational variability on structural health monitoring

BY HOON SOHN*

Civil and Environmental Engineering Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Stated in its most basic form, the objective of structural health monitoring is to ascertain if damage is present or not based on measured dynamic or static characteristics of a system to be monitored. In reality, structures are subject to changing environmental and operational conditions that affect measured signals, and these ambient variations of the system can often mask subtle changes in the system’s vibration signal caused by damage. Data normalization is a procedure to normalize datasets, so that signal changes caused by operational and environmental variations of the system can be separated from structural changes of interest, such as structural deterioration or degradation. This paper first reviews the effects of environmental and operational variations on real structures as reported in the literature. Then, this paper presents research progresses that have been made in the area of data normalization.

Keywords: data normalization; environmental and operational variation; damage detection

1. Introduction

Structural health monitoring (SHM) is a problem which can be addressed at many levels. Stated in its most basic form, the objective is to ascertain simply if damage is present or not. The basic premise of most damage detection methods is that damage will alter the stiffness, mass or energy dissipation properties of a system, which in turn will alter the measured dynamic response of the system. During the normal operation of a system or structure, measurements are recorded and features are extracted from data, which characterize the normal conditions. After training the diagnostic procedure in question, subsequent data can be examined to see if the features deviate significantly from the norm. That is, a simple damage classifier such as outlier analysis (Worden 1997; Worden et al. 2000) can be employed for deciding whether measurements from a system or structure indicate significant departure from the previously established normal conditions. Ideally, an alarm is signalled if features increase above a pre-determined threshold.

The basis for damage detection appears intuitive, but its actual application poses many significant technical challenges. Although many damage detection techniques are successfully applied to scale models or specimen tests in

*hsohn@andrew.cmu.edu

One contribution of 15 to a Theme Issue ‘Structural health monitoring’.
controlled laboratory environments, the performance of these techniques in field is still questionable and needs to be validated. SHM applied to rotating machinery, often referred to as condition monitoring, is somewhat mature, having made the transition from a research topic to actual practice. The observation of this technology transition in rotation machinery makes us wonder, ‘why is not the SHM technology penetrating into other market places?’ One of the main obstacles for deploying a SHM system for in-service structures is the environmental and operational variation of structures. In fact, these changes can often mask subtler structural changes caused by damage. Often the so-called damage-sensitive features employed in these damage detection techniques are also sensitive to changes in environmental and operational conditions of the structures.

For instance, Farrar et al. (1994) performed vibration tests on the I-40 Bridge over the Rio Grande in New Mexico, USA to investigate if modal parameters could be used to identify structural damage within the bridge. Four different levels of damage were introduced to the bridge by gradually cutting one of the bridge girders, as shown in figure 1. The change of the bridge’s fundamental frequency is plotted with respect to the four damage levels, as shown in figure 2. Because the magnitude of the bridge’s natural frequency is proportional to its stiffness, the frequency is expected to decrease as the damage progresses. However, the results in figure 2 belie the intuitive expectation. In fact, the frequency value increases for the first two damage levels, and then eventually decreases for the remaining two damage cases. Later investigation revealed that, beside the artificially introduced damage, the ambient temperature of the bridge played a major role in the variation of the bridge’s dynamic characteristics. Other researchers also acknowledged potential adverse effects of varying operational and environmental conditions on vibration-based damage detection (Cawley 1997; Ruotolo & Surace 1997; Roberts & Pearson 1998; Helmicki et al. 1999; Rohrmann et al. 1999; Williams & Messina 1999; Cioara & Alampalli 2000; Sohn et al. 2001a,b).

Figure 1. Damage detection study of the I-40 Bridge over the Rio Grande in New Mexico, USA. (a) Introduction of damage in one of the bridge girders by electric saw cutting. (b) Four levels of damage introduced at the girder (the shaded area represents reduced cross-section).
This paper focuses on data normalization issues of SHM related to in-service infrastructure. This paper starts by exposing data normalization issues reported in the literature and identifies technologies that seem promising for addressing the data normalization problems.

2. Environmental and operational variability

Many techniques have been proposed to identify the extent and location of damage in in-service structures using changes in the dynamic characteristics of the system. Damage detection is based on the premise that damage in the structure will cause changes in the measured vibration data. Many existing methods, however, neglect the important effect of changing environmental and operational conditions on the underlying structure. For in-service structures, the variability in dynamic properties can be a result of time-varying environmental and operational conditions. Environmental conditions include wind, temperature and humidity, while operational conditions include ambient loading conditions, operational speed and mass loading. In this section, a brief summary of reported examples of the environmental and operational variability is provided.

(a) Temperature effects

The effects of temperature variability on the measured dynamics response of structures have been addressed in several studies. It is intuitive that temperature variation may change the material properties of a structure. Wood (1992) reported that the changes of bridge responses were closely related to the structural temperature based on the vibration test of five bridges in the UK. Analyses based on the data compiled suggested that the variability of the asphalt elastic modulus due to temperature effects was a major contributor to the changes in the structural stiffness.

Temperature variation not only changes the material stiffness, but also alters the boundary conditions of a system. Based on a field test conducted on the Sutton Creek Bridge in Montana, USA, Moorty & Roeder (1992) reported that
the movements obtained from both the analytical model and the measured values showed a significant expansion of the bridge deck as temperature increased. Rohrmann et al. (1999) noted that when a bridge structure was obstructed from expanding or contracting, the expansion joints could be closed significantly altering the boundary conditions.

The temporal variation of the temperature also needs to be noted. Many structures exhibit daily and seasonal temperature variations. Based on the modal testing of Alamosa Canyon Bridge in New Mexico, USA, Doebling & Farrar (1997) showed that the first mode frequency of the bridge varied approximately 5% during the 24 h cycle. Askegaard & Mossing (1988) tested a three-span RC footbridge for a 3-year period, and about 10% seasonal changes in the frequencies were repeatedly observed for each year. The authors concluded that the changes were partially attributed to the variation of ambient temperature.

In reality, structures will be subjected to different operational and environmental variations simultaneously. Pirner & Fischer (1997) attempted to separate mechanically induced loadings from thermally induced ones in a TV tower in Prague, Czech Republic. The researchers concluded that it was impossible to directly distinguish these loadings using time histories alone. They noted, however, that temperature changes produced major and slowly changing stresses while wind loads produced faster changes. On the other hand, Helmicki et al. (1999) measured the effects of thermal stresses in a steel-stringer bridge while also recording traffic loads by means of a weigh-in-motion roadway scale. It was concluded that thermal stresses far exceeded stresses caused by the recorded traffic.

Finally, other studies on the influence of temperature variation on in-service structures have been reported (Churchward & Sokal 1981; Rohrmann et al. 1999; Cioara & Alampalli 2000; Peeters & De Roeck 2000).

(b) Boundary condition effects

Changes in the structure’s surroundings or boundary conditions such as thermal expansion can produce more significant changes in dynamic responses than damage. Using an analytical model of a cantilever beam, Cawley (1997) compared the effect of crack formation on the resonant frequency with the effect of the beam’s length on the resonant frequency. In this study, the crack was introduced at the fixed end of the cantilever beam, and the length of the beam was varied. His results demonstrated that the resonance frequency change caused by a crack, which was a 2% cut through the depth of the beam, was 40 times smaller than that caused by a 2% increase in the beam’s length.

Alampalli (1998) reported that, for a 6.76 by 5.26 m bridge span that they tested, the natural frequency changes caused by the freezing of the bridge supports were an order of magnitude larger than the variation introduced by an artificial saw cut across the bottom flanges of both girders. Peeters & De Roeck (2000) found a highly nonlinear or a piecewise linear relationship between the ambient temperature and the fundamental frequencies of the Z24 Bridge based on a yearlong monitoring of data. In particular, the influence of temperature variation became prominent when the temperature fell below the freezing point. Similar to Alampalli’s finding, it is possible that the piecewise linear relationship was mainly caused by the freezing of the bridge supports.
In particular, the changing boundary conditions would impose difficulties on the implementation of SHM systems, because it is often challenging to directly measure the boundary conditions of a structure.

(c) Mass loading effects

Mass loading such as traffic loading would be another operational variable that is difficult to precisely measure, but might be important for data normalization. The influence of traffic loading on modal parameters of bridge structures has been investigated (Kim et al. 1999; DeRoeck et al. 2002; Zhang et al. 2002). The researchers seem to agree that the mass loading effect of moving vehicles varies depending on the vehicles’ mass relative to the magnitude of the bridge. Kim et al. (1999) reported that the measured natural frequencies of a 46 m long simply supported plate girder bridge decreased by 5.4% as a result of heavy traffic. However, for the middle- and long-span bridges, changes in the measured natural frequencies due to different types of vehicle loading (heavy versus light) were hardly detectable. Zhang et al. (2002) found the damping ratios to be sensitive to the traffic mass, especially when the deck vibration exceeded a certain level. It is believed that the damping ratio increased because the energy dissipation capacity in the material and at the joint increased at higher traffic loading. In addition, it is speculated that the secondary structure–vehicle interaction effects could also influence the dynamic characteristic of bridge structures.

As part of an online monitoring system development for roller coasters, Sohn et al. (2004) studied the effect of passenger loading on their time-series-based damage detection algorithm. A specific type of damage investigated was the debonding between an inner aluminium wheel and an outer polymer layer of a roller coaster ride while the vehicle was on a rail (figure 3). The acceleration-time response signals were recorded from the roller coasters at one position of the roller coaster track. Data were subsequently acquired during a test operation from three different vehicles (trains 3, 4 and 6) with varying speeds, mass loading and damage conditions. Note that only train 3 was loaded with rocks on the seats to simulate passenger loading, while trains 4 and 6 were tested with the seats empty.

Then, the auto-regressive (AR) and moving average (MA) coefficients of the auto-regressive with exogenous input (ARX) model were plotted separately in figure 4a, b, respectively. In figure 4a, a clear separation between the AR coefficients from train 3 and those of trains 4 and 6 was observed. Although a
more thorough investigation is required, it seems that two distinctive clusters in figure 4a, one from train 3 and the other from trains 4 and 6, were caused by different mass loading conditions between train 3 and trains 4 and 6. A less obvious, but similar, conclusion can be drawn from figure 4b. The results from figure 4 clearly highlight that variation due to different mass loading was much bigger than that due to wheel debonding.

(d) Wind-induced variation effect

The wind-induced vibration plays an important role for long-span bridges. As a bridge vibrates in the wind, the energy input from the wind-induced vibration becomes larger than the energy dissipated by damping, causing flutter or buffeting. Fujino et al. (2000) observed that the dynamic behaviour of cable-stayed or suspension bridges is amplitude dependent. For instance, the fundamental frequency of a suspension bridge reduced as the wind speed increased. On the other hand, the modal damping increased when the wind velocity exceeded a certain level. Mahmoud et al. (2001) used continuously measured vibration data from the Hakusyo Suspension Bridge in Japan to study the dynamic behaviour of a suspension bridge. It was found that the vertical amplitude of the bridge response is almost a quadratic function of the wind speed and damping ratio is vibration amplitude dependent. In particular, the lower natural frequencies were significantly affected by wind speeds. However, the dependency of mode shapes on the vibration amplitude was mainly observed for higher modes and near the tower.

3. Data normalization methods

Since data can be measured under varying conditions, the ability to normalize the data becomes very important to the SHM process. This section contains new technical developments that attempt to tackle the aforementioned environmental and operational issues of SHM.

When the environmental or operating variability is an issue, there are three different situations for data normalization. First, when direct measurements of the varying environmental or operational parameters are available, various kinds

Figure 4. Separation of train 3 and trains 4 and 6 data using principal component analysis of AR-ARX (4,2,2) model coefficients. (a) Using AR coefficients. (b) Using MA coefficients.
of regression and interpolation analyses can be performed to relate measurements relevant to structural damage and those associated with environmental and operation variation of the system (Rohrmann et al. 1999; Peeters & De Roeck 2000; Worden et al. 2002; Fritzen et al. 2003). For example, the dependency of two-dimensional features on some environmental variable $T$ can be approximated by using regression analysis, as shown in figure 5a, when a large set of extracted features and measured environmental variables are available. Note that there might be some damage cases, which cannot be distinguished from the undamaged conditions unless the environmental variable is measured (e.g. the damage case shown in figure 5a).

On the other hand, there are situations in which direct measurements of these operational and environmental parameters are impractical or difficult to achieve and damage produces changes in the extracted features, which are ‘orthogonal’ to the changes caused by the operational and environmental variation of the system (figure 5b). For this situation, it may be possible to distinguish changes caused by damage from those caused by the operational and environmental variation of the system without measuring the operational and environmental parameters. Several researchers have tackled this situation by implicitly

Figure 5. Three conceptional situations exist for data normalization: (a) direct measurements of the varying environmental or operational parameters are available, (b) direct measurements of these parameters are impractical or difficult to achieve, but damage produces changes orthogonal to changes caused by damage, and (c) features that are mainly sensitive to damage but insensitive to operational and environmental variations can be extracted (not shown in the figure).
modelling the underlying relationship between the environmental variables and the damage-sensitive features (Ruotolo & Surace 1997; Manson 2002; Kullaa 2003; Sohn et al. 2003b). Others have attempted to divide baseline datasets into subsets, each corresponding to a different operational and environmental condition (Sohn & Farrar 2001; Sohn et al. 2001a,b).

Finally, there are other researchers who attempt to explicitly extract features that are mainly sensitive to damage but insensitive to operational and environmental variations (Manson 2002; Sohn et al. 2003a). In this section, these novel approaches to address the data normalization issue are briefly introduced.

(a) Regression analysis

Peeters (2000) performed a regression analysis of the natural frequencies of the Z24 Bridge on temperature to filter out the temperature effects from the measured frequencies. The Z24 Bridge in Switzerland was monitored during almost one year before it was artificially damaged. A bilinear relationship was observed between the measured frequencies and temperature. There was a clear linear relationship above the freezing temperature (0°C) and a different linear correlation below the freezing temperature. It was demonstrated that this bilinear behaviour was attributed to the asphalt layer on the deck. Although the asphalt layer did not contribute to the overall stiffness at warm temperatures, it added significant stiffness to the bridge at cold temperatures. In this study, only data corresponding to above freezing temperatures were used for the linear regression analysis.

In particular, to take into account the thermal inertia of the asphalt and the concrete, a dynamic linear regression model called ARX was fitted to the measured frequency–temperature data to reconstruct the dependence of the measured frequencies on temperature. Single-input and single-output (SISO) ARX models were constructed for the first four fundamental frequencies, describing the relation between one input temperature and one output frequency. Although multi-input and single-output (MISO) ARX models were also investigated, no significant improvement over the SISO model was observed.

Once the ARX model was constructed, the prediction error was used as a damage-sensitive feature. If the residual error exceeds a confidence interval established from the baseline training datasets, it is likely that the bridge undergoes abnormal conditions that it had not experienced during the training phase. In this study, the damage introduced was the incremental settlement of one of the bridge piers. As shown in figure 6, a drastic shift of the prediction error was observed in the fundamental frequency. Similar observations were reported for the next three fundamental frequencies as well.

A similar linear regression analysis was applied to vibration signals obtained from the Alamosa Canyon Bridge in New Mexico, USA to relate the change of the bridge’s fundamental frequency to the temperature gradient of the bridge (Sohn et al. 1998). The measured fundamental frequency of the Alamosa Canyon Bridge in New Mexico varied approximately 5% during a 24 h test period, and the change of the fundamental frequency was well correlated to the temperature difference across the bridge deck. Because the bridge was approximately aligned in the north and south direction, there was a large temperature gradient between the west and the east sides of the bridge deck throughout the day.
A simple linear filter using the temperature readings across the bridge as inputs was constructed and was able to predict the frequency variation, as shown in figure 7. The first dataset from 1996 was used to train the adaptive filter, while the second dataset from 1997 was used to test the prediction performance. A linear filter with two spatially separated and two temporally separated temperature measurements reproduced the variation of the frequencies of the first dataset. Based on the trained filter system, a prediction interval of the frequency for a new temperature profile was computed and the prediction performance was tested using the second dataset. Results indicated that a linear adaptive filter with four temperature inputs (two temporal and two spatial

Figure 6. Damage diagnosis using ARX model, which models the relation between the measured fundamental frequency and temperature (Peeters 2000).

Figure 7. Reproduction of the first mode frequency using a linear filter.

A simple linear filter using the temperature readings across the bridge as inputs was constructed and was able to predict the frequency variation, as shown in figure 7. The first dataset from 1996 was used to train the adaptive filter, while the second dataset from 1997 was used to test the prediction performance. A linear filter with two spatially separated and two temporally separated temperature measurements reproduced the variation of the frequencies of the first dataset. Based on the trained filter system, a prediction interval of the frequency for a new temperature profile was computed and the prediction performance was tested using the second dataset. Results indicated that a linear adaptive filter with four temperature inputs (two temporal and two spatial
dimensions) could reproduce the natural variability of the frequencies with respect to time of day reasonably well. Then, the regression model was also used to establish confidence intervals of the frequencies for a new temperature profile.

It is worthwhile to compare the regression approaches performed by Peeters (2000) and Sohn et al. (1998). While Peteers’ work emphasizes the thermal dynamics of the bridge by using a single temperature measurement with multiple time lags (the temporal variation of temperature), the work by Sohn et al. (1998) used temperature readings from multiple thermocouples to take into account the temperature gradient across the bridge (the spatial variation of temperature) as well as the temporal variation. The comparison of these two approaches clearly demonstrates that data normalization is problem-specific; one kind for each individual structure.

Because the Alamos Canyon Bridge examined by Sohn et al. (1998) is oriented in a north–south direction and it is located in southern New Mexico, where the temperature can easily go above 45°C, the bridge experiences a large temperature gradient across the bridge throughout the day. Therefore, it is feasible that the spatial variation of the bridge temperature might have been the main driving factor for the frequency variation. On the other hand, the magnitude of the Z24 Bridge, which had a 30 m long main span and two 20 m long side spans with 8.6 m width, is certainly larger than the 7.3 m wide and 15.2 m long span of the Alamos Canyon Bridge tested. Therefore, it is possible that it took a longer time before the temperature affected the dynamic properties of the bridge and the time-lag information of the temperature was more important for the Z24 Bridge than the Alamos Canyon Bridge.

(b) Subspace-based identification method

Fritzen et al. (2003) modified an existing subspace-based identification method (Peeters & De Roeck 1999) for temperature compensation. For damage diagnosis based on the subspace-based identification method, the following residual error is first computed:

$$z = \text{vec}(s_0^T H_1),$$

where $z$ is the residual error; $H_0$ and $H_1$ are the Hankel matrices of the baseline and test structures obtained from either impulse time signals or frequency response functions, respectively; $s_0$ is the left singular vector of the baseline Hankel matrix $H_0$; and vec(...) is the stack operator rearranging a matrix into a column vector. The residual error basically compares the response spaces spanned by the baseline system, $H_0$, and the system in question, $H_1$. Therefore, if damage occurs, this residual error will increase because the response spaces spanned by the damaged system will be different from those spanned by the baseline system. Therefore, statistical analysis can be performed on the residual error for damage diagnosis.

To address the temperature calibration issue, the left singular vectors corresponding to various temperature values are stored in a reference database. Then, when a new Hankel matrix is constructed at any measured temperature value, the residual error is computed by selecting the left singular vector corresponding to the same temperature from the previously established reference database.

This method has been tested using only numerical simulations. Note that this approach is based on the assumption that damage produces changes orthogonal to the response spaces spanned by the baseline data. However, other operational
and environmental conditions may produce changes orthogonal to the existing baseline conditions, and damage may produce changes along the response space spanned by the baseline signal.

**(c) Novelty detection**

Novelty detection first builds an internal representation of the system’s normal condition, and then examines subsequent data to see if they significantly depart from a normal condition. The discordancy of a candidate outlier is measured by the following squared Mahalanobis distance measure $D$ (Barnett & Lewis 1994; Farrar & Worden 2007; Worden & Manson 2007):

$$D = (x - \bar{x})^T C (x - \bar{x}),$$

where $x$ is a vector of the potential outlier; $\bar{x}$ and $C$ are the mean vector and covariance matrix of the baseline system, respectively. Once the Mahalanobis distance is computed, it is checked against a threshold value. Note that this novelty detection addresses only the simplest level of damage identification, i.e. whether damage is present or not.

One major advantage of this novelty detection is that statistical model building for damage classification is based only on data from the undamaged system. However, the success or failure of the novelty detection is contingent on the accuracy of the description of the normal condition. In reality, the normal condition of the system may experience a wide range of variation due to operational and environmental changes of the system.

Worden et al. (2002) tackled this issue via the construction of a reference set parameterized by environmental and operational variables. That is, new data are evaluated only with respect to reference data obtained from the same environmental conditions. It is assumed that the environment is parameterized by a single measurable variable $T$. Thus, the mean vector and the covariance matrix become functions of the underlying environmental parameter, and the functional relationships are approximated by polynomial regression models

$$\bar{x}_i = \sum_{j=0}^{n} a_i^j T^j \quad \text{and} \quad C_{ij} = \sum_{k=0}^{n} a_{ij}^k T^k,$$

where $\bar{x}_i$ and $C_{ij}$ are the $i$th entity of the mean vector and the $ij$th entity of the covariance matrix, respectively; $n$ is the polynomial order; and $a_i^j$ and $a_{ij}^k$ are the regression coefficients estimated by least-squares regression.

Now the monitoring strategy becomes clear. When a new set of data is measured at an environmental value at $T$ and is tested for novelty, the appropriate $\bar{x}$ and $C$ are estimated from the regression model at the same environmental value, and they are used to compute the Mahalanobis distance. The effectiveness of this approach is demonstrated by using simulated data from a lumped-mass system. It was shown that the regression approach was more sensitive to damage than the conventional novelty detection approach, where the mean vector and covariance matrix were constructed spanning all environmental conditions of interest. However, it was reported that when the training data were not properly sampled, the covariance matrix may not be positive semi-definite producing negative Mahalanobis distances. This problem was addressed by
introducing an interpolation approach to data normalization. The interpolation approach guaranteed that the estimated covariance matrix be positive semi-definite.

Manson (2002) used the same novelty detection for damage identification, but took an alternative approach for data normalization. His approach for data normalization is based on the premise that there might be a subset of features, which are more sensitive to damage yet insensitive to environmental variations. Therefore, his approach focuses on isolating those feature components that are sensitive to damage. In this study, Lamb-wave propagation data from a composite plate were used similar to the work by Worden (2002). The instrumented composite plate was placed in an environmental chamber and Lamb-wave signals were recorded every minute. Initially, the chamber temperature was held at a constant temperature of 25°C, and the temperature was fluctuated between 10 and 30°C. At the end of the test, a 10 mm hole was drilled in the plate between two sensors.

The first half of the data measured at 25°C was stored as a training dataset, and the rest of the data was tested against the training dataset for novelty. The diagnosis using the conventional novelty detection method is shown in figure 8. Most of the data points from the second half of the constant temperature set (the first segment in figure 8) were below the threshold. The datasets from the damage case (the last segment in figure 8) were all substantially over the threshold, so were the data from the temperature-cycled case (the middle segment in figure 8). This is certainly an undesirable situation.

To address this issue, two approaches were used. The first method calculated a univariate novelty index for each feature variable instead of the multivariate Mahalanobis distance, while the second method made use of the minor components taken from a principal component analysis (PCA) of the multivariate feature space. The first approach was straightforward. By examining individual feature variable, the first method identified a subset of the feature variables which were insensitive to temperature variation yet sensitive to damage. The second approach first performed a PCA on the training data from

Figure 8. Diagnosis using the conventional novelty detection (Manson 2002).
different temperature conditions. Then, assuming that the variance in the data was primarily produced by temperature variation, the original feature space was projected on to a reduced feature space using minor principal components corresponding to the smallest singular values. An improved diagnosis based on this PCA is shown in figure 9. All data from the temperature-cycled case without damage (the middle segment in figure 9) have been classified as undamaged, while the actual damage cases are flagged as such.

(d) Singular value decomposition

Ruotolo & Surace (1997) note that test structures can be subjected to alterations during normal operating conditions, such as changes in mass. To handle this situation, a damage detection method based on singular value decomposition (SVD) is proposed to distinguish between changes in the working conditions and the onset of damage. Let $v_i$ be the feature vector collected at $n$ different normal configurations ($i=1, 2, ..., n$). When a new feature vector $v_c$ is collected, the whole feature vectors can be arranged in a matrix $M$,

$$M = [v_1, v_2, ..., v_n, v_c].$$

If the structure is intact, the new feature vector $v_c$ will be close to one of the existing feature vectors $v_i$, and the rank of the matrix $M$ estimated by SVD should remain unchanged by adding $v_c$ to $M$. On the other hand, if the structure experiences damage, the rank of the matrix $M$ will increase by 1. This approach is based on the premise that any structural damage will produce orthogonal changes to the feature space expanded by the existing baseline data.

(e) Auto-associative neural network

Sohn et al. (2003b) developed a combination of time-series analysis, neural networks and statistical inference techniques for damage classification, explicitly taking into account ambient variation of the system. First, a time prediction model called an auto-regressive and auto-regressive with exogenous inputs (AR-ARX) model is fit to vibration signals measured during normal operating...
conditions of the structure. Through this AR-ARX model fitting, the AR and ARX coefficients corresponding to various operational conditions are obtained. Note that the parameters of the AR-ARX model should be constant if these parameters are obtained using time signals from a time-invariant system. However, when there are operational and environmental variations of the system, these AR-ARX coefficients are not constant anymore and they will be a function of these ambient variations.

Next, data normalization is performed based on an auto-associative neural network, where target outputs are simply inputs to the network. Using the extracted features, which are the parameters of the AR-ARX model previously extracted from the normal conditions of the structure, as inputs, the auto-associative neural network is trained to characterize the underlying dependency of the extracted features on the unmeasured environmental and operational variations by treating these environmental and operational conditions as hidden intrinsic variables in the neural network. Once the network is trained properly, the output layer of the auto-associative neural network will reproduce the extracted features in the input layer. That is, the dependence of the extracted features on environmental and operational variations is registered in the trained neural network.

The auto-associative network consists of mapping, bottleneck and de-mapping layers (figure 10). The bottleneck layer contains fewer nodes than input or output layers forcing the network to develop a compact representation of the input data. Kramer (1991) shows that this auto-associative neural network is a realization of a general nonlinear principal component analysis (NLPCA). While PCA is restricted to mapping only linear correlations among variables, NLPCA can reveal the nonlinear correlations present in data. If nonlinear correlations exist among variables in the original data, NLPCA can reproduce the original data with greater accuracy and/or with fewer factors than PCA.

Finally, when a new time signal is recorded from an unknown state of the system, the parameters of the time prediction model are computed for the new signal and fed to the trained neural network. When the structure undergoes

![Figure 10. Schematic of an auto-associative neural network.](http://rsta.royalsocietypublishing.org/)

Phil. Trans. R. Soc. A (2007)
structural degradation, it is expected that the prediction errors of the neural network will increase for a damage case. A more accurate statement might be that the prediction errors will grow when the neural network encounters a new signal that has not been registered into the network before. Therefore, it is very important to capture a wide range of operational and environmental variations of the baseline system for damage detection applications. Based on this premise, a statistical classifier can be developed to identify damage.

The usefulness of the proposed approach is demonstrated using an experimental study of an eight-degree-of-freedom (DOF) spring-mass system, as shown in figure 11. Damage is simulated by placing a bumper between two adjacent masses, so that the movement of one mass is limited relative to an adjacent mass. Figure 12 shows the hardware used to simulate nonlinear damage. When one end of a bumper, which is placed on one mass, hits the other mass, impact occurs. This impact simulates damage caused by the impact from the closing of a crack during vibration. Random excitation was accomplished with a 215 N peak force electro-dynamic shaker. The root mean square (r.m.s.) amplitude level of the input was varied from 3 to 7 V. The response of an individual mass is recorded by an accelerometer attached to each mass. The challenge in this example is to distinguish signal changes caused by the insertion of the bumper from those caused by varying excitation levels.

Figure 13 shows a typical relationship between the output of the bottleneck layer and the excitation level obtained from the network corresponding to mass 2. Here, the output of the bottleneck layer is an averaged output of multiple time-series corresponding to each input level. The bottleneck output is closely related to the excitation level, in that the relationship is linear and this is sufficient to reconstruct the input at the output layer. Similar results are also observed from the networks associated with the other measurement points. This result implies that the auto-associative neural network is trained properly to capture the underlying dependence of the measured time signals on the excitation levels. Finally, a hypothesis testing technique called a sequential probability ratio test is successfully performed on the normalized features to automatically infer the damage state of the system. Additional details can be found in Sohn et al. (2003b).
It should be noted that auto-associative neural network is not the only way to realize NLPCA. Malthouse (1998) suggested principal curves (Hastie & Stuetze 1989) as an alternative strategy for NLPCA, and Jia et al. (2000) use an input-training neural network for the same purpose. In particular, Worden (2002) showed that the principal curves could be used to infer a measurement of an environmental parameter in the context of damage detection.

(f) Factor analysis

Kullaa (2003) proposed a data normalization technique based on factor analysis (Johnson & Wichern 1998). Similar to the previous two approaches, it is assumed that data normalization can be performed without measuring environmental parameters.

A factor analysis is one form of multivariate analyses, which attempt to reveal the correlation among multiple sets of variables. In particular, the factor analysis assumes that there are a smaller number of underlying factors that describe the variation of the measured variables and the measured variables are subjected to
random errors as shown in the following:

\[ y = Ax + \epsilon, \]

where \( y \) is a \( n \times 1 \) vector of the measured variables; \( A \) is a \( n \times m \) factor loading matrix (\( n > m \)); \( x \) is a \( m \times 1 \) vector of unobservable common factors; and \( \epsilon \) is a \( n \times 1 \) vector of unique factors. By using numerical simulations, Kullaa attempted to model the dependency of the first four natural frequencies (the measured variables) on the temperature (the latent common factor) by using factor analysis. Note that if the latent common factor represents the temperature parameter, then the unique factors, \( \epsilon \), are variables independent of the common factors and can be used for damage diagnosis. That is, if the structural condition deteriorates, the previously trained factor analysis cannot explain the multivariate correlation of the newly collected data, causing an increase in the unique factors. This paradigm is similar to the previous auto-associative neural network approach. One major difference is that while the auto-associative neural network can model nonlinear relationships among the measured variables, this factor analysis is limited to linear relationships among variables. The author also attempted to model nonlinear relationships by mixing different linear factor models over different data spaces. At this point, this approach was tested only with a finite element model of a vehicle crane. Operational variation was simulated by changing the configuration of the crane vehicle, and damage was simulated by stiffness reduction. The first five modal parameters of the crane were used as the damage-sensitive features. All damage cases were identified using factor analysis, whereas damage detection was unsuccessful without factor analysis (Kullaa 2004).

\( (g) \) Lamb-wave propagation method

Since the 1960s, the ultrasonic research community has studied Lamb waves for the non-destructive evaluation of plates (Bourasseau et al. 2000). Lamb waves are mechanical waves with wavelength of the same order of magnitude as the thickness of the plate. Because the Lamb waves travel long distances and can be applied with conformable piezoelectric (PZT) actuators/sensors that require little power, it may prove suitable for online SHM. In particular, the recent adaptation of wavelet analysis to Lamb-wave techniques has allowed the application of Lamb-wave techniques to composite structures to flourish (Okafor et al. 1994; Staszewski et al. 1999 a,b; Badcock & Birt 2000; Monnier et al. 2000; Lemistre & Balageas 2001; Kessler 2002). The advances in sensor and hardware technologies for efficient generation and detection of Lamb waves and the increased usage of solid composites in load-carrying structures, particularly in aircraft industries, have led to the explosion of studies that use Lamb waves for detecting defects in composite structures. For online continuous monitoring, it is important to demonstrate that the Lamb-wave-based methods are robust when used under varying environmental and operational conditions.

Sohn et al. (2003a) developed a damage detection algorithm based on wavelet transform to extract features that are less sensitive to changing boundary conditions and temperature variations. In this study, a Morlet wavelet with a narrowband driving frequency is designed as the input waveform (figure 14a). Figure 14b shows the time response of one PZT patch when the Morlet input waveform is generated at another PZT patch. The solid line represents the
baseline signal, and the dashed line shows the response time signal when delamination is within a direct line of the actuator and sensor path.

The response signal is composed of several wave modes owing to wave scattering at the boundaries. In figure 14b, the first mode, which looks like a sine wave modulated by a cosine function, is the first arrival of the $A_o$ mode associated with the direct path of the wave propagation. The second mode is another $A_o$ mode that is reflected from the edge of the plate. The observation of figure 14b clearly reveals that the first $A_o$ mode is the most sensitive to the delamination damage. Based on these observations, a damage index is defined as the function of a signal’s attenuation at a limited time span (a signal portion corresponding to the first $A_o$ mode) and at a specific frequency (the input frequency of the signal). Note that the attenuation is correlated to the amount of energy dissipated by damage. In other words, the proposed damage index measures the degree of the test signal’s energy dissipation compared with the baseline signal, especially at the first $A_o$ mode and the input frequency value.

Next, the insensitivity of the proposed damage index to the boundary condition of the panel was investigated. One edge of the plate is clamped using an aluminium plate, as shown in figure 15a, and the associated response signal for the wave path of patches 2 and 3 is shown in figure 15b. It is clearly demonstrated that the clamped boundary condition only changes the second $A_o$ wave of the signal reflected from the clamped edge of the plate (the second modulated sine wave in figure 15b). However, it does not affect the performance of the proposed damage detection algorithm at all, because the damage index is based on the signal’s attenuation only at the first $A_o$ mode.

Although the data normalization issue is explicitly taken into account in this study, the procedure developed has only been verified on relatively simple laboratory test specimens. To fully verify that the proposed approach is truly robust, it will be necessary to test the proposed approach for a wide range of operational and environmental cases and for different representative damage types, such as ply crack, fibre breakage and through-hole penetrations.

From a similar Lamb-wave inspection study, the short- and long-term stability of the normal condition data is investigated by Manson et al. (2000). For the short-term study, the Lamb-wave responses were recorded once every minute for a period of just over 18 h. For the long-term test, signals were recorded every 10 min over a period of 11 days. During the first test, no artificial temperature variation was introduced. However, temperature was fluctuated by increasing room temperature using a heater during the second test period. Based
on this long-term stability test, the authors confirmed that the effect of temperature upon the normal condition dataset is one of great importance. For instance, observations, which were flagged as coming from a damaged structure, may actually have been flagged due to a shift in the normal condition set. Conversely, actual damage may not be flagged for the same reason.

The authors also investigated the long-term stability against humidity variation (Manson et al. 2001). The authors report that while temperature variation mainly produced a phase shift of the signal with a slight amplitude change, humidity changed the amplitude of the Lamb-wave response. However, the authors concluded that Lamb-wave propagation characteristics were more sensitive to temperature variations than changes in humidity.

### 4. Conclusions and discussions

Currently, various sensors are available to measure parameters relevant to ambient conditions of a structure and, in fact, several in-service structures are reported to have been instrumented with thermocouples, anemometers and humidity sensors. However, the information gathered from these sensors is still not directly applied to existing damage detection algorithms. It is the author’s opinion that SHM systems will not be accepted in practical applications unless robust techniques are developed to explicitly account for environmental and operational constraints/conditions of the systems to be monitored. However, there are few proven techniques that are able to address these issues properly.

The developments presented here will allow some progress in in-service monitoring of aerospace, automotive, civil and mechanical systems, which are subject to various operational and environmental conditions. Such a monitoring system will be less prone to false-positive indication of damage. To minimize this false indication of damage and to develop a more robust monitoring system, it is critical that training datasets are collected over a wide range of environmental and operational conditions of the system. Otherwise, the damage classifiers presented may not be able to make any definite statement regarding the existence of damage because abnormal operational conditions can have similar effects as that of damage on the monitoring system. In addition, for real world applications, a hybrid approach from all three categories will most probably be necessary.
Overall, it is the opinion of the author that sufficient evidence exists to promote the need for data normalization procedures for damage detection in structures. It is clear, however, that the development of data normalization procedures needs to be more focused on specific applications and industries that would benefit from this technology, such as the health monitoring of bridges, offshore oil platforms, airframes and other structures in fields. Additionally, research should be focused more on tests of real structures in their operating environment, rather than laboratory tests of representative structures. Owing to the magnitude of such projects, more cooperation will be required between academia, industry and government organizations. If specific techniques can be developed to quantify and extend the life of structures, the investment made in this technology will clearly be worthwhile.

References


Manson, G. 2002 Identifying damage sensitive, environmental insensitive features for damage detection. In 3rd Int. Conf. Identification in Engineering Systems, University of Wales Swansea, UK.


Phil. Trans. R. Soc. A (2007)


Worden, K. 2002 Inferential parameterisation using principal curves. 3rd International Conference on Identification in Engineering Systems, University of Wales Swansea, UK.


