Laory, I., Trinh, T.N. and Smith, I.F.C. "Evaluating two model-free data interpretation methods for measurements that are influenced by temperature" Advanced Engineering Informatics, Vol 25, No 3, 2011, pp 495-506, doi 10.1016/j.aei.2011.01.001

Evaluating two model-free data interpretation methods for measurements that are influenced by temperature

Irwanda Laory, Thanh N. Trinh, Ian F. C. Smith

Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 18,CH-1015 Lausanne, Switzerland

Abstract

Interpreting measurement data to extract meaningful information for damage detection is a challenge for continuous monitoring of structures. This paper presents an evaluation of two model-free data interpretation methods that have previously been identified to be attractive for applications in structural engineering: moving principal component analysis (MPCA) and robust regression analysis (RRA). The effects of three factors are evaluated: (a) sensor-damage location, (b) traffic loading intensity and (c) damage level, using two criteria: damage detectability and the time to damage detection. In addition, the effects of these three factors are studied for the first time in situations with and without removing seasonal variations through use of a moving average filter and an ideal low-pass filter. For this purpose, a parametric study is performed using a numerical model of a railway truss bridge. Results show that MPCA has higher damage detectability than RRA. On the other hand, RRA detects damages faster than MPCA. Seasonal variation removal reduces the time to damage detection of MPCA in some cases while the benefits are consistently modest for RRA.

Keywords: Moving principal component analysis; robust regression analysis; damage detection; damage detectability; time to damage detection; seasonal temperature variation.

1 Introduction

Recently, the collapse of civil structures such as the I-35W bridge (USA, 2007) [1-2] and the Paris Airport (France, 2004) [3] has decreased public confidence in the safety of structures. Thus, to ensure that structures behave according to design criteria, it is useful to monitor their performance. Monitoring for possible damage is referred to as structural health monitoring (SHM). Due to advances in sensor technology, data acquisition systems and computational power, the number of structures that are monitored is growing. Thus, large quantities of measurement data are retrieved every day and much more will be available in the future. Extracting useful information from this data to detect damage is a challenge for SHM. This task is even more difficult when measurement data are influenced by environmental variations, such as temperature, wind and humidity. Brownjohn et al [4] studied the thermal effects on performance on Tamar Bridge and showed that thermal effects dominate the measured bridge behaviour. Catbas et al. [5] observed that the peak-to-peak strain differential due to temperature over a one-year period is more than ten times higher than the strain due to observed maximum daily traffic.

Generally, there are two classes of data interpretation methods in SHM: model-based methods and model-free methods. These two classes are complementary since they are appropriate in different contexts. Strengths and weaknesses of both classes have been summarized in the ASCE state-of-the-art report on structural identification of constructed systems [6].

Model-based data interpretation methods typically utilize measurement data to identify models that are able to reflect the real behavior of structures. This is done through comparing structural responses with predictions of behavior models [7-8]. Thus, such methods involve the development and use of detailed models to validate the results. However, for civil infrastructures, creating such models is often difficult and expensive, and may not always

reflect real behavior due to the presence of uncertainties in complex civil-engineering structures [9]. Furthermore, model-based methods are not necessarily successful in identifying the right anomaly [10].

Alternatively, model-free data interpretation methods involve analyzing data without geometrical and material information. They evaluate data statistically and thus, they do not require knowledge of structural behavior. Since these methods involve only tracking changes in time-series signals, they are well-suited for analyzing measurements during continuous monitoring of structures. Omenzetter et al. [11] used a discrete wavelet transform method (DWT) [12-13] to detect changes in strain. Omenzetter and Brownjohn [14] proposed an autoregressive integrated moving average model method (ARIMA) to detect damage from measurements. Lanata and Grosso [15] applied a proper orthogonal decomposition method for continuous static monitoring of structures. Yan et al. [16-17] proposed local PCA-based damage detection for vibration-based SHM. The method involves a two-step procedure: a clustering of data space into several sub-regions and the application of PCA in each local region. All these studies are limited to a single methodology without comparison to other methods.

While much research has been performed, in general, no methodology for detection of anomalous behavior from measurement data can be reliably applied to complex structures in practical situations [9]. Most recently, a comparative study of many data interpretation methods for continuous monitoring has been performed [18]. Posenato et al. [9, 18] proposed two model-free data interpretation methods, MPCA and RRA, to detect and localize anomalous behavior for the particular context of civil-engineering structures, and compared their performance with many other methods: DWT [12-13], ARIMA [14], auto regressive with moving average [19-21], Box-Jenkins method [13], wavelet packet transform [22-23], instance

based method [24] and correlation anomaly scores analysis [25]. These comparative studies demonstrated that the performances of (MPCA) [9] and (RRA) [26-28] for anomaly detection were superior to other methods when dealing with civil-engineering challenges such as high noise levels, missing data and outliers. Both methods were also observed to require low computational resources to detect anomalies, even when there were large quantities of data. However, these methods were not evaluated in terms of detectability and time to detection. For example, in some cases, damage is not detectable for several weeks after it occurs.

This paper includes a study of strategies to reduce time to detection. It is carried out through investigating the influence of data processing for removing seasonal temperature-variations on these two model-free data-interpretation methods. Also, two criteria, damage detectability and time to detection, are evaluated with respect to changes in sensor-damage location, traffic loading, and damage level in a truss bridge. The paper is organized as follows: Section 2 describes moving principal component analysis and robust regression analysis. Section 3 introduces two data processing techniques for removing seasonal variations. Section 4 presents the results of the study. Section 5 discusses the effects of sensor-damage location, traffic loading and damage level in a case study of a railway bridge in Zangenberg, Germany. This section also compares MPCA and RRA in situations where seasonal variations are removed or not.

2 Model-free data interpretation methods for continuous monitoring of structures

Model-free data interpretation methods involve analysing measurement data and tracking changes in a structure without use of geometrical and material information. This section describes two model-free data interpretation methods, MPCA and RRA.

2.1 MPCA for continuous monitoring of structures

MPCA is a modified version of principal component analysis (PCA) [29]. PCA is a mathematical process of transforming a number of possibly correlated variables into a smaller number of uncorrelated variables, called principal components; the first few retain most of the variation present in the original variables. In the context of structural health monitoring, it is used to enhance the discrimination between features of undamaged and damaged structures and to limit calculation time. Nevertheless, for continuous monitoring when the number of measurements increases, an anomaly may be detected late due to the measurements that define the undamaged state. At the time when damage occurs, the influence of old measurements (undamaged state) is much higher than that of new measurements (damaged state). Additional time is required to have the eigenvector values dominated by new measurements so that they change values enough to indicate damage. Also, the computational time to compute principal components increases with the number of measurements. To address this, Posenato et al. [9] proposed "moving" PCA (MPCA) that computes the principal components inside a moving window of constant size. MPCA is applied to anomaly detection for continuous monitoring of structures as shown in Figure 1.

MPCA is carried out by observing the evolution of principal components. Damage is identified when there is a change in the values of principal components. Principal components are not employed to reconstruct data sets. MPCA starts with the construction of a matrix $\mathbf{U}$ that contains all time histories measured at all sensors as follows

$$\mathbf{U}(t) = \begin{bmatrix} u_1(t_1) & u_2(t_1) & \dots & u_{N_s}(t_1) \\ u_1(t_2) & u_2(t_2) & \dots & u_{N_s}(t_2) \\ \dots & \dots & \dots & \dots \\ u_1(t_{N_m}) & u_2(t_{N_m}) & \dots & u_{N_s}(t_{N_m}) \end{bmatrix} \quad (1)$$

N_m is the total number of observations of a time history, and N_s is the total number of sensors on the structure. Thus, each column of the matrix $\mathbf{U}$ is the time history of each sensor.

A fixed-size window moves along the columns of $\mathbf{U}$ to extract datasets at each time step k as

$$\mathbf{U}_k(t) = \begin{bmatrix} u_1(t_k) & u_2(t_k) & \dots & u_{N_s}(t_k) \\ u_1(t_{k+1}) & u_2(t_{k+1}) & \dots & u_{N_s}(t_{k+1}) \\ \dots & \dots & \dots & \dots \\ u_1(t_{k+N_w}) & u_2(t_{k+N_w}) & \dots & u_{N_s}(t_{k+N_w}) \end{bmatrix} \quad \text{for } k = 1, \dots, (N_m - N_w) \qquad (2)$$

N_w is the total number of observations within the moving window. To apply PCA to this dataset $\mathbf{U}_k$, each time series inside the active window is first normalized by subtracting its mean value. At time t_j, the vector of the normalized measurements is:

$$\mathbf{u}(t_j) = \begin{bmatrix} u_1(t_j) - \overline{u}_1 & u_2(t_j) - \overline{u}_2 & \dots & u_{N_s}(t_j) - \overline{u}_{N_s} \end{bmatrix} \qquad (3)$$

$\mathrm{u}(t_j)$ is the vector of normalized measurements at time t_j and $\overline{u}_i$ is the mean values for sensor i. Then, the $N_s \times N_s$ covariance matrix $\mathbf{C}_k$ for all measurement locations, summed over all time samples from time step k to $k + N_w$, is given by

$$\mathbf{C}_k = \sum_{j=k}^{k+N_w} \mathbf{u}\left(t_j\right)^T \mathbf{u}\left(t_j\right) \qquad (4)$$

The eigenvalues λ_i and eigenvectors $\boldsymbol{\psi}_i$ of the covariance matrix satisfy

$$\left(\mathbf{C}_k - \lambda_i \mathbf{I}\right)\boldsymbol{\psi}_i = 0 \qquad \text{for } i = 1, \dots, N_s \qquad (5)$$

Here, an eigenvector $\boldsymbol{\psi}_i$ is also called a principal component. Sorting the eigenvectors by eigenvalue in decreasing order, the components are arranged in order of significance. The first few principal components contain most of the variance of time series while the remaining components are defined by measurement noise. Thus, MPCA is used for anomaly detection by analyzing only the eigenvectors that are related to the first few eigenvalues.

Application of MPCA for anomaly detection during continuous monitoring includes two phases: training and monitoring. In the training phase, the structure is assumed to behave normally (no damage). The aim of this initialization phase is to estimate the variability of the time series and to define the thresholds of normal measurements in order to detect anomalous behavior in the monitoring phase. To do this, each eigenvector $\boldsymbol{\psi}_i$ during the reference period with time step $k = 1$ to $k = N_{ref} - N_w$, is stored and values for mean $\boldsymbol{\mu}_i$ and standard deviation $\boldsymbol{\sigma}_i$ are determined. The thresholds of normal measurements are defined to be $\pm 6\boldsymbol{\sigma}$. If there is no anomaly (damage), the eigenvectors remain within these thresholds.

In the monitoring phase, the window continues moving along time series to compute new eigenvalues and eigenvectors at each time step. When damage occurs, the mean values of time series and the components of covariance matrix change and as consequence, so do values of eigenvalues and eigenvectors. If the value ψ_{ji} of vector $\boldsymbol{\psi}_i$ exceeds the threshold bounds $\left(\pm \mathbf{6}\sigma_{ji}\right)$, an anomaly is flagged by sensor j at time step $k + N_w$.

An advantage of the use of a moving window rather than all measurements is speed of calculation of process parameters, thereby detecting the presence of anomalies in structures more rapidly since very old measurements do not bias results. However, there is always some delay in detection; this paper examines this aspect.

Another advantage is adaptability. Once new behavior is identified, adaptation allows detection of further anomalies. This is done by defining a new training phase corresponding to the new state of structure after each anomaly [9]. Note that the new training phase starts immediately after anomaly detection.

A key parameter of MPCA is the size of the moving window N_w. This parameter should be sufficiently large so that it is not influenced by variations in measurements due to

environmental effects and small enough to provide rapid anomaly detection. As stated earlier, if the time series has periodic variability, the window size should be at least as long as the longest period. This choice ensures that mean values of the time series within a window are stationary and that eigenvalues of the covariance matrix do not have periodic behavior.

2.2 RRA for continuous monitoring of structures

Similar to MPCA, application of RRA for anomaly detection for continuous monitoring also includes two phases: training and monitoring, as shown in Figure 2. However, RRA does not use a window moving along time series, but rather the whole time series in the training phase from first time step $k = 1$ to $k = N_{ref}$ to define the thresholds of confidence intervals for detecting anomalous behaviors in the monitoring phase. The idea behind RRA for continuous monitoring of structures is to find all sensor pairs that have a high correlation in the training phase, and then to focus on the correlation of these couples to detect anomalies in the monitoring phase.

To find sensor pairs with a high correlation, the correlation coefficient $r_{si,sj}$ between two sensors s_i and s_j $\left(i, j = [1 - N_s]\right)$ are computed and compared with the correlation coefficient threshold (r_t). All sensor pairs having a correlation coefficient greater than the threshold are contained in matrix $\mathbf{S}$ in order to compute the robust regression line regression. The linear relation between s_i and s_j is written as

$$s'_j = as_i + b \tag{6}$$

where s'_j represents the value of s_j calculated according to the linear relation. a and b are the coefficients of the robust regression line estimated from measurements in the training phase. These coefficients are estimated using iteratively reweighted least squares. Then, the

standard deviation σ_{ij} of the difference $\left|s_j - s'_j\right|$ in the training phase is computed to define the threshold of confidence intervals for each sensor pair s_i and s_j. The threshold bounds of normal measurements (no anomaly) is defined to be $6\sigma_{ij}$.

In the monitoring phase, s'_j is computed for each time step based on the linear relation above. If the differences $\left|s_j - s'_j\right|$ in N_{out} consecutive measurements from time step k to $k + N_{out}$ are out of the threshold bounds $\left(6\sigma_{ij} < \left|s_j - s'_j\right|\right)$, the anomaly is detected by the sensor pair s_i and s_j at time step k. Aside from the advantage of being insensitive to outliers and missing data, RRA is capable of adapting to the new state of a structure for identifying further anomalies by redefining a new training phase after an anomaly is identified.

2.3 Comparison of methods

MPCA and RRA are essentially based on correlation between measurements. However, the derivation of the correlation for each method is different. The correlation for MPCA is determined using a covariance matrix within a moving window. On the other hand, the correlation for RRA is determined through the use of robust regression lines. Another difference is that MPCA performs analysis of the correlation for all measurement data, while RRA does the analysis for all measurement pairs that have high correlations, i.e. the coefficient of correlation beyond a defined threshold. Also, MPCA performs PCA within a moving window in order to reduce the size of the buffer in the training phase. For RRA, on the other hand, the identification is performed for every measurement. Therefore, a new data point is not buffered by previous measurements. Three principal differences between MPCA and RRA are summarized in Table 1.

3 Data processing for removing seasonal temperature variations

Temperature variations have a dominant effect on the response and performance of a structure when compared with the influence of factors such as wind and traffic. Temperature variation generally includes seasonal and daily variations as shown in Figure 3. Daily variation is the periodic change in temperature over 24 hours while seasonal variation is the periodic change from spring to winter every year. The frequency of daily variation is thus much higher than that of seasonal variation. The seasonal temperature variation can be simply filtered from measured temperature data by using low-pass filter methods.

The aim of using low-pass filter methods is to remove seasonal temperature effects from measurement data before implementing MPCA and RRA. To do this, it is assumed that the relation between seasonal temperature variation and its effect on structural responses is linearly expressed as

$$\varepsilon_s = \alpha \times \Delta t_s \tag{7}$$

where ε_s is the structural response due to the seasonal temperature variation Δt_s.

In order to calculate the coefficient α, the extraction of seasonal variations from measurements in the first year (training phase) is studied for two filters: moving average [30] and ideal low-pass [31]. The moving average filter used in this study is a center moving average with a flat weighting. It smoothes measured signals through replacing each data point with the average of neighboring data points inside a moving window. Alternatively, an ideal low-pass filter allows for passing low frequency components in a signal, while attenuating frequency components which are higher than a specified cutoff frequency. This study utilizes an ideal low-pass filter of the signal processing toolbox in MATLAB.

Given the calculated coefficient α , the structural response due to seasonal temperature variation in subsequent years is predicted as

$$\varepsilon'_s = \alpha \times \Delta t'_s \tag{8}$$

where ε'_s is the predicted structural response due to the seasonal temperature variation $\Delta t'_s$. The structural response ε_d due to daily temperature variation in subsequent years is achieved by decomposing structural response due to seasonal temperature variation from response measurement ε as

$$\varepsilon_d = \varepsilon - \varepsilon'_s \tag{9}$$

4 Parametric study

As discussed in the introduction, the aim of this paper is to evaluate two model-free data interpretation methods: MPCA and RRA. The evaluation is carried out through a parametric study based on a framework as shown in Figure 4. MPCA and RRA are evaluated in term of two features, damage detectability and time to damage detection, with respect to the change in sensor-damage locations (distances from sensors to damage locations), traffic loading and damage levels.

For the purposes of this paper, damage detectability is defined as follows:

$$\text{Damage detectability (\%)} = 100\% - \text{Minimum detectable damage level (\%)} \tag{10}$$

where the minimum detectable damage level is the smallest percentage loss of stiffness in a member that can be detected. Damage detectability is the capability to interpret measurement data in order to identify damage. For continuous monitoring of structures, damage detectability may be affected by sensor-damage locations and traffic loading levels.

Time to damage detection is the time interval from the moment when damage occurs to the one when damage is detected. In addition to sensor-damage locations and traffic loading intensity, damage levels may also affect the time to damage detection of these methods. Finally, the influence of data processing to remove seasonal variations on these two features is also investigated in the case study that is described next.

5 Case study

To study the performance of model-free data interpretation methods for continuous monitoring of structures, a railway truss bridge in Zangenberg, Germany has been selected. This 80-m steel bridge is composed of two parallel trusses each having 77 members. Their properties are summarized in Table 2. The truss members are made of steel having an elastic modulus of 200 GPa and a density of 7870 kg/m^3. A finite element analysis under traffic loading and temperature variation provides responses (strains) at 15 members (marked by black bars on the truss in Figure 5. Figure 5 also shows damage locations marked as black dots. Only one truss of the bridge is modeled and the truss model is fixed at both ends.

Traffic loading is simulated by applying a randomly generated vertical load (0-19 tonnes) at each node in the bottom chords. A load of 19 tonnes is equivalent to an axle load of a railway locomotive. In this example, damage is assumed to result in a loss of axial stiffness. Damage is introduced at three locations. The first damage is at a sensor location while the other two occur away from sensors. Damage scenarios are used to evaluate the effects of sensor-damage location on the damage detectability and time to anomaly detection of the two methods. Furthermore, varying damage levels and traffic loading are simulated to evaluate the effects of

traffic loading and damage levels on anomaly detection. All effects are evaluated with and without data processing for seasonal temperature effects.

Figure 6 shows a time history of strain measurement at the top chord of the truss. Measurements are taken four times a day. Due to seasonal temperature effects, periodic behavior in structural response is visible. It is seen that although 20% damage is introduced at 2190 days at a sensor location (location 1), it is not recognizable from the plot due to temperature variations. However, this damage is detectable using MPCA and RRA as shown in Figure 7. The time of damage occurrence is identified by observing the variation of the eigenvector values. Damage is detected when values of the eigenvector exceed the threshold bounds ($\pm 6\sigma$). In a similar way, for RRA, the identification is performed by observing the evolution of the difference between measurements and the regression line.

5.1 Damage detectability with and without seasonal variation removal

5.1.1 Sensor-damage location

This section evaluates the effect of sensor-damage location on damage detectability of MPCA and RRA with and without data processing. Detection studies for three damage scenarios with one scenario at a sensor (location 1) and two scenarios away from sensors (location 2 and 3) are carried out. Damage levels from 3% to 100% (complete damage) with traffic loading of 100% (percentage of 19 tons) are applied at each node in the bottom chords) are investigated. Figure 8 shows the damage detectability of MPCA and RRA for all damage scenarios. As expected, for both methods, when damage occurs close to sensors, damage detectability is highest. In addition, by comparing minimum detectable damage levels without seasonal variation removal, Figure 8 demonstrates the damage detectability of MPCA is higher than that

of RRA, particularly for damage far away from sensors. For damage at location 3, the damage detectability decreases from 90% for MPCA to 68% for RRA.

It is also observed in Figure 8 that the seasonal variation removal has a significant influence on the damage detectability of MPCA when compared with RRA. Indeed, at damage location 3 after removing seasonal variations, the damage detectability when using MPCA decreases 30% (from 90% to 60%) while that using RRA is unchanged (about 1%). The discussion at the end of this paper provides an explanation.

5.1.2 Traffic loading

The effect of traffic loading with respect to position, magnitude and speed on the damage detectability of MPCA and RRA is studied. Traffic loading contributes to signal noise. Five traffic loading levels presented in terms of the percentage of 19 tons (from 20% to 100%) are analyzed for damage at location 2. Figure 9 shows the effect of traffic loading on the damage detectability of both methods with and without seasonal variation removal.

The damage detectability of MPCA and RRA decreases as traffic loading increases. It is also found that the effect of traffic loading on the damage detectability of MPCA is less than that of RRA. In addition, Figure 9 indicates that removal of seasonal variations reduces the damage detectability of MPCA while it does not affect that of RRA. For example, at traffic loading level of 40%, the damage detectability of MPCA without data processing is 95%, but decreases to 75% when removing seasonal variation. At the same traffic loading, however, the damage detectability of RRA changes only 2% after removing seasonal variation.

5.2 Time to detection with and without seasonal variation removal

5.2.1 Sensor-damage location

The effect of sensor-damage location on time to detection for MPCA and RRA is also studied for two cases: with and without seasonal variation removal. The same three damage scenarios, as introduced in the previous section, are examined. This study is performed for a damage level of 60% and a traffic level of 50%.

The time to damage detection of MPCA and RRA with and without seasonal variation removal is presented in Figure 10 . It is seen that for MPCA, damage taking place close to sensors is detected faster than that away from sensors. For example, the time to detect the damage at location 1 is 23 days, and this increases to 42 days for damage at location 2. In contrast to MPCA, RRA can detect damage away from a sensor as fast as damage at the sensor. This demonstrates that for these scenarios, time to damage detection is not affected by sensor-damage location when using RRA.

It is also observed in Figure 10 that for MPCA, removing seasonal variation reduces significantly the time to damage detection for damage at a sensor location. Indeed, time to detect damage at location 1 decreases from 23 days to one day after removing seasonal variation. Eigenvector time histories computed from this damage scenario at location 1 before and after removing seasonal variation are presented in Figure 11 . For RRA, nevertheless, the influence of seasonal variation removal on time to damage detection is negligible. As shown in Figure 12 , the time to detect damage remains the same before and after removing seasonal variation.

5.2.2 Traffic loading

Similarly to the study on traffic loading in Section 5.1, this section analyzes the effect of traffic loading on the time to damage detection. All parameters used here are the same as those in the previous section, except that the analyses are performed for 60% damage at location 2. The results of the time to detection for both methods with and without seasonal variation removal are presented in Figure 13. It is seen that as the traffic loading level increases, time to damage detection of MPCA increases while that of RRA does not change. The increase of time to detection of MPCA is 40 days without seasonal variation removal and is 160 days when removing seasonal variation. On the other hand, the time to detection of RRA remains at one day for traffic loading from 20% to 80%. Therefore, it can be concluded that RRA is unaffected by traffic loading. It is also noted in Figure 13 that the influence of removing seasonal variation on time to damage detection of MPCA is not stable. This removal only reduces time to detection for low traffic loading level (less than 40%).

Figure 14 presents eigenvector time histories when using MPCA with and without data processing at a traffic loading of 10%. It shows that seasonal variation removal is able to reduce time to detection from 23 days to 9 days.

5.2.3 Damage level

In order to evaluate the effect of damage level on the time to damage detection of both methods, a range of damage severity from 20% to 100% (complete damage) are considered at a traffic loading level of 50%. Figure 15 presents the time to damage detection using both methods before and after removing seasonal variation. As anticipated, high damage level results in short times to detection. For MPCA, without use of seasonal variation removal, the time to detection reduces from 131 days to only one day for damage level changes from 20% to 100%. Figure

15 also indicates that although seasonal variation removal results in increasing the minimum detectable damage level of MPCA to 40%, it can improve times to detection when damage level is greater 60%. Figure 16 shows the eigenvector time histories related to the main eigenvalue at damage level 80% before and after the data processing. The time to damage detection improves by 18 days.

For RRA, with the same sensor-damage location and traffic loading, changes in damage levels do not affect time to damage detection, even when removing seasonal variations. For damage levels larger than 50%, RRA can detect damage immediately.

5.3 Summary of observations

Results of the parametric study are summarized in Table 3. The following observations are made within the context of the definitions of damage detectability and time to detection described in this paper:

- The time to detection of RRA is shorter than that of MPCA.
- MPCA provides greater damage detectability than RRA.
- Removing seasonal variations decreases the damage detectability of MPCA.
- Removing seasonal variations has a small influence on damage detectability of RRA.
- Decreases in traffic loading increase damage detectability of both MPCA and RRA.
- When damage detectability is high (in case of high damage levels and low traffic loading), Removing seasonal variations may shorten the time to damage detection of MPCA.
- The negative effect of traffic loading and damage level on time to detection using RRA is less than that using MPCA.

5.4 Discussion

MPCA provides greater damage detectability than RRA. A possible reason for this observation is that MPCA performs the analysis of correlation of all measurements while RRA does so only for highly correlated measurement pairs (as mentioned in Section 2.3). Therefore, MPCA is more sensitive to change (or damage) of structures than RRA.

RRA detects damage faster than MPCA. This can be attributed to another difference between RRA and MPCA. MPCA evaluates eigenvectors within a moving window, resulting in a buffer from old measurements, whereas RRA evaluates directly the difference between a measured value and a robust regression line for each of the measurement points.

Another observation is that removing seasonal variation has a *negative* influence on damage detectability of MPCA while it has a very small influence on that of RRA. Seasonal temperature variations can be considered as a load case which has a different period from other load cases such as daily temperature variation and traffic loading. Such load cases establish a correlation of measurements. For MPCA, removing seasonal variation causes the loss of major correlation within data sets, resulting in fluctuation in eigenvector time histories. As a result, the threshold bounds also increase. This is seen in Figure 17 with eigenvector time histories using MPCA with and without removing seasonal variations for a damage scenario (20% damage at location 1). Consequently, seasonal variation removal reduces damage detectability of MPCA.

Unlike MPCA, which is based on correlations of all measurements, RRA performs the analysis only for measurement pairs which have strong correlations within the training phase. Therefore, the absolute difference between measurement data and the regression line will always be relatively small before and after removing seasonal variations; and the threshold bounds will be similar with and without seasonal variation removal. Figure 18 shows the plots

of the absolute value of the difference between strain measurement and robust regression line before and after data processing. The sizes of threshold bounds are similar.

6 Conclusions

From the observations made in the previous section and considering the scope of this study, the following conclusions are drawn:

- Moving principal component analysis (MPCA) is generally better than robust regression analysis (RRA) in terms of damage detectability. On the other hand, RRA is better than MPCA in terms of time to damage detection.
- RRA is not sensitive to seasonal variations.
- For MPCA, there is a trade-off between damage detectability and time to damage detection when removing seasonal variations.

From evaluating the performance of MPCA and RRA, it is recommended that these two methods should be considered complementary since they are most appropriate in different contexts. Therefore, synergies between MPCA and RRA could result in better damage detection strategies for structural health monitoring.

Acknowledgement

This work was partially funded by the Swiss Commission for Technology and Innovation. The authors are grateful to Professor Alain Nussbaumer (EPFL) for providing details of the case study and Professor Robert Amor (University of Auckland) for his comments on the preliminary version of this paper. Dr. Daniele Posenato, Dr. Nizar Bel Hadj Ali and Dr. Prakash Kripakaran are recognized for their contribution to the research that motivated this study.

References

[1] M. Liao, T. Okazaki, R. Ballarini, A.E. Schultz, T.V. Galambos, Analysis of Critical Gusset Plates in the Collapsed I-35W Bridge, in: L. Griffis, T. Helwig, M. Waggoner, M. Hoit (Eds.) Proceedings of the 2009 Structures Congress - Don't Mess with Structural Engineers: Expanding Our Role, ASCE, Austin, Texas, 2009, pp. 237-237.

[2] N. Subramanian, I-35W Mississippi river bridge failure - Is it a wake up call?, Indian Concrete Journal, 82 (2008) 29-38.

[3] P. Sparaco, Airport collapse, in: Aviation Week and Space Technology 2004, pp. 37.

[4] J.M.W. Brownjohn, K. Worden, E. Cross, D. List, R. Cole, T. Wood, Thermal effects on performance on Tamar Bridge, in: The Fourth International Conference on Structural Health Monitoring of Intelligent Infrastructure Zurich, Swirzerland, 2009, pp. 152.

[5] F.N. Catbas, M. Susoy, D.M. Frangopol, Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data, Engineering Structures, 30 (2008) 2347-2359.

[6] ASCE, Structural Identification of Constructed Systems, in, Structural Identification Comittee, American Society of Civil Engineers. In Press, 2010.

[7] C.G. Koh, T.N. Thanh, Challenges and Strategies in Using Genetic Algorithms for Structural Identification, in: B.H.V. Topping, Y. Tsompanakis (Eds.) Soft Computing in Civil and Structural Engineering, Saxe-Coburg Publications, Stirlingshire, UK, 2009, pp. 203-226.

[8] C.G. Koh, T.N. Thanh, Output-only Substructural Identification for Local Damage Detection, in: The Fifth International Conference on Bridge Maintenance, Safety and Management, Philadelphia, Pennsylvania, USA, 2010.

[9] D. Posenato, F. Lanata, D. Inaudi, I.F.C. Smith, Model-free data interpretation for continuous monitoring of complex structures, Advanced Engineering Informatics, 22 (2008) 135-144.

[10] S. Saitta, B. Raphael, I.F.C. Smith, Data mining techniques for improving the reliability of system identification, Advanced Engineering Informatics, 19 (2005) 289-298.

[11] P. Omenzetter, J.M.W. Brownjohn, P. Moyo, Identification of unusual events in multi-channel bridge monitoring data, Mechanical Systems and Signal Processing, 18 (2004) 409-430.

[12] Z. Hou, M. Noori, R. St. Amand, Wavelet-based approach for structural damage detection, Journal of Engineering Mechanics, 126 (2000) 677-683.

[13] P. Moyo, J.M.W. Brownjohn, Detection of Anomalous Structural Behaviour Using Wavelet Analysis, Mechanical Systems and Signal Processing, 16 (2002) 429-445.

[14] P. Omenzetter, J.M.W. Brownjohn, Application of time series analysis for bridge monitoring, Smart Materials and Structures, 15 (2006) 129-138.

[15] F. Lanata, A.D. Grosso, Damage detection and localization for continuous static monitoring of structures using a proper orthogonal decomposition of signals, Smart Materials and Structures, 15 (2006) 1811-1829.

[16] A.M. Yan, G. Kerschen, P. De Boe, J.C. Golinval, Structural damage diagnosis under varying environmental conditions--Part I: A linear analysis, Mechanical Systems and Signal Processing, 19 (2005) 847-864.

[17] A.M. Yan, G. Kerschen, P. De Boe, J.C. Golinval, Structural damage diagnosis under varying environmental conditions--part II: local PCA for non-linear cases, Mechanical Systems and Signal Processing, 19 (2005) 865-880.

[18] D. Posenato, P. Kripakaran, D. Inaudi, I.F.C. Smith, Methodologies for model-free data interpretation of civil engineering structures, Computers & Structures, 88 (2010) 467-482.

[19] K.K. Nair, A.S. Kiremidjian, K.H. Law, Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure, Journal of Sound and Vibration, 291 (2006) 349-368.

[20] E. Peter Carden, J.M.W. Brownjohn, ARMA modelled time-series classification for structural health monitoring of civil infrastructure, Mechanical Systems and Signal Processing, 22 (2008) 295-314.

[21] H. Sohn, J.A. Czarnecki, C.R. Farrar, Structural Health Monitoring Using Statistical Process Control, Journal of Structural Engineering, 126 (2000) 1356-1363.

[22] J.-G. Han, W.-X. Ren, Z.-S. Sun, Wavelet packet based damage identification of beam structures, International Journal of Solids and Structures, 42 (2005) 6610-6627.

[23] Z. Sun, C.C. Chang, Structural Damage Assessment Based on Wavelet Packet Transform, Journal of Structural Engineering, 128 (2002) 1354-1361.

[24] M. Berry, G. Linoff, Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, Inc., 1999.

[25] T. Ide, S. Papadimitriou, M. Vlachos, Computing correlation anomaly scores using stochastic nearest neighbors, in: The 2007 Seventh IEEE International Conference on Data Mining, IEEE, Omaha NE, USA, 2007, pp. 523-528.

[26] R. Andersen, Modern methods for robust regression, SAGE Publications, Inc, 2008.

[27] N. Jajo, A Review of Robust Regression and Diagnostic Procedures in Linear Regression, Acta Mathematicae Applicatae Sinica (English Series), 21 (2005) 209-224.

[28] J. Maeck, G. De Roeck, Damage Assessment Using Vibration Analysis on The Z24-Bridge, Mechanical Systems and Signal Processing, 17 (2003) 133-142.

[29] M. Hubert, P.J. Rousseeuw, K.V. Branden, ROBPCA: A new approach to robust principal component analysis, Technometrics, 47 (2005) 64-79.
[30] S.W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Pub, 1997.
[31] M. Owen, Practical Signal Processing, in, Cambridge university press, 2007.

List of figures

(left), and with removing seasonal variation using moving average filter (middle) and ideal low-pass filter (right).

Figure 13. Time to damage detection at location 2 using MPCA (left) and RRA (right) for traffic loading levels.

Figure 14. Eigenvector time histories related to the main eigenvalues computed using MPCA for damage level of 60% at location 2 and traffic loading 20% % without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low-pass filter (right).

Figure 15. Time to damage detection using MPCA (left) and RRA (right) for damage scenarios with the range of damage level from 20% to 80% at location 2.

Figure 16. Eigenvector time histories related to the main eigenvalues computed using MPCA for damage level of 80% at location 2 and traffic loading 50% without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low pass filter (right).

Figure 17. Plots of eigenvector time histories related to the main eigenvalues computed using MPCA without (left) and with (right) data processing for removing seasonal temperature effects.

Figure 18. Plots of absolute value of the difference between strain measurement and regression line computed using RRA without (left) and with (right) data processing for removing seasonal temperature effects.

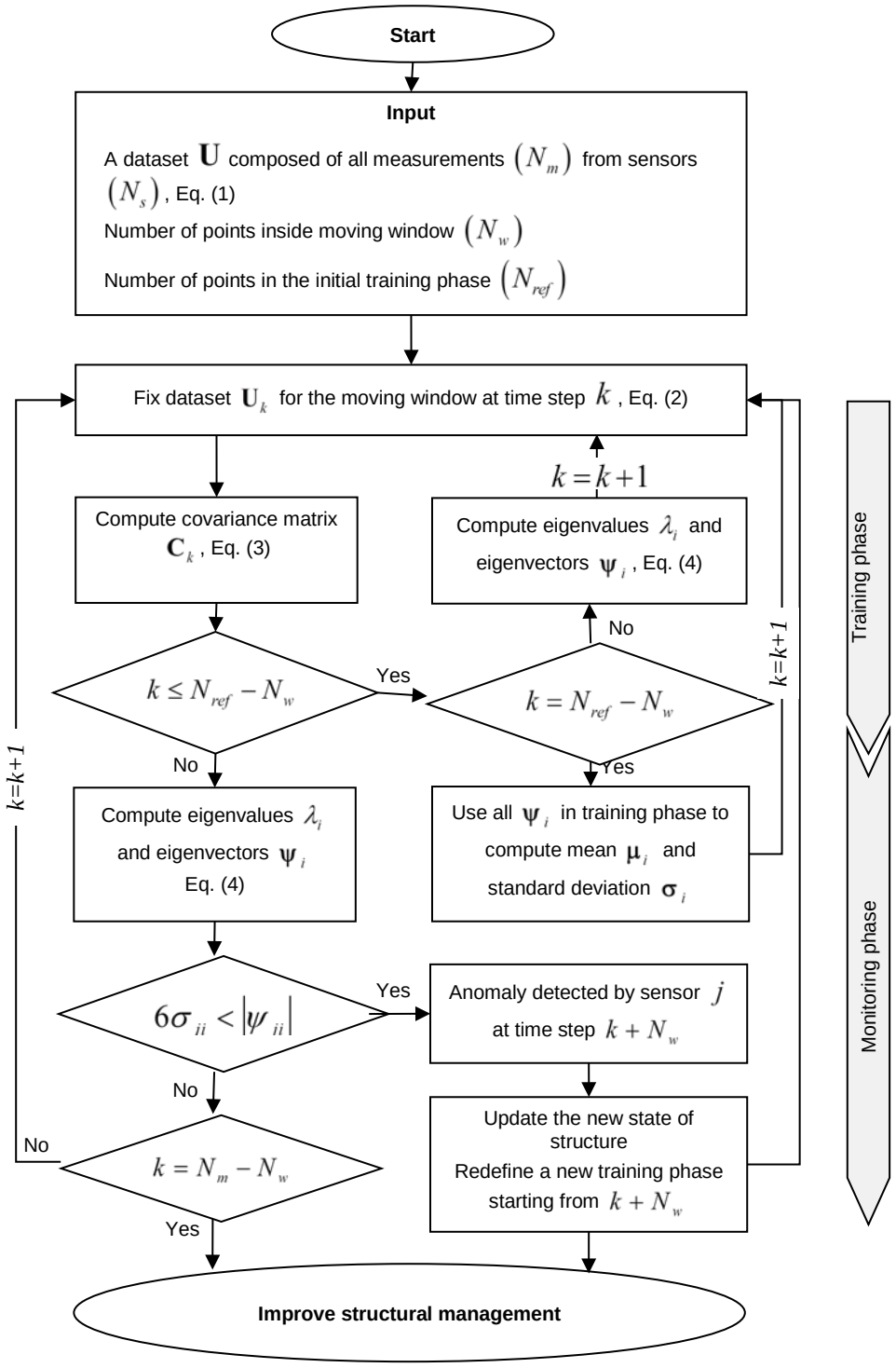

Figure 1. MPCA is applied to anomaly detection for continuous monitoring of structures according to this flowchart.

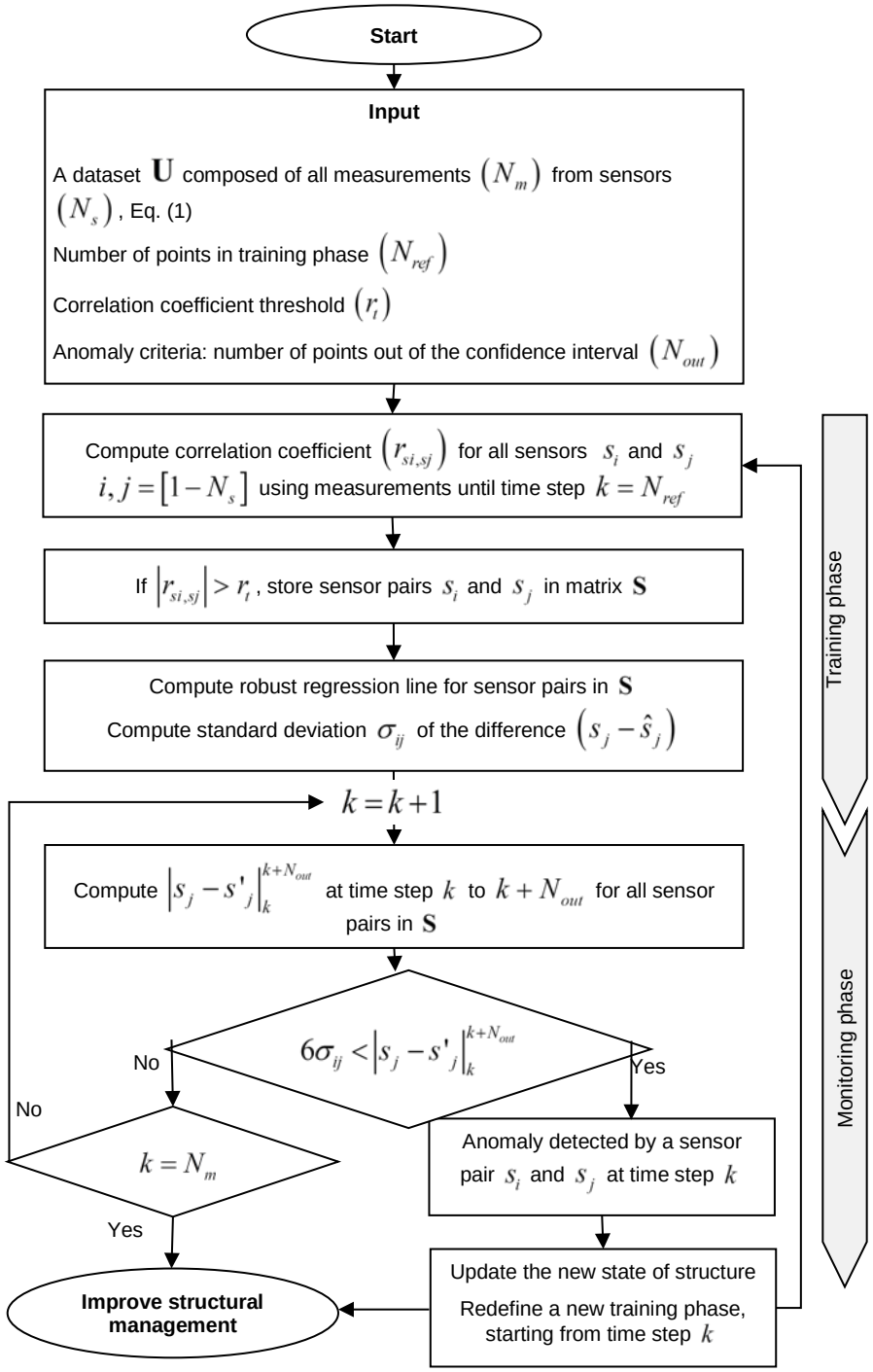

Figure 2. RRA is applied to anomaly detection for continuous monitoring of structures

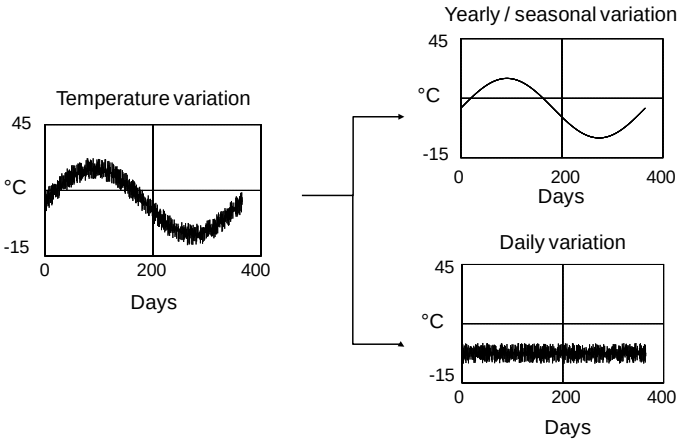

Figure 3. Temperature variation (seasonal and daily variation)

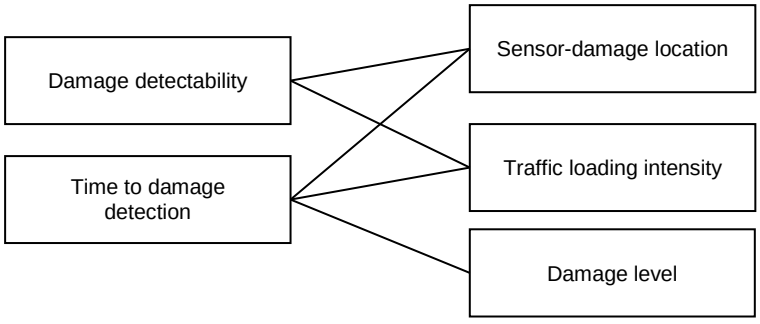

Figure 4. The framework of the parametric study for MPCA and RRA methods. The connection lines present the relations which are studied

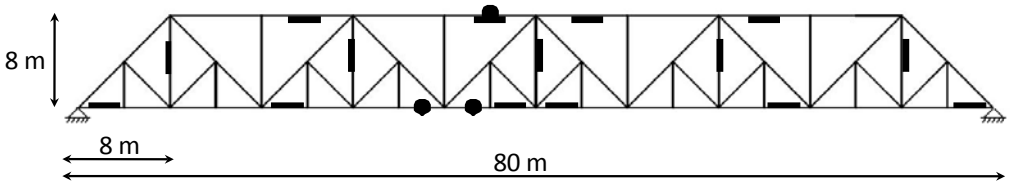

Figure 5. A truss structure of a 80-m railway bridge with sensor locations marked as black bars and damage locations marked as black dots.

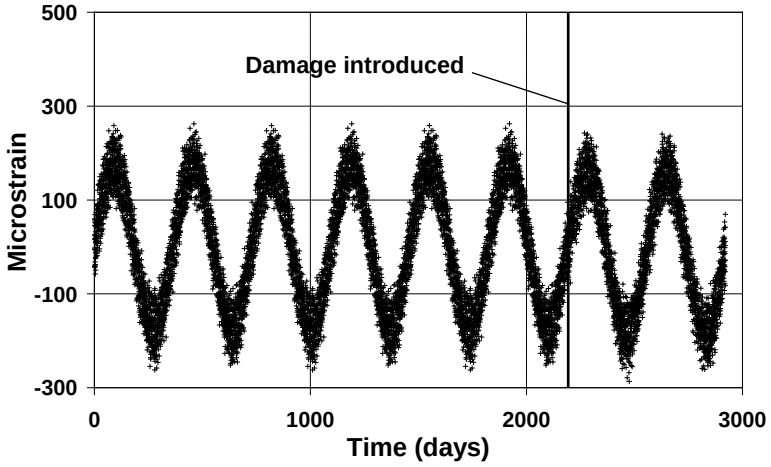

Figure 6. Strain time history measured by a sensor at damage location 1.

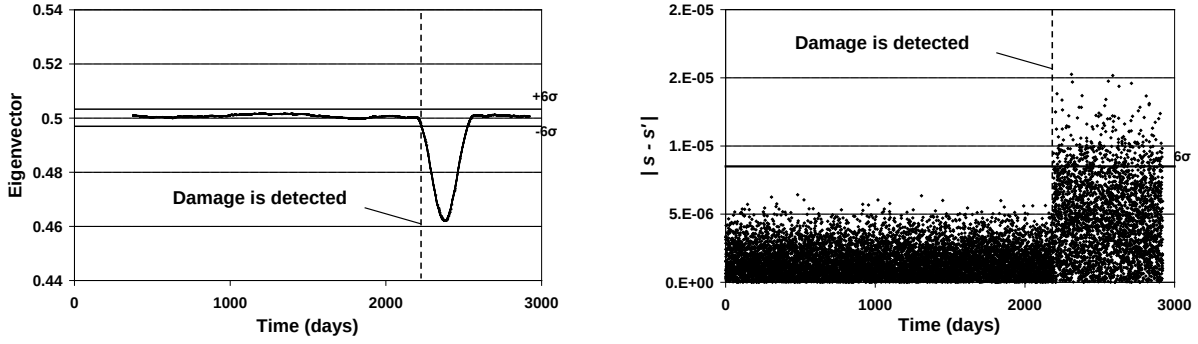

Figure 7. Plots of eigenvectors related to the main eigenvalue computed using MPCA (left) and the absolute value of the difference between strain measurement and regression line computed using RRA (right) for the corresponding damage scenarios in Figure 6.

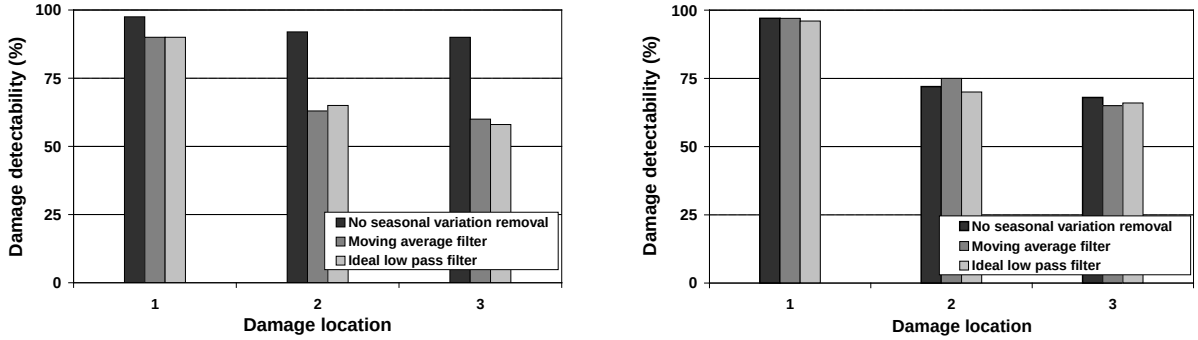

Figure 8. Damage detectability at three locations using MPCA (left) and RRA (right) with and without seasonal variation removal.

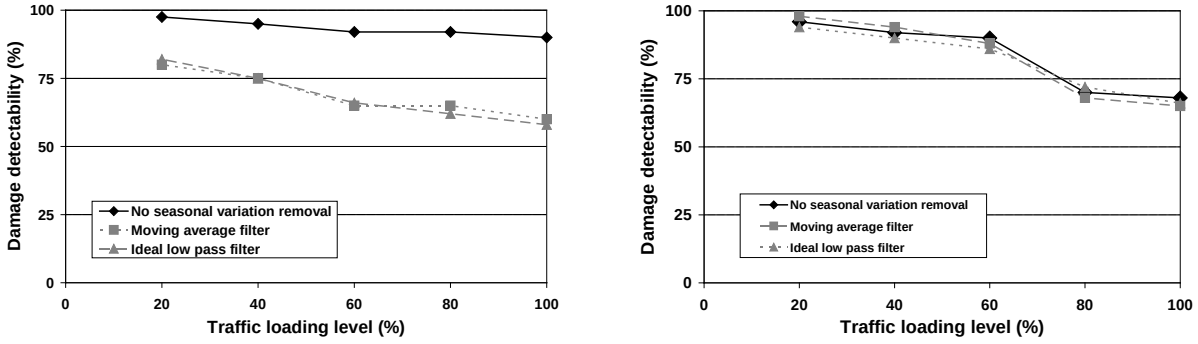

Figure 9. Damage detectability at location 3 using MPCA (left) and RRA (right) for traffic loading levels.

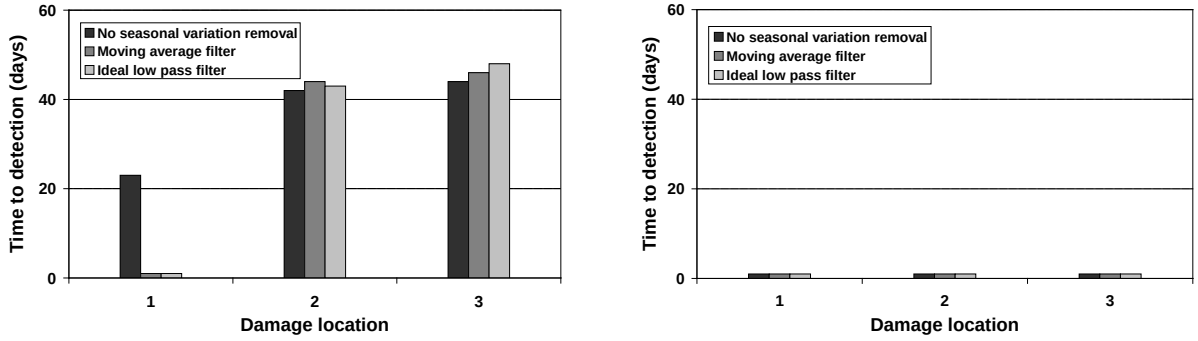

Figure 10. Time to detection for the three damage locations using MPCA (left) and RRA (right). This study is performed at a damage level of 60% and a traffic loading of 50%.

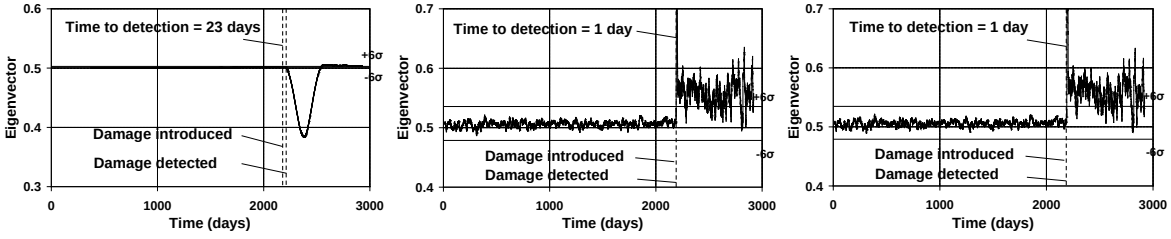

Figure 11. Eigenvector time histories related to the main eigenvalues computed using MPCA for a damage level of 60% at location 1 and a traffic loading of 50% without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low-pass filter (right).

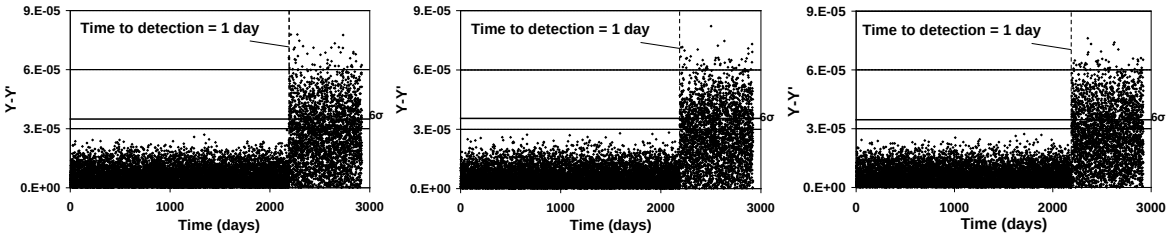

Figure 12. The absolute value of the difference between strain measurements to robust regression line for a damage level of 60% at location 1 and a traffic loading of 50% without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low-pass filter (right).

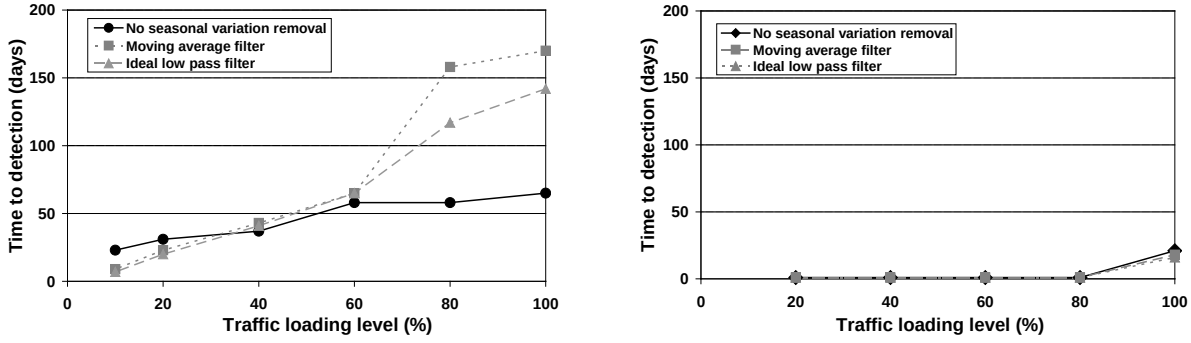

Figure 13. Time to damage detection at location 2 using MPCA (left) and RRA (right) for traffic loading levels.

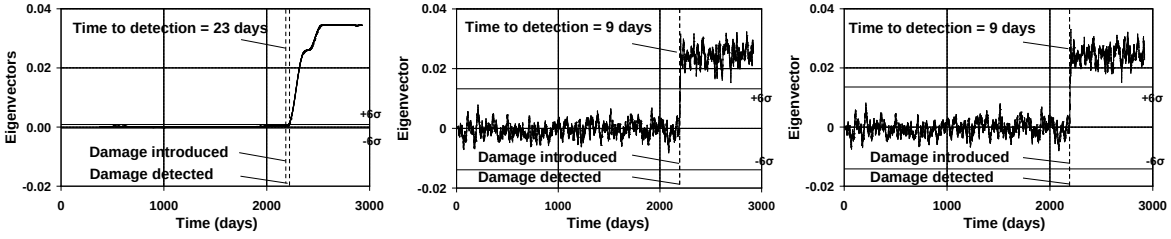

Figure 14. Eigenvector time histories related to the main eigenvalues computed using MPCA for damage level of 60% at location 2 and traffic loading 20% % without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low-pass filter (right).

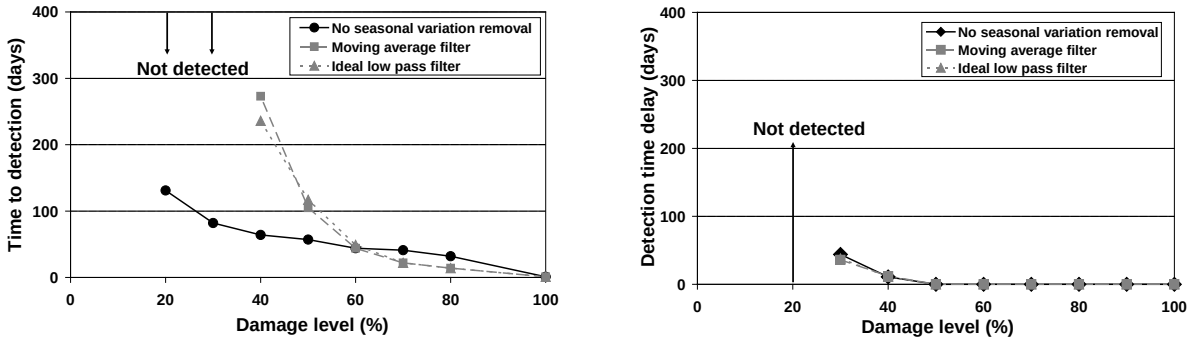

Figure 15. Time to damage detection using MPCA (left) and RRA (right) for damage scenarios with the range of damage level from 20% to 80% at location 2.

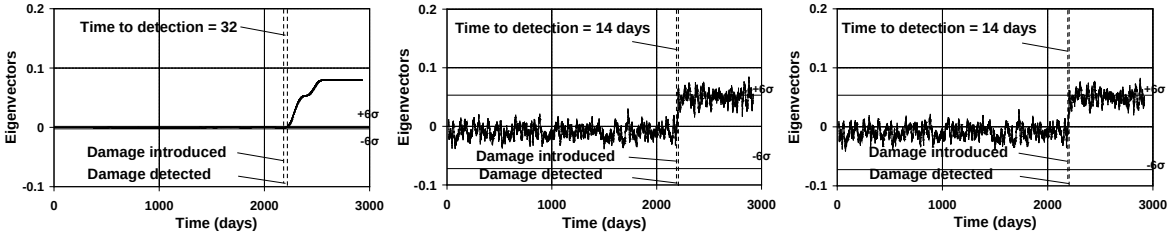

Figure 16. Eigenvector time histories related to the main eigenvalues computed using MPCA for damage level of 80% at location 2 and traffic loading 50% without removing seasonal variation (left), and with removing seasonal variation using moving average filter (middle) and ideal low pass filter (right).

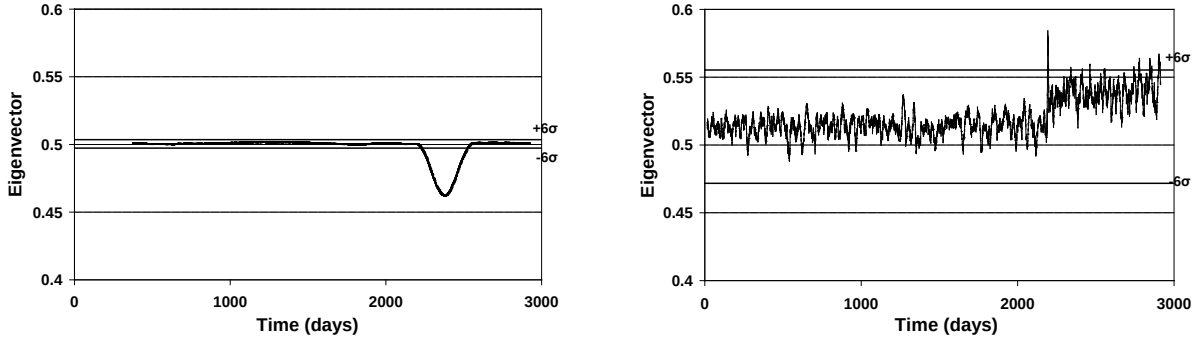

Figure 17. Plots of eigenvector time histories related to the main eigenvalues computed using MPCA without (left) and with (right) data processing for removing seasonal temperature effects.

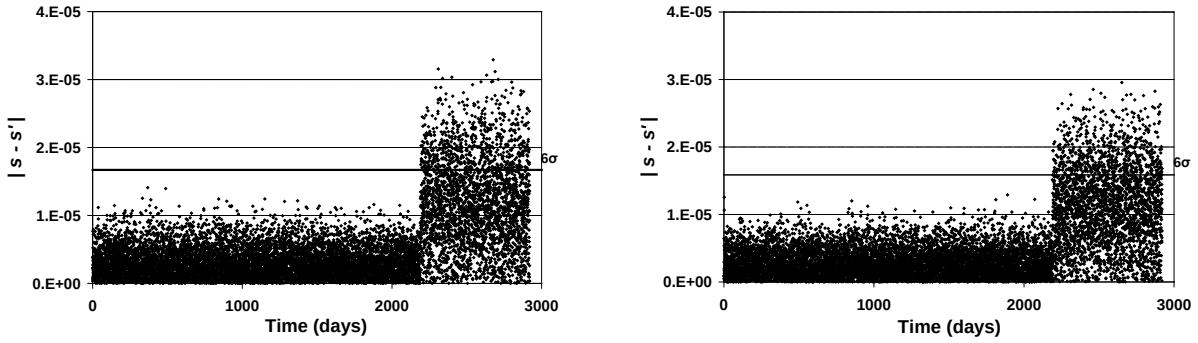

Figure 18. Plots of absolute value of the difference between strain measurement and regression line computed using RRA without (left) and with (right) data processing for removing seasonal temperature effects.

List of tables

Table 1 Differences between MPCA and RRA in the context of continuous monitoring of structures

MPCA	RRA
• Analyze all measurement data.	• Analyze only measurement pairs which are strongly correlated in the training phase.
• Observe changes in eigenvectors calculated from a covariance matrix.	• Observe changes in distance between measurement points and regression line.
• Performing PCA within a moving window.	• No window is used.

Table 2. Properties of truss members of a railway bridge in Zangenberg, Germany

Member type	Area (m^2)	I_x (m^4)	I_y (m^4)	Length (m)
Top chord	5.15×10^{-2}	2.267×10^{-3}	2.586×10^{-3}	4.00
Bottom chord	3.03×10^{-2}	1.467×10^{-3}	1.458×10^{-3}	2.00
Vertical	2.19×10^{-2}	1.215×10^{-3}	4.245×10^{-5}	4.00
Diagonal	3.69×10^{-2}	9.704×10^{-4}	4.164×10^{-3}	5.66
Small diagonal	2.19×10^{-2}	1.215×10^{-3}	4.245×10^{-5}	5.66

Table 3. Summary of the parametric study

Parameter	Damage detectability		Time to damage detection	
	MPCA	RRA	MPCA	RRA
Removal of seasonal variations	↓	*	↕	*
Proximity of the sensor location to the damage location	↑	↑	↓	*
Decreases in traffic loading	↑	↑	↓	*
Decreases in damage level	…	…	↓	*

↑ = positive influence
↓ = negative influence
↕ = influence may be positive but only for specific situation
* = influence is modest
… = influence is not studied (damage detectability is defined by the minimum detectable damage level)