Laory, I., Trinh, T.N., Posenato, D. and Smith, I.F.C. "Combined Model-Free Data-Interpretation Methodologies for Damage Detection during Continuous Monitoring of Structures" J of Computing in Civil Engineering, Vol. 27, No. 6, 2013, pp 657-666. http://cedb.asce.org

Combined model-free data-interpretation methodologies for damage detection during continuous monitoring of structures

Irwanda Laory[1], Thanh N. Trinh[2], Daniele Posenato[3], Ian F. C. Smith[4]

[1] Graduate Student, Applied Computing and Mechanics Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Station 18,CH-1015 Lausanne, Switzerland (corresponding author). E-mail: Irwanda.Laory@epfl.ch

[2] Postdoctoral Researcher, Applied Computing and Mechanics Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Station 18,CH-1015 Lausanne, Switzerland. E-mail: NgocThanh.Trinh@epfl.ch

[3] Project Engineer, SMARTEC SA, CH-6928 Manno, Switzerland. E-mail: Daniele.Posenato@smartec.ch

[4] Professor, Applied Computing and Mechanics Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Station 18,CH-1015 Lausanne, Switzerland. E-mail: Ian.Smith@epfl.ch

Abstract

Despite the recent advances in sensor technologies and data acquisition systems, interpreting measurement data for structural monitoring remains as challenge. Furthermore, due to the complexity of the structures, materials used and uncertain environments, behavioral models are difficult to build accurately. This paper presents novel model-free data-interpretation methodologies that combine MPCA with each of four regression-analysis methods – Robust Regression Analysis (RRA), Multiple Linear Analysis (MLR), Support Vector Regression (SVR) and Random Forest (RF) – for damage detection during continuous monitoring of structures. The principal goal is to exploit the advantages of both MPCA and regression-analysis methods. The applicability of these combined methods is evaluated and compared with individual applications of MPCA, RRA, MLR, SVR and RF through four case studies. Result showed that the combined methods outperformed non-combined methods in terms of damage detectability and time to detection.

Keywords: Model free, data interpretation, regression analysis, robust regression analysis, support vector regression, random forest, damage detectability, time to detection

1 Introduction

The performance of civil engineering structures under operational and environmental actions may decrease over time due to factors such as deterioration of structural materials, extreme and other actions that were not adequately taken into account during design. In the USA, it has been estimated that more than two trillion dollars are needed to bring America's infrastructure up to an acceptable performance level. Current infrastructure budgets are only a fraction of this amount and future deficit reduction plans will widen the gap (ASCE 2009). Structural Health Monitoring (SHM) has the potential to save money through early detection and this may lead to cheaper repairs and replacement avoidance. SHM is a process aimed at providing accurate and real-time information concerning structural condition and performance (Glisic and Inaudi 2008). It consists of periodic or continuous monitoring that measures quantities such as structural responses and environmental variations for the evaluation of structural performance.

Recent advances in sensor technologies and data acquisition systems allow complex structures to be equipped with hundreds of sensors that measure quantities such as structural responses (acceleration, deformation rotation etc) and environmental variations (temperature, humidity, wind, etc.). Despite the continuous evolution and development of measurement technologies, interpreting a large amount of measurement data to obtain useful information on structural conditions remains a challenge. This task falls into the field of Structural Identification (St-Id) which is an application of System Identification (Sys-Id) to civil structural systems. The Sys-Id concept (originated in electrical engineering) was first studied in engineering mechanics by Hart and Yao (1977) and in structural engineering by Liu and Yao (1978).

Classifying according to the presence or absence of physics-based behavioral models, there are generally two types of data interpretation methods: model-based and model-free methods. Strengths and weaknesses of both types have been summarized in the ASCE State-of-the-art Report on Structural Identification of Constructed Systems (ASCE 2011). Both types are complementary since they are appropriate in different contexts. Model-based data interpretation methods are typically performed through comparing predictions of behavior models with measured structural responses (Okasha et al. 2012; Koh and Thanh 2010; Ren and Chen 2010; Reynders et al. 2010; Koh and Thanh 2009; Strauss et al. 2008). Behavior models are used to support decisions related to long-term structural management such as estimation of reserve capacity and repair. However, behavior models are expensive to build and identifying a unique model is difficult due to the intrinsic ambiguity of inverse tasks as well as uncertainties. Furthermore many model predictions might approximately match observations and due to compensating and systematic errors, the best matching model may not be the correct model (Goulet et al. 2010; Robert-Nicoud et al. 2005; Saitta et al. 2005; Raphael and Smith 1998).

Model-free data-interpretation methods involve analyzing measurement time series only; they do not require geometrical and material information of a structure. These methods are well-suited for analyzing measurements during continuous monitoring of structures since they involve only tracking changes in time-series signals. Omenzetter et al. (2004), Hou et al (2000), Moyo and Brownjohn (2002) used wavelet-based methods for damage detection. Omenzetter and Brownjohn (2006) proposed an autoregressive integrated moving average method (ARIMA) to detect damage from measurements. Lanata and Grosso (2006) applied a proper orthogonal decomposition method for continuous static monitoring of structures. Yan et al. (2005a; 2005b) proposed a local PCA-based damage–detection method for vibration-based SHM. All these studies are limited to a single methodology without comparison to other methods. Gul and Catbas (2011) employed Auto-

Regressive models with eXogenous input (ARX) for different sensor clusters by using the free response of a structure to assess damage.

Posenato et al. (2010; 2008) proposed two methods, MPCA and RRA for damage detection during continuous structural monitoring and performed a comparative study of these methods with several other model-free data-interpretation methods (Wavelet packet transform, Discrete wavelet transform, ARIMA, Box-Jenkins, Instance based method, Short Term Fourier Transform and correlation anomaly scores analysis) . Results demonstrated that the performances of MPCA and RRA for damage detection were superior to other methods when dealing with civil-engineering challenges such as significant noise, missing data and outliers. Both methods were observed to require low computational resources to detect anomalies, even when there were large quantities of data.

Many studies have shown that structural responses due to temperature variation have a significant effect on the overall system reliability. For example, the magnitude of thermal stresses was found to be comparable to live and dead load stresses (Peng and Qiang 2007). Catbas and Aktan (2002) observed that the magnitude of strains due to daily temperature variations far exceed those due to traffic. Bell et al. (2008) found that temperature effects mask the load applied to Rollins Road Bridge over the duration of load test. Brownjohn et al (2009) studied the thermal effects on performance on the Tamar suspension bridge and showed that thermal effects dominate the bridge behavior. The task of data interpretation is even more difficult in such situations. Laory et al. (2011) evaluated the performance of MPCA and RRA under traffic and temperature variations. The study showed that although MPCA is better than RRA in terms of damage detectability, RRA is better than MPCA in terms of time to detection. Hence, both methods were considered to be complementary and it was noted that synergies between both methods may result in a better overall methodology for damage detection.

Building on these previous studies, this paper presents a new methodology that combines MPCA and regression-analysis methods - Robust Regression Analysis (RRA), Multiple Linear Analysis (MLR), Support Vector Regression (SVR) and Random Forest (RF) - for damage detection during continuous monitoring of structures. Applications of SVR in the field Structural Health Monitoring have provided good results (Zhang et al. 2012; Ni et al. 2005; Loutas et al.). In addition, RF has been successfully employed for classification, prediction, studying variable importance, variable selection, and outlier detection (Rodriguez-Galiano et al. 2012; Verikas et al. 2011; Breiman 2001).

In this paper, the objective of combining MPCA with such regression analyses is to exploit the advantages of individual methods through appropriate combinations. The performance of the combined methods are evaluated and compared with single applications of MPCA, RRA, MLR, SVR and RF through four case studies. Comparison criteria are damage detectability, time to detection and performance in the presence of non-linear behavior. The next section includes description of several methodologies for damage detection. Four combined methods are also presented. This is followed by a section that evaluates effectiveness on four case studies.

2 Model-free data-interpretation methodologies for damage detection

2.1 Moving Principal Component Analysis (MPCA)

Moving Principal Component Analysis (MPCA) was first proposed for interpreting measurements from continuous monitoring for damage detection by Posenato et al. (2008). MPCA essentially applies Principal Component Analysis (PCA) (Hubert et al. 2005) to enhance the discrimination between features of undamaged and damaged state. In order to reduce computational time, PCA is applied to a sliding fixed-sized window of measurements instead of the whole dataset.

MPCA is carried out by first constructing a matrix that contains the history of all measured parameters. The second step is to iteratively extract datasets corresponding to a sliding window. The principal components are then computed by solving the eigenvalue problem of the covariance matrix of the extracted datasets. The components are arranged in order of significance by sorting the eigenvectors by eigenvalues in decreasing order. MPCA is conducted by analyzing only the eigenvectors that are related to the first few eigenvalues. When damage occurs, mean values and components of the covariance matrix change and as consequence, so do values of eigenvalues and eigenvectors.

A key issue is selecting the dimension of the moving window. It is necessary to select a value that is sufficiently large to minimize the influence of variations in measurements due to changes that are not related to damage (environmental effects, noise, etc.). If the time series has a periodic behavior, the choice of the window size should be at least as long as the longest period. This ensures the stationary behavior of the mean values over time and that eigenvalues of the covariance matrix do not have periodic behavior.

2.2 Robust Regression Analysis (RRA)

The application of RRA for damage detection in continuous monitoring is based on the distance of measurement points to computed regressions lines estimated during the undamaged state. The analysis is carried out by pairing sensors that are highly correlated and then focusing on these couples to detect anomalies. These sensor pairs are identified by computing correlation coefficients between measurement data and comparing them with a pre-defined correlation coefficient threshold. All sensor pairs having a correlation coefficient greater than the threshold are selected in order to formulate the robust regression model. The linear relation between y_i and y_j is written as

$$y'_j = \beta_0 + \beta_1 y_i \tag{1}$$

where β_0 and β_1 are the coefficients of the robust regression line estimated from measurements using iteratively reweighted least squares. y'_j represents the value of y_j computed according to the linear relation. The robust regression analysis is carried out by observing the regression residuals (discrepancies between the measurements y_j and the prediction by linear regression-line y'_j). Damage is identified when the value exceeds a confidence interval that is defined using standard deviation of the difference in the undamaged state. The advantage of RRA is that it is insensitive to outliers and missing data. It is thus suitable for civil engineering applications since all measurements of civil-engineering structures contain outliers and most have missing data (Posenato et al. 2010).

2.3 Multiple Linear Regression (MLR)

The aim of multiple linear regression is to evaluate the relationship between several independent (predictor) variables and a dependent (criterion) variable by fitting a linear equation to observed data. Given n observations, the multiple linear regression is formulated as

$$y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots \beta_p x_p \qquad for\ i = \ 1, 2, \ \ldots n \tag{2}$$

where β is a regression coefficient associated with the i^{th} input variable $(x_i,\ \ i=1,\ldots,n)$. Using the dataset of observations in measurement time series, the unknown coefficients are determined using the least squares method. In the application for damage detection, similar to robust regression, detection is based on the regression residual.

2.4 Support Vector Regression (SVR)

Support vector machines is a new class of learning algorithms that are derived from statistical learning theory (Vapnik and Lerner 1963). These algorithms can be used for regression analysis and thus known as Support Vector Regression (SVR). SVR builds a linear regression function in a high dimensional new space (i.e. feature space in machine learning) where the input data in the original space is mapped using a transformation function. A distinctive characteristic of SVR is that instead of minimizing the observed training error such as MLR, SVR conducts the minimization of the generalization error bound in order to obtain generalized performance. The generalization error bound is the combination of the training error and a regularization term that controls the complexity of the prediction functions. The linear regression function in the new space is given by

$$y(x) = w^T \varphi(x) + b \tag{3}$$

where input values $x \in R^n$ and output (or response) values $y \in R$; w is a weight vector; b is a constant and $\varphi(x)$ is a transformation function that maps the input vector x into the high dimensional space. Given a training set $\{x_i, y_i\}$ $(i = 1,...,N)$, a regression function is formulated by minimizing the following objective function (Suykens et al. 2002)

$$\begin{aligned} &\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 \\ &\text{subject to } y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1,...,N. \end{aligned} \tag{4}$$

where e_i is the error and γ is the regularization parameter that determine the trade-off between the training error minimization and the complexity of the function. The optimization task is solved by constructing the Lagrangian function

$$L(w,b,e,\alpha) = J(w,e) - \sum_{i=1}^{N} \alpha_i \left\{ w^T \varphi(x_i) + b + e_i - y_i \right\} \tag{5}$$

where α_i are Lagrange multipliers.

A kernel function is employed to compute inner-products in the new space using only the original input data. The advantage of using kernels for inner products is that if a kernel function is known, it is not necessary to define the explicit form of the transformation function $\varphi(x)$ as well as the new space. The selection of the kernel function generally depends on the application domain. It has been shown that Gaussian radial-basis function (RBF) is a reasonable first choice of kernel functions since it has only a single parameter (standard deviation, σ) to be determined (Saitta et al. 2010). The Gaussian RBF is expressed as

$$K(x_i, x_j) = e^{-\left\|x_i - x_j\right\|^2 / 2\sigma^2} \tag{6}$$

When using the RBF kernel function, only two tuning parameters, γ and σ, need to be determined to formulate a prediction function and their optimal values could be determined using grid search method.

2.5 Random Forest (RF)

Random forest is a nonparametric statistical regression method that offers an alternative to parametric regression methods (Breiman 2001). The prediction is achieved by constructing an ensemble of regression trees. Given a training dataset $L = (x_{ij}, y_i)$ where $i = 1, ..., N$ is the number of observations and $j = 1, ..., p$ is the number of input variables.

The first step is to generate B training sub-datasets by continuously copying observations randomly from the original training dataset L until each sub-dataset L_b $(b = 1, ..., B)$ has the same number of observations as the original training dataset. Thus, some of the observations from the original dataset can be repeatedly copied into each sub-dataset, while others are not copied at all. The set of non-copied observations corresponding to each sub-dataset functions as a validation dataset.

The second step involves building B regression trees using the generated B training sub-datasets. A regression tree T_b is built by recursively splitting each sub-dataset into more and more homogeneous groups. From the decision-tree point of view, the entire training sub-dataset is represented by a *root* node and the splitting groups are represented by nodes, as shown in Figure 1.

When B regression trees are built from B sub-datasets, an ensemble of these trees is called a random forest. For each individual tree, the prediction of the response for a new observation x is determined by following the path from the root node down the appropriate terminal node and the prediction value is the average response in that terminal node. Finally, the overall prediction of the forest for a new observation is the average of prediction values from individual trees.

3 Combined methodologies

The methodologies presented in the previous section have advantages and limitations for data interpretation in the field of Structural Health Monitoring (SHM). The objectives of combining methodologies are to exploit the advantages of such methodologies and overcome limitations associated to each of them through an appropriate combination. For example, MPCA and RRA are found to be complementary methods for damage detection of continuously monitored structures under environmental variations since they are most appropriate in different situations. While MPCA is better than RRA in terms of damage detectability, RRA detects damage faster than MPCA. Hence, the combination of these methods is able to improve the detection performance in terms of damage detectability and time to detection.

This study proposes methodologies that combine MPCA with the previously presented four regression analysis methods: RRA, MLR, SVR and RF, for damage detection during continuous monitoring. Figure 2 shows the layout of the combined methodologies that is composed of two main steps. The first step is to transform measurement data into main eigenvector time series (main principal components) using MPCA. As mentioned in Section 2.1, MPCA is carried out by using a fixed-size window that moves along the measurement time series to extract specified datasets. The data within the window are used to compute a covariance matrix and then solve the eigenvalue problem of the covariance matrix to obtain the eigenvector time series.

The second step involves analyzing the correlations between eigenvector time series to detect damage in structures. The idea behind this step is built on an assumption that when damage occurs in structures, the correlations between the principal components will be changed. Thus, damage can be detected by tracking changes in these correlations over time. This step explores the correlations between PCs by developing regression functions using RRA, MLR, SVR and RF. For long-term monitoring of structures, the regression functions are then used to predict an eigenvector corresponding to a measurement location based on the known eigenvectors of other locations. If the difference between the predicted eigenvector obtained from regression functions and the known values (regression residual) exceeds a defined threshold bound, then damage is detected. The distinctive feature of these combined methods from the application of each individual regression analysis method for damage detection is that instead of tracking directly the correlations between measurement time series, the combined methods analyze the correlations between eigenvector time series. Therefore, it is capable of taking full advantages of high damage detectability from MPCA and small time to detection from regression analysis methods. The performance of such methods is demonstrated in the following case studies.

4 Case studies

4.1 Numerical studies

4.1.1 A railway truss bridge

A railway truss bridge in Zangenberg, Germany has been selected for a case study. The bridge is composed of two parallel trusses each having 77 members. A numerical model inspired by this bridge is used to provide responses (strain) under traffic loading and temperature variations. These responses are taken as measurement data from continuous monitoring. Only one truss of the bridge is modeled (Figure 3). Truss members are made of steel having an elastic modulus of 200 GPa and a density of 7870 kg/m^3. Their properties are summarized in Table 1. Supports of the truss are restrained in vertical and horizontal directions. Although this is not the boundary conditions that were designed for this bridge, these supports represent an upper-bound worst case when the supports have deteriorated with age. 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. Daily and seasonal temperature variations are simulated as thermal loads. Temperature differences between top and bottom chords due to solar radiation are also taken into account in the simulations. In this example, damage is represented by a loss of member axial stiffness. Damage scenarios are used to evaluate the damage detectability and time to detection for all methods. In this study, damage detectability is represented by the minimum detectable damage level that is the smallest percentage loss of axial stiffness in a member that can be detected.

Data-interpretation methods ,includes single applications and combined method are employed for damage detection. Four years of undamaged data is simulated and treated as training period. A window size of a year is chosen for MPCA. Figure 4 shows the minimum detectable damage-level using 9 model-free data-interpretation methods, including the applications of single individual method as well as combined methods. The figure demonstrates that combined methods are better in terms of damage detectability than individual methods. As expected, MPCA is better than RRA in terms of damage detectability. MPCA also shows a superior performance in comparison to other individual methods. Generally, while RRA performs the least, combined MPCA-RRA shows the best performance and it is able to detect a damage of 3% stiffness loss. Such small damage can be caused by many sources such as cracks and localized corrosion.

To evaluate performance in terms of time to detection, a damage scenario of 50% loss of axial stiffness in a member is chosen. Figure 5 shows the time to detection for all methods. In comparison to other methods, MPCA requires the longest time to detect damage. Another expected observation, RRA shows a better performance than MPCA, that is RRA can detect damage earlier than MPCA. However, it is seen that other regression analysis methods such as MLR, SVR and RF, are able to detect damage instantly. For the combined methods, the combination of MPCA with regression analysis methods performs better than the use of MPCA alone. Figure 5 also shows that although both RRA and MLR are based on linear regression analysis, MPCA-RRA can detect damage earlier than MPCA-MLR. Indeed, MPCA-RRA detects damage instantly while MPCA-MLR takes about 20 days. A plausible reason is that while RRA only performs analysis on high correlated measurement pairs, MLR analyzes all measurements regardless of the correlations within measurement data.

These results show that not all combinations lead to better performance in damage detection. For this case study, the combination of MPCA and RRA outperforms other methods in terms of damage detectability and it is as good as the individual application of MLR, SVR and RF in terms of time to detection.

4.1.2 A concrete frame

This case study takes structural responses from a numerical model of a concrete frame (Figure 6) as measurement data. This case study revisits numerical simulation data that was performed by Cavadas (2011). The model was used for evaluation of damage detection approach using influence lines of moving loads. It is a simply supported concrete frame with Young's modulus of 15 GPa. Four responses – vertical displacement at mid-span, horizontal displacement at roller support, rotation over the left support bearing and rotation over the right support bearing – are measured for damage detection. Taking into account sensor accuracy of available sensors with $\pm$ 0.01 mm for displacements and $\pm$ 1°x10^{-3} for rotations, a uniform distributed noise is added to measurement data. Damage is introduced as stiffness reduction along 30 cm of the beam element (Figure 6).

For this case study, 500 influence lines are used as a training period and the window size for MPCA is defined as 200 influence lines. Figure 7 shows the minimum detectable damage-level of individual and combined methods. Indeed, the combined methods are able to detect lower damage level than the minimum detectable damage-level when using single methods. Thus, it is concluded that the combination of MPCA with regression analysis methods results in a better methodology in terms of damage detectability. As shown in Figure 8, the best performance is achieved when MPCA is combined with SVR.

For time to detection, a damage scenario of 35% stiffness reduction is used. Times to detections of all methods are shown in Figure 8. For this scenario, all combined methods are able to detect damage instantly. Therefore, it is concluded that for this case study, combining MPCA with regression analysis improves the performance in terms of time to detection.

While the previous case study shows that combination of MPCA and RRA (MPCA-RRA) performs best in terms of damage detectability, this case study demonstrates that MPCA-SVR outperforms other methods. These results indicate that the selection of regression analysis to be combined with MPCA is case-dependent. A reason for this is that since combined methods conduct damage detection based on the correlation of eigenvector time histories, detection is dependent on the characteristics of these time series.

Figure 9 shows the plots of the relationship of two eigenvector time series for both case studies above. For the first case study (Figure 9 left), the relationship between eigenvector components of sensor 13 and 14 is shown to be linear and thus linear regression is well-suited to analyze such relationship. On the other hand, for the second case study (Figure 9 right), the relationship between eigenvector components of sensors 1 and 3 is shown to be non-linear. Therefore the combination of MPCA with non-linear regression is more appropriate for the second case study. Results from both case studies of damage detectability demonstrate that the most appropriate regression-analysis methods to be combined with MPCA are those that are compatible with eigenvector-correlation characteristics.

4.2 A full-scale test on the Ricciolo viaduct

The applicability of the combined methods for damage detection under environmental variations is also assessed in this paper using measurements from a full-scale test on the Ricciolo viaduct that was conducted by Posenato et al. (2010). The Ricciolo viaduct was built in 2004 - 2005 at the Lugano North exit of Swiss motorway A2. This bridge was continuously monitored at a rate of one measurement session per hour. The monitoring system includes parallel and crossed sensor topologies and inclinometers in order to monitor axial strain, horizontal and vertical curvature changes, torsion, average shear strain and rotations in both vertical plans. The configuration of the measurement system is given in Figure 10 and Figure 11.

During the first four and a half months of the monitoring period, the bridge was under construction. Several important stages in the construction process are listed in Table 2.

During the monitoring period, the bridge is in a good condition and there are no damage events that could generate anomalous behavior. Therefore, the time scale of the monitoring data is inverted so that events during the construction period appear as anomalous events. Previously, Posenato et al. (2010) demonstrated the successful application of MPCA and RRA in detecting construction stages. Following this study, this section compares the performance of MPCA and RRA with the combined methods in terms of detectability and time to detection.

As mentioned in Section 2.1, MPCA is carried out by observing the eigenvector time histories. For this method, a window size of four months is used for data analysis. Figure 12 presents the resulting eigenvector time histories and shows that detection is visible. One notable observation from this figure is that the eigenvector time histories are highly correlated before event 6 occurs and this correlation is suddenly changed when the anomalous event occurs. This is a good example that verifies the data interpretation methods proposed in this study where detection is based on the correlation of the eigenvectors (main principal components).

Figure 13 presents an eigenvector time history through-out the monitoring period. Confidence interval $\left(\pm 6\sigma\right)$ is used as detection criteria and an anomalous event is detected when the eigenvector falls out of the confidence interval. It is seen that the event is successfully detected using MPCA. However, it requires a period of 11 days to detect this event.

Figure 14 presents the comparison of RRA and MPCA-RRA. Both methods detect damage based on the linear correlation. However, the difference is that RRA analyzes the correlations between measurement data while MPCA-RRA analyzes the correlations between the eigenvectors (main principal component) time histories. Figure 14 shows that these two methods are better than MPCA in terms of time to detection. They are able to detect anomalous event almost instantaneously.

In addition, there is a remarkable observation in Figure 14. Although the magnitude of the changes in the regression residuals due to Event 6 is almost the same, there is a significant difference in the amount of scatter. The threshold size within the reference period for MPCA-RRA is thus much smaller than that of RRA. Unlike abrupt structural changes due to construction stages, in most cases, structural degradation occurs gradually starting from small damage. Thus, the figure implies that it will be more difficult for RRA to detect a change that is smaller than the change of Event 6 due to the size of the threshold. Obviously, this is not the case for MPCA-RRA since the size of the threshold is so small that detection is possible for relatively small changes. This shows that MPCA-RRA has the potential for higher damage detectability than RRA alone.

Similar to MPCA-RRA, the combination of MPCA with other regression analysis are able to detect Event 6 instantaneously. Thus, it can be concluded that combined methodologies are better than each individual method in terms of damage detectability and time to detection.

5 Conclusions

Results of four case studies lead to the following conclusions:

- The combination of Moving Principal Component Analysis (MPCA) and regression-analysis methods, including Robust Regression Analysis (RRA), Multiple Linear Analysis (MLR), Support Vector Regression (SVR) and Random Forest (RF) performs better than each individual method in terms of damage detectability and time to detection.
- For the combined data-interpretation methods, the most appropriate regression analyses are those that are compatible with eigenvector-correlation characteristics. For example, RRA and MLR are appropriate when eigenvector correlations are linear while SVR and RF are appropriate when eigenvector correlations are non-linear.
- Correlation-based methods are useful tools for damage detection of civil engineering structures. These methods are notably suitable for continuous monitoring of structures where there are large quantities of measurement data that are influenced by traffic load and environmental parameters such as temperature.

Acknowledgements

This work was funded by Swiss National Science Foundation under contract no. 200020-126385. Authors would like to thank Mr. Filippe Cavadas, Prof. Alain Nussbaumer, Dr. Daniele Inaudi and SMARTEC SA for their contributions related to the case studies.

References

ASCE. (2009). "Report card for America's infrastructure." American Society of Civil Engineers.

ASCE. (2011). "Structural Identification (St-Id) of Constructed Facilities : Approaches, Methods and Technologies for Effective Practice of St-Id." A State-of-the-Art Report by ASCE SEI Committee on Structural Identification of Constructed Systems.

Bell, E. S., Sipple, J., and Yost, J. (2008). "Long-Term Thermal Performance of a CFRP-Reinforced Bridge Deck." Vancouver, BC, Canada, 242-242.

Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.

Brownjohn, J. M. (2009). "Thermal effects on performance on Tamar Bridge." *4th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-4)*, Zurich, Swirzerland.

Catbas, F. N., and Aktan, A. E. (2002). "Condition and Damage Assessment: Issues and Some Promising Indices." *Journal of Structural Engineering*, 128(8), 1026-1036.

Cavadas, F. (2011). "Structural Health Monitoring Applied to the Life-Cycle Management of Bridges - Approaches for Damage Identification," Research Thesis Project, Universidade do Porto.

Glisic, B., and Inaudi, D. (2008). *Fibre Optic Methods for Structural Health Monitoring*, John Wiles & Sons.

Goulet, J.-A., Kripakaran, P., and Smith, I. F. C. (2010). "Multimodel Structural Performance Monitoring." *Journal of Structural Engineering*, 136(10), 1309-1318.

Gul, M., and Catbas, F. N. (2011). "Structural health monitoring and damage assessment using a novel time series analysis methodology with sensor clustering." *Journal of Sound and Vibration*, 330(6), 1196-1210.

Hart, G. C., and Yao, J. T. P. (1977). "System Identification in structural dynamics." *Journal of Engineering Mechanics, Division ASCE*, 103(EM6), 1089-1104.

Hou, Z., Noori, M., and Amand, R. S. (2000). "Wavelet-based approach for structural damage detection." *Journal of Engineering Mechanics*, 126(7), 677-683.

Hubert, M., Rousseeuw, P. J., and Branden, K. V. (2005). "ROBPCA: A new approach to robust principal component analysis." *Technometrics*, 47, 64-79.

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

Koh, C. G., and Thanh, T. N. (2010). "Output-only Substructural Identification for Local Damage Detection." *The Fifth International Conference on Bridge Maintenance, Safety and Management*, Philadelphia, Pennsylvania, USA.

Lanata, F., and Grosso, A. D. (2006). "Damage detection and localization for continuous static monitoring of structures using a proper orthogonal decomposition of signals." *Smart Materials and Structures*, 15(6), 1811-1829.

Laory, I., Trinh, T. N., and Smith, I. F. C. (2011). "Evaluating two model-free data interpretation methods for measurements that are influenced by temperature." *Advanced Engineering Informatics*, 25(3), 495-506.

Liu, S. C., and Yao, J. T. P. (1978). "Structural identification concept." *Journal Structural Division,ASCE*, 104(12), 1845-1858.

Loutas, T. H., Panopoulou, A., Roulias, D., and Kostopoulos, V. "Intelligent health monitoring of aerospace composite structures based on dynamic strain measurements." *Expert Systems with Applications*(0).

Moyo, P., and Brownjohn, J. M. W. (2002). "Detection of Anomalous Structural Behaviour Using Wavelet Analysis." *Mechanical Systems and Signal Processing*, 16(2-3), 429-445.

Ni, Y. Q., Hua, X. G., Fan, K. Q., and Ko, J. M. (2005). "Correlating modal properties with temperature using long-term monitoring data and support vector machine technique." *Engineering Structures*, 27(12), 1762-1773.

Okasha, N. M., Frangopol, D. M., and Orcesi, A. D. (2012). "Automated finite element updating using strain data for the lifetime reliability assessment of bridges." *Reliability Engineering & System Safety*, 99(0), 139-150.

Omenzetter, P., Brownjohn, J. M. W., and Moyo, P. (2004). "Identification of unusual events in multi-channel bridge monitoring data." *Mechanical Systems and Signal Processing*, 18(2), 409-430.

Omenzetter, P., and Brownjohn, J. M. W. (2006). "Application of time series analysis for bridge monitoring." *Smart Materials and Structures*, 15(1), 129-138.

Peng, Y., and Qiang, S. (2007). "Analytical Solution to Temperature Variations in Highway Concrete Bridges Due to Solar Radiation." Chengdu, China, 252-252.

Posenato, D., Lanata, F., Inaudi, D., and Smith, I. F. C. (2008). "Model-free data interpretation for continuous monitoring of complex structures." *Advanced Engineering Informatics*, 22(1), 135-144.

Posenato, D., Kripakaran, P., Inaudi, D., and Smith, I. F. C. (2010). "Methodologies for model-free data interpretation of civil engineering structures." *Computers & Structures*, 88(7-8), 467-482.

Raphael, B., and Smith, I. (1998). "Finding the right model for bridge diagnosis." Artificial Intelligence in Structural Engineering in Computer Science, Springer, Heidelberg, 308-319.

Ren, W.-X., and Chen, H.-B. (2010). "Finite element model updating in structural dynamics by using the response surface method." *Engineering Structures*, 32(8), 2455-2465.

Reynders, E., Teughels, A., and De Roeck, G. (2010). "Finite element model updating and structural damage identification using OMAX data." *Mechanical Systems and Signal Processing*, 24(5), 1306-1323.

Robert-Nicoud, Y., Raphael, B., Burdet, O., and Smith, I. F. C. (2005). "Model Identification of Bridges Using Measurement Data." *Computer-Aided Civil and Infrastructure Engineering*, 20(2), 118-131.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). "An assessment of the effectiveness of a random forest classifier for land-cover classification." *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(0), 93-104.

Saitta, S., Raphael, B., and Smith, I. F. C. (2005). "Data mining techniques for improving the reliability of system identification." *Advanced Engineering Informatics*, 19(4), 289-298.

Saitta, S., Kripakaran, P., Raphael, B., and Smith, I. F. C. (2010). "Feature Selection using Stochastic Search: An Application to System Identification." *Journal of Computing in Civil Engineering*, 24(1), 3-10.

Strauss, A., Frangopol, D. M., and Kim, S. (2008). "Use of monitoring extreme data for the performance prediction of structures: Bayesian updating." *Engineering Structures*, 30(12), 3654-3666.

Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Moor, B. D., and Vandewalle, J. (2002). *Least Squares Support Vector Machines*, World Scientific, Singapore.

Vapnik, V., and Lerner, A. (1963). "Pattern Recognition using Generalized Portrait Method." *Automation and Remote Control*, 24, 774-780.

Verikas, A., Gelzinis, A., and Bacauskiene, M. (2011). "Mining data with random forests: A survey and results of new tests." *Pattern Recognition*, 44(2), 330-349.

Yan, A. M., Kerschen, G., De Boe, P., and Golinval, J. C. (2005a). "Structural damage diagnosis under varying environmental conditions--Part I: A linear analysis." *Mechanical Systems and Signal Processing*, 19(4), 847-864.

Yan, A. M., Kerschen, G., De Boe, P., and Golinval, J. C. (2005b). "Structural damage diagnosis under varying environmental conditions--part II: local PCA for non-linear cases." *Mechanical Systems and Signal Processing*, 19(4), 865-880.

Zhang, X., Liang, D., Zeng, J., and Asundi, A. (2012). "Genetic algorithm-support vector regression for high reliability SHM system based on FBG sensor network." *Optics and Lasers in Engineering*, 50(2), 148-153.

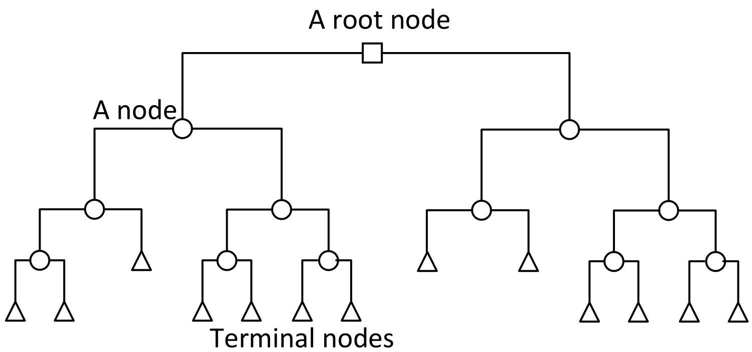

Figure 1 An example of a regression tree with 12 splits and 13 terminal nodes. A node represents a group of data.

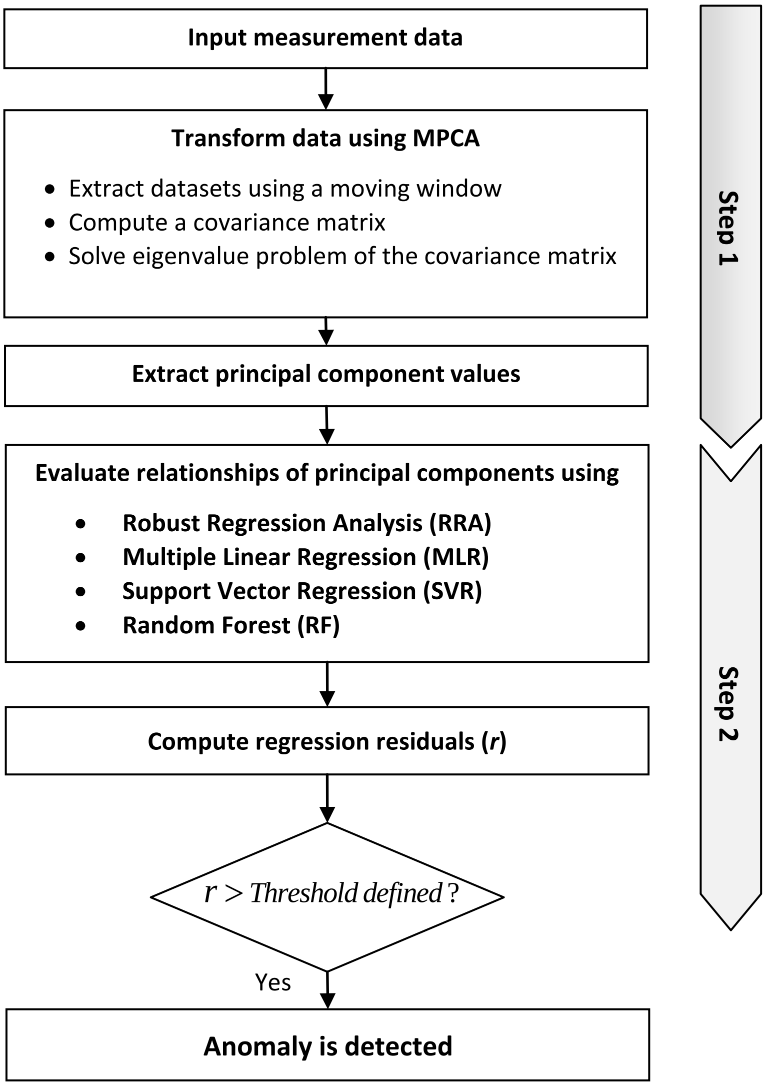

Figure 2 Flowchart of the combined model-free data-interpretation methodologies

476

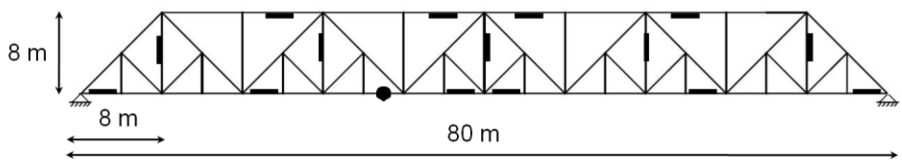

Figure 3 . An 80-m railway steel truss bridge with sensor locations marked as black bars and the damage location marked as a black dot.

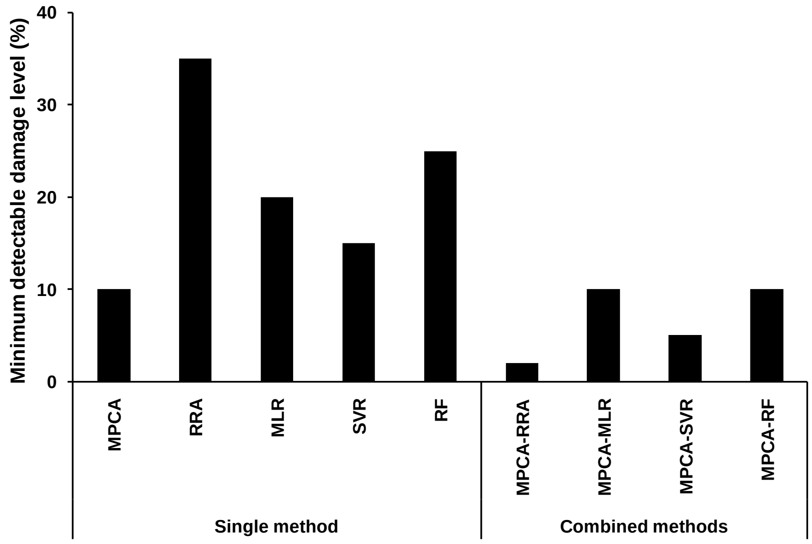

Figure 4 Minimum detectable damage-level for a truss bridge using 9 model-free data interpretation methods.

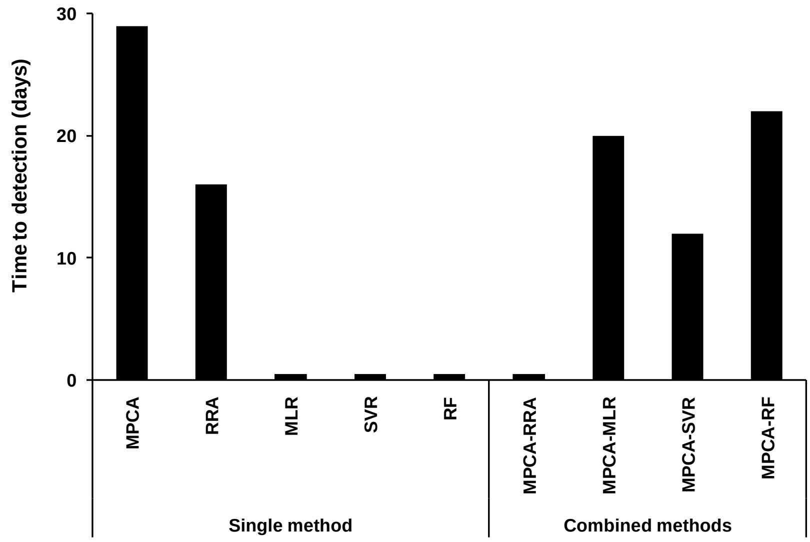

Figure 5 Time to damage detection for a truss bridge using 9 model-free data interpretation methods for 50% damage level

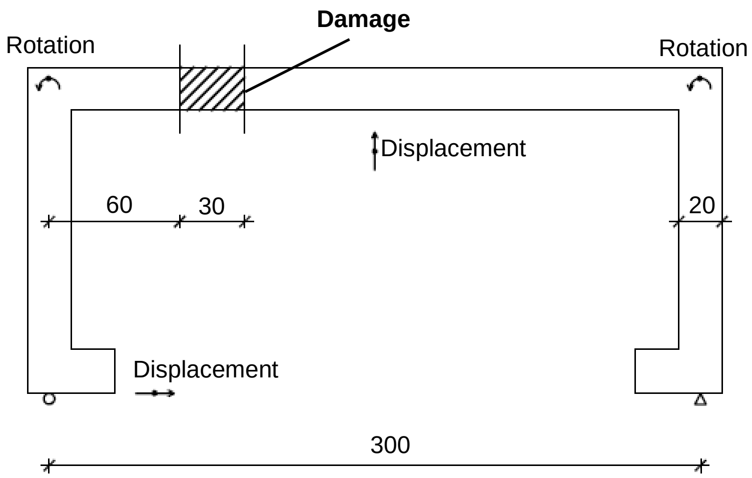

Figure 6 A concrete frame model with four measured responses including vertical displacement at midspan, horizontal displacement at roller support, rotation over the left support bearing and rotation over the right support bearing.

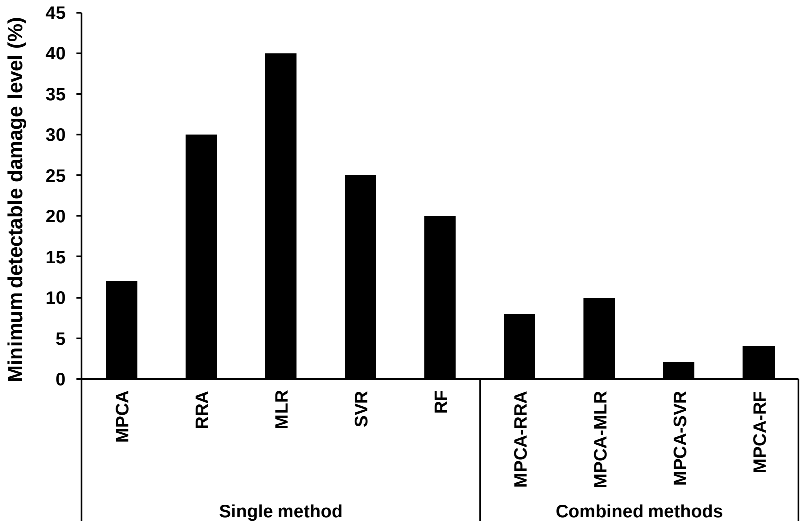

Figure 7 Minimum detectable damage level for a concrete frame using 9 model-free data interpretation methods

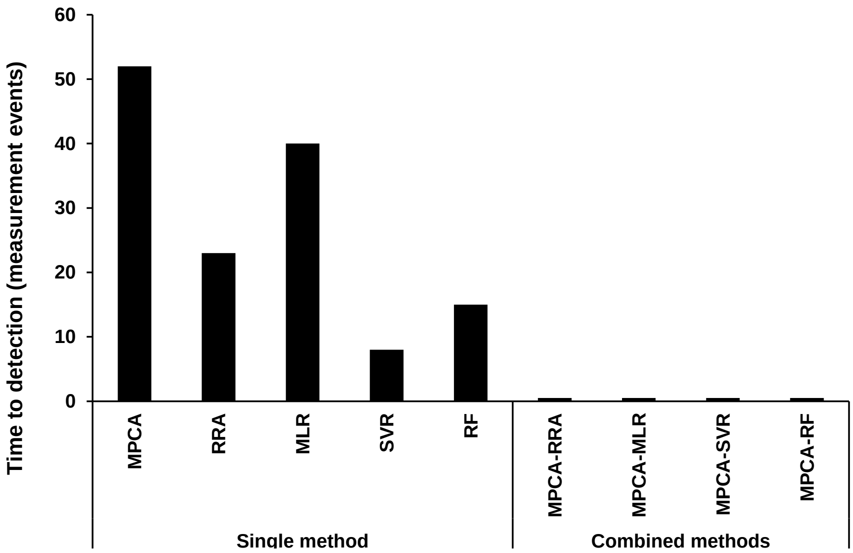

Figure 8 Time to detection for a concrete frame using 9 model-free data interpretation methods for 35% damage level

485

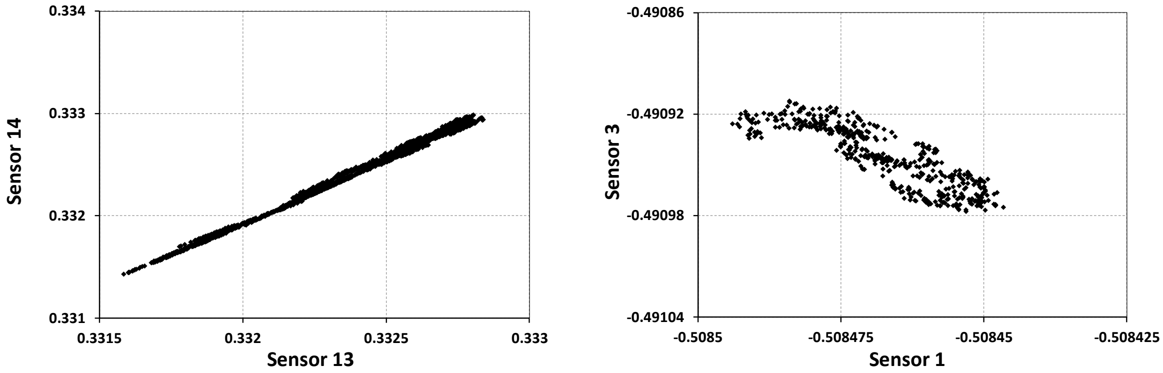

Figure 9 High correlated principal component pairs for the truss model (left) and the frame model (right).

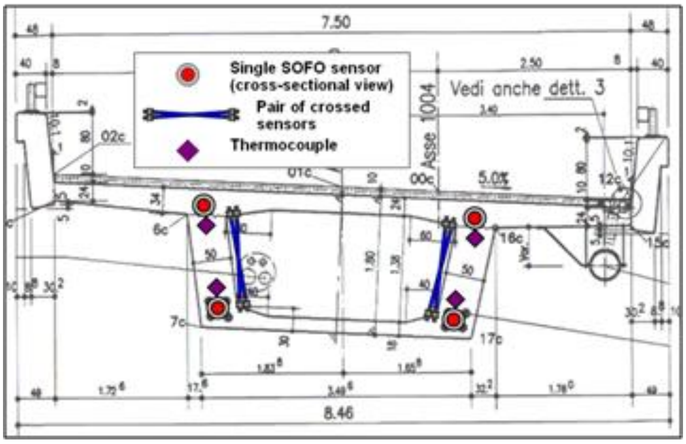

Figure 10 Measurement configuration for the Ricciolo viaduct: a cross-section view (Posenato et. al., 2010)

488

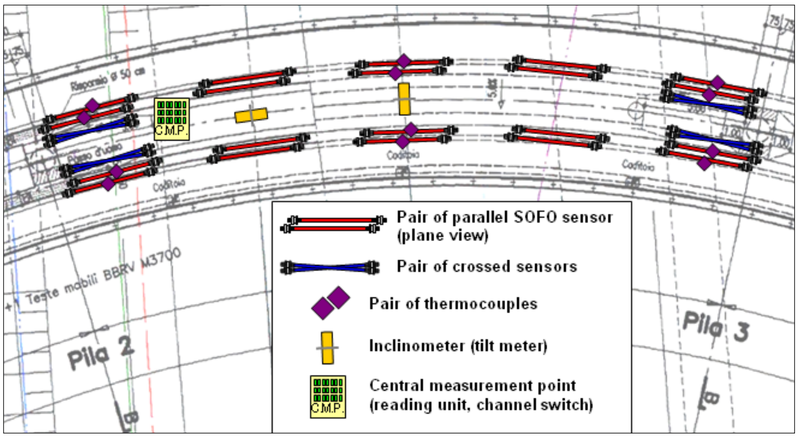

Figure 11 Measurement configuration for the Ricciolo viaduct: a plane-view (Posenato et. al., 2010)

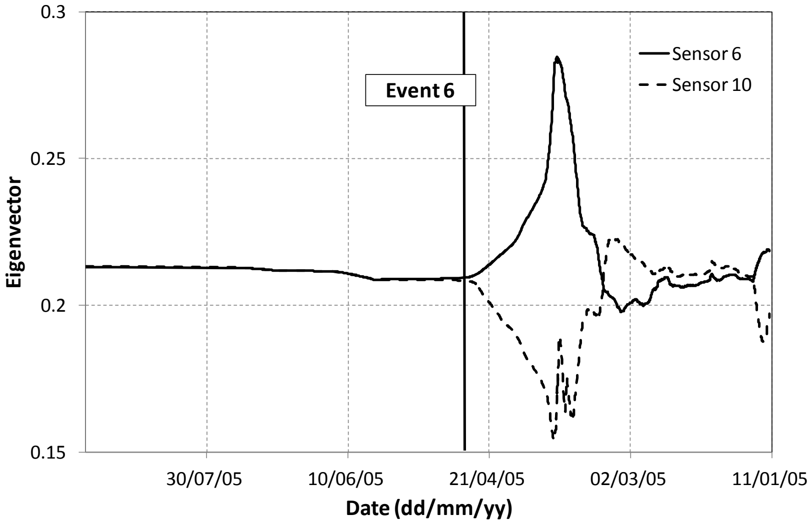

Figure 12 Plot of the first eigenvector during the monitoring period recalculated from Posenato et al. (2010). Detection of construction stages as events that simulate anomalous structural behavior (time scale is inverted)

490

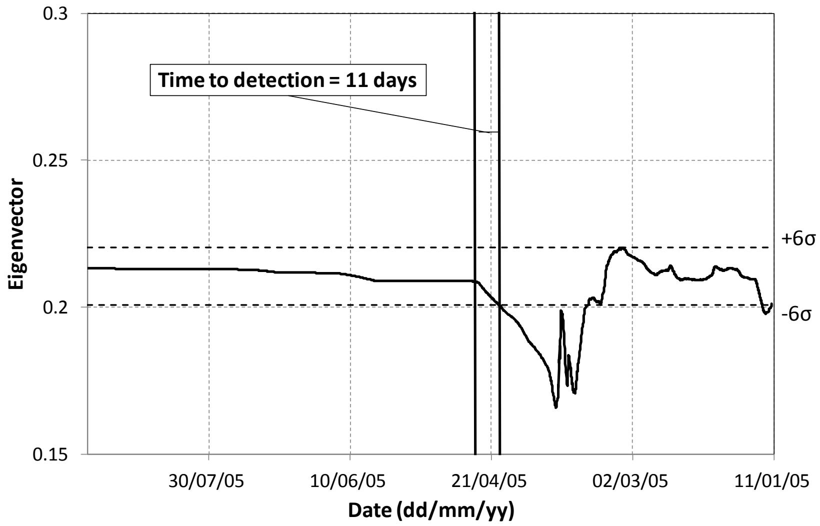

Figure 13 Time to detection for event number 6, recalculated from Posenato et al. (2010)

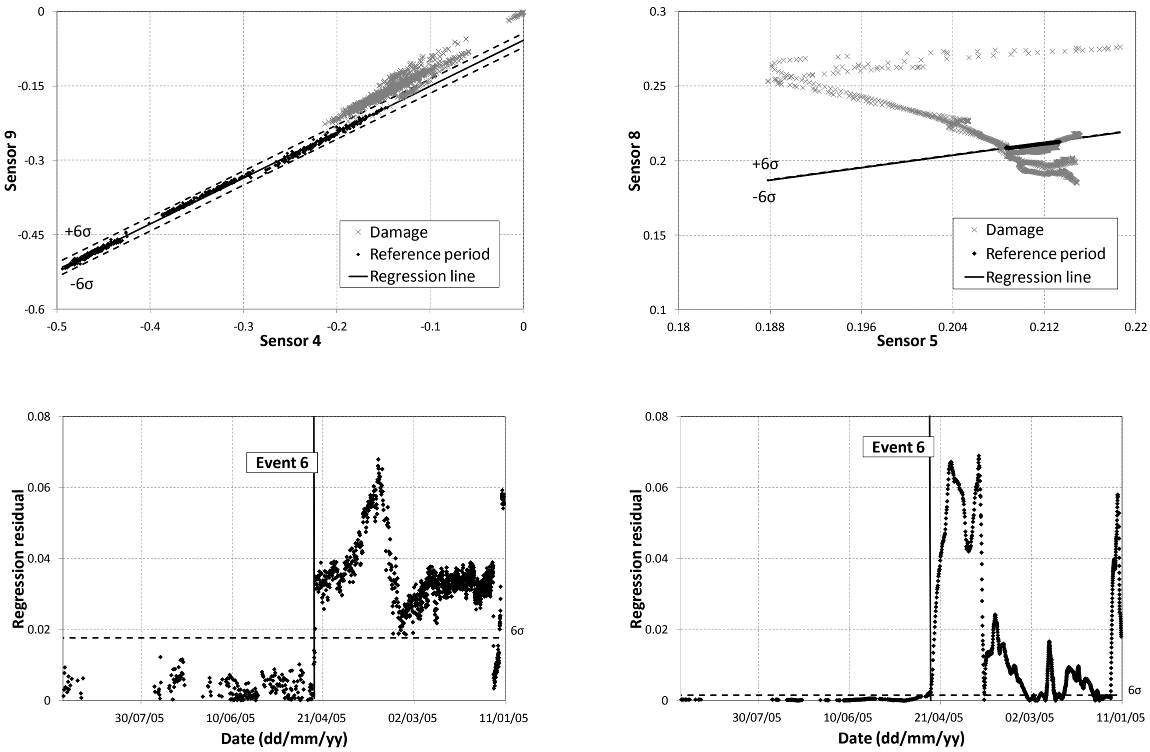

Figure 14 Results of correlation-based methods for damage detection using RRA (left) and MPCA-RRA (right)

Table 1. Properties of truss members of a railway bridge in Zangenberg, Germany

Member type	Area (m^2)	I_x (m^4)	I_y (m^4)
Top chord	5.15×10^{-2}	2.267×10^{-3}	2.586×10^{-3}
Bottom chord	3.03×10^{-1}	1.467×10^{-3}	1.458×10^{-3}
Vertical	2.19×10^{-2}	1.215×10^{-3}	4.245×10^{-5}
Diagonal	3.69×10^{-2}	9.704×10^{-4}	4.164×10^{-3}
Small diagonal	2.19×10^{-2}	1.215×10^{-3}	4.245×10^{-5}

Table 2 List of events

Event number	Period	Description
1	January 12-14	Post-tensioning from 30 to 70%
2	January 17	Partial lowering formworks
3	January 17 - April 22	Construction of lateral protection walls
4	April 25-26	Post-tensioning from 70 to 100%
5	April 25-27	Cast of left side wing
6	April 25-27	Removal of external formworks