

Detection during Continuous Monitoring of Structures" J of Computing in Civil Engineering, Vol. 27, No. 6, 2013, pp 657-666. <http://cedb.asce.org> Copyright ASCE

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

Irwanda Laory¹, Thanh N. Trinh², Daniele Posenato³, Ian F. C. Smith⁴

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: NqocThanh.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

1

2 **Abstract**

3

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

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

17 **1 Introduction**

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

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

38 Classifying according to the presence or absence of physics-based behavioral models, there are
39 generally two types of data interpretation methods: model-based and model-free methods.
40 Strengths and weaknesses of both types have been summarized in the ASCE State-of-the-art Report
41 on Structural Identification of Constructed Systems (ASCE 2011). Both types are complementary since
42 they are appropriate in different contexts. Model-based data interpretation methods are typically
43 performed through comparing predictions of behavior models with measured structural responses
44 (Okasha et al. 2012; Koh and Thanh 2010; Ren and Chen 2010; Reynders et al. 2010; Koh and Thanh
45 2009; Strauss et al. 2008). Behavior models are used to support decisions related to long-term
46 structural management such as estimation of reserve capacity and repair. However, behavior models
47 are expensive to build and identifying a unique model is difficult due to the intrinsic ambiguity of
48 inverse tasks as well as uncertainties. Furthermore many model predictions might approximately
49 match observations and due to compensating and systematic errors, the best matching model may
50 not be the correct model (Goulet et al. 2010; Robert-Nicoud et al. 2005; Saitta et al. 2005; Raphael
51 and Smith 1998).

52 Model-free data-interpretation methods involve analyzing measurement time series only; they do
53 not require geometrical and material information of a structure. These methods are well-suited for
54 analyzing measurements during continuous monitoring of structures since they involve only tracking
55 changes in time-series signals. Omenzetter et al. (2004), Hou et al (2000), Moyo and Brownjohn
56 (2002) used wavelet-based methods for damage detection. Omenzetter and Brownjohn (2006)
57 proposed an autoregressive integrated moving average method (ARIMA) to detect damage from
58 measurements. Lanata and Grosso (2006) applied a proper orthogonal decomposition method for
59 continuous static monitoring of structures. Yan et al. (2005a; 2005b) proposed a local PCA-based
60 damage-detection method for vibration-based SHM. All these studies are limited to a single
61 methodology without comparison to other methods. Gul and Catbas (2011) employed Auto-

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

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

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

85 Building on these previous studies, this paper presents a new methodology that combines MPCA and
86 regression-analysis methods - Robust Regression Analysis (RRA), Multiple Linear Analysis (MLR),

87 Support Vector Regression (SVR) and Random Forest (RF) - for damage detection during continuous
88 monitoring of structures. Applications of SVR in the field Structural Health Monitoring have provided
89 good results (Zhang et al. 2012; Ni et al. 2005; Loutas et al.). In addition, RF has been successfully
90 employed for classification, prediction, studying variable importance, variable selection, and outlier
91 detection (Rodriguez-Galiano et al. 2012; Verikas et al. 2011; Breiman 2001).

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

99 **2 Model-free data-interpretation methodologies for damage detection**

100 **2.1 Moving Principal Component Analysis (MPCA)**

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

106 MPCA is carried out by first constructing a matrix that contains the history of all measured
107 parameters. The second step is to iteratively extract datasets corresponding to a sliding window. The
108 principal components are then computed by solving the eigenvalue problem of the covariance matrix
109 of the extracted datasets. The components are arranged in order of significance by sorting the
110 eigenvectors by eigenvalues in decreasing order. MPCA is conducted by analyzing only the

111 eigenvectors that are related to the first few eigenvalues. When damage occurs, mean values and
112 components of the covariance matrix change and as consequence, so do values of eigenvalues and
113 eigenvectors.

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

120 **2.2 Robust Regression Analysis (RRA)**

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

$$128 \quad y'_j = \beta_0 + \beta_1 y_i \quad (1)$$

129 where β_0 and β_1 are the coefficients of the robust regression line estimated from measurements
130 using iteratively reweighted least squares. y'_j represents the value of y_j computed according to the
131 linear relation. The robust regression analysis is carried out by observing the regression residuals
132 (discrepancies between the measurements y_j and the prediction by linear regression-line y'_j).

133 Damage is identified when the value exceeds a confidence interval that is defined using standard

134 deviation of the difference in the undamaged state. The advantage of RRA is that it is insensitive to
135 outliers and missing data. It is thus suitable for civil engineering applications since all measurements
136 of civil-engineering structures contain outliers and most have missing data (Posenato et al. 2010).

137 **2.3 Multiple Linear Regression (MLR)**

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

$$141 \quad y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad \text{for } i = 1, 2, \dots, n \quad (2)$$

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

146 **2.4 Support Vector Regression (SVR)**

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

$$156 \quad y(x) = w^T \varphi(x) + b \quad (3)$$

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

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2$$

161 subject to $y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, \dots, N.$ (4)

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

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

165 (5)

166 where α_i are Lagrange multipliers.

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

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

174 (6)

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

178 **2.5 Random Forest (RF)**

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

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

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

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

197 **3 Combined methodologies**

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

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

215 The second step involves analyzing the correlations between eigenvector time series to detect
216 damage in structures. The idea behind this step is built on an assumption that when damage occurs
217 in structures, the correlations between the principal components will be changed. Thus, damage can
218 be detected by tracking changes in these correlations over time. This step explores the correlations
219 between PCs by developing regression functions using RRA, MLR, SVR and RF. For long-term
220 monitoring of structures, the regression functions are then used to predict an eigenvector
221 corresponding to a measurement location based on the known eigenvectors of other locations. If the

222 difference between the predicted eigenvector obtained from regression functions and the known
223 values (regression residual) exceeds a defined threshold bound, then damage is detected. The
224 distinctive feature of these combined methods from the application of each individual regression
225 analysis method for damage detection is that instead of tracking directly the correlations between
226 measurement time series, the combined methods analyze the correlations between eigenvector time
227 series. Therefore, it is capable of taking full advantages of high damage detectability from MPCA and
228 small time to detection from regression analysis methods. The performance of such methods is
229 demonstrated in the following case studies.

230 **4 Case studies**

231 **4.1 Numerical studies**

232 **4.1.1 A railway truss bridge**

233 A railway truss bridge in Zangenberg, Germany has been selected for a case study. The bridge is
234 composed of two parallel trusses each having 77 members. A numerical model inspired by this bridge
235 is used to provide responses (strain) under traffic loading and temperature variations. These
236 responses are taken as measurement data from continuous monitoring. Only one truss of the bridge
237 is modeled (Figure 3). Truss members are made of steel having an elastic modulus of 200 GPa and a
238 density of 7870 kg/m^3 . Their properties are summarized in Table 1. Supports of the truss are
239 restrained in vertical and horizontal directions. Although this is not the boundary conditions that
240 were designed for this bridge, these supports represent an upper-bound worst case when the
241 supports have deteriorated with age. Traffic loading is simulated by applying a randomly generated
242 vertical load (0-19 tonnes) at each node in the bottom chords. A load of 19 tonnes is equivalent to an
243 axle load of a railway locomotive. Daily and seasonal temperature variations are simulated as
244 thermal loads. Temperature differences between top and bottom chords due to solar radiation are
245 also taken into account in the simulations. In this example, damage is represented by a loss of
246 member axial stiffness. Damage scenarios are used to evaluate the damage detectability and time to
247 detection for all methods. In this study, damage detectability is represented by the minimum

248 detectable damage level that is the smallest percentage loss of axial stiffness in a member that can
249 be detected.

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

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

271 These results show that not all combinations lead to better performance in damage detection. For
272 this case study, the combination of MPCA and RRA outperforms other methods in terms of damage

273 detectability and it is as good as the individual application of MLR, SVR and RF in terms of time to
274 detection.

275 **4.1.2 A concrete frame**

276 This case study takes structural responses from a numerical model of a concrete frame (Figure 6) as
277 measurement data. This case study revisits numerical simulation data that was performed by
278 Cavadas (2011). The model was used for evaluation of damage detection approach using influence
279 lines of moving loads. It is a simply supported concrete frame with Young's modulus of 15 GPa. Four
280 responses – vertical displacement at mid-span, horizontal displacement at roller support, rotation
281 over the left support bearing and rotation over the right support bearing – are measured for damage
282 detection. Taking into account sensor accuracy of available sensors with ± 0.01 mm for
283 displacements and $\pm 1 \times 10^{-3}$ for rotations, a uniform distributed noise is added to measurement data.
284 Damage is introduced as stiffness reduction along 30 cm of the beam element (Figure 6).

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

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

296 While the previous case study shows that combination of MPCA and RRA (MPCA-RRA) performs best
297 in terms of damage detectability, this case study demonstrates that MPCA-SVR outperforms other

298 methods. These results indicate that the selection of regression analysis to be combined with MPCA
299 is case-dependent. A reason for this is that since combined methods conduct damage detection
300 based on the correlation of eigenvector time histories, detection is dependent on the characteristics
301 of these time series.

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

311 **4.2 A full-scale test on the Ricciolo viaduct**

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

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

322 During the monitoring period, the bridge is in a good condition and there are no damage events that
323 could generate anomalous behavior. Therefore, the time scale of the monitoring data is inverted so
324 that events during the construction period appear as anomalous events. Previously, Posenato et al.
325 (2010) demonstrated the successful application of MPCA and RRA in detecting construction stages.
326 Following this study, this section compares the performance of MPCA and RRA with the combined
327 methods in terms of detectability and time to detection.

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

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

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

344 In addition, there is a remarkable observation in Figure 14. Although the magnitude of the changes in
345 the regression residuals due to Event 6 is almost the same, there is a significant difference in the
346 amount of scatter. The threshold size within the reference period for MPCA-RRA is thus much smaller

347 than that of RRA. Unlike abrupt structural changes due to construction stages, in most cases,
348 structural degradation occurs gradually starting from small damage. Thus, the figure implies that it
349 will be more difficult for RRA to detect a change that is smaller than the change of Event 6 due to the
350 size of the threshold. Obviously, this is not the case for MPCA-RRA since the size of the threshold is
351 so small that detection is possible for relatively small changes. This shows that MPCA-RRA has the
352 potential for higher damage detectability than RRA alone.

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

356 **5 Conclusions**

357 Results of four case studies lead to the following conclusions:

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

370 **Acknowledgements**

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

374 **References**

- 375 ASCE. (2009). "Report card for America's infrastructure." American Society of Civil Engineers.
- 376 ASCE. (2011). "Structural Identification (St-Id) of Constructed Facilities : Approaches, Methods and
377 Technologies for Effective Practice of St-Id." A State-of-the-Art Report by ASCE SEI
378 Committee on Structural Identification of Constructed Systems.
- 379 Bell, E. S., Sipple, J., and Yost, J. (2008). "Long-Term Thermal Performance of a CFRP-Reinforced
380 Bridge Deck." Vancouver, BC, Canada, 242-242.
- 381 Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.
- 382 Brownjohn, J. M. (2009). "Thermal effects on performance on Tamar Bridge." *4th International
383 Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-4)*, Zurich,
384 Switzerland.
- 385 Catbas, F. N., and Aktan, A. E. (2002). "Condition and Damage Assessment: Issues and Some
386 Promising Indices." *Journal of Structural Engineering*, 128(8), 1026-1036.
- 387 Cavadas, F. (2011). "Structural Health Monitoring Applied to the Life-Cycle Management of Bridges -
388 Approaches for Damage Identification," Research Thesis Project, Universidade do Porto.
- 389 Glisic, B., and Inaudi, D. (2008). *Fibre Optic Methods for Structural Health Monitoring*, John Wiles &
390 Sons.
- 391 Goulet, J.-A., Kripakaran, P., and Smith, I. F. C. (2010). "Multimodel Structural Performance
392 Monitoring." *Journal of Structural Engineering*, 136(10), 1309-1318.
- 393 Gul, M., and Catbas, F. N. (2011). "Structural health monitoring and damage assessment using a
394 novel time series analysis methodology with sensor clustering." *Journal of Sound and
395 Vibration*, 330(6), 1196-1210.
- 396 Hart, G. C., and Yao, J. T. P. (1977). "System Identification in structural dynamics." *Journal of
397 Engineering Mechanics, Division ASCE*, 103(EM6), 1089-1104.
- 398 Hou, Z., Noori, M., and Amand, R. S. (2000). "Wavelet-based approach for structural damage
399 detection." *Journal of Engineering Mechanics*, 126(7), 677-683.
- 400 Hubert, M., Rousseeuw, P. J., and Branden, K. V. (2005). "ROBPCA: A new approach to robust
401 principal component analysis." *Technometrics*, 47, 64-79.
- 402 Koh, C. G., and Thanh, T. N. (2009). "Challenges and Strategies in Using Genetic Algorithms for
403 Structural Identification." *Soft Computing in Civil and Structural Engineering*, B. H. V. Topping
404 and Y. Tsompanakis, eds., Saxe-Coburg Publications, Stirlingshire, UK, 203-226.
- 405 Koh, C. G., and Thanh, T. N. (2010). "Output-only Substructural Identification for Local Damage
406 Detection." *The Fifth International Conference on Bridge Maintenance, Safety and
407 Management*, Philadelphia, Pennsylvania, USA.

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

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

414 Liu, S. C., and Yao, J. T. P. (1978). "Structural identification concept." *Journal Structural Division,ASCE*,
415 104(12), 1845-1858.

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

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

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

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

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

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

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

433 Posenato, D., Lanata, F., Inaudi, D., and Smith, I. F. C. (2008). "Model-free data interpretation for
434 continuous monitoring of complex structures." *Advanced Engineering Informatics*, 22(1),
435 135-144.

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

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

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

442 Reynders, E., Teughels, A., and De Roeck, G. (2010). "Finite element model updating and structural
443 damage identification using OMAX data." *Mechanical Systems and Signal Processing*, 24(5),
444 1306-1323.

445 Robert-Nicoud, Y., Raphael, B., Burdet, O., and Smith, I. F. C. (2005). "Model Identification of Bridges
446 Using Measurement Data." *Computer-Aided Civil and Infrastructure Engineering*, 20(2), 118-
447 131.

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

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

453 Saitta, S., Kripakaran, P., Raphael, B., and Smith, I. F. C. (2010). "Feature Selection using Stochastic
454 Search: An Application to System Identification." *Journal of Computing in Civil Engineering*,
455 24(1), 3-10.

456 Strauss, A., Frangopol, D. M., and Kim, S. (2008). "Use of monitoring extreme data for the
457 performance prediction of structures: Bayesian updating." *Engineering Structures*, 30(12),
458 3654-3666.

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

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

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

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

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

471 Zhang, X., Liang, D., Zeng, J., and Asundi, A. (2012). "Genetic algorithm-support vector regression for
472 high reliability SHM system based on FBG sensor network." *Optics and Lasers in Engineering*,
473 50(2), 148-153.

474

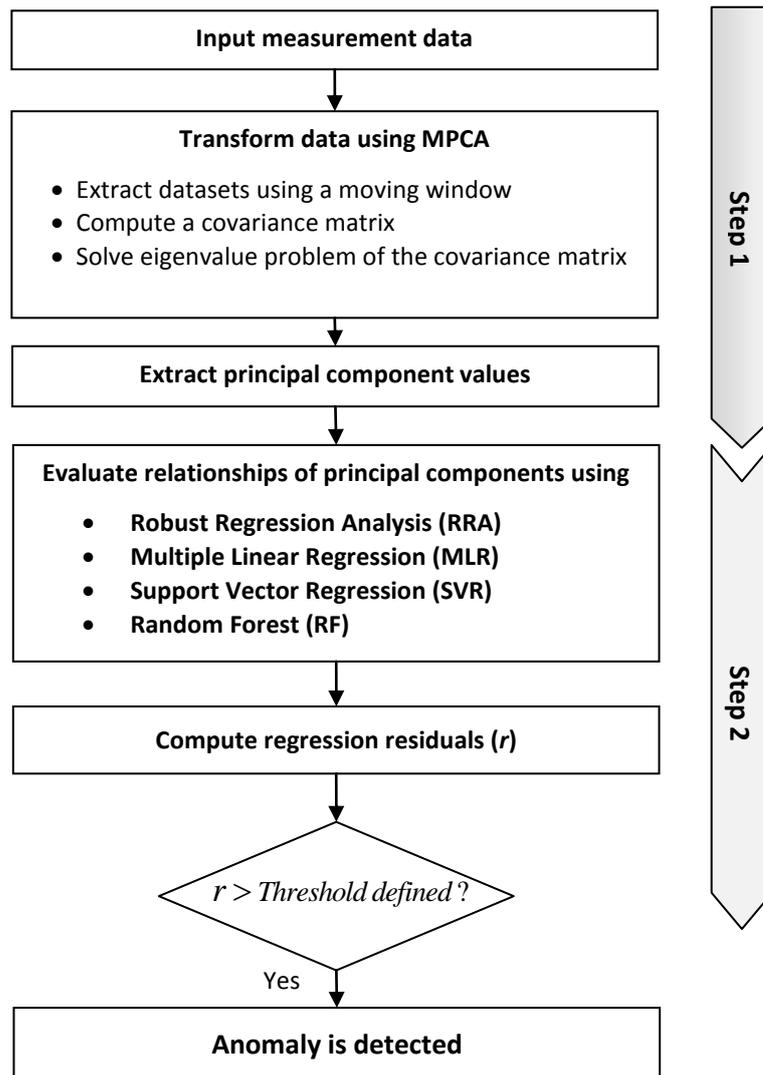


Figure 2 Flowchart of the combined model-free data-interpretation methodologies

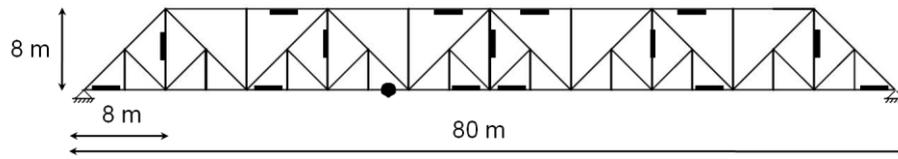


Figure 3 . An 80-m railway steel truss bridge with sensor locations marked as black bars and the damage location marked as a black dot.

477
478

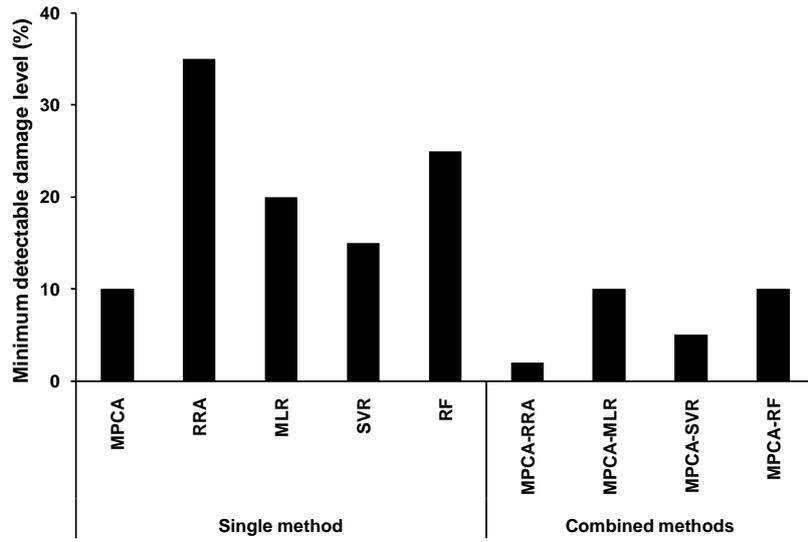


Figure 4 Minimum detectable damage-level for a truss bridge using 9 model-free data interpretation methods.

479

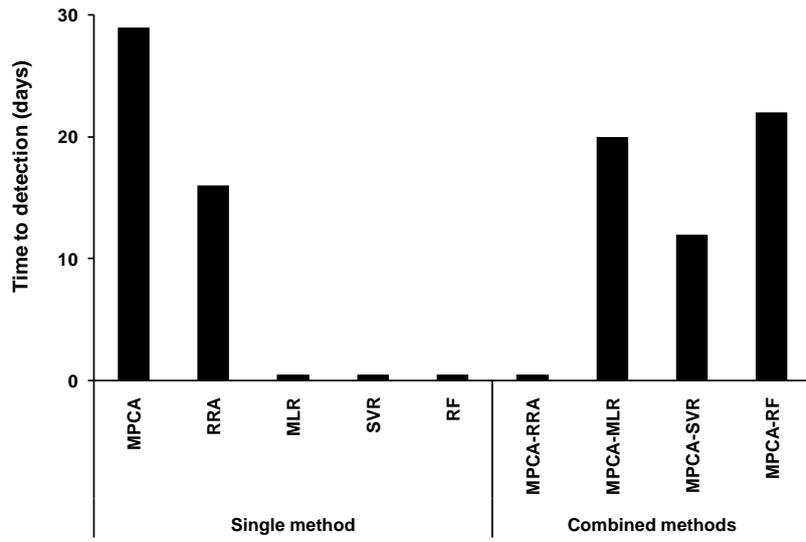


Figure 5 Time to damage detection for a truss bridge using 9 model-free data interpretation methods for 50% damage level

480

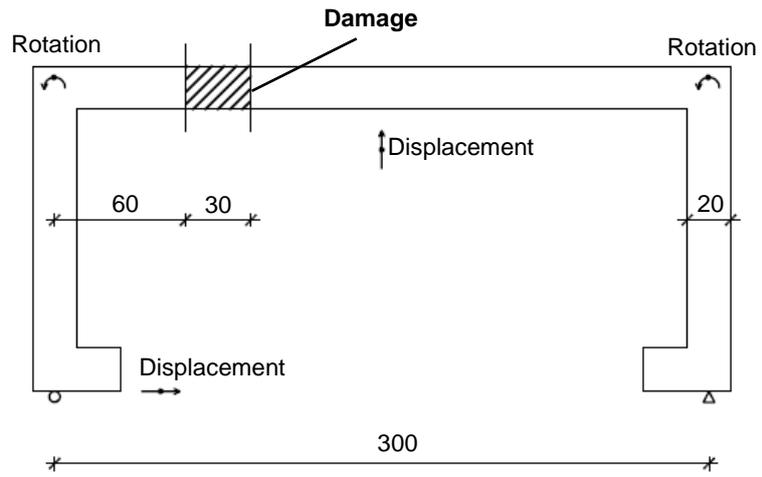


Figure 6 A concrete frame model with four measured responses including vertical displacement at midspan, horizontal displacement at roller support, rotation over the left support bearing and rotation over the right support bearing.

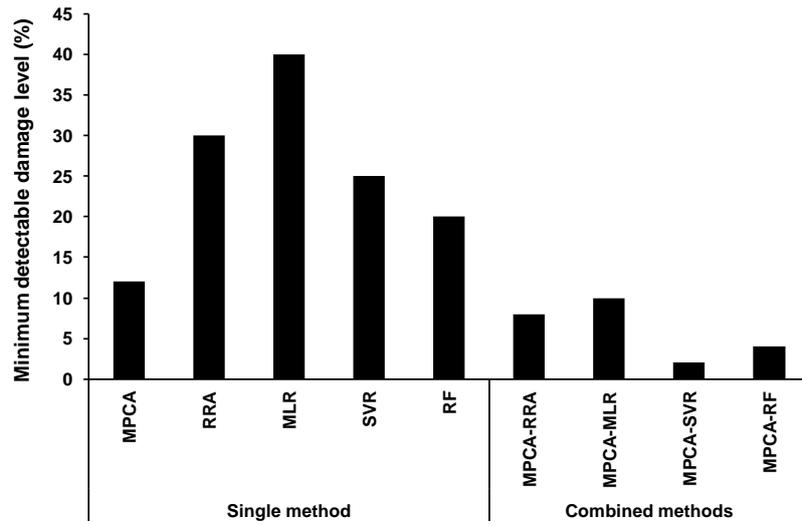


Figure 7 Minimum detectable damage level for a concrete frame using 9 model-free data interpretation methods

482
483
484

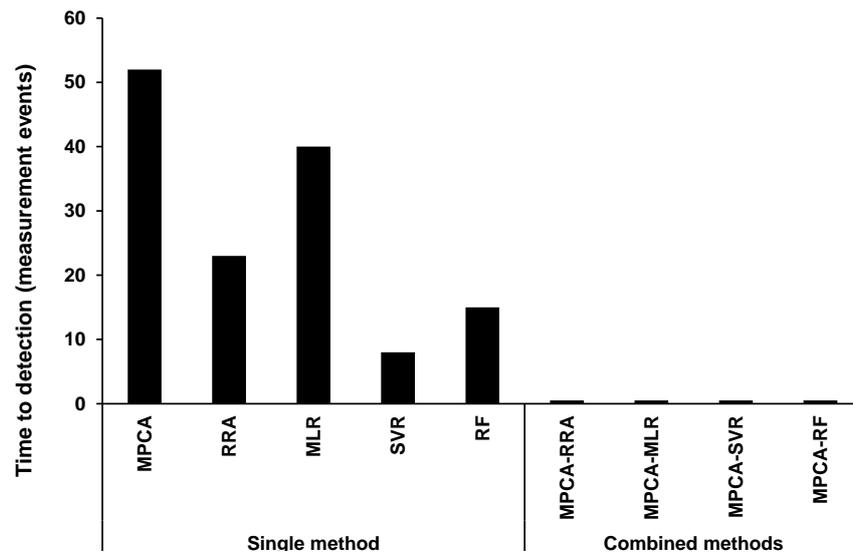


Figure 8 Time to detection for a concrete frame using 9 model-free data interpretation methods for 35% damage level

485

486

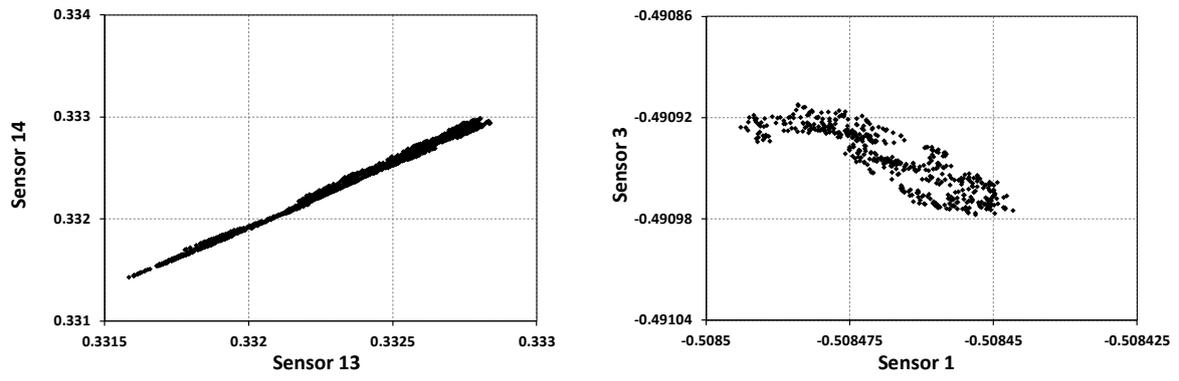


Figure 9 High correlated principal component pairs for the truss model (left) and the frame model (right).

487

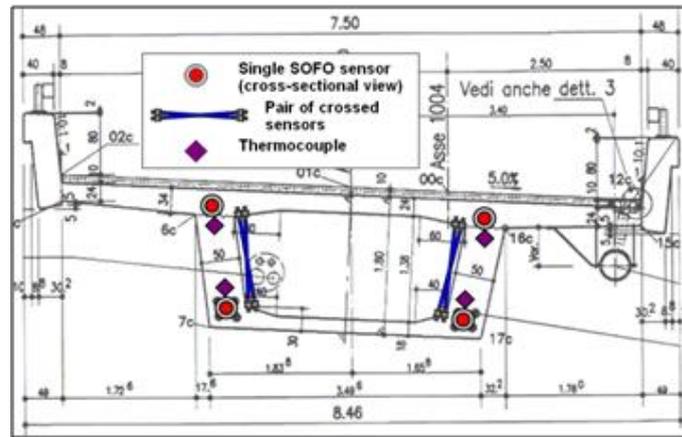


Figure 10 Measurement configuration for the Ricciolo viaduct: a cross-section view (Posenato et. al., 2010)

488

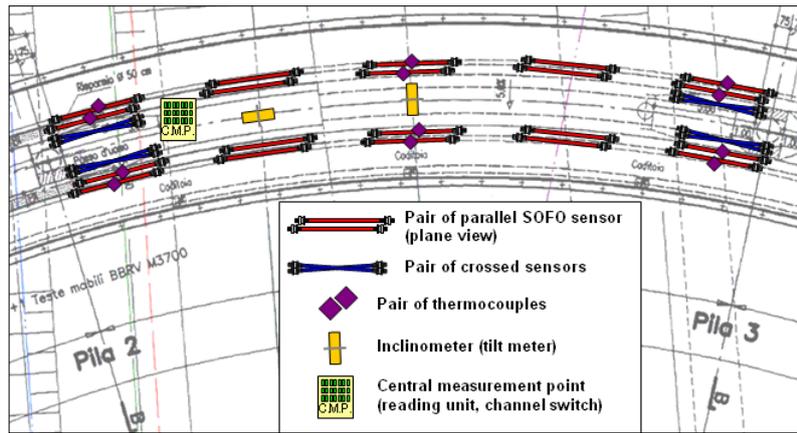


Figure 11 Measurement configuration for the Ricciolo viaduct: a plane-view (Posenato et. al., 2010)

489

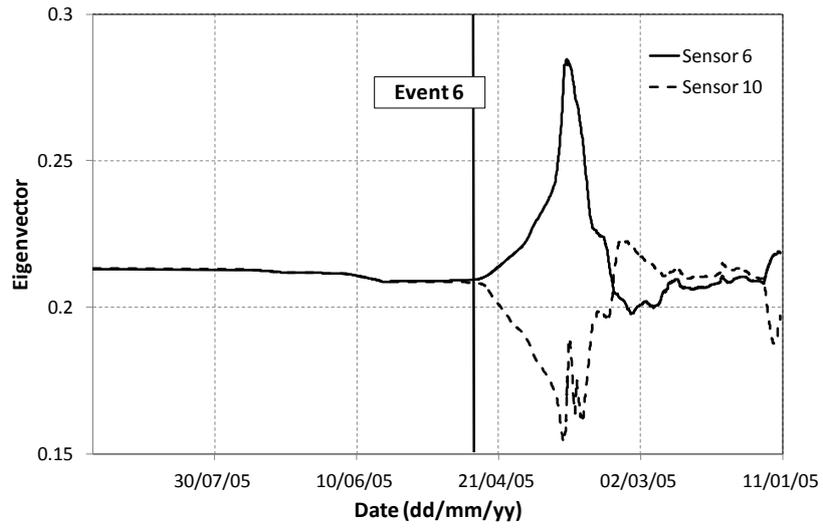


Figure 12 Plot of the first eigenvector during the monitoring period recalculated from Posenato et al. (2010). Detection of construction stages as events that simulate anomalous structural behavior (time scale is inverted)

490

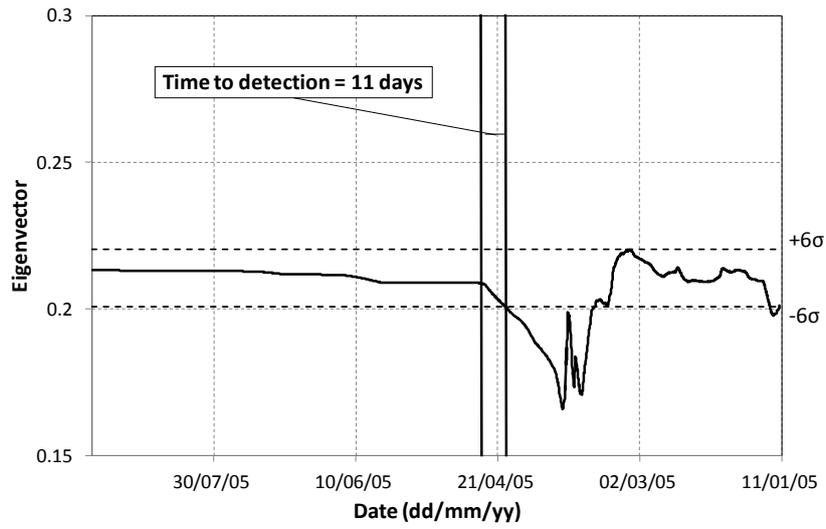


Figure 13 Time to detection for event number 6, recalculated from Posenato et al. (2010)

491

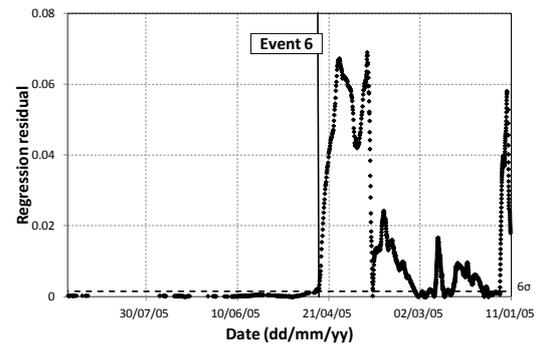
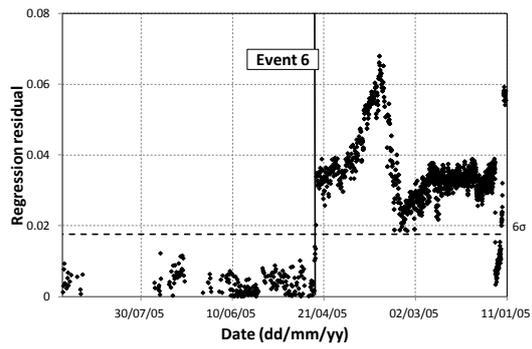
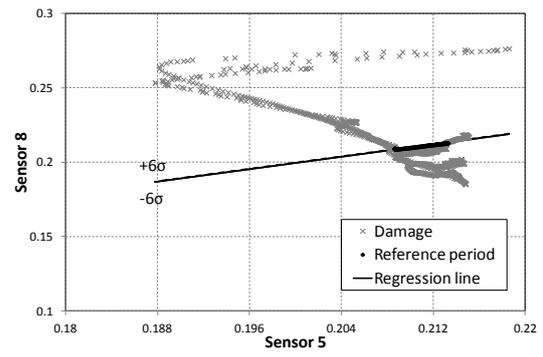
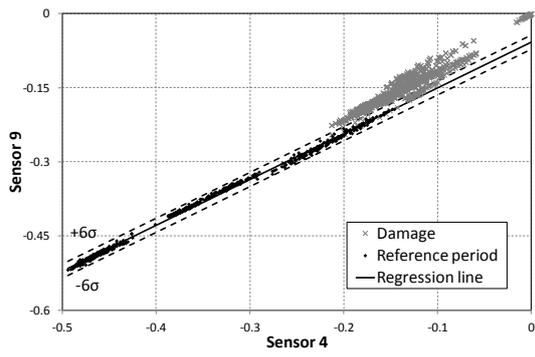


Figure 14 Results of correlation-based methods for damage detection using RRA (left) and MPCA-RRA (right)

492
493
494
495

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

Member type	Area (m²)	I_x (m⁴)	I_y (m⁴)
Top chord	5.15 x 10 ⁻²	2.267 x 10 ⁻³	2.586 x 10 ⁻³
Bottom chord	3.03 x 10 ⁻¹	1.467 x 10 ⁻³	1.458 x 10 ⁻³
Vertical	2.19 x 10 ⁻²	1.215 x 10 ⁻³	4.245 x 10 ⁻⁵
Diagonal	3.69 x 10 ⁻²	9.704 x 10 ⁻⁴	4.164 x 10 ⁻³
Small diagonal	2.19 x 10 ⁻²	1.215 x 10 ⁻³	4.245 x 10 ⁻⁵

497

498

499 **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

500

501