

# A clustering approach for structural health monitoring on bridges

Alberto Diez<sup>1</sup> • Nguyen Lu Dang Khoa<sup>2</sup> • Mehrisadat Makki Alamdari<sup>2</sup> •  
Yang Wang<sup>2</sup> • Fang Chen<sup>2</sup> • Peter Runcie<sup>2</sup>

Nguyen Lu Dang Khoa khoa.nguyen@nicta.com.au, Alberto Diez [alberto.diez@tecnalia.com](mailto:alberto.diez@tecnalia.com), Mehrisadat Makki Alamdari mehri.makkialamdari@nicta.com.au, Yang Wang yang.wang@nicta.com.au, Fang Chen fang.chen@nictacom.au, Peter Runcie peter.runcie@nictacom.au

**Abstract** Structural health monitoring is a process for identifying damage in civil infrastructures using sensing system. It has been increasingly employed due to advances in sensing technologies and data analytic using machine learning. A common problem within this scenario is that limited data of real structural faults are available. Therefore, unsupervised and novelty detection machine learning methods must be employed. This work presents a clustering based approach to group substructures or joints with similar behaviour on bridge and then detect abnormal or damaged ones, as part of efforts in applying structural health monitoring to the Sydney Harbour Bridge, one of iconic structures in Australia. The approach is a combination of feature extraction, a nearest neighbor based outlier removal, followed by a clustering approach over both vibration events and joints representatives. Vibration signals caused by passing vehicles from different joints are then classified and damaged joints can be detected and located. The validity of the approach was demonstrated using real data collected from the Sydney Harbour Bridge. The clustering results showed correlations among similarly located joints in different bridge zones. Moreover, it also helped to detect a damaged joint and a joint with a faulty instrumented sensor, and thus demonstrated the feasibility of the proposed clustering based approach to complement existing damage detection strategies.

**Keywords:** Structural health monitoring, Damage detection, Novelty detection, Unsupervised learning, K-means clustering

## 1 Introduction

Ageing and damage in transport infrastructures, such as roads, bridges and tunnels, are becoming a big issue nowadays. In order to improve safety and reduce costs derived from ageing and damage, early identification of damage in a structure is important to avoid further risks, both in life-safety and economic losses. Structural Health Monitoring (SHM) based approaches have been increasingly used during recent years to address this problem [13, 30].

With regard to bridges, and despite the advances in abstract analysis and controlled testing, failures have the most conspicuous influence on their design, construction and management. Many failures are mainly caused by inappropriate design and poor maintenance (corrosion, scour, etc.) [8]. As established by Sibly and Walker [29], each innovative bridge form is developed by trial and error method until its limits are surpassed and spectacular failure occurs. Only then does theory catch up with the practice and fully explains the structural behaviour of the

<sup>1</sup> Tecnalia Research & Innovation, Donostia-San Sebastián, Gipuzkoa, Spain

<sup>2</sup> National ICT Australia (NICTA), Sydney, Australia

infrastructure. Two popular examples are Tacoma Bridge, in 1940, due to a poor design [18], and Silver Bridge, in 1967, due to the use of bad material [22]. The more complex dynamic behaviour of these structures resulted in greater lapses in the designer's knowledge, and hence, many failures under wind loads, as well as those caused by rushing crowds.

For this reason maintenance should be considered as a fundamental pillar to face ageing and damage effects in bridges [15]. Maintenance strategies are essentially based on nonintrusive sensing, monitoring and analysis techniques to provide flexible decision support, normally in the form of inspection recommendations (when, where and why to act). To this concern, data-driven or machine learning based analysis [35] and model-driven or finite element analysis [19] are most commonly used to generate data models (from an data analytical perspective) and physical models (from a numerical or mechanical perspective) that aim to represent the structure. Generated models are then used to identify damage in the structures [34].

This work is an effort to apply SHM to the Sydney Harbour Bridge by a data-driven machine learning approach. It presents a clustering based technique to group bridge joints with similar behaviour and then detect abnormal or damaged ones. Vibration events caused by passing vehicles are acquired by accelerometers located in several joints of the structure. Since there are not clear evidences of anomalies, data are analysed in an unsupervised fashion. A combination of feature extraction and outlier removal is performed, and then similar events and joints are grouped by the clustering technique. This will allow isolating possible damaged joints on the bridge.

## 2 Related work

Statistical pattern recognition has been widely used to identify damage in the structure. As can be seen in [12, 21, 37], group classification, regression analysis, neural networks and density estimation were popular methods. Nevertheless, in most of the cases success was restricted to simulations, laboratory studies and well-controlled experiments, such as in the Z24 Bridge [23]. Thus, limited evidences of real structural faults are available and in most of the cases the effectiveness of the approaches still remains to be validated for operational civil structures.

In an effort of overcoming this limitation and addressing the novelty detection challenge, many authors have proposed clustering and unsupervised learning approaches for damage detection in SHM [16]. When there is no data available from damaged states, one way to proceed is to estimate the density of feature vectors, assuming the

structure is in good condition, and then quantify the consistency of new feature vectors with relation to the estimated density. *K*-means, Gaussian Mixture Models (GMM), Support Vector Clustering (SVC) and Self-Organizing Maps (SOM) algorithms are representative examples of clustering-based approaches for statistical modeling and feature classification in SHM [28]. Yu et al. [39] used a fuzzy *c*-means (FCM) clustering approach for structural damage detection. Several feature extraction techniques were applied in their study such as data projection algorithms, principal component analysis (PCA) and kernel principal component analysis (KPCA). Toivola et al. [31] used different dimensionality reduction methods such as Curvilinear Component Analysis, PCA and Random Projection to reduce high-dimensional vibration measurements in conjunction with nearest neighbor algorithm for novelty detection. In [7] a hierarchical parameter clustering technique was used to group acceleration data from a cable-stayed bridge, based on the similarity of parameters involved in the process. Other clustering approaches for SHM focused on grouping information coming from sensor networks distributed throughout the structure. Yin et al. [38] proposed a clustering-based multi-hop routing protocol to group similar nodes in each span of the bridge. Nie et al. [26] also studied the use of cluster-based data aggregation architectures for SHM to transform the raw data into useful information.

Several signal processing and feature extraction techniques have been applied to extract features which are sensitive to changes in the structure. Some of these techniques can be found in [5, 6, 11]. The most popular techniques for signal processing are wavelet transform (WT) and Fourier transform (FT), as discussed by Newland et al. [25]. Alternatively, time series analysis is commonly applied in order to extract damage sensitive features from SHM data. Time series models are used to fit to the vibration data and then damage indicators are obtained by comparing new data against the learnt models. Gul et al. [17] employed time series modelling in conjunction with Mahalanobis distance to identify structural changes on different case studies.

## 3 Proposed approach

This section describes the proposed clustering approach for SHM of bridges. The proposed algorithm is based on an unsupervised classification of vibration events and joints. The main steps of the methodology can be seen in Fig. 1, which has been applied for the presented work on the Sydney Harbour Bridge.

The vibration responses of the structure excited by passing vehicles, also named as events, are measured from

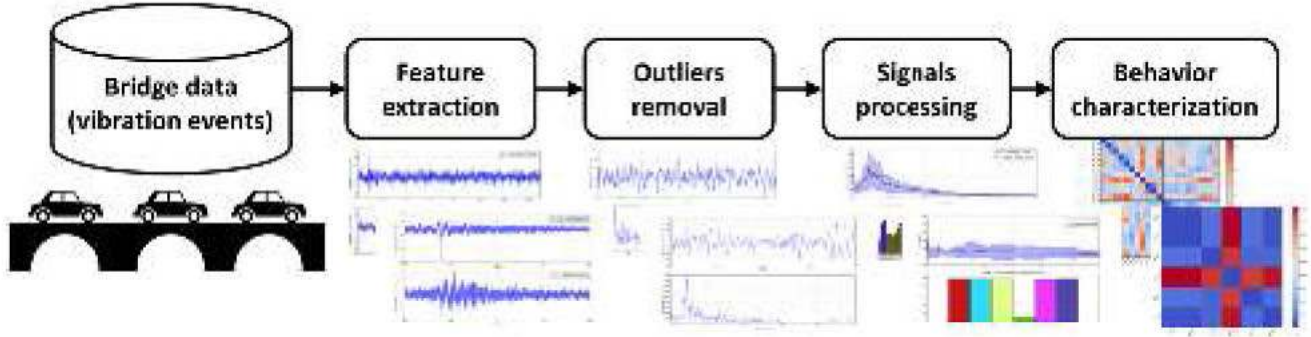


Fig. 1 Flowchart of the proposed clustering based approach for SHM

hundreds of joints within the structure using tri-axial accelerometers. The measured time history of vibration responses, corresponding to each event, is stored for further analysis. For every event, raw acceleration data are transformed into a unique feature in time domain which is explained in detail in Sect. 3.1. Outliers are then removed based on density estimation of bridge joints and the energy of the signals, allowing reducing dimension of the data, as can be seen in Sect. 3.2. Having removed the outliers, the remaining events are then transformed into the frequency domain using fast Fourier transform (see Sect. 3.3). Finally, clustering techniques are employed to train models from available data that are able to characterize normal structural behaviour. Joints with similar behaviour are grouped and damage joints are detected. This last phase is described in Sect. 3.4.

### 3.1 Data preprocessing for feature extraction

An event is defined as a time period in which a vehicle is driving across an instrumented joint. Vibrations caused by passing vehicles are recorded by tri-axial accelerometers positioned at the joints located in different parts of the structure. For every event sample, accelerometer data in the three axis,  $(x, y, z)$ , are transformed in a unique feature,  $V_i = |A_i| - |A_r|$ , where  $A_i$  is the instantaneous acceleration at  $i$ th sample and  $A_r$  is the rest vector or the average of the three readings  $(x, y, z)$  in the first 100 samples. They are collected before the event is triggered for ensuring that all the events have a similar wave form. Events that are triggered within the first 100 samples are filtered out (these events are mainly caused by vehicles that were driving close to each other). Data standardisation is then applied to every  $V$  as,

$$X = \frac{V - \mu(V)}{\sigma(V)} \quad (1)$$

where  $V = V_1, \dots, V_n$  are extracted features from an acceleration event.

### 3.2 $k$ -Nearest neighbors for outliers removal

$k$ -Nearest Neighbors (kNN) searches for the  $k$  closest neighbors in the feature space. It can be mainly used for classification [10] and regression [2], predicting the label or the value of a new data sample as the majority of labels or values among its  $k$  closest neighbors.

kNN-based approach proposed within this study performs an iterative process that allows removing outliers and noisy signals incrementally, until the convergence criteria is met. Outliers are joints signals that are far from their joint representative, calculated as the mean of all joint events. This also allows resampling the number of events for every joint in order to balance the whole data set.  $K$  Dimensional-Trees (KD-Trees) are used for optimising the  $(k)$  nearest neighbors searching process [24].

For every event,  $X = (x_1, \dots, x_n)$ , the sum of the energy in time domain is calculated as,

$$E(X) = \sum_{i=1}^n |x_i|^2. \quad (2)$$

Then, for each iteration, the  $k$  closest neighbors to the mean of the energy of the joint signals,  $\mu_{\text{joint}}$ , are taken. The distance of the  $k$  closest signals,  $d(E(X_j), \mu_{\text{joint}})$ ,  $j = 1, \dots, k$ , for every joint is calculated.  $X_j$  is marked as outlier and is removed from current joint if the condition is met,

$$d(E(X_j), \mu_{\text{joint}}) > \mu(D) + 2 * \sigma(D), \quad (3)$$

where  $\mu(D) = (1/k) * \sum_{j=1}^k d(E(X_j), \mu_{\text{joint}})$  is the mean of the distance of every event energy to the joint mean and  $\sigma(D)$  is the standard deviation of such distances.

This process is repeated until any of the previously established stopping criteria is met. Namely, a maximum number of iterations and a distance threshold, calculated during the first iteration as,

$$\text{threshold} = \mu(D) + (0.5 * \sigma(D)). \quad (4)$$

If the maximum distance obtained from the  $k$  nearest points to the mean of the joint at any iteration is below this threshold, the process is stopped.

### 3.3 Fourier transform for vibration signals processing

Fourier transform is a signal processing technique that decomposes a time domain signal into the series of frequencies (amplitudes and frequencies). It was first discussed by Joseph Fourier [14], and since then it has been further developed, becoming a robust frequency domain method in modal analysis [9]. The basic idea of spectral analysis is to represent the original vibration signal as a new sequence, which determines the importance of each frequency component in the dynamics of the signal. This is achieved by using the discrete version of the Fourier transform as,

$$X(f) = \sum_{-\infty}^{\infty} x(t) e^{-2\pi i f t}, \quad (5)$$

where  $f$  denotes the frequency at which  $X(f)$  is evaluated.

Within proposed approach, Fast Fourier Transform (FFT) is employed. The frequency spectrum of each vibration signal in the time domain,  $X = x_1, \dots, x_n$ , is thus computed as,

$$A(f_x) = \sum_{j=1}^{j=n} x_j \omega_n^{(j-1)(f_x-1)}, \quad (6)$$

where  $f_x$  represents frequency,  $|A(f_x)|$  is signal amplitude in frequency domain,  $x_j$  is one of the  $n$  time domain sampling points of the vibration signal and  $\omega_n = e^{-\frac{2\pi i}{n}}$ .

This signal analysis technique provides a powerful spectral based diagnostic method in stationary conditions, when there is no transient signals involved.

### 3.4 K-means clustering for behaviour characterization

K-means clustering algorithm performs a partition of data space into  $K$  clusters. Each cluster is represented by an object, named centroid or the mean point, whose initial value can be randomly set or estimated by applying some kind of heuristic. In an iterative process, each element is assigned to the partition with the least distance between it and the centroid of the partition. Once all elements have been assigned into clusters, cluster centroids are recalculated using the information of the elements that belong to each partition. The process converges to a solution with a linear complexity  $O(n)$ , which is not always a global optimum. The success of K-means method is strongly

determined by the choice of the  $K$  value, the distance metric employed and the initial centroids values [36].

Once the FFT is applied over remaining events after the outlier removal process, Euclidean metric is computed in order to determine the distance between events [1]. It is calculated as follows,

$$\text{dist}(X, Y) = \|X - Y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (7)$$

where  $X = (x_1, \dots, x_n)$  and  $Y = (y_1, \dots, y_n)$  are vibration signals after applying FFT.

K-means clustering for damage detection allows detecting anomalies like damage which are usually isolated in small clusters. They can be removed and further analysed. Therefore, remaining data set can be used for training accurate normality models. Classification of new events is then performed by computing their distances to previously learnt cluster centroids that represent behaviour of interest.

## 4 Experimental results

### 4.1 Field test measurement

The Sydney Harbour Bridge was opened in 1932. It is a steel through arch bridge operated by Roads and Maritime Services (RMS) the road authority of the state of New South Wales, Australia. The bridge carries lanes of 8-road traffic and two railway lines. Traffic lane 7 is a dedicated bus and taxi lane on the eastern side of the bridge. Lane 7 consists of an asphalt road surface on a concrete deck supported by concrete and steel jack arches as in Fig. 2. There are approximately 800 jack arches over a total distance of 1.2 km. The jack arches are difficult to access and are inspected visually every two years according to standard inspection practices. Figure 2 illustrates one single jack arch in the structure.

Acceleration data from each jack arch are collected using three tri-axial accelerometers, labeled sensors 1, 2 and 3 where, respectively, are mounted to the joint in left, middle and right positions as shown in Fig. 3. The sensitivity of the accelerometers have been set to  $\pm 2$  g but they are capable of measuring acceleration response up to  $\pm 8$  g. Each tri-axial accelerometer measures acceleration responses in the direction of passing traffic ( $x$ ), in the direction of gravity ( $z$ ) and in the direction of ( $y$ ). In this study, only measured responses in ( $z$ ) direction of sensor 2 is utilized since ( $z$ ) direction accounts for most of the acceleration of the middle sensor 2. Acceleration data are collected after the vehicles, usually buses, drive over the deck where the sensors are located. A pre-set threshold is used to trigger the recording in each sensor, which corresponds to an event.



Fig. 2 Illustration of a the bus lane on the Sydney Harbour Bridge and b one of the concrete jack arches underneath the bus lane

Fig. 3 6 Joints experiment, schematic of the evaluated joints

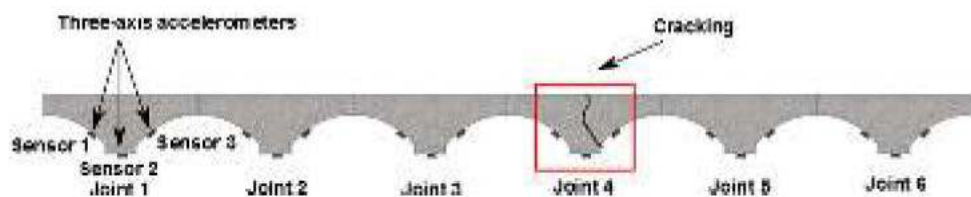


Fig. 4 The Sydney Harbour Bridge schema

**6-Joint case** In the first case study, 6 joints from North pylon and North main span were monitored during the first week of August 2012, namely Joints 1 to 6 as shown in Fig. 3. It is known that a crack was present in the 4th joint at that time. Vibration signals were sampled at a frequency of 375 Hz and during 1.6 s, resulting in 600 samples per event. Frequency range was set from 0 to 187.5 Hz.

**71-Joint case** In a separate case study, 71 joints were monitored from the first week of October 2014. Monitored joints belong to the following bridge zones: span 6, span 7, span 8, North pylon and North main span (see Fig. 4 for locations of these monitored areas). It is known that the sensor mounted on Joint 135, located in the second bay of span 7, was faulty at that time. Vibration signals were sampled at a frequency of 250 Hz during 2 s, resulting in 500 samples per event. Frequency range is set from 0 to 125 Hz.

Figures 5 and 6 show plots of typical vectors  $V = |A| - |A_r|$  obtained from raw acceleration data of considered data sets. Corresponding FFT signals are also provided.

For both case studies, Frequency Domain Decomposition technique (FDD) [3] was implemented to investigate the natural frequencies of the jack arches. For an arbitrary joint within 71-joint data, FDD technique was applied on 10 randomly selected event data and the results for the first normalized singular value was presented in Fig. 7a (for each event a different line pattern was used). As shown in the figure, within the frequency range of 16 to 18 Hz, a significant mode can be extracted. Interestingly, a variation is observed between natural frequencies obtained from different events, which may be caused by different environmental and traffic conditions. The same behaviour was observed from all the other joints.

FDD technique was also applied on the damaged joint in the 6-joint case. Compared to the healthy state (after repair), there is a considerable reduction in the dominant mode. Figure 7b compares the first normalized singular values for the damaged joint before and after repair obtained from 30 randomly selected events. In Fig. 7b, the results for healthy state and damaged state are, respectively, shown by black line and grey line.

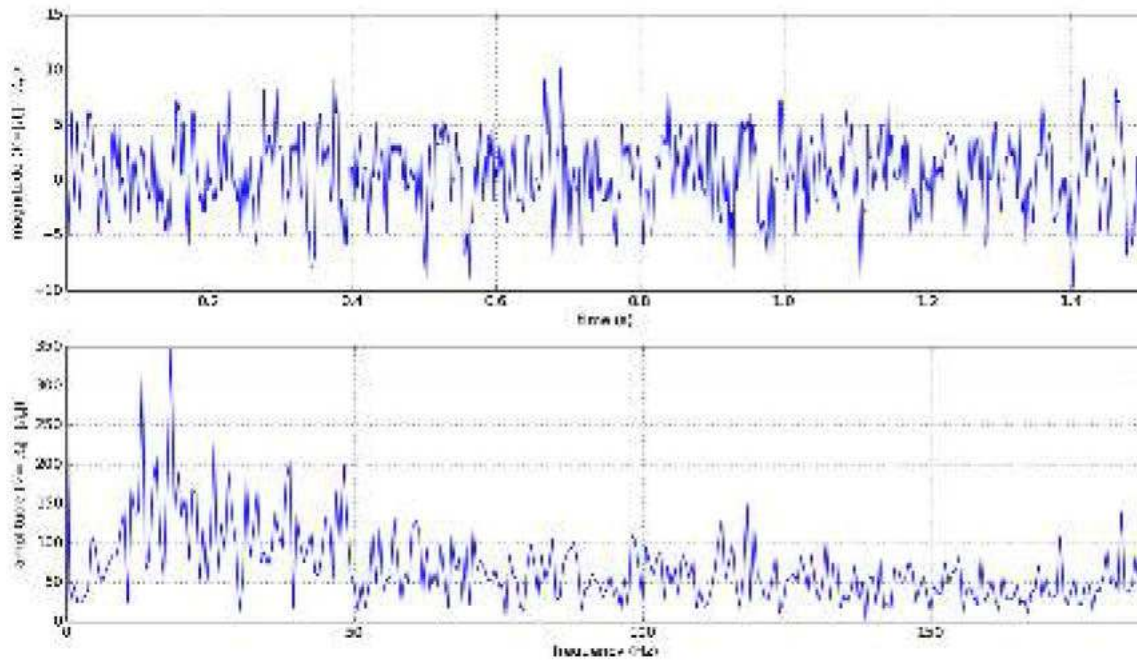


Fig. 5 6-Joint case: an example of vibration signal (*above*) and its corresponding FFT (*below*)

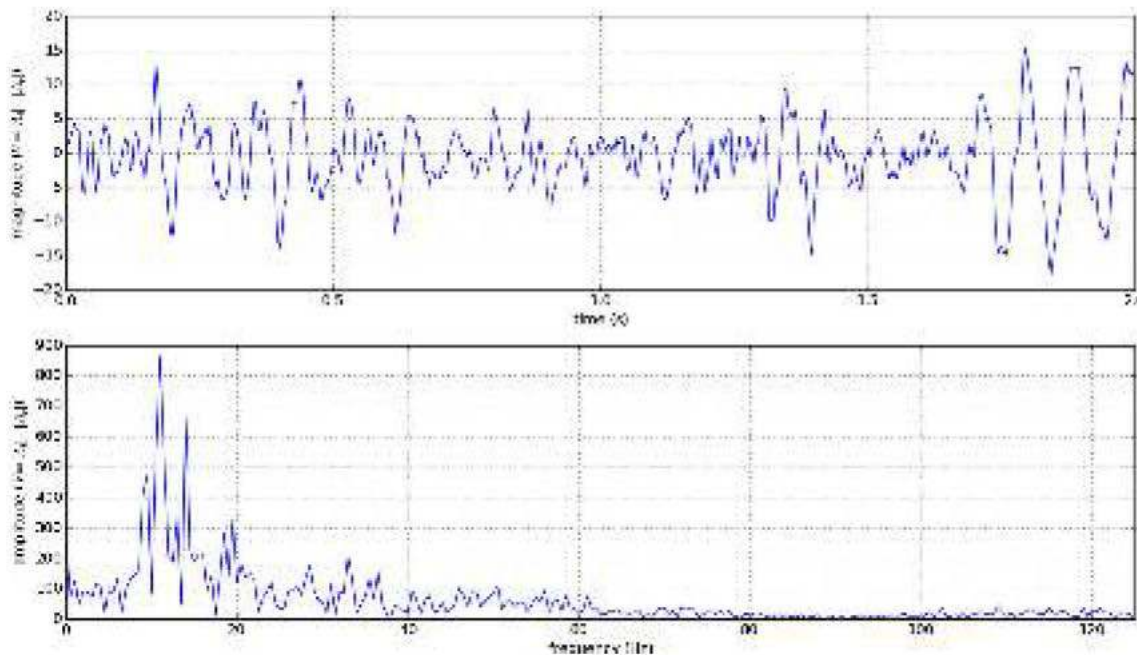


Fig. 6 71-Joint case: an example of vibration signal (*above*) and its corresponding FFT (*below*)

## 4.2 Event-based clustering

Event-based clustering is performed to group similar vibration signals, and thus capturing the structural behaviour. Clusters formed are shown as graphs containing the pattern (centroid or mean values) and the variance of grouped events for each frequency. Clusters distribution is

also provided as the percentage of events belonging to each joint that are grouped in a cluster.

### 4.2.1 6-Joints case

In this experiment, comprising Joints 1, 2 and 3 in North main span and Joints 4, 5 and 6 in North pylon, the main

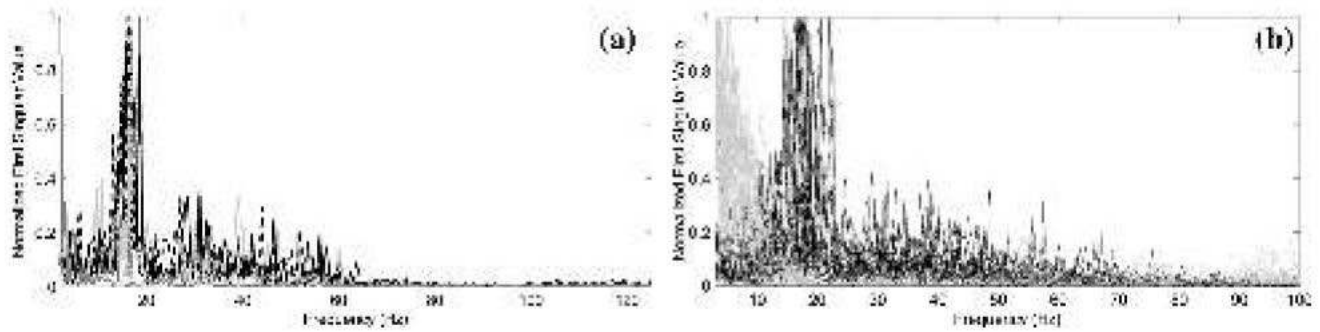


Fig. 7 Frequency analysis of a an arbitrary joint among 71 studied joints and b damaged joint before repair (*grey line*) and after repair (*black line*)

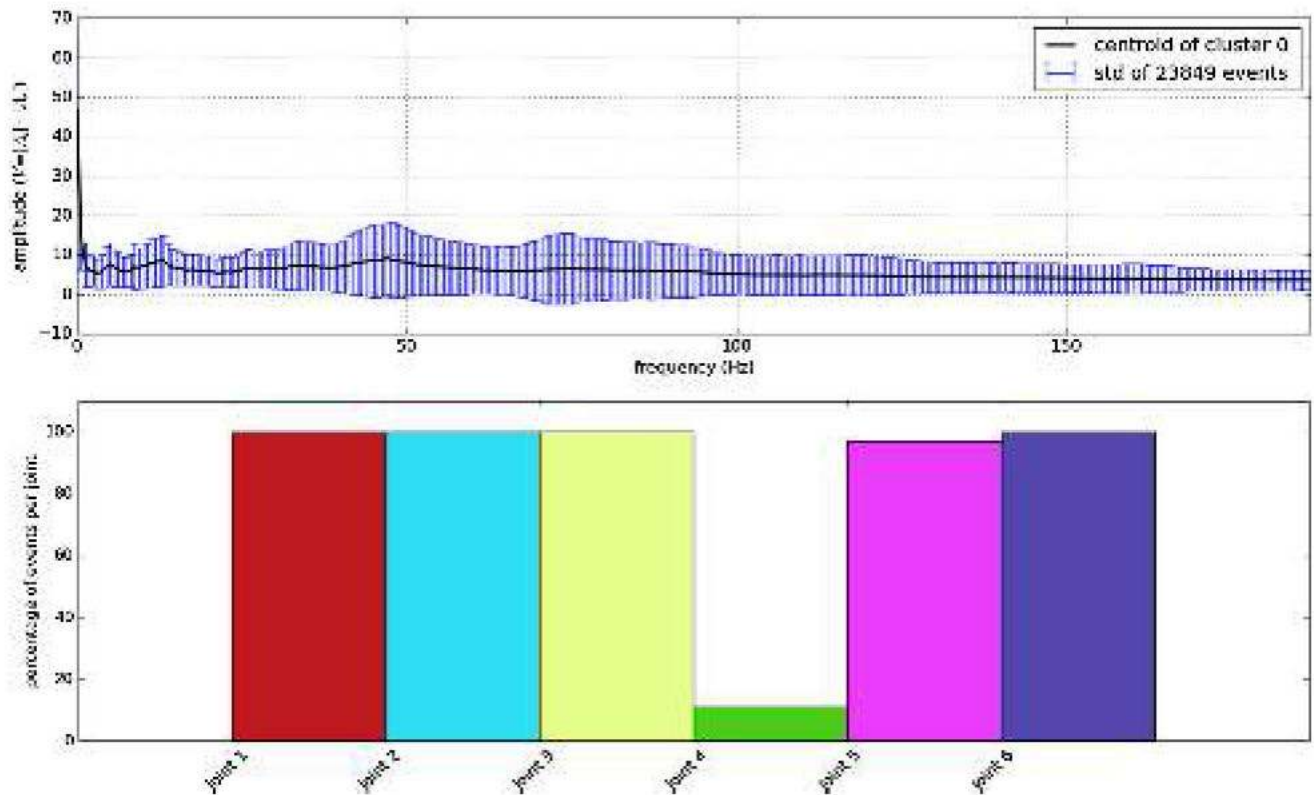


Fig. 8 6-Joint case: cluster 0 showed a normal behaviour (centroid and standard deviation of cluster events *above* and joints distribution *below*)

motivation was to characterize normality from joints events. When there is no previous knowledge about the presence of any kind of abnormal behaviour, the goal is to try to isolate outliers from high-density regions, containing the majority of the data. In this case, notably, with  $K = 2$  normal events of all joints were grouped in a big cluster, containing a total of 23,849 events, whereas 4662 events related to the damage, mostly located in joint 4, were isolated in a small cluster. Clusters formed can be seen in Figs. 8 and 9. Events belonging to the same joint are equally coloured.

#### 4.2.2 71-Joint case

This section shows results achieved when clustering events from 71 joints. A quick test was first conducted only considering 5 joints located in the second bay of span 7. As in previous case  $K$ -means was executed with  $K = 2$  in order to find a small cluster defining abnormal behaviour in the data.

As it can be seen in Figs. 10 and 11, several events from Joint 135 were located in a small cluster (together with a few events of Joint 131, located in the same bay and span),

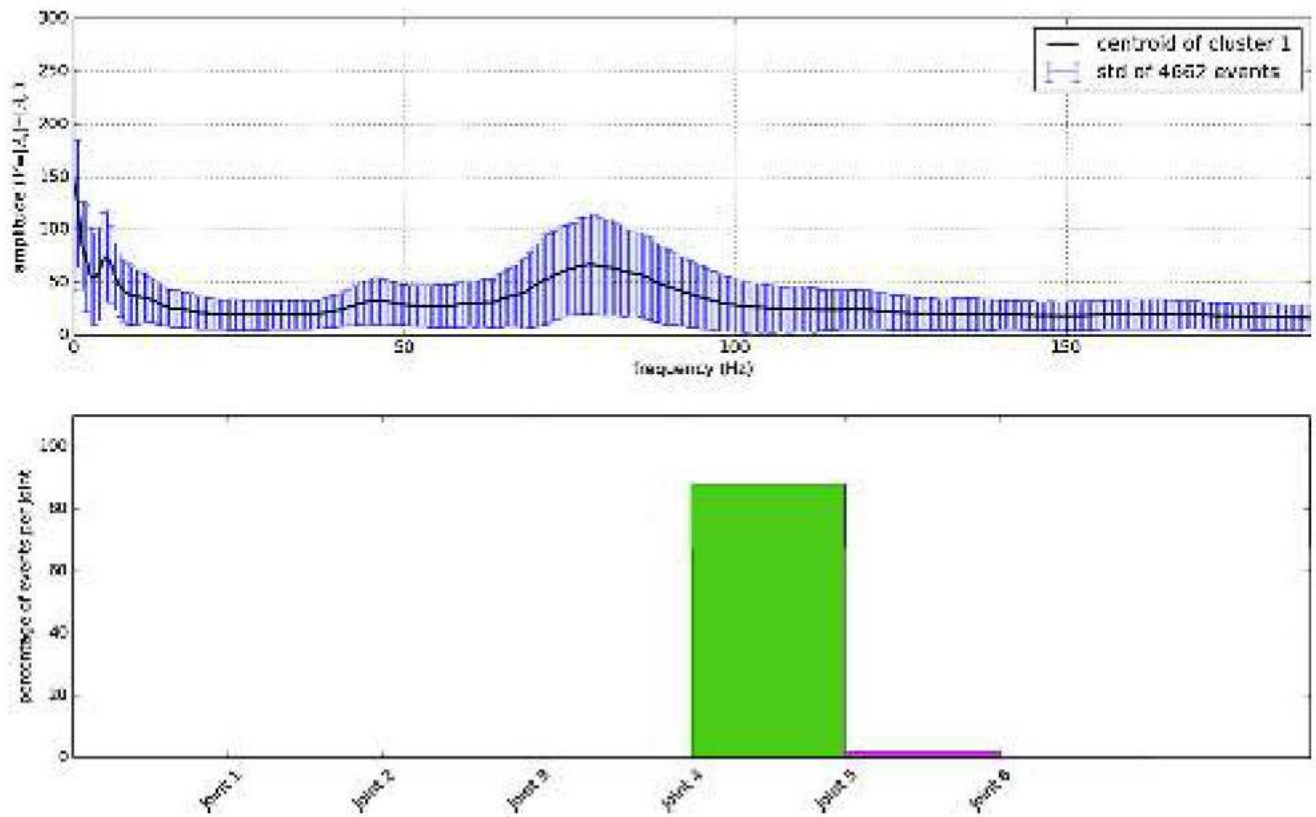


Fig. 9 6-Joint case: cluster 1 showed a damaged behaviour

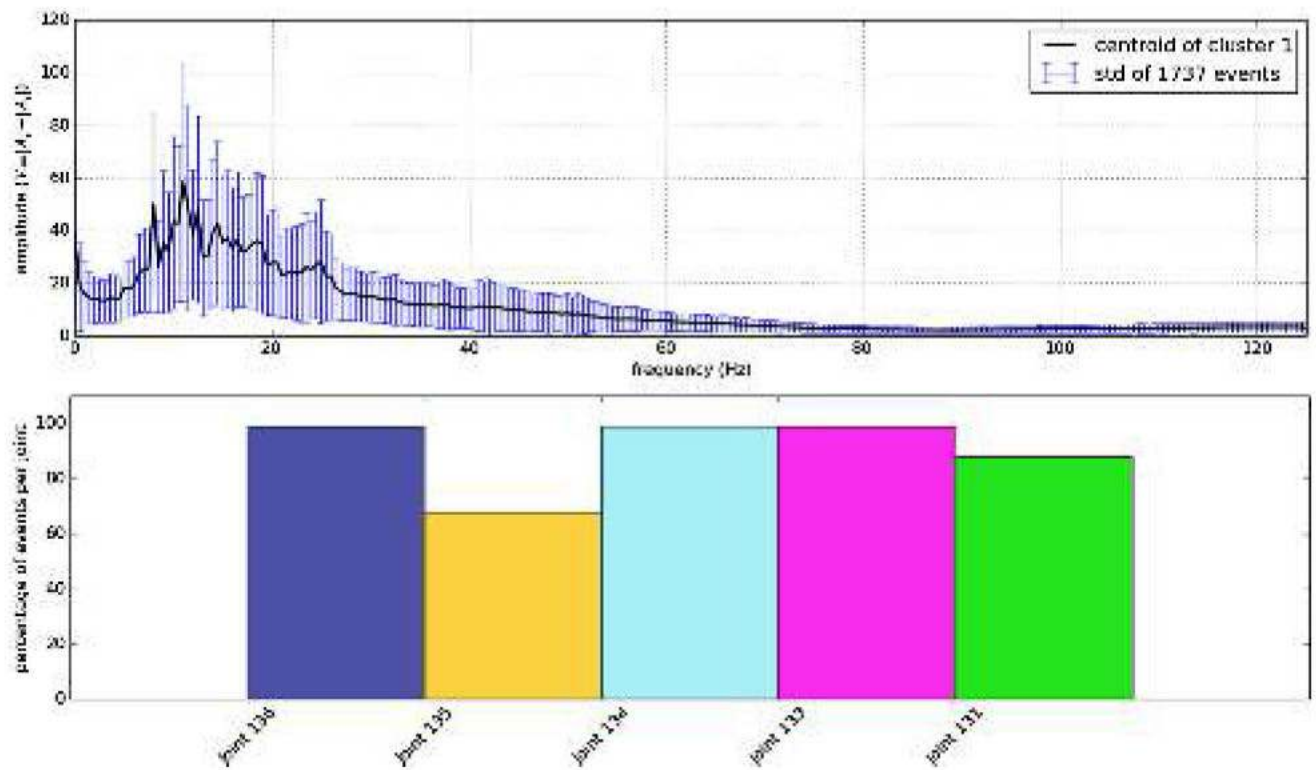


Fig. 10 Analysis of 5 joints located in the second bay of span 7: cluster 0 showed a normal behaviour



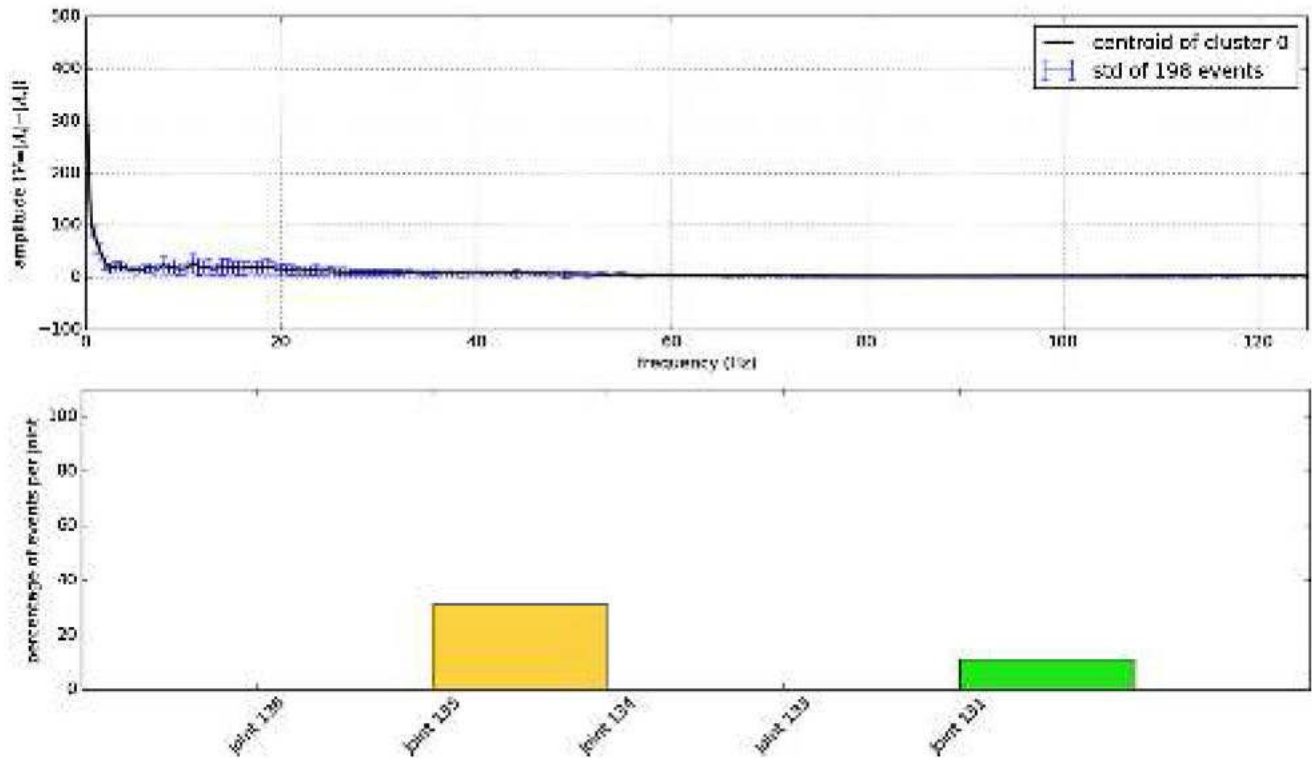


Fig. 11 Analysis of 5 joints located in the second bay of span 7: cluster 1 with events from a faulty sensor

showing an abnormal pattern characterized by centroid values. The other cluster grouped the majority of the events, representing a normal behaviour.

In the next experiment, all 71 joints were clustered. The goal was to group similar joints located in similar relative positions along the bridge. *K*-means clustering was executed with  $K = 5$  since 5 different bridge areas were involved:

- Span 6: 6 bays, including 33 joints.
- Span 7: 5 bays, including 21 joints.
- Span 8: 4 bays, including 12 joints.
- Span North pylon: 2 bays, including 7 joints.
- Span North main span: only 1 joint.

Figure 12 shows the clustering result in one cluster. Joints belonging to the same bay were equally colored. It can be seen that no useful information can be acquired given the fact that events from one joint can appear in several clusters. Therefore, it is sometime not straightforward to group events from the same joint together using event-based clustering.

### 4.3 Joint-based clustering

To overcome the weakness of the event-based clustering mentioned above, joint-based clustering was utilized. In joint-based clustering, a map of pairwise distances among

representatives of all joints was generated. A joint representative is calculated as the mean values of all events of each joint, after outlier removal phase. As in the case of event-based clustering, the distance metric used was Euclidean distance.

#### 4.3.1 6-Joint case

For the experiment related to the analysis of 6 joints, it can be appreciated that the distance between the representative of joint 4 and the others is significantly high (see Fig. 13). The more red the colour the higher the dissimilarity between joints representatives. Therefore, the damage behaviour in this joint is detected.

#### 4.3.2 71-Joint case

Since only spans 6 and 7 had sufficient instrumented joints covering most of the span, we only calculate the map of pairwise distances for these two spans. Missing joints in spans 6 and 7 are:

- In span 6: 1 joint in bay 4; 3 joints in bay 5 and 2 joints in bay 6.
- In span 7: 1 joint in bay 1; 1 joint in bay 2; 2 joints in bay 3; 2 joints in bay 4; all joints in bay 5 and 2 joints in bay 6.

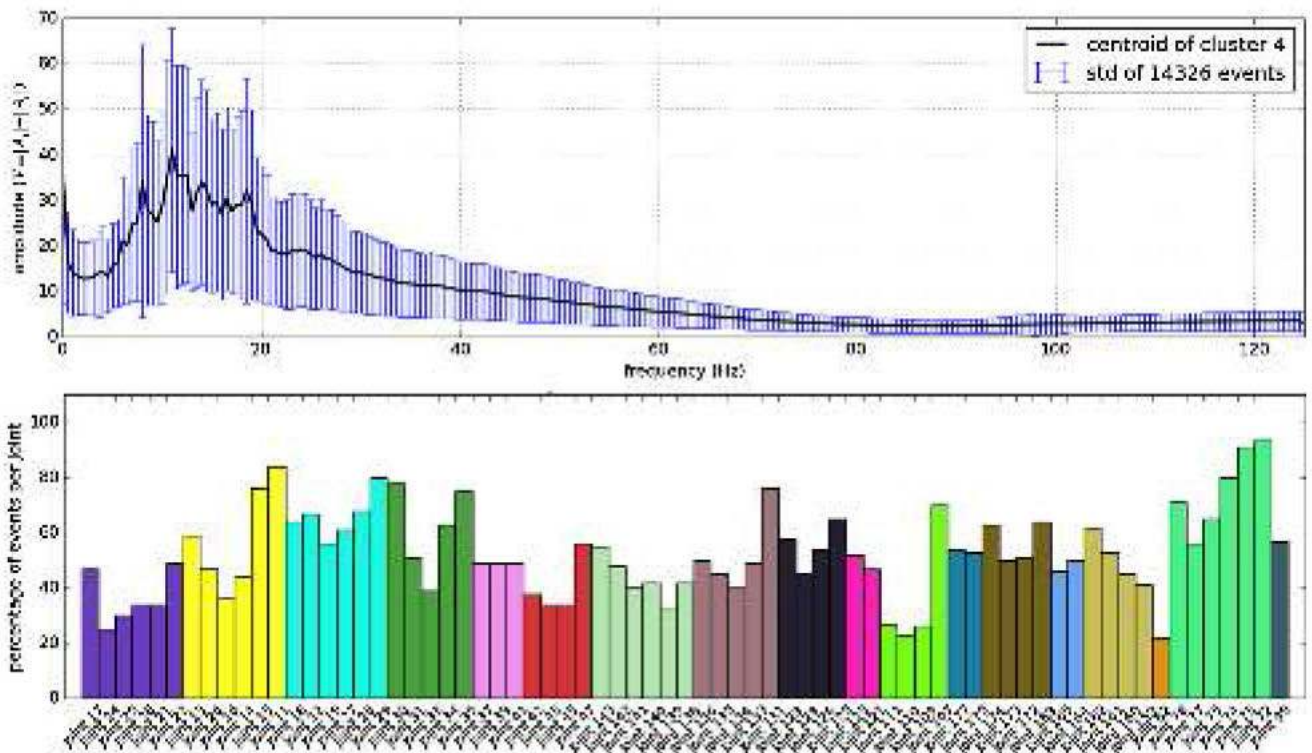
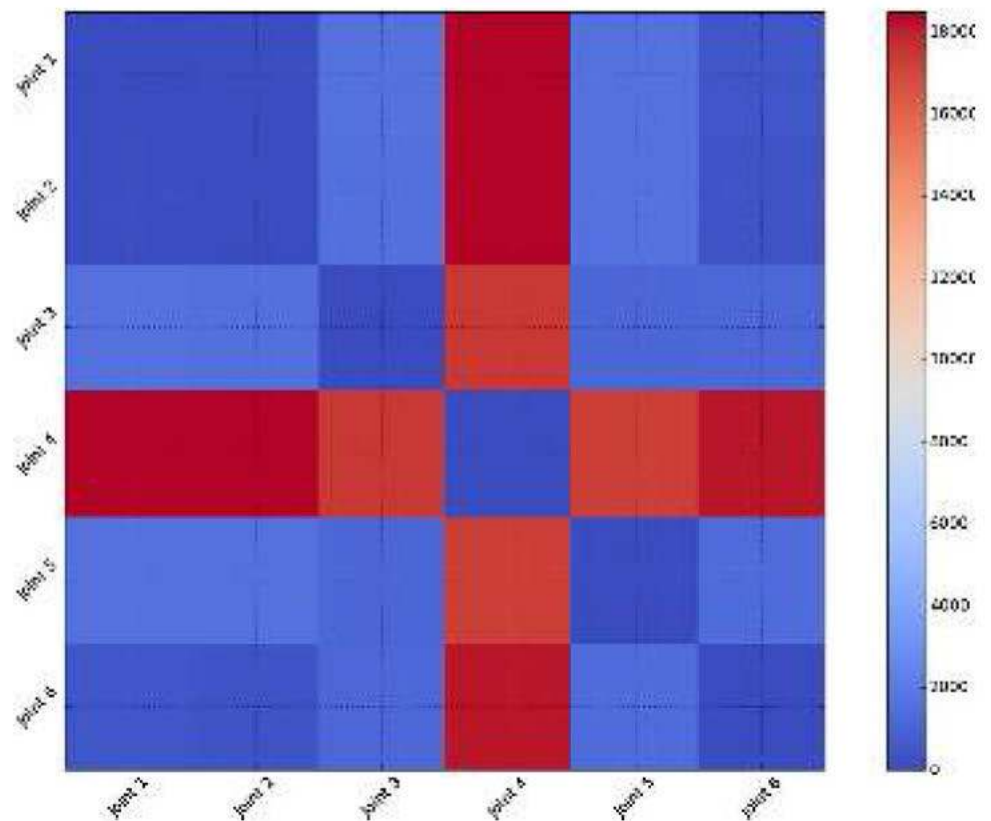
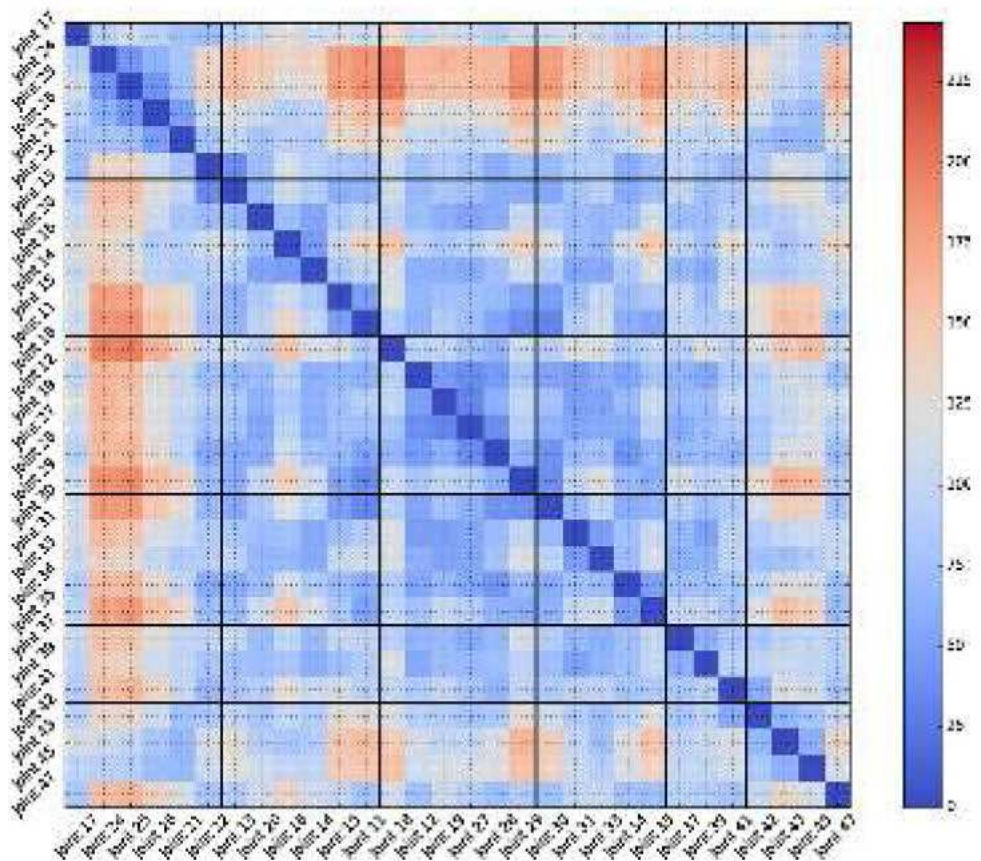


Fig. 12 71-Joint case: events from one joint can appear in several clusters

Fig. 13 6-Joint case: damage in joint 4 was captured



**Fig. 14** 71-Joint cases, span 6: joints at middle of a bay/span behave similarly and are different from joints at the two ends



Resulting maps of pairwise distances for spans 6 and 7 can be seen in Figs. 14 and 15. The black lines within the map in each figure delimit regions that correspond to different bays of the span. Each small cell with a color shows a pairwise distance between two corresponding joints in the span. The higher the distance, the less similarity between two related joints.

Overall, the similarities between joints located in different bridge parts are found out. In span 6, it is shown that joints at middle of a bay/span behave similarly and are different from joints at the two ends. It may suggest an indication of a bridge global behaviour in a span. In span 7, although a global view is not available due to the missing joints, Joint 135 located in bay 2 of span 7 had considerably large distances to other joints. It is known that the sensor in Joint 135 was faulty at the time the data were collected. Joint 131 was also behaving differently compared to the others, which implies a possible problem that should be checked.

#### 4.3.3 Cross correlation

In this section the performance of the proposed approach was compared with the one obtained from applying cross-correlation instead of Euclidean distance to measure the

similarity between signals. The correlation coefficient of two signals measures the level of similarity or linear dependence [33]. Comparing two sequences  $X$  and  $Y$ , the Pearson product-moment correlation is defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (8)$$

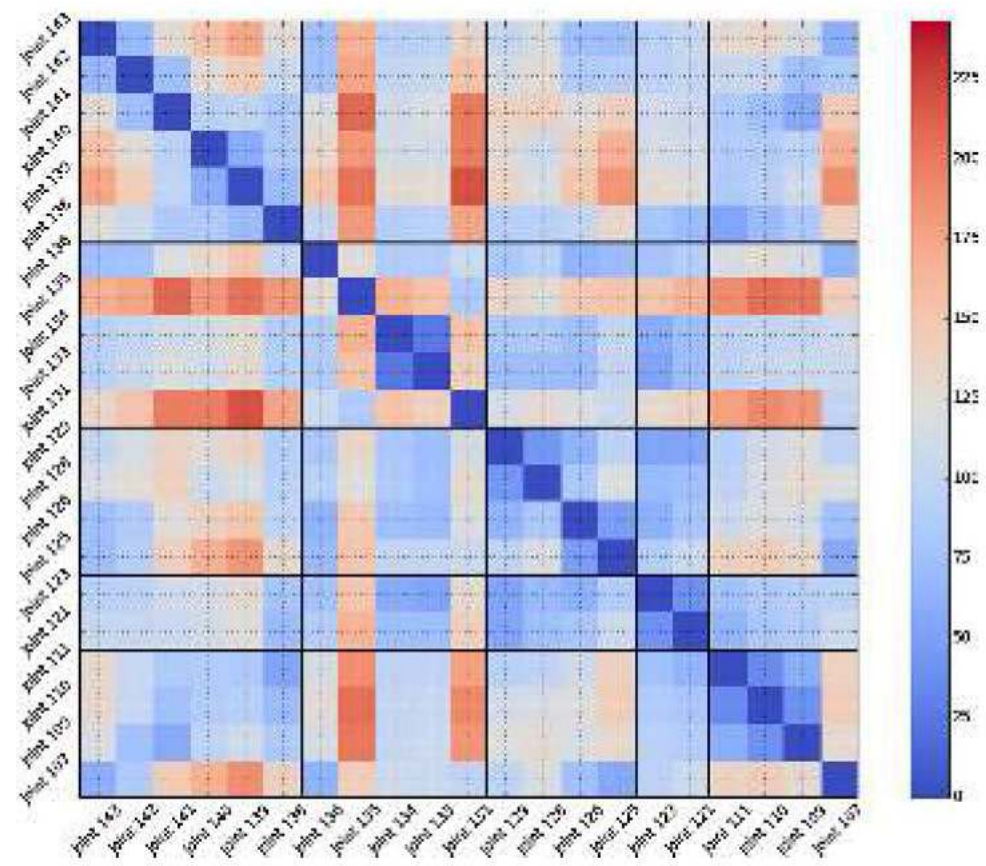
where  $\text{cov}(X, Y)$  is the covariance of  $(X, Y)$ ,  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ , respectively.

The correlation between two signals is thus the covariance divided by the product of the standard deviations of both signals. Resulting coefficient will vary between  $-1$ , if there is a total negative correlation,  $0$ , when there is no correlation, and  $1$  for the case of total correlation. It can be expressed as a dot product between the normalized signals. For damage detection, it has been widely used to compare frequency domain signals in combination with Hilbert transform [32].

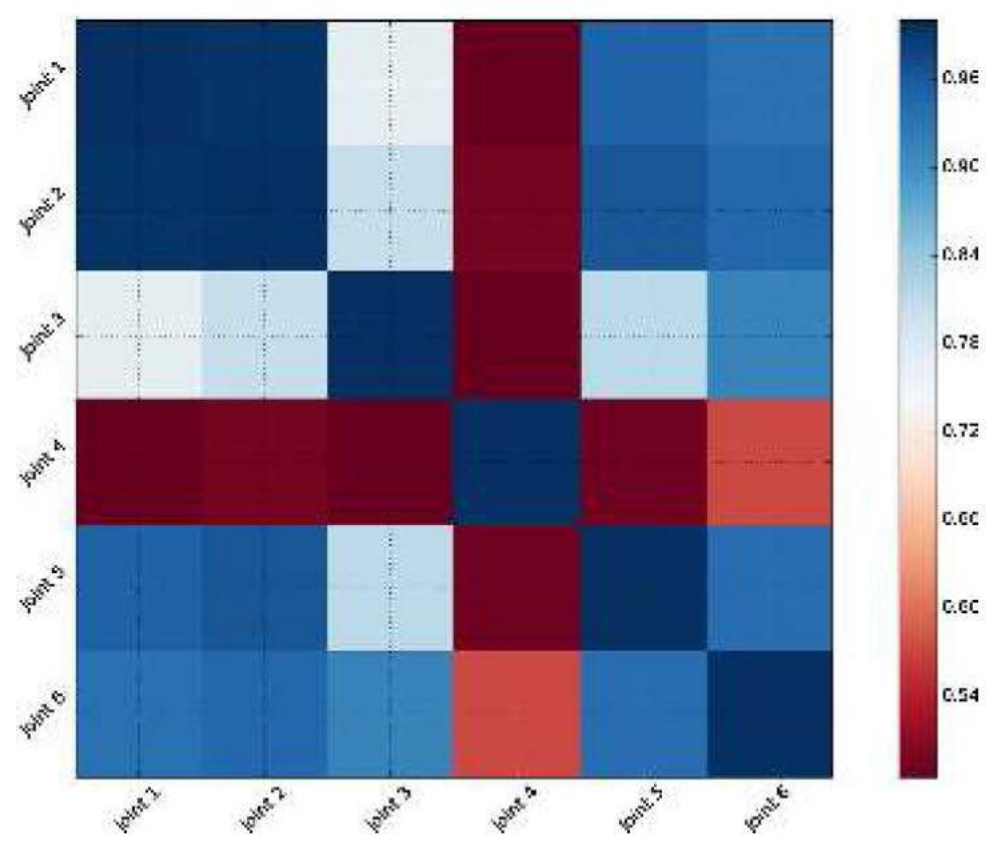
In our approach, instead of Euclidean distance, cross correlation can be used as a similarity measure between signals. Results obtained in the case of joint-based clustering can be seen in Figs. 16, 17 and 18.

Concerning 6-joint experiment, the results of Euclidean distance and cross-correlation were similar. However, differences between healthy joints and the damaged one were

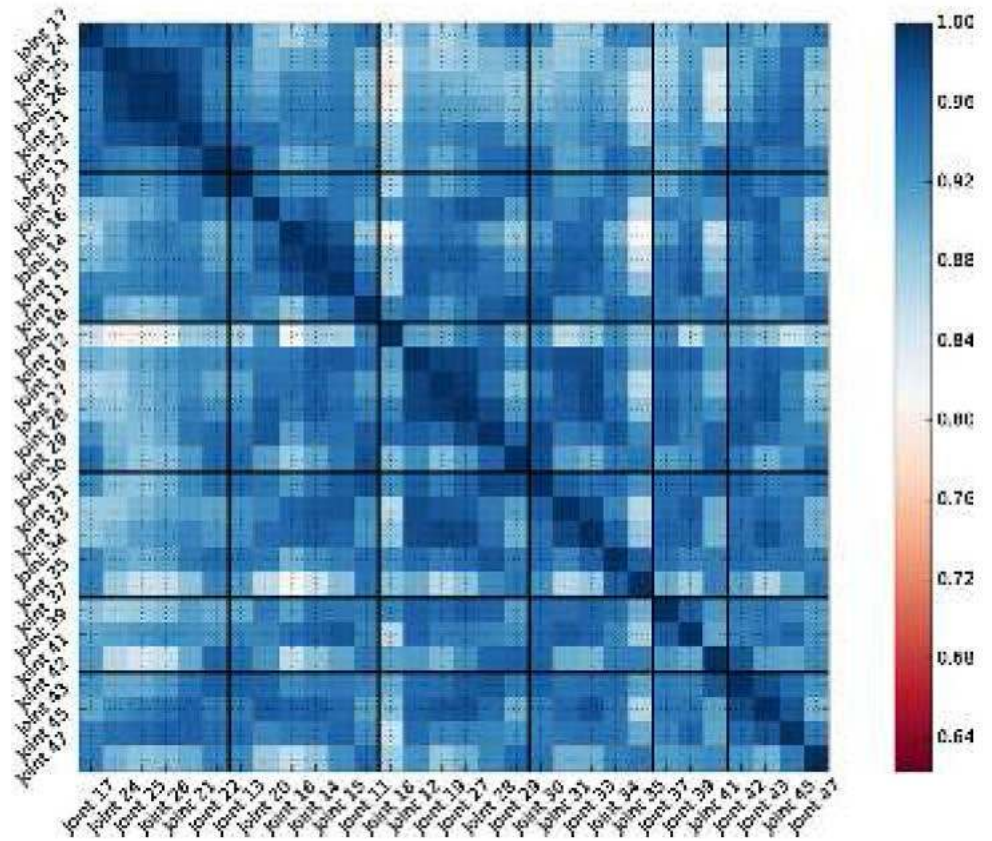
**Fig. 15** 71-Joint case, span 7: faulty sensor at Joint 135 was detected



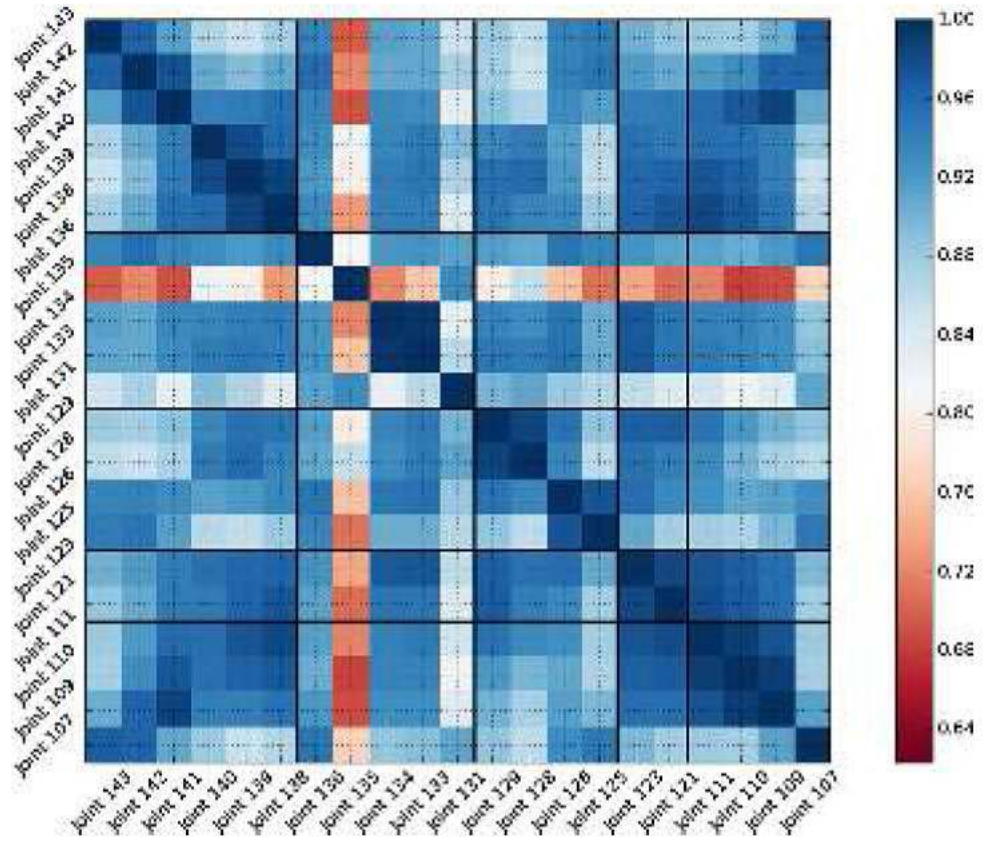
**Fig. 16** 6-Joint case: map of pairwise distances using cross correlation



**Fig. 17** 71-Joint case, span 6:  
map of pairwise distances using  
cross correlation



**Fig. 18** 71-Joint case, span 7:  
map of pairwise distances using  
cross correlation



less pronounced. Whereas in the case of 71-joint experiment the cross-correlation detected the faulty sensor in Joint 135 but showed a correlation coefficient with the other joints over 0.5, which indicated they can be considered moderately correlated. Cross-correlation also estimated worse the location dependency among joints, as it can be appreciated in Fig. 17. Joints in all bays of span 6 showed a correlation over 0.8. It therefore seems that cross correlation coefficient is less sensitive to small changes in joint behaviour than Euclidean distance. It must be also noticed that due to the computational cost of covariance computation when dealing with high-dimensional signals, it is an inappropriate distance estimator in its basic form [4].

#### 4.4 Effect of kNN-based outlier removal

kNN-based outlier removal process was executed with a different value of  $k$  for each experiment. This is due to the fact that the number of available events per joint in each case varied significantly.  $k$  must be specified according to the least amount of joint events that are being analysed. This allows keeping a balance between joints cardinality, avoiding any of them to dominate the clustering process. In

6-joint case,  $k$  was set to 5000. Regarding the 71-joint case,  $k$  was set to 500, since less events per joint were available. At every iteration and for each joint,  $k$  is the number of closest events (in terms of a distance metric) to the mean of the energy of all joint events. In case the number of events in joint  $i$ ,  $|\text{events}_i| < k$ ,  $k$  is set to the number of events in joint  $i$ . The maximum number of iterations was set to 10.

Results obtained in terms of the number of events filtered for each experiment are summarized in Table 1. Additionally, the standard deviation of signals filtered and signals selected for training are shown in Figs. 19 and 20. The results show that filtered events had larger standard deviation compared to events used for training.

#### 4.5 Online learning and computational time

For online damage detection, previously generated models from historical data can be used. When new excitation events are obtained, feature extraction and signal processing steps are applied as explained above. Then, in the case of event-based classification, distances to behaviour models are computed and similarities are established. Events far from the normality can be further studied for any

Table 1 Results obtained by kNN outlier removal process

	6-Joint case 36,947 events	71-Joint case 45,818 events
Number of events for training	28,511	27,407
Number of filtered events	8436	18,411

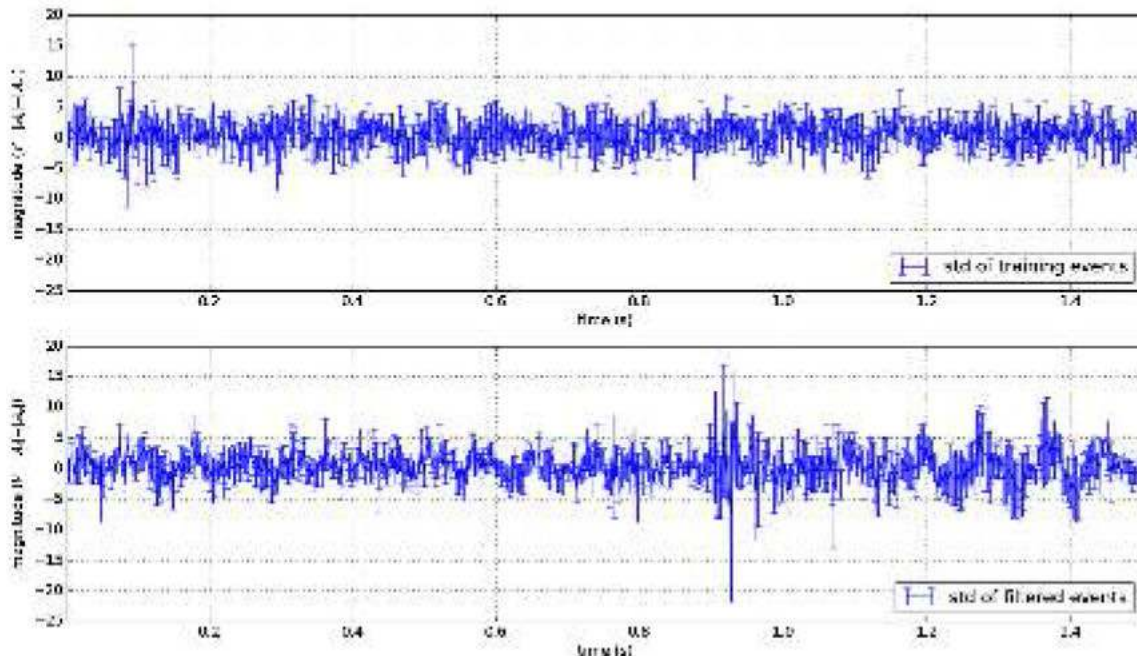


Fig. 19 6-Joint experiment. Standard deviation of the training events (above) and the filtered events (below) after performing the kNN-based outlier removal step

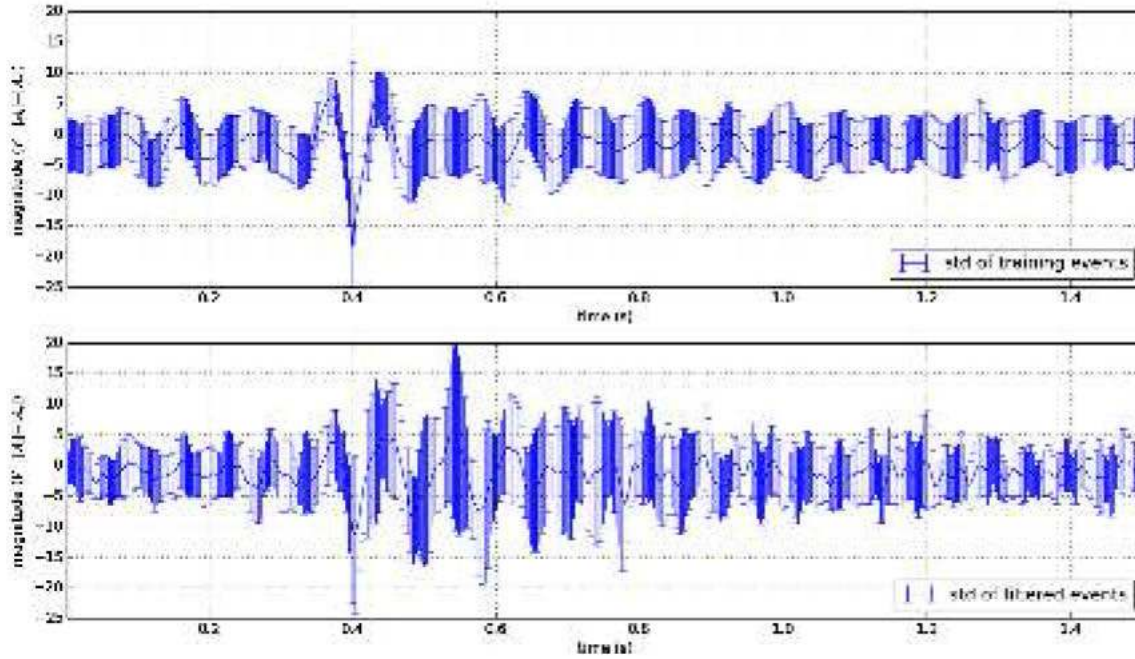


Fig. 20 71-Joint experiment: standard deviation of the training events (*above*) and the filtered events (*below*) after performing the kNN-based outlier removal step

Table 2 Computational time (in seconds) for the proposed approaches

	Event-based clustering		Joint-based clustering	
	6 Joints 36,947 events	71 Joints 45,818 events	6 Joints 36,947 events	71 Joints 45,818 events
Feature extraction	35.45	17.18	29.79	20.88
kNN	11.1 ( $k = 5000$ )	10.59 ( $k = 500$ )	9.17 ( $k = 5000$ )	10.47 ( $k = 500$ )
FFT	12.55	0.53	12.57	3.18
K-means	10.6 ( $K = 2$ )	55.61 ( $K = 5$ )	0.12 (pairwise map)	0.63 (pairwise map)
Total time	69.7	83.91	51.65	35.16
Instant warning	0.002	0.002	0.001	0.001

abnormal behaviour such as cracking and sensor failure. Instant warning can be also performed by defining membership rules based on patterns of event-based models. Regarding joint-based classification, joint-based models are updated so that any change in the behaviour of the structure over time can be determined. Therefore, the overall status of the structure can be estimated and any deviation from the normality can be detected.

Experiments were conducted in Python 2.7.6 64bits on Ubuntu-Linux, on a CORE i5 desktop PC with 4 GB of RAM memory. We used scikit-learn and scipy packages, available at <http://scikit-learn.org/stable/> [20, 27], respectively. Using mentioned software and hardware processing specifications, computational time for conducted experiments is presented in Table 2. This includes the required time for feature extraction, signal processing, distance calculation and the total time (in seconds). Having

constructed the clustering models, a random signal was investigated to calculate the required computational time for online damage detection. The result of this exercise is tabulated in the last row of Table 2. According to the obtained results, it can be seen that the required computational time for online warning is considerably short which demonstrates the fact that combination of off-line learning and online testing using the proposed approach can be applied for real-time damage identification.

## 5 Conclusions

This work presents a clustering based approach to group joints with similar behaviour on bridge and then detect abnormal or damaged joints. Two case studies involving vibration data monitored from the Sydney Harbour Bridge

were implemented to demonstrate the validity of proposed approach. On the basis of an iterative kNN-based outlier and noisy signal removal and a Fourier transform of resulting joint events, the *K*-means based clustering of both joint events and joint representatives were performed. The clustering results indicated similarity between joints located in different bridge locations and areas and helped to group those with similar behaviour. In addition, a damaged joint and a joint with a faulty sensor were also captured as a result of clustering, demonstrating the robustness of the approach when detecting a structural damage or a sensor problem. It is thus proved the validity of this approach for complementing existing damage detection techniques.

The proposed approach can also be easily deployed to provide a real-time health score of the structure. Once the behaviour models that best represent the structure are learnt from historical data, they can be used to check new events using distances to normality models obtained, namely clusters centroids and joints representatives. Thus, any potential damage in the structure can be detected. In order to improve the clustering result, further experiments can be conducted focusing on vehicle tracking to reduce variance among different events.

**Acknowledgments** The main author would like to thank National ICT Australia for the great support to this work and during his internship, as part of the work on his diploma thesis. The authors also wish to thank the Road and Maritime Services (RMS) in New South Wales for provision of the support and testing facilities for this research work.

## References

- Aggarwal CC, Yu PS (2013) Proximity-based outlier detection. In: *Outlier analysis*. Springer, New York, pp 101–133
- Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 46(3):175–185
- Brincker R, Zhang L, Andersen P (2000) Modal identification from ambient responses using frequency domain decomposition. In: *Proceedings of 18th international modal analysis conference—IMAC*, pp 625–630
- Cao G, Bouman C (2009) Covariance estimation for high dimensional data vectors using the sparse matrix transform. In: *Advances in neural information processing systems*, vol 21, pp 225–232
- Carden EP, Fanning P (2004) Vibration based condition monitoring: a review. *Struct Health Monit* 3(4):355–377
- Chang PC, Flatau A, Liu S (2003) Review paper: health monitoring of civil infrastructure. *Struct Health Monit* 2(3):257–267
- Cho S-J (2011) Structural health monitoring of cable-stayed bridge using wireless smart sensors. Ph. D. Dissertation, KAIST, Daejeon, Korea
- Collings D (2008) Lessons from historical bridge failures. *Proc Inst Civ Eng-Civ Eng* 161(6):20–27
- Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex fourier series. *Math Comput* 19(90):297–301
- Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
- Doebbling S, Farrar C, Prime M, Shevitz D (1996) Damage identification in structures and mechanical systems based on changes in their vibration characteristics: a detailed literature survey. Los Alamos National Laboratory, Rep No LA-13070-MS
- Farrar CR, Duffey TA, Doebbling SW, Nix DA (1999) A statistical pattern recognition paradigm for vibration-based structural health monitoring. *Struct Health Monit* 2000:764–773
- Farrar CR, Worden K (2007) An introduction to structural health monitoring. *Philos Trans R Soc A Math Phys Eng Sci* 365(1851):303–315
- Fourier J (1820) Mémoire sur le refroidissement séculaire du globe terrestre. *Ann Chim Phys* (2) 13:418–437
- Frangopol DM, Liu M (2007) Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. *Struct Infrastruct Eng* 3(1):29–41
- Fugate ML, Sohn H, Farrar CR (2000) Unsupervised learning methods for vibration-based damage detection. In: *Proceedings of 18th international modal analysis conference—IMAC*, p 18
- Gul M, Necati Catbas F (2009) Statistical pattern recognition for structural health monitoring using time series modeling: theory and experimental verifications. *Mech Syst Signal Process* 23(7):2192–2204
- Irwin PA, Stoyanoff S, Xie J, Hunter M (2005) Tacoma narrows 50 years later wind engineering investigations for parallel bridges. *Bridg Struct* 1(1):3–17
- Jafarkhani R, Masri SF (2011) Finite element model updating using evolutionary strategy for damage detection. *Comput Aided Civ Infrastruct Eng* 26(3):207–224
- Jones E, Oliphant T, Peterson P et al (2001) SciPy: open source scientific tools for Python. Retrieved from <http://www.scipy.org/>
- Lederman G, Wang Z, Bielak J, Noh H, Garrett J, Chen S et al (2014) Damage quantification and localization algorithms for indirect SHM of bridges. In: *Proceedings of the international conference on bridge maintenance, safety management*, Shanghai, China
- Lichtenstein AG (1993) The silver bridge collapse recounted. *J Perform Constr Facil* 7(4):249–261
- Maeck J, Peeters B, De Roeck G (2001) Damage identification on the z24 bridge using vibration monitoring. *Smart Mater Struct* 10(3):512
- Moore AW (1991) An introductory tutorial on kd-trees. Technical Report No. 209, Computer Laboratory, University of Cambridge
- Newland DE (2012) An introduction to random vibrations, spectral & wavelet analysis, 3rd edn. Dover Publications Inc, New York
- Nie P, Li B (2011) A cluster-based data aggregation architecture in WSN for structural health monitoring. In: *7th International wireless communications and mobile computing conference*, 2011. IWCMC 2011, pp 546–552
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Santos A, Figueiredo E, Costa J (2015) Clustering studies for damage detection in bridges: a comparison study. *Struct Health Monit* 2015
- Sibly P, Walker A, Stephenson R, Moisseiff L et al (1977) Structural accidents and their causes, I. *Proc Inst Civ Eng* 62:191–208
- Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR et al (2004) A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory, Los Alamos
- Toivola J, Prada MA, Hollméen J (2010) Novelty detection in projected spaces for structural health monitoring. In: *Advances in intelligent data analysis ix*. Springer, Berlin, pp 208–219



32. Tsai D-M, Lin C-T, Chen J-F (2003) The evaluation of normalized cross correlations for defect detection. *Pattern Recognit Lett* 24(15):2525–2535
33. Wei WW-S (1994) *Time series analysis*. Addison-Wesley, Reading
34. Wenzel H (2008) *Health monitoring of bridges*. Wiley, New York
35. Worden K, Manson G (2007) The application of machine learning to structural health monitoring. *Philos Trans R Soc A Math Phys Eng Sci* 365(1851):515–537
36. Wu W, Xiong H, Shekhar S (2004) Clustering and information retrieval, vol 11. Springer, Berlin
37. Yeung W, Smith J (2005) Damage detection in bridges using neural networks for pattern recognition of vibration signatures. *Eng Struct* 27(5):685–698
38. Yin A, Wang B, Hu X, Dai Z (2009) MHop-CL: a clustering protocol for bridge structure health monitoring system. In: *International symposium on computer network and multimedia technology*, 2009. CNMT 2009, pp 1–4
39. Yu L, Zhu J-H, Yu L-L (2013) Structural damage detection in a truss bridge model using fuzzy clustering and measured FRF data reduced by principal component projection. *Adv Struct Eng* 16(1):207–218