Mechanical Systems and Signal Processing 87 (2017) 384–400

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A spectral-based clustering for structural health monitoring of the Sydney Harbour Bridge

Mehrisadat Makki Alamdari*, Thierry Rakotoarivelo, Nguyen Lu Dang Khoa

CSIRO, Data61, 13 Garden Street, Eveleigh, NSW 2015, Australia

ARTICLE INFO

Keywords:
Structural Health Monitoring
Power Spectral Density
Spectral Moment
Clustering
Anomaly Detection
k-means−−

ABSTRACT

This paper presents the results of a large scale Structural Health Monitoring application on the Sydney Harbour Bridge in Australia. This bridge has many structural components, and our work focuses on a subset of 800 jack arches under the traffic lane 7. Our goal is to identify which of these jack arches (if any) respond differently to the traffic input, due to potential structural damages or instrumentation issues.

We propose a novel *non-model-based* method to achieve this objective, using a spectrum-driven feature based on the Spectral Moments (SMs) from measured responses from the jack arches. SMs contain information from the entire frequency range, thus subtle differences between the normal signals and distorted ones could be identified. Our method then applies a modified *k*-means−− clustering algorithm to these features, followed by a selection mechanism on the clustering results to identify jack arches with abnormal responses.

We performed an extensive evaluation of the proposed method using real data from the bridge. This evaluation included a *control* component, where the approach successfully detected jack arches with already known damage or issues. It also included a *test* component, which applied the method to a large set of nodes over a month of data to detect any potential anomaly. The detected anomalies turned out to have indeed system issues after further investigations.

1. Introduction

With time, civil structures such as bridges experience natural and human-induced damages, i.e. any change in the structure that adversely impacts its performance or safety (e.g. material deterioration, or boundary condition degradation). Many bridges have usually been monitored for such damages using visual inspections at set intervals (e.g. yearly) [1]. This approach has some major drawbacks, e.g. failures could happen between inspections, and incipient damages may go unnoticed during inspections. Thus monitoring the integrity of bridges in a systematic and continuous way is a challenge to the engineering community. Structural Health Monitoring (SHM) systems aim at implementing damage identification strategies for a structure. The information produced by SHM systems allows engineers and asset owners to improve the safety, serviceability, and operational cost of critical structures through their life cycle [2]. A SHM system combines multidisciplinary technologies and techniques, and usually involves the following main steps:

1. Data acquisition is the process of collecting measurements from a structure using an array of sensors. Several issues need to be

* Corresponding author.
E-mail addresses: MehrisadatMakki.Alamdari@data61.csiro.au (M.M. Alamdari), Thierry.Rakotoarivelo@data61.csiro.au (T. Rakotoarivelo), NguyenLuDang.Khoa@data61.csiro.au (N.L.D. Khoa).

http://dx.doi.org/10.1016/j.ymssp.2016.10.033
Received 24 May 2016; Received in revised form 25 October 2016; Accepted 30 October 2016

addressed at this step, such as the type or the location of the sensors [3], the sampling frequency, the data transmission and storage, or the energy cost of the systems [4].
2. Feature extraction is the process of selecting the attributes from the data, that are most sensitive to the damages and least sensitive to the operational and environmental variations. The research in this area has focused on time or frequency domain features [5,6], through signal processing techniques, such as data filtering, fusion, or pattern recognition [7–9].
3. Data analysis establishes a statistical framework for decision making to distinguish between damaged and healthy states of the structure. For this purpose, statistical pattern recognition methods using unsupervised or supervised learning algorithms can be implemented [10,11].

Most of the research in SHM has focused on developing and verifying algorithms in a controlled laboratory environment, or on small scale structures [12]. Indeed, deploying and using a SHM system on a large scale structure is prone to significant uncertainties [13]. Furthermore, many of these contributions have been based on short or medium term monitoring, and thus did not capture potential long term variabilities (e.g. sensor failure over time, temperature).

The objective of this paper is to present a novel approach to detect anomalies such as structural damages and system issues in continuously monitored large scale infrastructures. This approach was successfully evaluated on a real iconic bridge, i.e. the Sydney Harbour Bridge (SHB) in Australia. We instrumented that bridge with a unique SHM system, which has monitored more than 400 substructures (i.e. jack arches) since early 2014 [14]. The contribution of this paper is three-fold. First, it presents the recent additional data collection capabilities deployed on SHB. Second, it describes in detail a novel approach to detect structural damages and system issues on this large scale SHM system. This approach uses a novel clustering-based scheme to correlate the location of substructures with their behaviours. Finally, it demonstrates through several case studies the performance and benefits of this approach. While our SHM deployment has collected data for over a year, studying the effect of environmental variability on the system is outside the scope of this paper, and will be presented in future works.

A previous description of the SHB system [15] presented a whole high level overview for infrastructure managers. It briefly reviewed each system component from the on-site nodes to the remote web-based interface for managers and engineers. In contrast, the deployment update in this paper focuses on the data acquisition component only, and provides a technical description of its new functions, which enabled the collection of the data used in this paper.

The remainder of this paper is organized as follows. Section 2 presents some related research works. Then Section 3 provides an update of the new data acquisition functions of the SHB system. Section 4 introduces the theoretical background for the proposed approach to detect structural damages and system issues. Section 5 then presents the results of the successful applications of the proposed method to several case studies. Finally, Section 6 discusses some limits and potential improvements, which are followed by concluding remarks in Section 7.

2. Related works

There have been many contributions related to SHM applications on bridge structures. Many of them studied modal parameters (e.g. natural frequencies, mode shapes and modal damping) for damage identification in bridges. One of the first study proposing such a technique identified some progressive cracks in a three-span highway bridge during cyclic fatigue loading [16]. Other studies focused on the changes in the modal parameters for damage detection [17,18]. They concluded that the identification of higher modes was required to reliably identify damages using modal parameters. Thus, the first few lower modes might not provide useful information about any existing damages [18].

The Los Alamos National Laboratory (LANL) conducted large scale studies on the Interstate 40 (I-40) and the Alamosa Canyon bridges in New Mexico [19,20]. They extracted modal parameters using ambient and forced vibration testing, and reported some of the experienced difficulties. First, the use of eigen-parameters did not provide information about damages, which were located in the nodal points. Second, the identification of the modal parameters was difficult when the modes were not well-separated, which is often the case for bridges. These studies concluded that eigen-parameters were highly dependent on uncertainties in the system, e.g. environmental conditions, traffic loads, excitation source, the identification algorithm and the measurement precision. To account for such variabilities, they suggested the collection of data over a year, in various weather conditions, and with different traffic conditions.

While most of these previous works were based on short-term monitoring, some recent studies reported on long-term SHM applied to bridges with in-service traffic loads [21,22]. They deployed different sensors to monitor various variables and structural responses, e.g. temperature, humidity, wind, corrosion level and input loading using Weight-in-Motion stations. These contributions emphasized the necessity for statistical data analysis to quantify and account for the uncertainties over long time periods [22]. However, they focused on either small structures or had small number of instrumented locations.

SHM systems often only have data from the healthy states of structures, Thus many contributions proposed damage detection methods based on unsupervised or one-class approaches. For example, Worden et al. used a Mahalanobis distance to find data anomalies, which are likely to be structural damages [23]. Other authors proposed clustering approaches for damage detection, which deal with operational and environmental effects, such as Gaussian Mixture Models (GMM), Support Vector Clustering (SVC) and Self-Organizing Maps (SOM) algorithms [24]. Law et al. used a fuzzy c-means clustering approach for structural damage detection [25]. Other methods focused on grouping information coming from sensors distributed throughout the structure. Yin et al. proposed a clustering-based routing protocol to group similar nodes in each span of the bridge [26]. Diez et al. discussed the use of k-means to group bridge substructures with similar behaviour and then detect anomalies [27].

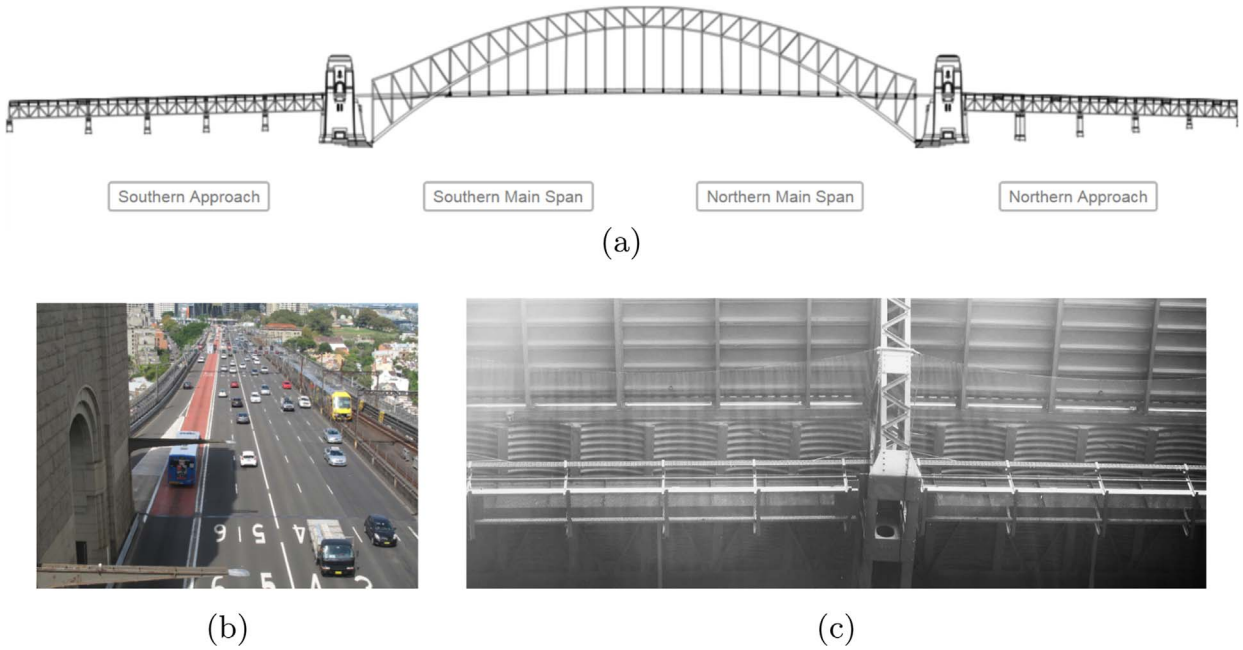


Fig. 1. (a) Schema of the SHB with its high-level substructures; (b) the monitored Traffic Lane 7 in red, with only buses and taxis, and (c) a series of 9 jack arches under Lane 7.

Other works incorporated finite element analysis of a bridge with measured SHM data to better understand and verify its behaviour [28,29]. These contributions mainly involved parametric studies and model updating of the bridge using ambient vibration testing to validate the established model.

More recently, some researchers proposed alternative approaches based on indirect measurements [30]. They hypothesized that changes in a bridge structure caused variation in the vehicle-bridge interaction. Thus, they determined the state of a bridge using advanced signal processing on data collected from sensors mounted on vehicles driving over the bridge. Finally, other SHM research works proposed the use of advanced sensing systems, wireless transmission, or embedded sensors in a bridge during its construction [31].

3. A large scale SHM on the sydney harbour bridge

3.1. Bridge description

The Sydney Harbour Bridge (SHB) is the main connection between the city's northern suburbs and its center. It supports an eight-lane road and two railway lines, and carries a typical daily traffic of many thousands of vehicles. SHB is a steel through arch bridge operated by the Roads and Maritime Services (RMS) of New South Wales, Australia. Fig. 1a presents a schema of SHB, which can be divided in high level substructures, i.e. the Southern and Northern Approach and the Southern and Northern Main Span.

Traffic lane 7 is dedicated to buses and taxis on the eastern side of the bridge, see Fig. 1b. This lane consists of an asphalt road surface on a concrete deck supported by concrete and steel jack arches. There are 800 jack arches over a total distance of 1.2 km, as illustrated in Fig. 1c. The jack arches are substructures, which are difficult to access and thus are inspected typically at two yearly intervals according to standard visual inspection practices.

Our team at Data61 developed and deployed a SHM system to continuously monitor each of the 800 jack arches and alert the asset manager and inspector when it detects any issue. An inspection can then be scheduled, and preventative maintenance carried out without disrupting road users. This comprehensive SHM system includes sensing, data transport, storage, and analysis and a secure web-based user dashboard. This dashboard provides the bridge manager and inspector relevant information about the structural health of the jack arch components, and recommendations about any required actions (e.g. inspection advised for component X, attend to detached sensor for component Y) [15].

3.2. Data acquisition

Each jack arch was instrumented with 3 low cost tri-axial MEMS accelerometers, as shown in Fig. 2a. One is glued at the bottom of the jack arch (i.e. sensor 2), and the 2 others (i.e. sensor 1 and 3) are glued on each side. The maximum sampling rate when a single sensor is in operation is 1500 Hz. This drops to 250 Hz when the 3 sensors are operated simultaneously. Continuously operating all of the 2400 sensors would produce approximately 1TB of data per day. Transmitting that amount of data for remote

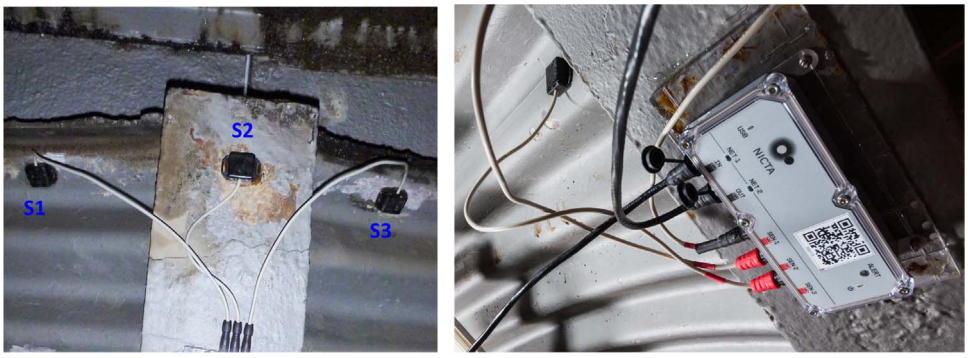


Fig. 2. Illustration of (a) the 3 tri-axial accelerometers glued underneath an individual jack arch; (b) a Data61 collection/processing node.

processing would be technically challenging and expensive. Therefore, a weatherproof small computing node was also glued to each jack arch and connected to the sensors, as illustrated in Fig. 2b. Each node is a programmable Linux-based embedded computer, which runs data filtering, processing and analysis software according to a configurable schedule. This software architecture allows new analytical techniques and operation schedules to be developed remotely and then deployed to each jack arch over a data network [15].

3.3. Data collection

The daily data collection schedule consists typically of 2 concurrent phases. Within the *production* phase, the sensor data are processed directly on the node to produce short structural health information, which is transmitted with low latency to the web dashboard for the bridge manager and inspector. Within the *research* phase, fixed amount of sensor data are stored locally on the node, then forwarded to a central larger storage at a later time, where they are archived and made available for researchers. These systems allow 2 types of data collection, i.e. *event-based* and *continuous*.

3.3.1. Event-based data collection

The event-based data collection is used in both the production and research phases. It captures a 2-second window of data that overlaps with the period when a vehicle (e.g. a bus) drives over the jack arch, from 4:00 to 23:59 everyday. Within this process, a triggering algorithm continuously monitors the sensor data. It computes the average acceleration value when no vehicle drives over the jack arch, this is the acceleration at rest A_R. When a vehicle drives over the jack arch, the difference in the measured acceleration compared to A_R triggers the collection of 2 s of data (i.e. 0.4 s before the trigger and 1.6 s after the trigger) at 250 Hz from sensors 1, 2, and 3. After that, the collection process stops for a 2-min period, then starts again. Thus a maximum of 20 event-based data sets could be collected for a group of 10 jack arches for a period of 4 min. In practice when there is no technical issue in the data collection, the actual number of events collected per day per jack arch is about 200, i.e. 73k events in a year.

The event-based data collection was the first implemented method. Thus the oldest available data sets from early 2014 were collected using this method. The structural health information produced in the *production* phase are computed only from these event-based data.

3.3.2. Continuous data collection

The continuous data collection was implemented later (early 2015) and is used currently only in the research phase. For each jack arch, it captures a non-stop 10-min window of data at 5 specific times during the day, i.e. at 1500 Hz for Sensor 2 only at 00:15, then at 250 Hz for all 3 sensors at 00:30, 6:00, 12:00, and 18:00. Such large amount of data are stored in local storage and manually regularly collected at later time. They provide *raw* information to researchers for future analysis and algorithm developments. An optional process may be performed on this data to remove any events, which may be detected by the previous triggering method. This filtering isolates the data related to the response of the bridge to the *ambient* background vibrations. When there is no technical problem with the data collection, each jack arch produces 5 set of continuous data per day, i.e. 1.825k sets in a year.

4. Proposed approach

One of the key difference between the SHB system and other proposed systems, is that it has collected data from a large number of substructures (>400 jack arches) over a large period of time (i.e. since early 2014). This data set allows the study of novel damage detection methods, which harvest its unique spatial and temporal characteristic (i.e. many locations over long period).

This paper proposes an approach based on the following hypothesis: within the population of substructures composing a large bridge, the ones which are structurally similar and located close to each other *should* respond similarly to the same excitation. In other words, nearby jack arches should produce similar responses for a given time window. As presented in Section 2, previous SHM

related works proposed many cluster- based methods following a similar assumption. However, they only had access to data from a limited set of nodes over limited time period compared to the SHB system.

The proposed approach follows the following steps, which are further detailed in the remainder of this section:

- a structurally meaningful feature is extracted from the measured acceleration for each jack arch for many time windows,
- a modified *k*-means−− clustering algorithm is applied to this feature data to identify groups of similar substructures and potential anomalies,
- a multi-indices criterion is used to select the best grouping outcome,
- based on the previous assumption, near-by substructures should have similar behaviours and thus should belong to the same cluster groups. Therefore, any substructure which is identifid as an outlier or which belongs to a one-member group, is then marked as an anomaly.

4.1. Feature extraction

We propose to use the Spectral Moments (SMs) of the measured acceleration responses as the feature in our approach. The SM of a signal is computed from its Power Spectral Density (PSD). For a stationary random process, the PSD contains some major characteristics of the system (e.g. eigenvalues and eigenvectors) that can be extracted. Furthermore, it is capable of characterizing both spatial and temporal signals. It quantifies the energy contents of the signal within a frequency bandwidth of interest and hence, it does not suffer lack of information using a limited number of eigen- frequencies.

Other SHM applications have used PSD-based methods to identify damages in structures [32]. Liberatore and Carman established the power spectral energy near the resonance frequencies using the output PSD. They then compared the current energy with a benchmark to detect and localize damages in a simply-supported beam [33]. A similar study used spectral strain energy as a sensitive feature to localize damages in a plate-like structure [34]. In another work, the PSD of the dynamic strain response was first generated and the ratio change between the peak values of the PSD was used as damage indicator [35]. Fang and Perera [36] introduced a new concept of PSD-derived power mode shapes, and proposed a damage indicative feature using the curvature of these power mode shapes. Masciotta et al. [37] studied the use of the second order spectral properties of the nodal response processes for damage identification. Moreover, Zheng et al. [38] proposed a method based on sensitivity analysis of the PSD in conjunction with finite element model updating to identify single and multiple damages in a frame structure. While these previous studies used the PSD in SHM context, in this paper the PSD is used as a computational component of the SMs. To the best of our knowledge no other previous contributions have studied the use of SMs in SHM systems.

The SM of a random stationary signal provides some important information about its statistical properties. It explicitly depends on the frequency content and the entire frequency range of a signal, which make it suitable for SHM applications. Although SM has non-dimensional nature, it captures information from the entire spectra and hence it can distinguish any subtle differences between normal and distorted signals. Thus, it has been used in areas, such as speech analysis, meteorology, or fatigue life prediction [39,40]. The high sensitivity of SM to identify damages has been reported in our previously published work [32]. Numerical simulations on an Euler-Bernoulli beam demonstrated that the proposed feature is sensitive enough to identify a very small damage. Indeed it could reliably provide an indication of damage which resulted in 0.06%, 2.05%, 2.17% and 0.38%, reduction in the first four modes of the structure, respectively. This investigation was carried out in the presence of 5% randomly generated noise which demonstrated the robustness of the method to identify small damages under 5% uncertainty. Thus using all frequency spectra to construct a SM-based feature allows the detection of subtle difference between a normal signal and a distorted one. This paper proposes to use SM as a damage indicative feature for SHM application. This subsection provides an analytical description of this SM-based feature for a random process.

For a structure under some ambient excitation (e.g. wind or traffic loadings), it has been demonstrated that the structural system follows a white Gaussian stochastic process [41]. For such a system, the response process can be characterized by a correlation function $R_{xx}(\tau)$ in the time domain as,

$$R_{xx}(\tau) = E[x(t)x(t+\tau)] \tag{1}$$

where x is the measured time response, $E[]$ is the probabilistic expected value operator, and τ refers to the lag operator. This $R_{xx}(\tau)$ function defines how a signal is correlated with itself, with a time separation τ.

The frequency characterization of such random stationary process can be computed using the PSD function which is calculated by taking the Fourier transform of the autocorrelation function as,

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)\mathrm{e}^{-iw\tau}\mathrm{d}\tau \tag{2}$$

where $S_{xx}(\omega)$ is the PSD of the response at frequency ω. For a given PSD, the nth-order SM can then be computed as,

$$\lambda_x^n = \int_{-\infty}^{\infty} |\omega|^n S_{xx}(\omega)\mathrm{d}\omega \tag{3}$$

where n is the order of SM. Finally, for a discretized signal x, the nth-order SM $\lambda_x{}^n$ can be obtained using,

$$\lambda_x^n = \frac{2}{N^{n+1}} \sum_{0}^{\lfloor N/2 \rfloor} S_{xx}(j)\left(\frac{j}{\Delta t}\right)^n \quad j \in [1{:}\ N/2] \tag{4}$$

where S_{xx} is the discrete N-point spectral density obtained by the discrete Fourier transform of the autocorrelation function and Δt is the sampling period.

The SMs can be viewed as a way to use statistics to analyse the power spectrum. For instance, the most common SM is λ^0, which determines the variance of a zero mean process PSD. The different moments have different statistical interpretations. The zero-order SM is basically the average of the values in the frequency interval and is proportional to the mean energy of the signal in that interval. Basically, the higher the order of the spectral moment, the more weight is given to higher frequencies. If the quality of the signal and the level of response at high-frequency range is acceptable, it should improve the sensitivity of the spectral moment to identify damage since it is well-established that high- frequency spectral components are more informative compared to low-frequency components to highlight the local structural changes as a result of damage occurrence. However, since there might not be much energy in the high frequency components it might have an adverse effect if the order exceeds certain limit. Previous research studies have concluded that SMs with orders 1–4 provide useful information about the system, whereas higher order moments usually do not provide further information [42].

In the SHB system, the studied response signals are the tri-axial (x, y, z) accelerations from a given sensor, which are normalized with respect to their mean and standard deviation as,

$$\overline{x}(t) = \frac{x(t) - \mu}{\sigma} \tag{5}$$

where μ and σ are the mean and standard deviation of the acceleration, respectively. For the first, second and the third orders, the SMs are calculated separately for each directions (x, y, z) using Eq. (4), and a resulting 3-dimensional vector is constructed as,

$$\Lambda^n = [\lambda_x^n \lambda_y^n \lambda_z^n] \tag{6}$$

The next steps of our proposed approach use this Λ^n vector as the indicative feature to identify any structural or system issues with the SHB system.

4.2. Identification of groups and anomalies

The objective of this step is to group substructures (jack arches) and detect anomalies using the feature from Eq. (6). Based on the previous assumption, the group membership of a given jack arch should strongly correlate with its physical and/or structural location along the SHB structure. The contrary would then suggest a structural or system issue. This step proposes a modified k-means−− clustering algorithm to perform this grouping process, and a selection mechanism to then detect potential anomalies.

4.2.1. k-means and k-means++ clustering

Data clustering has been studied extensively in data mining research [43]. Clustering techniques try to group data objects into different clusters so that objects in the same cluster are more similar to each other than to those in other clusters. k-means clustering is a popular clustering technique. It partitions data into k clusters $C = \{C_1, \ldots, C_k\}$ so that the within-cluster sum of squares are minimized, i.e.

$$\underset{C}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in C_i} \| x - \mu_i \|^2 \tag{7}$$

where μ_i is the center of cluster i (mean of data instance in C_i). This optimization problem is typically addressed using an iterative technique as described in Algorithm 1. Convergence is reached when the assignments no longer change.

Algorithm 1. k-means clustering. .

Input: Matrix X of data points, number of clusters k
Output: Cluster ID for each data point in X

1: **Initialization**: randomly select k centers
2: **Assignment**: assign each data point to the nearest center
3: **Update**: recompute the center for each cluster
4: **Iterate**: repeat steps 2 and 3 until the algorithm converges, i.e. no further changes in the data point assignments

The k-means algorithm is sensitive to the center initialization and the algorithm may converge to a sub-optimal solution. Arthur and Vassilvitskii [44] proposed k-means++, a variant of the algorithm, which intelligently initializes the cluster centers before applying the standard k-means procedure. Instead of choosing all the centers uniformly at random, it only selects the first center uniformly at random from all data instances. Then each subsequent cluster center is chosen from the remaining data instances with probability proportional to its squared distance to the closest existing center.

4.2.2. k-means−− clustering

Besides the initialization sensitivity problem, *k*-means clustering is also very sensitive to outliers since they can skew the right cluster centers. Chawla and Gionis [45] proposed a method, named *k*-means−, as a unified approach for simultaneously clustering and detecting outliers (e.g. potential anomalies) in data. These anomalies are then isolated before the iterative cluster update process. Thus in this method potential anomalies are explicitly detected, instead of being a by-product results of the clustering technique (e.g. anomalies are likely located in significantly small clusters).

Our proposed approach adds the following modification to the original *k*-means−− method. After convergence is reached, existing clusters with only one member are removed from the final cluster set and their data points are instead added to the final anomaly set. This additional step removes biases for the selection of the best cluster outcomes, which is described in the next subsection. Our modified *k*-means− is described in Algorithm 2. Based on the iterative framework of *k*-means, it first finds o anomalies in the data before grouping the remaining data points into k clusters. Thus after the assignment step, the o data points that are furthest from their closet centers are isolated and are not used to recompute the centers in the update step.

Algorithm 2. A modified *k*-means−− clustering.

Input: Matrix of X data points, number of clusters k, number of anomalies o
Output: o anomalies in L, cluster ID for each data point in $X - L$

1: **Initialization**: using *k*-means++ to find k initial centers
2: **Assignment**: assign each data point to the nearest center
3: **Anomaly isolation**: find top o anomalies (furthest from their cluster centers)
4: **Update**: recompute the center for each cluster (excluding the found anomalies)
5: **Iterate**: repeat steps 2–4 until the algorithm converges, i.e. no further changes in the data point assignments
6: **Finalise**: convert clusters with only one data point to anomalies

4.2.3. Selection of the clustering outcome

Given the random selection of the first center in our modified *k*-means− method, different runs of the algorithm on a data set will result in different clustering outcomes. Similarly, the multiple choice for the input parameters k and o will also lead to many different clustering results. Using a high number of replications (e.g. 50) mitigates the first issue. To address the second issues, our approach limits the choice for k to a value guided by the characteristics of the structure (e.g. the number of top-level structural components of a bridge, such as a span), while o may remain arbitrarily low (e.g less than 5). We then propose the following method to select the most informative clustering outcomes in terms of damage detection.

For each pair of input parameter (k, o) and its corresponding clustering result from a 50-replication run, we compute the values of 3 clustering performance indices, namely the Silhouette [46], the Davies- Bouldin [47], and the Dunn [48] index. These indices report on different characteristics of the computed cluster groups. The Silhouette method computes the averaged dissimilarity of each data point against its assigned cluster. The Silhouette index then compares these dissimilarities against the dissimilarity of the data points with their nearest neighbouring clusters. The Davies-Bouldin index measures the ratio between the similarities within a cluster and the differences between clusters, i.e. how *tight* and *distinct* the clusters are. Finally the Dunn index measures the ratio between the most similar points across different clusters and the less similar points within clusters. Appendix A provides the formal definitions of these indices.

For each of these computed indices, we select the cases with the extremum values, i.e. maximum value for Silhouette and Dunn; and minimum value for Davies-Bouldin. Within these selected cases, we take the logical intersection of all identified anomalies as the final set of anomalies, i.e. jack arches with potential structural damage or system issue. Since these 3 types of index measure different attributes of the produced cluster groups, using their intersection provides a potentially less biased and more accurate anomaly detection. The following Section 5.2 provides a practical example of this selection mechanism.

5. Case studies and results

This section presents 5 case studies, which illustrate the performance of our proposed approach. These case studies are organized in a *control* and a *test* categories. In this section a jack arch is also referred to as a *node*. The locations of all nodes in all 5 case studies

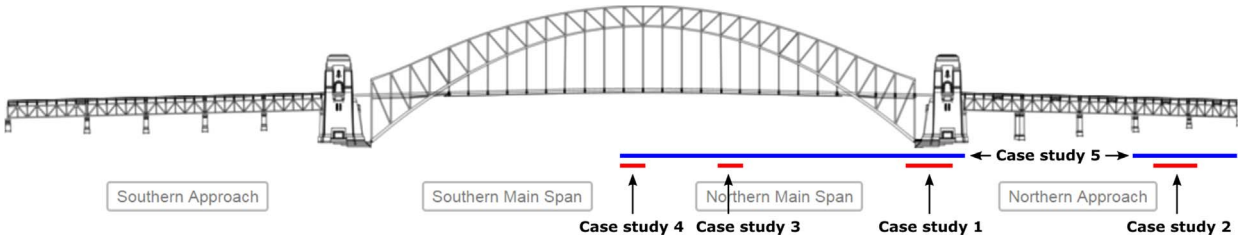


Fig. 3. Locations of nodes in 5 case studies on the SHB.

are described in Fig. 3. For each case study, a solid line indicates the location of the associated nodes. However, not all the nodes in an area covered by a line were monitored at the time of data collection.

The first *control* category (i.e. case studies 1–4) focuses on the following question: “Can our proposed approach identify an existing jack arch with known and verified issues? ” Each case study involves a small group of nodes with one having a known abnormal behaviour. Such a behaviour can be either due to a change in the structure (structural anomaly/damage) or to a fault in the sensing hardware or software (system anomaly). For these specific nodes, the exact cause is known a-priori and has been verified through inspections and system audits. Without this prior knowledge, the goal for the following case studies is to identify the single node with an abnormal behaviour using the proposed approach with k=1 and o=1.

The second *test* category (i.e. case study 5) focuses on the following question: “Can our proposed approach identify previously unknown jack arches which on further investigation do have issues? ” This example involves a large set of 85 nodes over a long period of 22 days in July 2015. The goal is to use the proposed approach to group nodes with similar behaviours. Following our working assumption (Section 4), any identified anomaly is then investigated further to find out why the related node is not within a group including its nearest physically located neighbours. In this case study, the number of nodes with abnormal behaviour are not known beforehand, thus it provides a real-life example of our proposed approach.

For each of these case studies, the SM features were computed on data sets with identical time window length. Thus the uncertainties in SM values within a single case study are comparable for each jack arch. It is noted that for case study 1 event data were used whereas ambient continuous data were used for the remaining case studies. The reason was that the data with known damage in case study 1 were collected in 2012, when continuous data collection had not been implemented in the system. The details of both event-based and continuous data collection procedures were described in Section 3.3.

5.1. Control case studies

5.1.1. Case study 1: damaged structure

This case study involved 6 jack arches, i.e. nodes 41–46, located close to the north pylon and north main span (see Fig. 3). An expansion joint separates nodes 41–43 from nodes 44–46, as schematically illustrated in Fig. 4. In 2012, a known crack existed on node 44 while all the others were in good condition. The damaged node 44 was repaired in February 2013. Fig. 5a shows a comparison between the normalized acceleration response collected from the healthy node 41 and the damaged node 44. The acceleration responses in the healthy and damaged jack arches are visually different.

In the first case study, the input excitation is traffic-induced vibration which is a narrow-band excitation and does not meet the Gaussian stationary input assumption; however, to address non-stationarity in the first case study a large number of event-data in a fairly long duration was considered. For each jack arch, 4000 sets of event data (Section 3.3) were randomly selected from August to October 2012 (i.e. when node 44 was damaged). For each set, the first, second, and third order SM, in the x, y and z directions were calculated according to Eq. (4). The average SM from the 4000 events was used as the feature vector following Eq. (6). Our modified k-means– method was then applied to the acquired feature vectors with the number of cluster and anomaly both set to 1. As a result, all the healthy substructures were grouped in a single cluster and the damaged substructure was identified as an anomaly. Figs. 5b to d shows the 3d scatterplot results using orders 1, 2 and 3, respectively. As illustrated, our proposed approach can successfully identify the jack arch with structural issue.

5.1.2. Case study 2: sensor with low signal to noise ratio

This second case study focused on 13 neighbouring jack arches located on the north approach (see Fig. 3). In October 2014, a remote software audit showed some abnormal behaviour for node 184. Fig. 6a compares the continuous acceleration response in z-direction obtained from the sensor 2 of node 184 and node 161 (i.e. a jack arch with no sensor issue). Further investigations revealed that the signal to noise ratio for sensor 2 of node 184 was quite low compared to the adjacent nodes. The noise level in MEMS sensors has both a mechanical and electrical part (e.g. vibration of the micro-machined elements, or heat from the integrated circuit) [49]. Its abnormal increase may be due to a failure from these sources. A thorough study of this failure would not be cost effective, due to the sensor's low cost and the required electronic expertise.

Fig. 6b shows the result of our proposed approach applied to this case, using 1 set of 10 min Continuous data (Section 3.3) for each of the 13 jack arches. The plot of the second order SM clearly identified the node with the sensor issue as an anomaly. The plots for the first and third order SMs were similar and thus not presented here. Replacing the faulty sensor 2 of node 184 was added to the list of maintenance tasks for the overall SHM system. However due to the physical difficulty in accessing that node, its sensor has not been replaced yet. In the meantime, its data have not been used for damage detection.

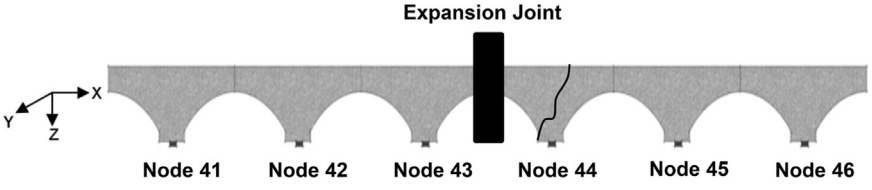


Fig. 4. Schematic of the evaluated nodes/jack arches.

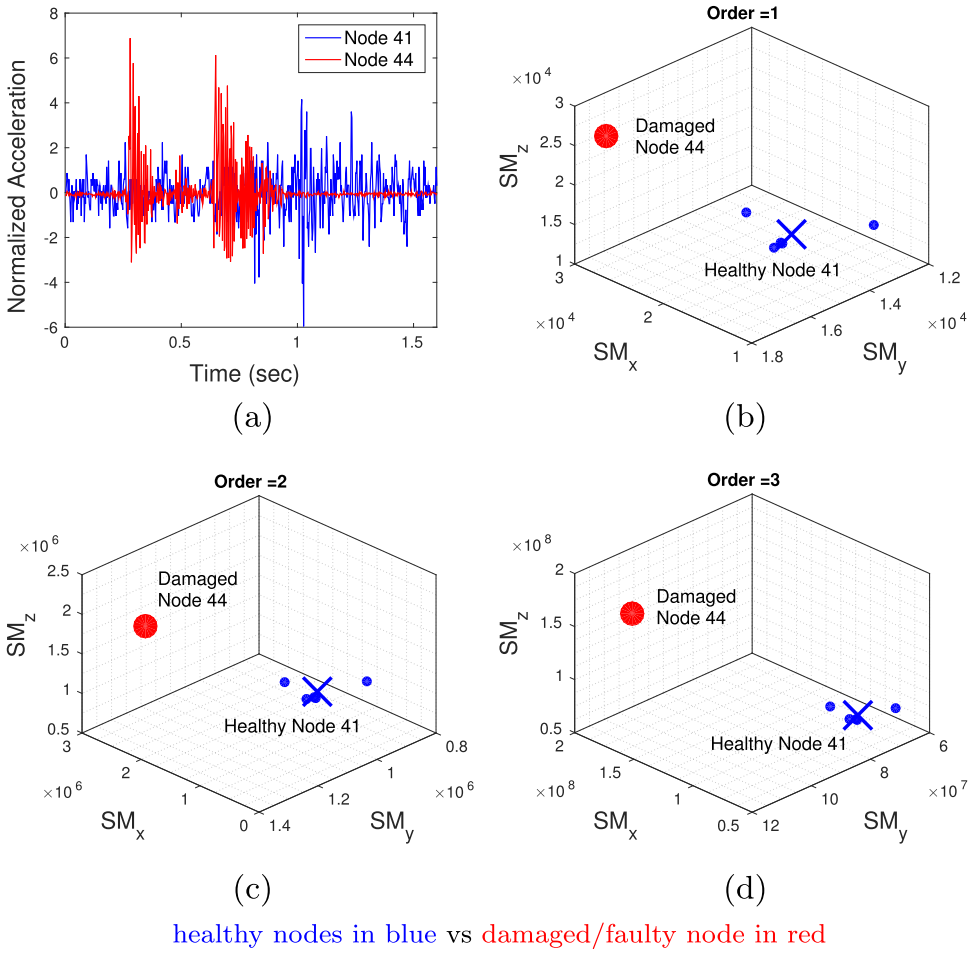


Fig. 5. (a) Comparison of acceleration response in z-direction between healthy node 41 and damaged node 44; and the anomaly detection results with k=1 and o=1, (b) using the first order SM, (c) using the second order SM, (d) using the third order SM.

5.1.3. Case study 3: sensor with communication error

This next example studied 9 nearby jack arches from the northern main span (see Fig. 3), using 1 set of 10 min Continuous data for each of the 9 jack arches. In January 2016, another remote software audit reported a high number of communication errors on the digital bus connecting to the sensor 2 of node 513. As a result, the continuous acceleration response in the z-direction for that

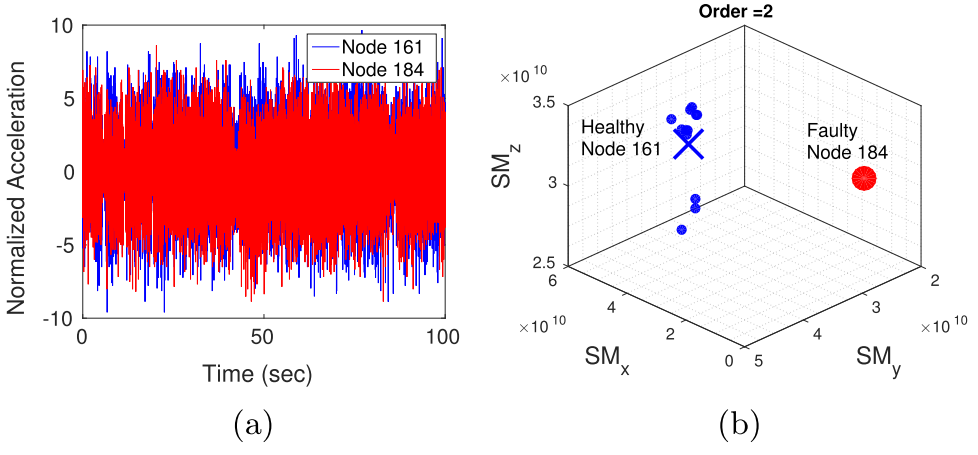


Fig. 6. (a) Comparison of acceleration response in z-direction between node 184 (with sensor issue) and node 161 (with no issue); and (b) the identification of node 184 as an anomaly, using the second order SM, k=1 and o=1.

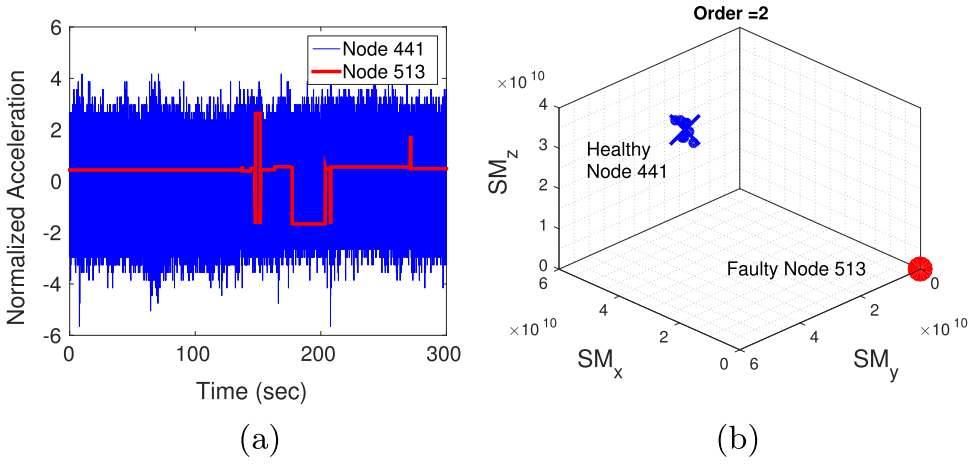


Fig. 7. (a) Comparison of acceleration response in z-direction between node 513 (with communication errors) and node 441 (with no issue); and (b) the identification of node 513 as an anomaly, using the second order SM, k=1 and o=1.

node was visually different compared to another node (node 441) as illustrated on Fig. 7a. This error was clearly identified, as there were numerous software driver messages stating I2C (Inter-Integrated Circuit) bus errors, which were printed in the kernel log file of the operating systems running on node 513. Similar to node 184, further work in the cause of these I2C errors would not be cost effective in the context of our SHB deployment. Fig. 7b shows the plot for the second order SM using our proposed approach, which identified node 513 as an anomaly. As for the previous example, the plots for the first and third order SMs were similar and thus not presented here. Hardware replacement for node 513 was subsequently added to the maintenance tasks.

5.1.4. Case study 4: sensor with periodic spikes

This fourth case study involves 10 close by jack arches in the southern main span (see Fig. 3), which were recently instrumented. In January 2016, the continuous acceleration data for node 506 showed numerous periodic spikes, as illustrated in Fig. 13b. There are potential causes to these spikes, such as clock-speed timing issue in read/write tasks[1] or fast thermal shifts in the electronics.[2] For this specific node, the problem was limited to sensor 2, as sensor 1 and 3 did not show any spikes. This sensor has been scheduled for replacement. As with the previous case studies, Fig. 8b shows that our proposed approach correctly identified node 506 as having some issues. This analysis used 1 set of 10 min Continuous data for each of the 10 jack arches.

5.1.5. Comparisons with a typical anomaly detection method

Distance-based approach is a popular technique to find anomalies in data [50]. It assumes that a data point is an anomaly if it is far away from other data points in the Euclidean space. For every data points, the average distances to k nearest neighbours are computed as an anomaly score (knn score), and the points with highest scores are marked as anomalies. In this subsection, we applied that technique to the aforementioned case studies using the same feature proposed in Eq. (6). The results in Fig. 9 shows that our modified k-means– method identified the same anomalies as a typical k-nearest- neighbour method (anomalies had the highest knn scores in Fig. 9a to d). However, the latter method does not allow the grouping of jack arches into clusters, which we would further correleate with their physical or structural locations.

5.2. Test case study

As opposed to the previous case studies, this final one aimed at identifying jack arches with abnormal responses, which were not a-priori known to have any issues. It involved a large set of 85 nodes located on 5 different structural locations, i.e. Span 6 and Span 7 which are part of the North Approach, the North Pylon, and the Spans N14-28 and N0-14 which are part of the North Main Span (see Fig. 3). For each node, the continuous acceleration data for 22 days in July 2015 were filtered to isolate 1 min of ambient responses (Section 3.3). This provided 22 sets of 1 min Continuous data at 1500 Hz per jack arch. We computed the SM for each set, and finally averaged them over the 22 days for each of the 85 jack arches.

The 85-node set in this case study do not include node 44 (case study 1) as it is structurally different from all the other jack arches, since its 2013 repair. This set does not include nodes from the southern main span (case study 4), as that area of the bridge has been instrumented last and data over a long time frame were not yet available at the time of our study. Node 184 (case study 2) is part of this test case, and was not replaced since the sensor problem detected by case study 2 on data from October 2014. Finally, node 513 (case study 3) is also part of this test case, but the data used here is from July 2015, i.e. prior to the January 2016 problem detected by case study 3.

[1] http://electronics.stackexchange.com/questions/109823
[2] http://www.wilcoxon.com/knowdesk/TN14_Troubleshooting_Accel_Installation.pdf

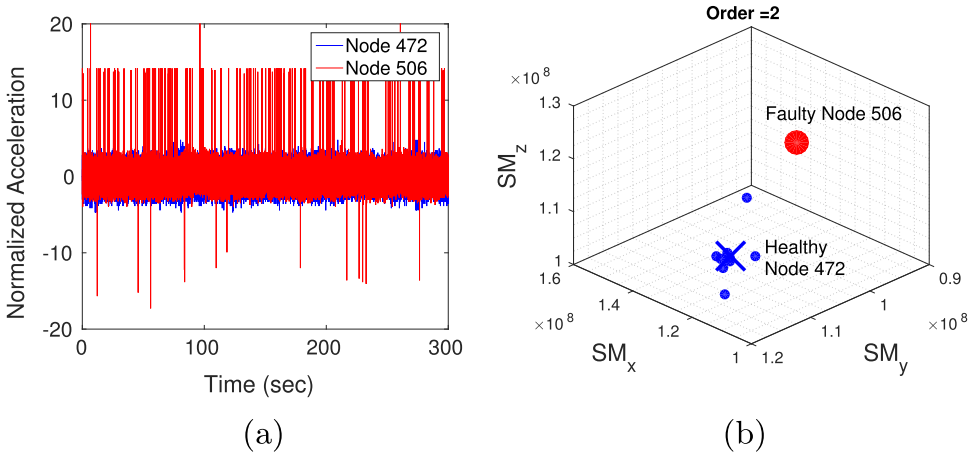


Fig. 8. (a) Comparison of acceleration response in z-direction between node 506 (with periodic spikes) and node 472 (with no issue); and (b) the identification of node 506 as an anomaly, using the second order SM, k=1 and o=1.

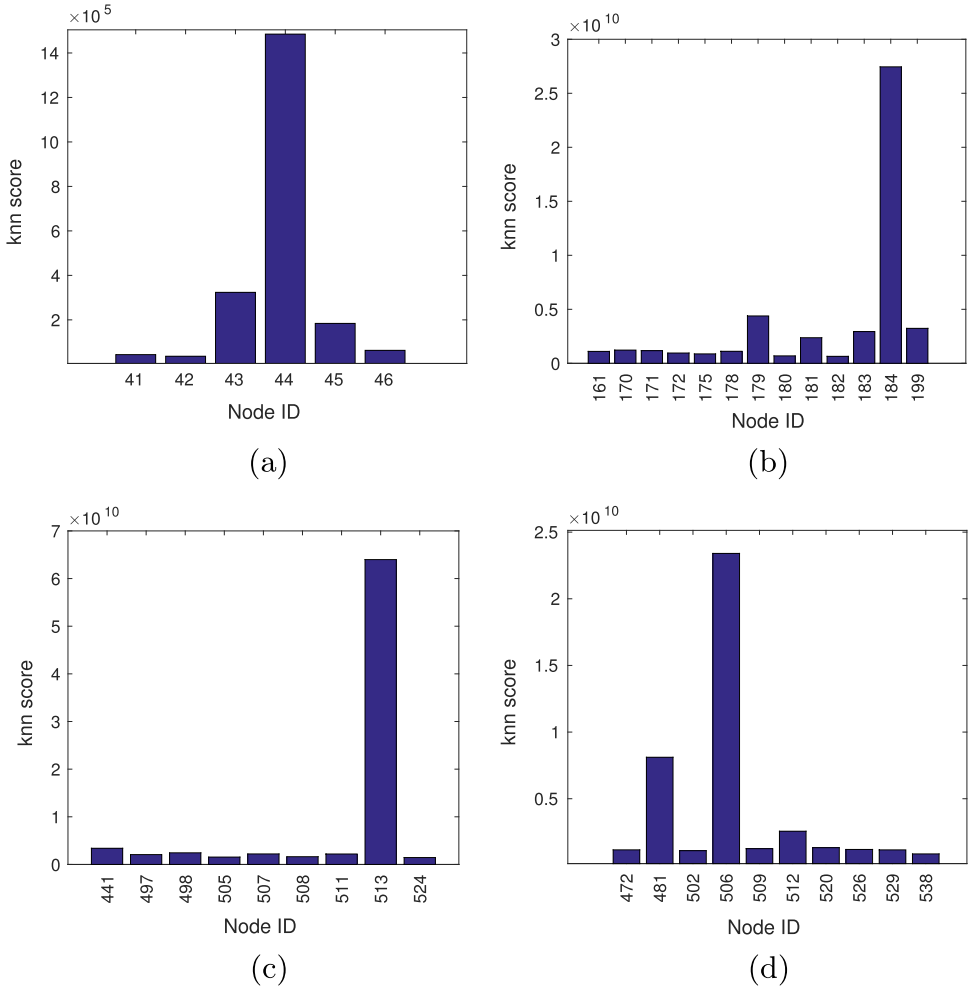


Fig. 9. Anomalies identified by a typical distance-based anomaly detection technique, for each of the previous case studies (a) 1 to (d) 4.

5.2.1. Anomaly identification

The complete method from Section 4 was applied to the studied data set using the second order SM. The number of clusters (k) varied from 2 to 6 as the 85 nodes were spread across 5 different structural locations, while the number of anomaly (o) varied from 0 to 4. This process was repeated 10 times. The first and third order SMs produced similar results and were not included below.

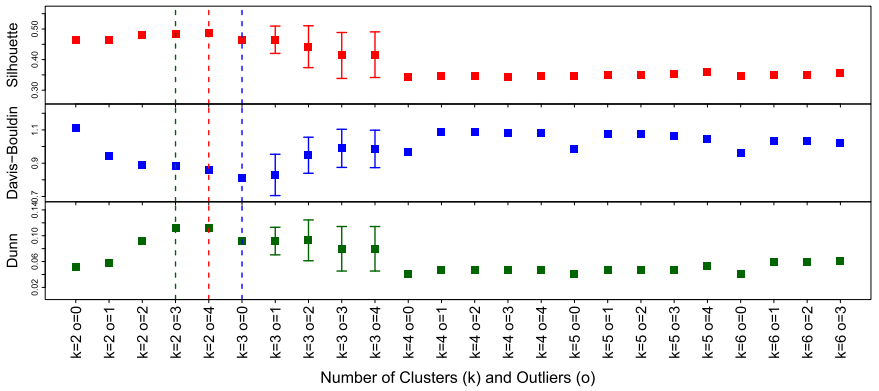
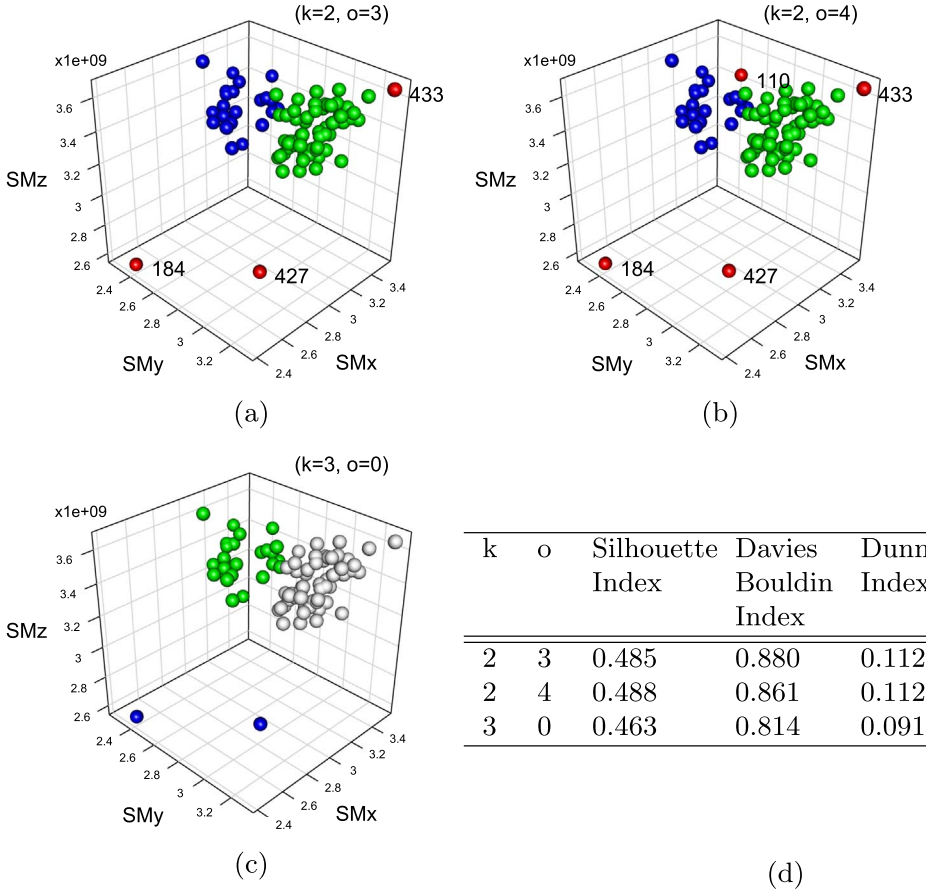


Fig. 10. Silhouette, Davies-Bouldin, and Dunn performance indices for different k-means−− clustering parameters.



k	o	Silhouette Index	Davies Bouldin Index	Dunn Index
2	3	0.485	0.880	0.112
2	4	0.488	0.861	0.112
3	0	0.463	0.814	0.091

(d)

Fig. 11. (a, b,c) Selected 3D scatter plots of SMs for each node, which are coloured based on their cluster membership for specific parameters, and (d) their corresponding performance index values. Cluster groups are coloured in blue, green, and grey, anomalies are coloured in red.

Fig. 10 shows the computed values of the Silhouette, Davies-Bouldin, and Dunn indices for each (k,o) pair. As described in Section 4.2.3, the pairs (k=2,o=3), (k=2,o=4), and (k=3,o=0) were selected as they had the highest Silhouette and Dunn values, and the lowest Davies-Bouldin value. Fig. 11 shows the scatterplots for each of these pairs (Fig. 11a,b, and c, respectively) and their index values (Table 11d). The intersection of the identified anomalies from (k=2,o=3) and (k=2,o=4) gives node 184, 427, and 433.

In the case of (k=3,o=0), the node 184 and 427 were singled out in a third cluster, which is well separated from the other two groups in the 3d feature space. Moreover, node 433 was included into one of the existing clusters. This result is due to the parameter setting o=0, which forced the clustering algorithm to reject any outright identified anomalies (i.e. step 3 of Algorithm 2 is not executed). Limiting the range of o to strictly positive integers (e.g. $o \in [1, 4]$) would provide an alternate result where node 433 would be identified as an anomaly. However in general, having $o > 1$ may provide potential false positives. Indeed, it will force the

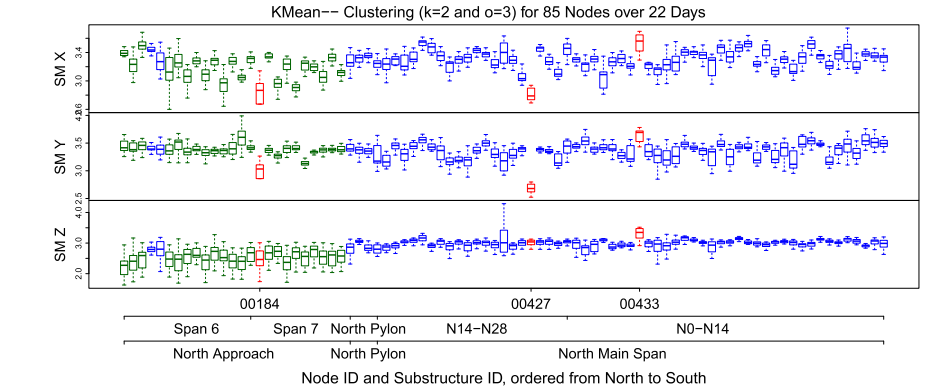


Fig. 12. Boxplot of SMs for each node and direction. The nodes are ordered by their location on the x-axis (from north to south), and coloured based on their cluster membership for (k=2, o=3). Cluster groups are in green and blue, anomalies are in red.

clustering process to mark the most distant point in a data set as an anomaly, even if that point is *well* matched to a cluster. From the point of view of the infrastructure managers and engineers, this trade-off may be more acceptable, as it could be safer to discard a false positive after an uneventful engineering inspection than letting a false negative remain undetected. In the remaining of this section, we decided to keep node 433 as an identified anomaly.

For the pair (k=2,o=3), Fig. 12 shows the boxplot of SMs over the study period for each direction and each node ordered on the x-axis according to their physical location from north (left) to south (right). The colour for a node indicates its membership to a group. These plots confirms our working assumption (Section 4). Indeed, most of the nodes located on the North Approach substructure are clustered in the same group in green, and all of the nodes located on the North Main Span substructure are clustered in the same group in blue. The anomalies are marked in red.

5.2.2. Anomaly investigation

We further investigated the identified node 184, 427 and 433 for any potential system issues. Fig. 13a and b compare the average voltage and acceleration at rest values for these nodes and a control node 170. Node 170 is located on the North Approach, and was correctly assigned to the same cluster as its direct neighbours in the previous step. The voltage plot 13a shows that node 427 did not return any value for most of July 2015 (i.e. the study period), thus this node was most probably powered off during that period. As a result only a few samples at the end of July 2015 were used to compute the SM feature for 427, potentially introducing a bias which may explain this node being an anomaly. The cause of this transient power failure may be linked to the power circuit within the computing node. This node has been scheduled for replacement, and further testing of its power unit will be performed once it has been retrieved.

The acceleration at rest plot 13b confirms the previous finding regarding node 184 (Section 5.1.2), i.e. it is subject to a very low signal to noise ratio. This plot also shows a visually less dense number of measurements reported for node 433, this prompted us to further look at the timing and sequence of the measured data for that node for any potential delay or loss.

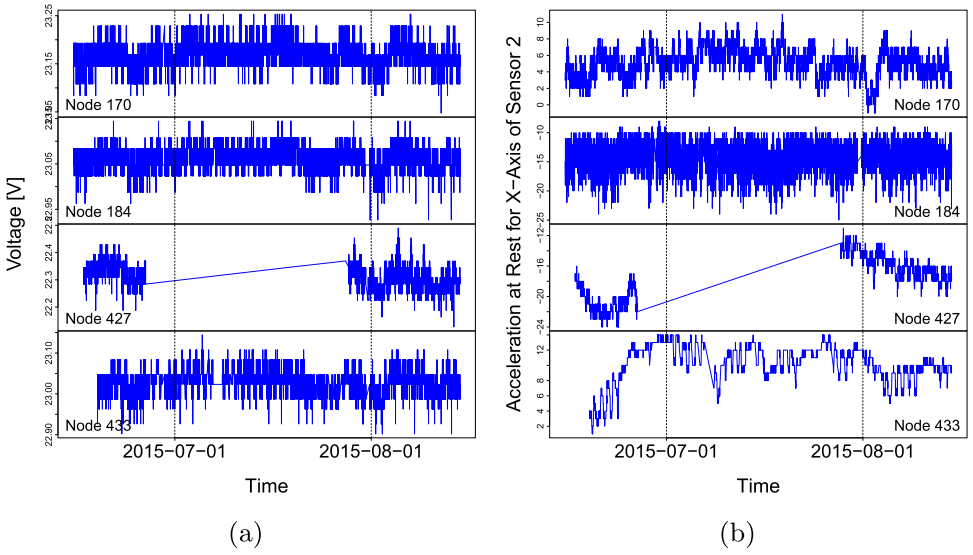


Fig. 13. (a) Measured voltage data and (b) X-axis acceleration at rest of sensor 2 for node 170, 184, 427 and 433, from mid June to mid August 2015.

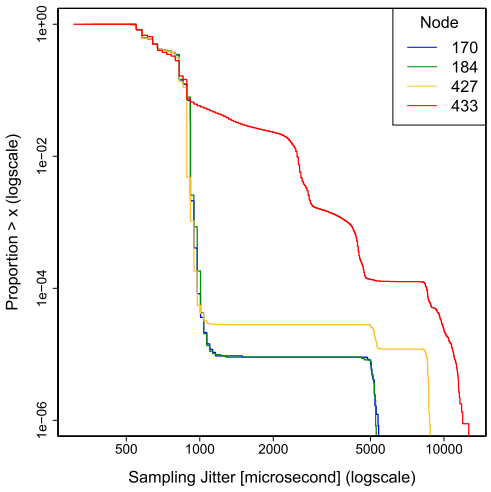


Fig. 14. Empirical Cumulative Distribution Function (ECDF) of the data sampling jitter for node 170, 184, 427, and 433 over July 2015.

For each node in different colour, Fig. 14 shows the empirical cumulative distribution function (ECDF) in log-scale for the variation of the time interval (i.e. duration) between two data sample, i.e. the data sampling jitter. For an ideal data collection, this jitter value should be constant at about 0 s. This Figure shows that for node 170, 184 and 427 only 0.01% of the sampled data points have a jitter above 1 ms. In contrast, the data for node 433 had a significantly different jitter distribution with more than 1% having a jitter above 2 ms, i.e. two order of magnitude difference and twice the duration. With such a high jitter in its data sampling, node 433 would have had a significantly different frequency spectrum distribution compared to the other nodes. This explains why node 433 was marked as an anomaly, as the SM feature is directly based on the frequency characteristics of a response signal. The cause for such a jitter in node 433 may be due to a failure of the oscillator-based clock on the sensor itself. This node has been scheduled for replacement.

6. Discussion

As presented in the previous Section 5, jack arches which are physically nearby under lane 7 of the SHB generally respond in a similar manner to vehicles driving on that lane. Moreover, the proposed approach from Section 4 is successful at detecting the specific jack arches (if any) that produce a response significantly different from their neighbours. This approach was evaluated using data from one tri-axial accelerometer per jack arch, which was a limitation of the physical hardware deployment on the SHB. As there is no dependency in the number of sensor in the proposed feature extraction (Section 4.1), this approach could be applied to other multi-sensor multi-axial deployments with the construction of a λ^n with higher dimension (Eq. (6)).

However, the proposed approach does not allow the distinction between the causes for such an abnormal response, i.e. it could be due to structural damage of the jack arch, or hardware/software fault of the monitoring system. The implicit workflow in Section 5.2 first tried to rule out any system issue before considering any potential structural damage for a given abnormal response. We have started to deploy such a workflow in the actual production deployment of our SHM system on the SHB. While in that Section 5.2 all of the data analysis for the system issue investigations were done manually, on the production SHB system we have started to implement automatic software-based audit routines. They periodically control for common system issues, such as node powered off, faulty sensor communication, or sensor detachment. When such an issue is detected, the data from that node are not used to detect structural damages.

The proposed approach operates on offline data sets, i.e. the sequences of steps from Section 4 were performed on the data many months after their collection. It would need to be adapted to operate on real-time streams of data in order to be used in a production system for damage detection. Although real-time online versions of clustering algorithms have been proposed [51], other challenges would need to be addressed, such as how to distribute the required computation among the embedded nodes on a bridge to minimize network communication and computing delay, or how to increase the resilience of the approach when multiple nodes have system issues.

One trade-off highlighted in Section 5.2.1 is the initial setting of the range for the anomaly parameter o of Algorithm 2. Having o=0 allows potential clustering results without any false positives (i.e. if a data set does not have any anomalies, then none would be reported). On the other hand, limiting $o > 0$ allows clustering results with potential false positives (i.e. a normal data point at the edge of a cluster may be forcibly marked as an anomaly). In the SHM context, the latter case could be a safer alternative as mentioned in Section 5.2.1. In our specific SHB deployment, we will further study and mitigate the effect of $o > 0$ on the rate of false positives in future works.

Following the initial work and results presented earlier, we have continued our study of the proposed damage detection method to address some of its limits. More specifically, we have been focusing on the following points:

- the estimation of the variability induced by daily and seasonal local changes in environmental conditions, such as wind and temperature,
- the study of the effect of long-term seasonal variations on the proposed approach (e.g. potential change in cluster memberships between summer and winter),
- the dynamic deployement of algorithms to the on-site computing nodes to automatically filter out environmental effects and known device issues,
- the application of our SHM system to bridges in rural and remote areas.

7. Conclusion

This paper presented the results of a large scale Structural Health Monitoring application on the Sydney Harbour Bridge (SHB) in Australia. This bridge has a large number of structural components, and this study focused on a subset of them, i.e. the 800 jack arches under the traffic lane 7. This paper first presented a short update on the deployed SHB system, which describes some new capabilities (e.g. continuous data collection) since the original deployment.

The main contribution of this work is then a novel approach to detect structural damages and system issues. This approach extracted spectrum-driven features, i.e. the SMs, from measured accelerations from the jack arches. SMs contain information from the entire frequency range, thus subtle differences between the normal signals and distorted ones could be identified. Our approach then used these SMs as the inputs to a modified k-means– clustering algorithm. Finally, our approach proposed a selection mechanism to process the clustering result, and identified jack arches with potential abnormal behaviours, i.e. structural damage or system issue.

This paper's final contribution is an extensive performance evaluation of the proposed approach. This evaluation included a *control* component, where the approach successfully detected jack arches with already known damage or issues, in a small scale context (i.e. small group of nodes and short time frame for the data). Furthermore, the evaluation also included a *test* component, where the approach was used on a large group of nodes with almost a month of data to detect any potential anomalies. It did detected 3 jack arches, which on subsequent further investigations turned out to have indeed system issues.

We further discussed some of the limits of the proposed approach, such as the use of real-time data streams. Finally, we concluded with a short overview of our current ongoing work on this SHB system, such as the study of the variabilities induced by daily and seasonal local environmental changes.

Acknowledgment

The authors wish to thank the Roads and Maritime Services (RMS) in New South Wales, Australia for provision of the support and testing facilities for this research work. NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program. CSIRO's Digital Productivity business unit and NICTA have joined forces to create digital powerhouse Data61.

Appendix A. Definitions of the clustering performance indices

This section provides the equations used to compute the clustering performance indices used by our approach, as described in Section 4.2.3.

The Silhouette [46] index s for a clustering of a data set including n points is defined as,

$$s = \frac{1}{n}\sum_{x=1}^{n}\frac{b(x)-a(x)}{\max\{a(x),\,b(x)\}} \tag{A.1}$$

where $a(x)$ is the average dissimilarity of a data point x to all the other data points within its assigned cluster, and $b(x)$ is the lowest average dissimilarity of x to all the other data points within any other clusters not including x. A value of s close to 1 indicates that on average all data points are *well* assigned to their respective clusters.

The Davies-Bouldin [47] index d for a data set, which is partitioned into N clusters, is defined as,

$$d = \frac{1}{N}\sum_{i=1}^{N}\max_{j\neq i}\left(\frac{S_i+S_j}{M_{i,j}}\right) \tag{A.2}$$

where S_i is a dispersion measure for cluster i, and $M_{i,j}$ is a separation measure between cluster i and j. Following [47], we used S_i and $M_{i,j}$ as,

$$S_i = \left(\frac{1}{T_i}\sum_{k=1}^{T_i}|X_k - A_i|^p\right)^{1/p} \quad \text{and} \quad M_{i,j} = \left(\sum_{l=1}^{N}|a_{li}-a_{lj}|^q\right)^{1/q} \tag{A.3}$$

where T_i is the number of data points in cluster i, X_k is the kth data point in cluster i, A_i is the centroid of cluster i, and a_{li} is the lth

component of A_i. Based on these definitions, a lower value of d indicates a *better* clustering result.

The Dunn [48] index D for a partition of N clusters is defined as,

$$D = \min_{1\le k\le N}\left(\min_{i+1\le l\le N}\left(\frac{dist\,(c_k,\, c_l)}{\max_{1\le m\le N} diam\,(c_m)}\right)\right) \tag{A.4}$$

where c_k is the kth cluster in the partition, $dist\,(c_k,\, c_l)$ is a distance measure between cluster c_k and c_l, and $diam\,(c_m)$ is a diameter measure of cluster c_m, as defined by,

$$dist\,(c_k,\, c_l) = \min_{(x_k\in c_k, x_l\in c_l)} d\,(x_k,\, x_l) \quad \text{and} \quad diam\,(c_m) = \max_{(x_{m1}, x_{m2}\in c_m)} d\,(x_{m1},\, x_{m2}) \tag{A.5}$$

where x_k is a data point in cluster c_k, and $d\,(x_k,\, x_l)$ is the Euclidean distance between data points x_k and x_l. Based on these definitions, a large value of D indicates compact and well separated clusters.

References

[1] M. Biswas, A. Pandey, M. Samman, Diagnostic experimental spectral/modal analysis of a highway bridge, Int. J. Anal. Exp. Modal Anal., 5, 1990, pp. 33–42.
[2] C.R. Farrar, K. Worden, An introduction to structural health monitoring, in: Philos. Trans. R. Soc. London A: Math. Phys. Eng. Sci., Vol. 365, 2007, pp. 303–315.
[3] W. Staszewski, K.Worden, Overview of optimal sensor location methods for damage detection, in: SPIE's Proceedings of the 8th Annual International Symposium on Smart Structures and Materials, 2001, pp. 179–187.
[4] G. Park, T. Rosing, M. Todd, C. Farrar, W. Hodgkiss, Energy harvesting for structural health monitoring sensor networks, J. Infrastruct. Syst., Vol. 14, 2008, pp. 64–79.
[5] M. M. Alamdari, Vibration-based structural health monitoring,(Ph.D. thesis), University of Technology, Sydney, 2015.
[6] M. M. Alamdari, J. Li, B. Samali, A comparative study on the performance of the damage detection methods in the frequency domain, in: Australasian Conference on the Mechanics of Structures and Materials, 2013.
[7] M. M. Alamdari, J. Li, B. Samali, Frf-based damage localization method with noise suppression approach, J. Sound Vib., Vol. 333, 2014, pp. 3305–3320.
[8] L. Mujica, J. Vehi, M. Ruiz, M. Verleysen, W. Staszewski, K. Worden, Multivariate statistics process control for dimensionality reduction in structural assessment, Mech. Syst. Signal Process., Vol. 22, 2008, pp. 155–171.
[9] D. Tibaduiza, M. Torres-Arredondo, L. Mujica, J. Rodellar, C. Fritzen, A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring, Mech. Syst. Signal Process., Vol. 41, 2013, pp. 467–484.
[10] M. M. Alamdari, N. Khoa, P. Runcie, S. Mustapha, U. Dackermann, V. Nguyen, L. Jianchun, X.Gu, Application of unsupervised support vector machine for condition assessment of concrete structures, in: Symposium in Proceedings of the Second International Conference on Performance-based and Lifecycle Structural Engineering (PLSE), 2015.
[11] Q. Jiang, M. Jia, J. Hu, F. Xu, Machinery fault diagnosis using supervised manifold learning, in: Mech. Syst. Signal Process., Vol. 23, 2009, pp. 2301–2311.
[12] H.Sohn, C. Farrar, F. Hemez, D. Shunk, D. Stinemates, B. Nadler, J. Czarnecki, A review of structural health monitoring literature: 1996–2001, Los Alamos National Laboratory, USA.
[13] M. Chandrashekhar, R. Ganguli, Uncertainty handling in structural damage detection using fuzzy logic and probabilistic simulation, Mech. Syst. Signal Process., Vol. 23, 2009, pp. 384–404.
[14] P. Runcie, S. Mustapha, T. Rakotoarivelo, Advances in structural health monitoring system architecture, in: International Symposium on Life-Cycle Civil Engineering, 2014, pp. 1064 –1071.
[15] P. Runcie, An update on a large scale shm deployment on sydney's harbour bridge and associated research activities, in: International Workshop on Structural Health Monitoring 2015 (IWSHM 2015), 2015.
[16] H. Salane, J.W. Baldwin, R. Duffield, Dynamics approach for monitoring bridge deterioration, Transp. Res. Rec. 832 (1981) 21–28.
[17] M. Kato, S. Shimada, Vibration of pc bridge during failure process, in: Journal of Structural Engineering, Vol. 112, 1986, pp. 1692–1703.
[18] M. Raghavendrachar, A.E. Aktan, Flexibility by multireference impact testing for bridge diagnostics, J. Struct. Eng. 118 (8) (1992) 2186–2203.
[19] C. Farrar, W. Baker, T. Bell, Dynamic Characterization and Damage Detection in the I-40 Bridge Over the Rio Grande, Los Alamos National Laboratory, 1994.
[20] C. Farrar, P. Cornwell, S. Doeblingand, M. Prime, Structural Health Monitoring Studies of the Alamosa Canyon and I-40 Bridges, Los Alamos National Laboratory, 2000.
[21] S. Li, S. Zhu, Y. Xu, Z. Chen, H. Li, Longterm condition assessment of suspenders under traffic loads based on structural monitoring system: Application to the tsing ma bridge, in: Structural Control and Health Monitoring, Vol. 19, 2012, pp. 82–101.
[22] M. Abe, Y. Fujino, M. Yanagihara, M.Sato, Monitoring of hakucho suspension bridge by ambient vibration measurement, in: SPIE Proceedings of the 5th Annual International Symposium on Nondestructive Evaluation and Health Monitoring of Aging Infrastructure, 2000, pp. 237–244.
[23] K. Worden, G. Manson, N. Fieller, Damage detection using outlier analysis, in: Journal of Sound and Vibration, Vol. 229, 2000, pp. 647–667.
[24] A. Santos, E. Figueiredo, J. Costa, Clustering studies for damage detection in bridges: A comparison study, in: Structural Health Monitoring, 2015.
[25] L. Yu, J. Zhu, L. Yu, Structural damage detection in a truss bridge model using fuzzy clustering and measured frf data reduced by principal component projection, Adv. Struct. Eng. 16 (2013) 207–217.
[26] A. Yin, B. Wang, X. Hu, Z. Dai, Mhop-cl: a clustering protocol for bridge structure health monitoring system, in: IEEE International Symposium on Computer Network and Multimedia Technology, IEEE, 2009, pp. 1–4.
[27] A. Diez, N. Khoa, M. M. Alamdari, Y. Wang, F. Chen, P. Runcie, A clustering approach for structural health monitoring on bridges, in: Journal of Civil Structural Health Monitoring, 2016, pp. 1–17.
[28] W. Ren, X. Peng, Baseline finite element modeling of a large span cable-stayed bridge through field ambient vibration tests, in: Computers and Structures, Vol. 2005, 2005, pp. 536–550.
[29] J. Brownjohn, F. Magalhaes, E. Caetano, A. Cunha, Ambient vibration re-testing and operational modal analysis of the humber bridge, Eng. Struct. 32 (8) (2010) 2003–2018.
[30] D. Siringoringo, Y. Fujino, Estimating bridge fundamental frequency from vibration response of instrumented passing vehicle: Analytical and experimental study, in: Advances in Structural Engineering, Vol. 15, 2012, pp. 417–433.
[31] J. Lynch, K. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, in: Shock and Vibration Digest, Vol. 38, 2006, pp. 91–130.
[32] M.M. Alamdari, B. Samali, J. Li, H. Kalhori, S. Mustapha, Spectral-based damage identification in structures under ambient vibration, in: Journal of Computing in Civil Engineering, 2015.
[33] S. Liberatore, G. Carman, Power spectral density analysis for damage identification and location, J. Sound Vib. 274 (3) (2004) 761–776.
[34] W.W. Bayissa, N. Haritos, Structural damage identification in plates using spectral strain energy analysis, J. Sound Vib. 307 (1) (2007) 226–249.
[35] W. Hong, Z. Wu, C. Yang, C. Wan, G. Wu, Investigation on the damage identification of bridges using distributed long-gauge dynamic macrostrain response

under ambient excitation, J. Intell. Mater. Syst. Struct. 23 (1) (2012) 85–103.
[36] S. Fang, R. Perera, Power mode shapes for early damage detection in linear structures, J. Sound Vib. 324 (1) (2009) 40–56.
[37] M.G. Masciotta, L.F. Ramos, P.B. Lourenço, M. Vasta, G.D. Roeck, A spectrum-driven damage identification technique: application and validation through the numerical simulation of the Z24 bridge, Mech. Syst. Signal Process. 70–71 (2016) 578–600.
[38] Z. Zheng, Z. Lu, W. Chen, J. Liu, Structural damage identification based on power spectral density sensitivity analysis of dynamic responses, Comput. Struct. 146 (2015) 176–184.
[39] F. Ralph, P. Neiman, D.V. de Kamp, D. Law, Using spectral moment data from noaa's 404-mhz radar wind profilers to observe precipitation, in: Bulletin of the American Meteorological Society, Vol. 76, 1995, pp. 1717–1740.
[40] L. Lutes, C. Larsen, Improved spectral method for variable amplitude fatigue prediction, J. Struct. Eng. 116 (4) (1990) 1149–1164.
[41] D. Baker, G. Manolis, S. Mohan, Hibic: expert system for highway bridge dynamics, J. Comput. Civ. Eng. 3 (4) (1989) 348–366.
[42] M. Di-Paola, G. Muscolino, On the convergent parts of high order spectral moments of stationary structural responses, in: Journal of Sound and Vibration, Vol. 110, 1986, pp. 233–245.
[43] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323.
[44] D. Arthur, S.Vassilvitskii, K-means++: The advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007, pp. 1027–1035.
[45] S. Chawla, A.Gionis, k-means-: A unified approach to clustering and outlier detection, in: Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA., 2013, pp. 189–197.
[46] P. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, in: Journal of Computational and Applied Mathematics, Vol. 20, 1987, pp. 53–65.
[47] D. Davies, D. Bouldin, A cluster separation measure, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 1, 1979, pp. 224–227.
[48] J. Dunn, A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters, in: Journal of Cybernetics, Vol. 3, 1973, pp. 32–57.
[49] F. Mohd-Yasin, N. Zaiyadi, D.J. Nagel, D.S. Ong, C.E. Korman, A.R. Faidz, Noise and reliability measurement of a three-axis micro-accelerometer, Microelectron. Eng. 86 (46) (2009) 991–995.
[50] S. Ramaswamy, R. Rastogi, K. Shim, Efficient algorithms for mining outliers from large data sets, in: SIGMOD '00: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, 2000, pp. 427–438.
[51] A.V.E. Pomponi, A real-time approach to acoustic emission clustering, in: Mechanical Systems and Signal Processing, Vol. 40, 2013, pp. 791–804.