 SDHM, vol.8, no.3, pp.271-293, 2012

Sensor Fault Detection in Large Sensor Networks using PCA with a Multi-level Search Algorithm

A. Rama Mohan Rao[1], S. Krishna Kumar[1] and K. Lakshmi[1]

Abstract: Current advancements in structural health monitoring, sensor and sensor network technologies have encouraged using large number of sensor networks in monitoring spatially large civil structures like bridges. Large amount of spatial information obtained from these sensor networks will enhance the reliability in truly assessing the state of the health of the structure. However, if sensors go faulty during operation, the feature extraction techniques embedded into SHM scheme may lead to an erroneous conclusion and often end up with false alarms. Hence it is highly desirable to robustly detect the faulty sensors, isolate and correct the data, if the data at faulty sensor locations are sensitive. Several sensor fault detection algorithms have been reported in the literature and among them the PCA based sensor fault detection algorithm [Kerschen et al (2005)] appears to be robust in isolating all types of sensor faults. However, in a large sensor network, the computational time in isolating the faulty sensor is prohibitive for online fault detection and isolation. In this paper we propose a multi-level search algorithm, which improves the performance of the PCA based sensor fault detection algorithm quite appreciably. Numerical simulation studies have been carried out to demonstrate the effectiveness of the proposed algorithm. The sensitivities of the proposed algorithm with parameter settings are also presented.

Keywords: Structural health monitoring, smart sensor networks, Principal component analysis, multi-level algorithm, singular value decomposition.

1 Introduction

Civil engineering infrastructure systems like bridges are spatially large and hence the structural health monitoring strategies are being equipped with large number of sensors for continuous online health monitoring and extracting features for robust damage diagnosis. As the number of sensors deployed for a SHM application increases, it is also likely that the sensors themselves represent a weak link in the

[1] CSIR-Structural Engineering Research Centre, CSIR Campus, Taramani, Chennai

SHM system. Since the structural health monitoring schemes are heavily based on the measurement data recorded during a long period, the sensor faults if goes undetected, may misrepresent as a structural fault. Hence we need to thoroughly investigate the long-term reliability, robustness, and calibration of these sensors. Robust algorithms need to be developed which can offer a reliable damage detection capability even under malfunction of some sensing nodes of the global SHM system. The dense sensor network constitutes a redundant system, which can be used to detect sensor malfunction, identify the faulty sensor and reconstruct the faulty sensor data, if required.

A sensor is declared faulty when it displays "a non-permitted deviation from the characteristic properties". This deviation may appear in four forms namely: bias, drift, complete failure, and precision degradation. A sensor reading is biased if the reading differs by a constant value from the actual value. If the difference between the sensor reading and the actual value changes linearly with time, the corresponding fault is referred to as drift. The sensor is considered to be completely faulty, if the sensor reading remains constant regardless of the changes in actual value. If the sensor reading is associated with an excessive-variance white noise, it is referred to as precision degradation. Gain fault is due to increase in the variance of the sensor and said to be uncommon [Kerschen et al (2005)].

The control and chemical engineering community have considered the sensor validation problem, and have used models and sensor redundancy to good effect. However, these approaches usually use the faulty sensor to predict the response and look for errors between predictions and measurement. Using the faulty sensor in the prediction process will propagate errors to the predicted responses. Often neural networks, or artificial intelligence approaches are used for the analysis.

Structural health monitoring schemes of civil engineering structures typically include several sensors at different locations of the structure in order to extract features for damage detection or damage localization. Such sensor network constitutes a redundant system, which can be used to detect sensor malfunction or failure, identify, and even correct the faulty sensor. Dunia et al (1996) have investigated the detection, isolation, and reconstruction of a faulty sensor using principal component analysis (PCA). They have reconstructed each sensor data making use of data from the remaining sensors. They have proposed sensor validity index (SVI), evaluated from the residuals obtained from the sensor data before and after reconstruction. This index is used for fault detection and isolation. They have also proposed two approaches to reconstruct a single sensor. The first one is an iterative method and the second one is an optimization based method. However, both these methods result in the same closed-form solution. Dunia et al (1996) have investigated different types of sensor faults, i.e., bias, complete failure, drifting, and precision degradation. Different residuals were investigated. Dunia and Qin (1998) have also derived conditions for fault detectability, reconstructability, and identifiability. Both sensor and process faults could be identified, but the fault direction must be known, which is easy for a sensor fault, but difficult for process faults.

Kerschen et al (2005, 2004) have also applied PCA to sensor validation and used the angle between the principal subspaces as a feature for fault detection. They also presented a closed-form equation to reconstruct data of a single faulty sensor. The faulty sensor is isolated by removing the data pertaining to one sensor at a time from both the reference and current data sets and the principal angle of the subspaces is evaluated. The angle should then be very small, when the faulty sensor is discarded.

Friswell and Inman (2000, 1999) have introduced a sensor validation method based on the modal model of the structure. This approach requires that the sensors are functioning correctly, should be continuous and automatically monitored. The two approaches (based on modal filtering and PCA) to sensor validation are developed, based on the assumption that a model of the structure is available. They have concluded that the modal filtering approach performs better, if an accurate modal model is available. They have also observed that a multiplicative sensor fault is more difficult to locate than an additive fault.

Abdelghani and Friswell (2004, 2007) have investigated model-based methods to validate sensors with additive or multiplicative faults. They have proposed parity space and the modal filtering approaches for identification of sensors with an additive fault. For the multiplicative fault, they have proposed a correlation index to isolate the faulty sensor.

Hernandez-Garcia and Masri (2008) compared PCA, independent component analysis (ICA), and modified ICA(MICA) for sensor fault detection. They have used Hotelling's T^2 statistic and the squared prediction error (SPE) for fault detection. It was concluded that the detection performance of ICA and MICA was higher than that of PCA.

Kullaa (2009) introduced factor analysis for sensor validation and compared it with PCA. A closed-form solution to reconstruct several sensors using the remaining sensors was also proposed.

Kullaa (2006) introduced the minimum mean square error (MMSE) estimation to sensor validation. It works directly in the data space and does not require to determine any model parameters. A spatio-temporal extension was introduced [Kullaa (2007)], which can be applied if the number of sensors is lower than the number of active modes in the structure. The faulty sensor was identified by removing one sensor at a time and performing MMSE estimation on the remaining sensors. The faulty sensor was the missing sensor in the analysis with the lowest mean-square residual compared to that of the training data. The main disadvantage is that the analysis needs to be carried out for each sensor and hence it is tedious.

Zhiling et al (2007) have developed a technique for sensor failure detection by formulating a sensor error function. In this approach, the sensors are divided into two groups; reference sensors and uncertain sensors. Reference sensors correctly measure the structural responses and uncertain sensors may fail to correctly measure the structural responses. A sensor error function is formulated to detect the instants of failure of the corresponding uncertain sensor, using the measurements from reference sensors and the uncertain sensor examined. The sensor error function is derived using indirect and direct approaches. The applicability of the sensor failure detection formulation has been demonstrated experimentally by a 4 m long eight bay truss structure.

Before presenting our investigations on the effectiveness of various sensor validation algorithms, it is worthwhile to define the ideal characteristics of a sensor fault detection algorithm. An ideal sensor validation technique should be capable of isolating the faulty sensors in the sensor network, with the following characteristics:

- The ambient loading on civil engineering structures is not measurable. Hence the algorithm should be capable of isolating faulty sensors with output only responses.
- The algorithm should be able to identify all forms of sensor faults i.e., bias, drift, complete failure, and precision degradation.
- The algorithm should be capable of working without the prior knowledge of the number of faulty sensors in the network.
- The algorithm should have ability to find instance of origination of fault.
- Ease of implementation and computational efficiency.

Numerical investigations have been carried out on some selected sensor fault diagnosis techniques and their characteristics with respect to the above five desirable features are listed in Table 1. The details provided in Table 1, clearly indicate that the technique based on principal component analysis (PCA), where the angle between the principal subspaces is used as a feature for fault detection is robust in terms of identifying all the four types of sensor faults. However in this algorithm, one need to remove the time history data of a sensor (or multiple sensors) each time and compute the subspace principle angle in order to identify the single or multiple sensor faults. This process of sensor validation is computationally very expensive and time consuming especially for spatially large civil engineering structures like bridges, which are instrumented with large number of smart sensors. In the state-of-the-art structural health monitoring schemes, it is quite common to use large sensor networks. Keeping this in view, we propose a multi-level algorithm, which substantially reduces the number of operations in identifying the multiple sensor faults. Numerical investigations have been carried out to validate and also verify the robustness of the proposed algorithm in identifying the sensor faults. The computational complexities with single and multiple sensor faults are also analyzed.

2 Computation of Principal component subspace

It is always more efficient to identify directly the principal components, also called principal directions, rather than performing an exact modal identification to compute the trajectories covered by the measurements. However, under certain assumptions, principal components may represent the vibration modes of the system [Feeny and Kappagantu (1998)].

In the present work, it is assumed that the number of sensors 'n' is greater than the number of structural modes (m+1) involved in order to maintain the redundancy of the data. Let Q denote a discrete block time-history of n x b (where b $\gg$ n) sampled responses

$$Q = \begin{bmatrix} x_1(t_{j+1}) & .... & x_1(t_{j+b}) \\ . & .... & . \\ x_n(t_{j+1}) & .... & x_n(t_{j+b}) \end{bmatrix} \tag{1}$$

The singular value decomposition (SVD) of the block data Q gives:

$$Q = U\Sigma V^T \tag{2}$$

where U is an orthonormal matrix (n X n) whose columns define the principal components (PCs) and form a subspace spanning the data. Each column of U is associated with the (b X b) time coefficient matrix V. The singular values, given by the (n X b) diagonal matrix Σ and sorted in descending order, can be related to the energy associated with the corresponding principal components of U. This means that the structure will react mainly in the directions of the principal components associated with the highest energies. We may note here that it is computationally more efficient to calculate the SVD of:

$$QQ^T = U\Sigma^2 V^T \tag{3}$$

Theoretically, only the first m + 1 eigenvalues of Q are nonzero. Nevertheless, we know that test data contains sensor noise. Since noise has much lower energy than

Table 1: Characteristics of selected sensor fault diagnosis techniques

S. No	Feature of the Algorithm	MMSE [Kullaa (2006)]	MF [Friswell and Inman (2000, 1999)]	SEF-I [Zhiling et al (2007)]	SEF-D [Zhiling et al (2007)]	PCA [Kerschen et al (2005)]
1	Output only responses	Yes	No	No	Yes	Yes
2	All forms of sensor Faults	Yes	No	Yes	Yes	Yes
3	Prior knowledge of number of sensor faults	Required	Not required	Required	Required	Not required
4	Ability to find the exact instance of origination of sensor fault	No	No	Yes	Yes	Yes
5	Ease of implementation	Yes	Yes	No	Yes	Yes
6	Computational efficiency	Yes	No	Yes	Yes	No

MMSE: Minimum mean square error method; MF: Modal filtering method; SEF-I: Sensor Error Function- in direct approach; SEF-D: Sensor Error Function- direct approach; PCA: Subspace angles using PCA the structural modes, the components of U associated with eigenvalues presenting an order of magnitude much lower than others have to be discarded from the principal component base. In the linear case, the principal directions extracted from test data, always lie in the subspace (or hyper-plane) generated by the participating modes

Mathematically speaking, this means that the so-called principal hyper-plane is invariant, even if the directions of the principal vectors are dependent on the structural excitation. Nevertheless, the principal hyper-plane is dependent on the structural characteristics. The PCA may be then considered as a powerful and straightforward approach to compute a modal metrics of test data and to detect a potential sensor failure by comparing reference and current structural states.

2.1 Angles between Subspaces

Given a set of data, the active principal components–PCs define a subspace (or hyper-plane) that characterizes the dynamic behavior of the system. A change in the system modifies consequently its dynamic state and affects the subspace spanned by the PCs. This change may be estimated using the concept of angles between two subspaces introduced by Golub and Van Loan (1996). This concept allows quantifying the spatial coherence between two time-history blocks of an oscillating system. Let $A \in \mathrm{R}^{n_s \times p}$ and $B \in \mathrm{R}^{n_s \times q}$ be two subsets, each with linearly independent columns. First a QR factorization allows computing the orthonormal bases of A and B:

$$\begin{array}{ll} A = Q_A R_A & \mathrm{Q}_A \in \mathrm{R}^{n_s \times p} \\ B = Q_B R_B & \mathrm{Q}_B \in \mathrm{R}^{n_s \times q} \end{array} \tag{4}$$

Thus, the singular values of $Q_A^T Q_B$ define the q cosines of the principal angles θ_i between A and B.

$$SVD(Q_A^T Q_B) \rightarrow \mathrm{Diag}(\cos(\theta_i))\ \mathrm{i} = 1, ...\mathrm{q} \tag{5}$$

The largest angle allows quantifying how the subspaces A and B are globally different.

The faulty sensor can be detected using these principal angles. The responses of the subspaces between reference data, i.e., when all sensors are functioning properly and the current data i.e., when some of the sensors have become faulty, are compared by computing the principal angles. If the principal angle between these two subspaces is appreciably high, it can be concluded that some sensors in the current response gives erroneous results. Theoretically, if all sensors are functioning correctly, the angle between the subspaces spanned by reference data and the current data should be zero. However, practically, it will not be zero due to environmental variances and also the noises present in the measurement process. Before applying the sensor validation process, number of reference data sets are collected by taking measurements at different time instants and partitioned into several sets. The principal angle between the subspace spanned by each of these sets and the subspace spanned by the whole data set is computed, which gives us a collection of different subspace angle values. When dealing with the current data set, an alarm is issued when the monitored angle exceeds the upper control limit (UCL) defined as the mean angle plus three times its standard deviation. This corresponds to a 99.7% confidence interval for a normal distribution.

When an alert for faulty sensor is given, we need to isolate the faulty sensors. In order to accomplish this, one has to remove one of the sensors each time from both reference and current data sets and check the subspace angle. If the subspace angle is minimum i.e. below the UCL specified, then the sensor isolated is faulty. Otherwise, the isolated sensor is considered as working fine. The PCA based sensor fault detection algorithm works very robustly and able to identify both additive and multiplicative faults associated with sensors. However, the major problem associated with the PCA based sensor fault detection is the computational complexity, especially while dealing with large smart sensor setup. It is not uncommon to employ large smart sensor networks in the state of the art structural health monitoring schemes. For example, if there are hundred sensors and assuming the upper limit of faulty sensors is five, the number of searches to identify the faulty sensors will be of the order of 75e06(100 C_5) , which will be computationally very tedious and might not be a good option for online sensor fault detection in continuous health monitoring schemes. In this paper, we propose a new search algorithm, which can identify the sensor faults with minimum number of searches. The details of the algorithm are described here.

2.2 Multi-level Search algorithm for PCA based sensor fault detection

The frame work of the proposed multi-level algorithm is developed keeping the following objectives in view

- The main objective of the search algorithm is to find any faulty sensor combination with least number of computations.

- The sensor subsets chosen must be as diverse as possible. Comparing subsets like ‘ 1 3 5 6 ..‘ and ‘1 4 5 6. . . ’ is unproductive since it would take lot of operations to single out the correct combination.(Samples should be like ‘ 1 3 5 6’ and ‘ 4 6 8 7’ , having little in common).

- The method should be as simple as possible. Also, it should be robust for any size of sensor network and also any size of faulty sensor subset.

The algorithm works on the principle of forming subspaces using several diverse combinations of spatial sensors (i.e., constructing subspaces using the time history data of a typical spatial sensor combination) and checking the subspace angles by comparing with reference subspace. Based on the subspace angles, the combination of spatial sensors chosen can be identified as faulty set or true set. Let the total number of sensors placed on a structure be n. Form a subspace from the time history data obtained from all n sensors. The search algorithm makes use of the subspace data and also the reference subspace data to identify robustly the faulty sensor/sensors using the principal angles concept discussed earlier. The sensor fault detection process runs in multiple levels. In the first level, all the sensors in the network are categorized as probable faulty sensors group (PFSG). From this, diverse subsets are formed at each level and the size of PFSG is reduced by identifying the subsets of working sensors using principal angles. In the next level, the same operations are repeated on the reduced PFSG. Thus at each level, the size of PFSG is reduced by identifying and eliminating the working sensor subsets formed at that level. The algorithm terminates when PFSG is reduced to true faulty sensor set. The details of forming subsets at each level are as follows:

Form new subspaces using the following subsets of time history data collected:

Subspace S_1:- 1, 3, 5, ...n

Subspace S_2:- 1, 4, 7, ...n

Subspace S_3:- 1, 5, 9,. . . ..n

. . . .

Subspace S_k:- 1, 1+(k+1), 1+2(k+1), ...n

The number of subsets, k, to be formed depends on the total number of sensors and also probable number of faulty sensors. The number of subspaces considered at each level is called the depth factor and is denoted by k. The issues related to the optimal parameter settings will be discussed later in this paper.

It can be observed from the subsets formed that sensor numbers like 2, 6, 8, ...etc., will be missing. To avoid this, we form further k number of following subsets:

Subspace R_1:- n, n-2, n-4, n-6, ...1

Subspace R_2:- n, n-3, n-6, n-9, ...1

Subspace R_3:- n, n-4, n-8, n-12, ...1

. . . .

Subspace R_k:- n, n-(k+1), n-2(k+1), n-3(k+1), ...1

Comparing the subspaces of all the above diversified subsets with the subspace of the reference data, we can find the combinations that are *fully* comprised of working (non faulty) sensors. The reduced set of probable faulty sensors group (PFSG) for the next level is formed, by eliminating the already established working sensors from the total sensors. The probable faulty sensor group (PFSG) need to be analysed for further reduction of this group. Similar approach described above is repeated for the current PFSG. Each combination is grouped with all or some of working sensors to provide spatial coverage. In each iteration, the number in PFSG reduces, finally converges to the exact faulty sensors.

The algorithm assumes that end sensors are working. If one or both end sensors happen to be faulty, all the combinations of the subsets formed at that level, will become faulty. The algorithm stalls at this instant. In order to avoid this rare case, a feature to handle this situation is included. When such an instant is encountered, the end sensors are left out and remaining sensors (2 to n-1) are sampled and evaluated. If 2 and n-1 are found to be working, they are used for spatial coverage. Otherwise sensors 3 to n-2 are sampled and the process goes on. This avoids stalling of the algorithm, due to faulty end sensors. We may encounter similar problem, while forming subsets with PFSG, i.e., the end sensors may be faulty. To overcome this, we can ensure that the end sensors are not faulty by using the sensors from working sensor group as end sensors. Further, while forming the S and R subsets, it is possible that, we may end up with the same sensors in both S and R subsets. To overcome this, the generated R and S subsets are verified for duplication and may be eliminated before proceeding for computing subspace angles.

3 Numerical studies

Numerical studies have been carried out by simulating numerically various types of sensor faults in order to ensure the robustness of the sensor fault detection technique. Later, we have employed the proposed multi-level search algorithm in conjunction with the sensor fault diagnosis technique based on subspace angles to investigate the efficiency, robustness in detection of sensor faults and also parameter settings.

3.1 Simply supported beam

The first numerical example considered is a simply supported beam with twenty one accelerometers placed at equal intervals on the beam, to measure the response. All four types of sensor fault are simulated. First, the acceleration measured at the ninth sensor is multiplied by 1.2 (gain fault). Secondly, it is replaced by a white-noise sequence of the same variance. Three different cases are considered for this numerical example. They are:

- case-A : one sensor at fault as indicated in Figure 1(a)
- case-B : three adjacent sensors are faulty as indicated in Figure 1(b)
- case-C : three sensors at different locations are faulty as indicated in Figure 1(c)

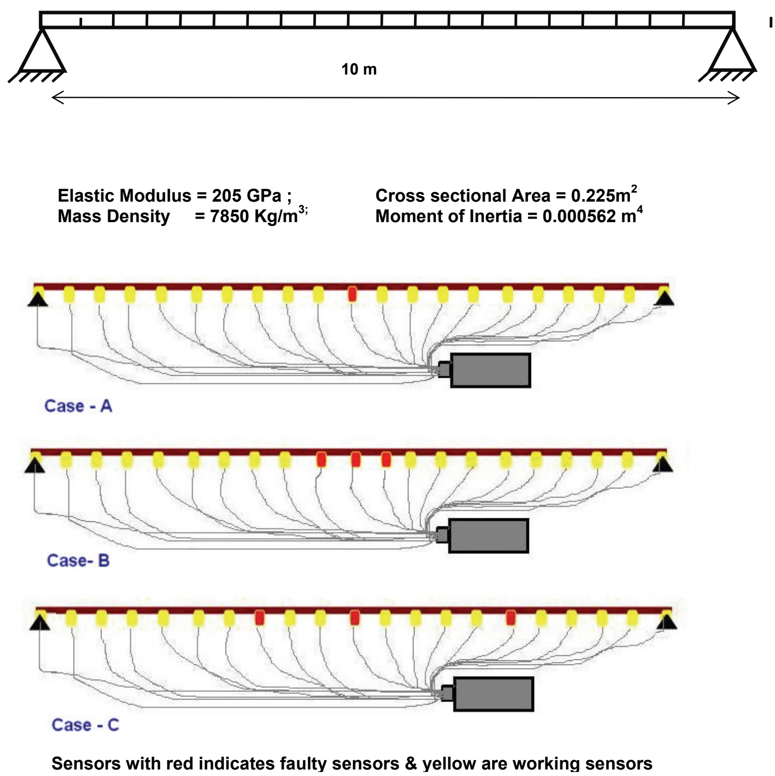


Figure 1: Simply supported beam with 21 accelerometers

For case-A, the reference data set contains 110000 samples from each of the twenty one channels. The first step is to estimate the upper control limit (UCL). In order to arrive at UCL value, the data are partitioned into 11 different subsets, each containing 10000 points. These 11 reference subsets will give us a collection of 11 different angle values. The first three PCA modes are considered in the analysis as they capture almost 99% of the energy. For a three-dimensional subspace, an alarm will thus be issued when the monitored angle will exceed the following angle (in degrees).

$$UCL_3 = \bar{\theta} + 3\sigma_\theta = 0.4150 + 3 \times 0.5261 = 1.9933 \tag{6}$$

Where $\bar{\theta} = \frac{1}{N}\sum_{k=1}^{N}\theta_k$ and N=11

$$\sigma_\theta = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\theta_k - \bar{\theta})^2}$$

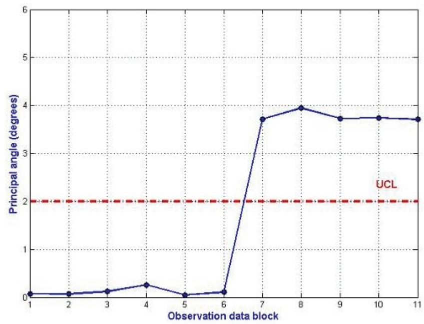
The current data set contains 110000 samples, whose last 50000 samples correspond to a sensor fault. The current data set is partitioned into 11 different sets containing 10000 points each. Figure 2 shows the principal angle between the subspace spanned by each of these subsets of current data and the subspace spanned by the whole reference set for both sensor faults. It can be observed that, the first six subsets (i.e., 60000 samples) are well below the UCL. However, the subspace angle computed for rest of the five subsets (from 60000 to 110000 samples) is well above the UCL clearly indicating that the some of the sensors have become faulty. An alarm will be issued to indicate the presence of faulty sensors in the network. Once the alarm has been issued, the faulty sensor needs to be identified using the methodology discussed earlier in this paper. Figure 3 presents all the subspace angles obtained when the sensors are removed one at a time. It is clear from Figure 3 that the computed subspace angle falls below UCL, when the ninth sensor is removed. In all the other cases, the computed subspace angle lies well above UCL. With this, one can clearly identify that ninth sensor is faulty.

Now that the ninth sensor has been identified as faulty, the next step is to retrieve its original response. The response can be obtained using the redundancy of measurement data, i.e., the number of sensors are more than the number of structural modes involved. We use the approach suggested by Kramer (1992) and is given by

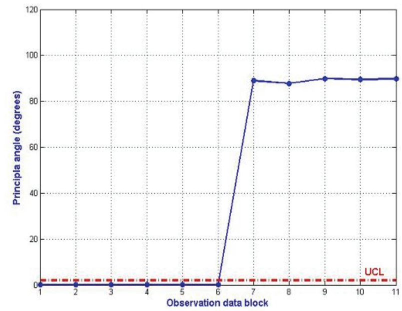
$$x^* = -\frac{\{a'_j\}^T \sum_{k\neq j=1}^{n_s} x_k\{a'_k\}}{\{a'_j\}^T\{a'_j\}} \tag{7}$$

a'_j, a'_k are the jth and kth columns of the matrix A, where A= P^{T}P. Matrix P contains the principal components. x_k and x^* are the acceleration vector at kth location (node) and the corrected acceleration data at the faulty sensor location respectively.

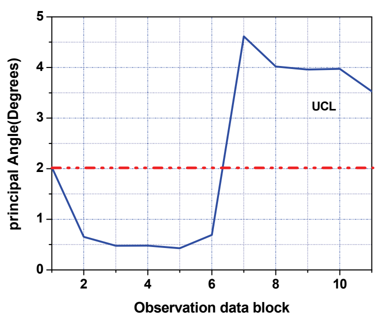
As underlined in the previous section, we have at our disposal 110000 reference data points. First, the optimum number of retained PCA modes is determined.



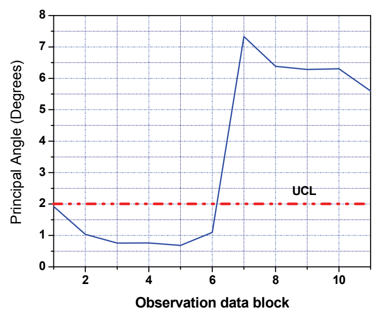
(a) Sensor Gain plot



(b) Sensor failure plot



(c) additive sensor failure plot

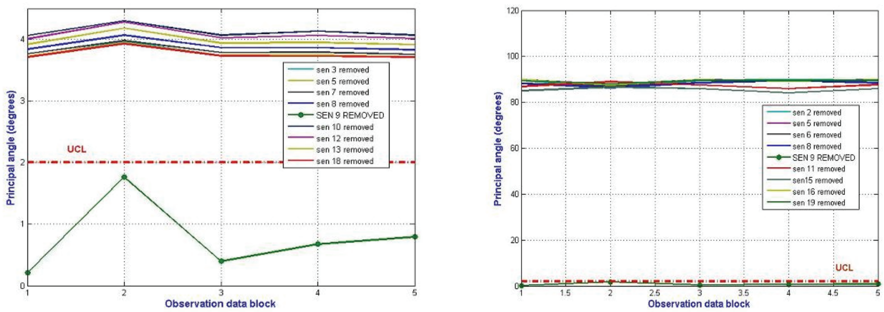


(d) Sensor drift failure plot

Figure 2: Monitoring the subspace angle- Simply supported beam

More precisely, the response of the ninth sensor for the last 50000 points is predicted using equation (7) and the modes of the first 60000 points. The number of modes which gives the minimum mean square error (MMSE) is chosen. The studies suggest three PCA modes. The reconstructed response of the faulty sensor is now compared to its original response in Figure 4. Even though, 50000 points are faulty, for the sake of clarity, only 500 points are represented in the figure. A close look at Figure 4 clearly indicates that the two curves compares very well. The difference between them is marginal and not visible.

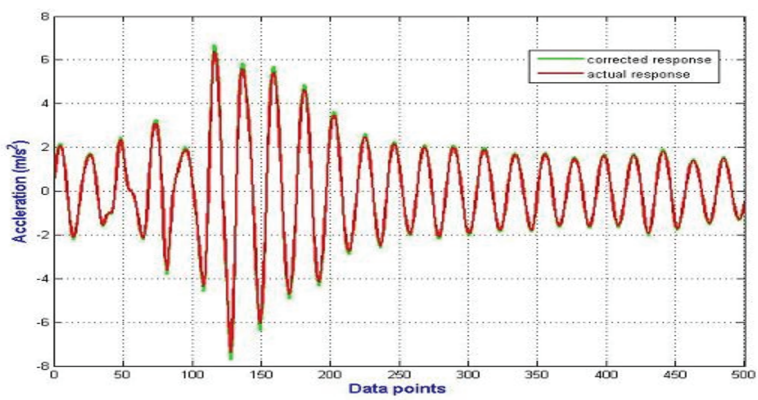
Similarly, we have considered the case-B and case-C, where three adjacent sensors and three arbitrary located sensors are faulty. The faulty sensors are identified using sequential elimination technique. The faulty sensor data is corrected using



(a) Gain plot (b) Sensor plot

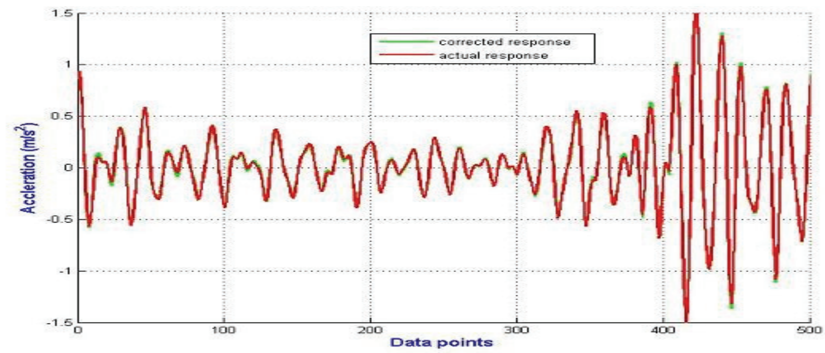
Figure 3: Faulty sensor isolation

multi variant approach discussed in the earlier section and also iterative refinement of the corrected data to improve the solutions. The results obtained for case-B and case-C are shown in figures 5 and 6 respectively



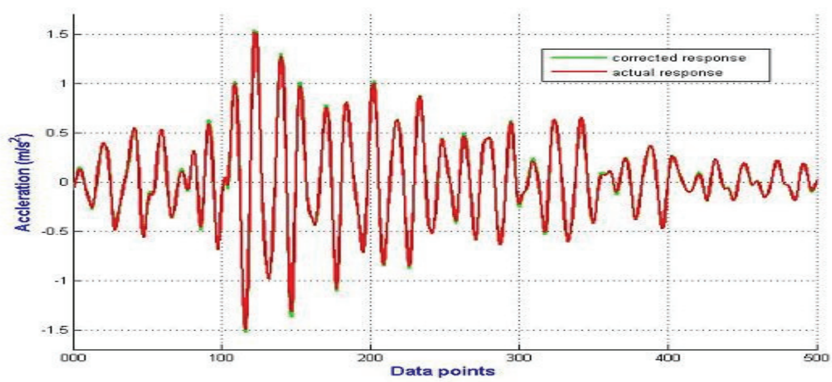
Case-A

Figure 4: Reconstruction of faulty data of 9^{th} sensor



Case-B

Figure 5: Reconstruction of faulty 9^{th} sensor data after identification of 9, 10, and 11 as faulty sensors



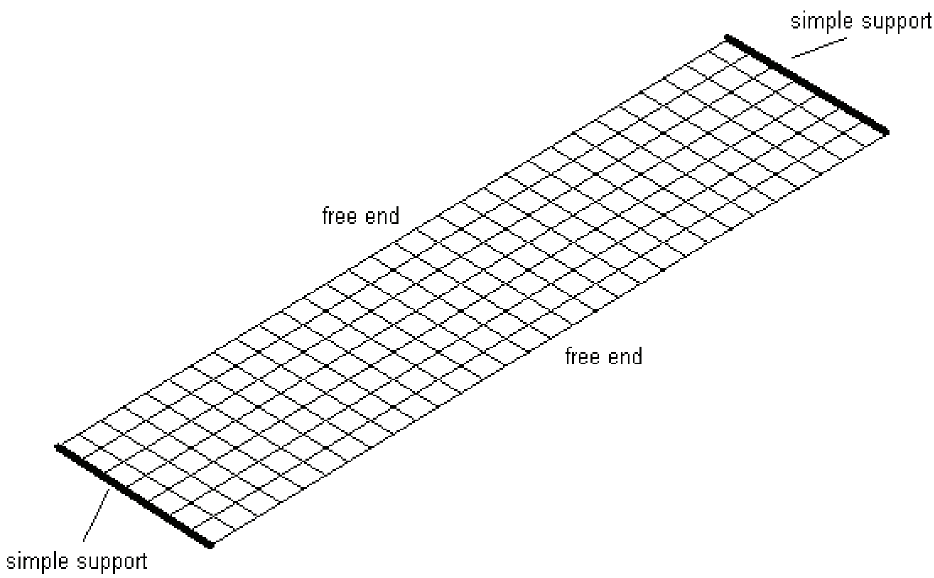
Case-C

Figure 6: Reconstruction of faulty 15^{th} sensor data after identification of 7, 10, and 15 as faulty sensors

3.2 Slab bridge

The second numerical example considered is a slab bridge shown in Figure 7. In this example, the number of accelerometers is considered as 24. The reference data contains 110000 samples collected from each of the 24 channels. The data is partitioned into 11 different sets containing 10000 points each, which gives us a collection of 11 different angle values. The number of PCA modes is considered as five. An alarm will thus be issued indicating presence of faulty sensors, when the monitored angle exceeds the following angle (in degrees):

$UCL_3 = 0.7450 + 3 \times 0.6630 = 2.734$

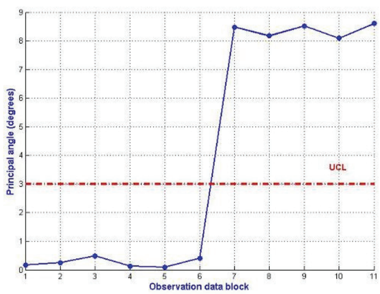


Dimensions of Slab Bridge: 15 m length by 3.5 m width

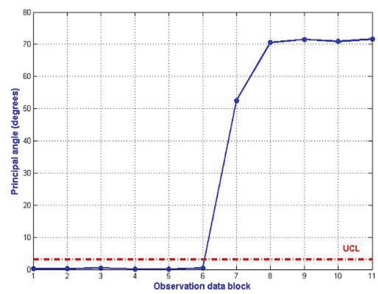
Number of active nodes: 232 Thickness: 0.25 m
Elastic modulus: 25.1 GPa Poisson's ratio, ν : 0.18

Figure 7: Slab bridge

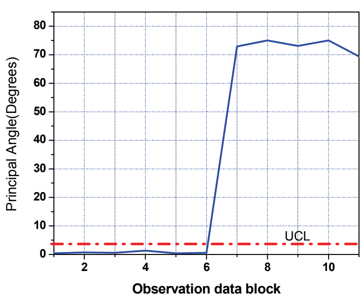
The current data set contains 110000 samples and the last 60000 samples correspond to a sensor fault. The current data set is partitioned into 11 different sets containing 10000 points each. Figure 8 displays the principal angle between the subspace spanned by each of these sets and the subspace spanned by the whole reference set for both sensor faults. In both cases, it can be observed that the first six subsets (60000 samples) are well below the UCL and the rest of of five data sets are above the UCL. Once the faulty sensor state is identified, the exact locations of faulty sensors need to be identified using the same methodology employed earlier for simply supported beam. Figure 9 presents all the subspace angles computed, when data pertaining to one sensor is removed at a time from the available current subset data. For both faults, it clearly appears that the angle approaches zero when the eleventh sensor is removed. The reconstruction of the faulty data for eleventh sensor is carried out using equation (7) and the reconstructed sensor data is compared with the original data at the sensor location and presented in Figure 10. It can be observed that the reconstructed data is almost same as the original data.



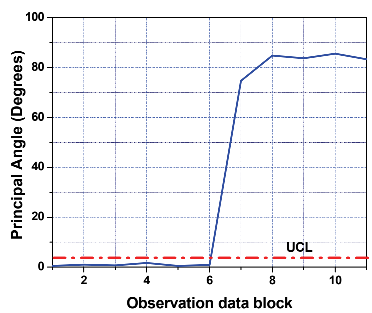
(a) Sensor Gain plot



(b) Sensor Failure plot



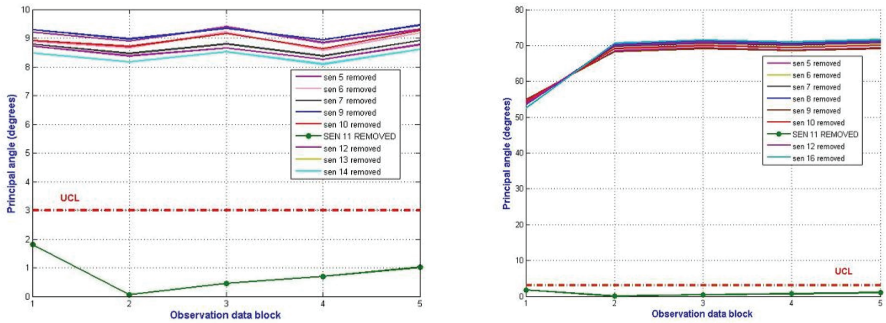
(c) Sensor Additive plot



(d) Sensor Drift Failure plot

Figure 8: Monitoring the subspace angle - Slab bridge

It can be observed from the above numerical simulation studies that the sensor correction scheme based on subspace angles can detect all possible sensor errors. As mentioned earlier, the major problem associated with the PCA based sensor



(a) Gain plot (b) Sensor failure plot

Figure 9: Faulty sensor isolation- Slab bridge

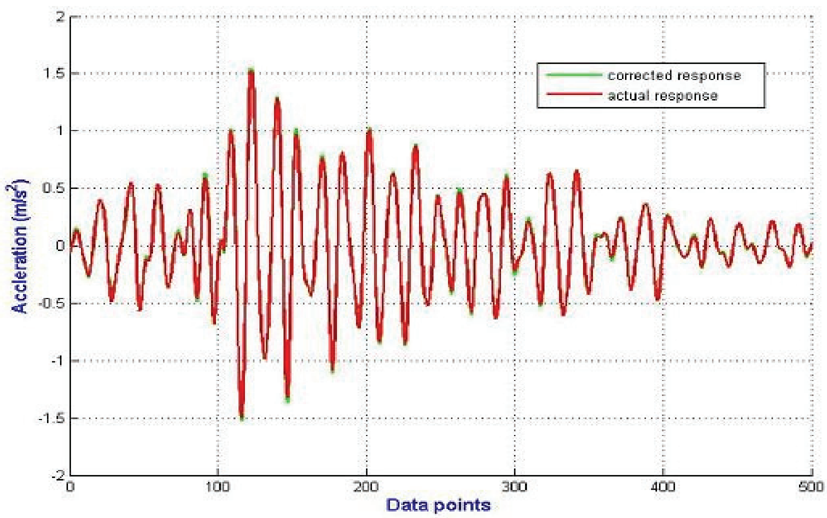


Figure 10: Reconstruction of faulty eleventh sensor data- slab Bridge

fault detection technique is the high computational cost associated in isolating the sensors.

The new multi-level search algorithm proposed in this paper can overcome this problem to a large extent. The gain in terms of computational efficiency increases with the size of sensor network and also with number of faulty sensors. In order to investigate the performance and robustness of the proposed search algorithm, we have considered the above two examples and also small to larger sensor networks. The smaller sensor network configuration consists of ten sensors on a beam and the larger sensor network configuration on slab bridge consisting of 1000 sensors. We have also tested with various sizes of faulty sensor subsets at various spatial locations to test the robustness of the algorithm in isolating the faulty sensors. Further, number of trials, has been conducted, choosing each time randomly a fixed subset of faulty sensors to test the robustness and also performance of the algorithm. The number of trails for each sensor depends on the total size of the sensor network and also size of the faulty sensor set. The details of the studies carried out are furnished in Table 2. The computational performance indicated in the table are the number of times the subspace angles need to be computed and the entries shown in the table are the average performance over specified number of trails made on each sensor set configuration with varied size of faulty sensor set. In order to test the reliability of the algorithm, we have computed the reliability index and shown in parenthesis for each of the entry in the table. The reliability index is defined as the ratio of number of successful trials to the total number of trials made for a given set of sensor network with the specified faulty sensor configuration. The following observations can be made from the details furnished in the table.

The reliability of the algorithm increases with the increase in the depth parameter k. This is obvious as k increases, the number of combinations of sensor sets increases and there by the chances of identifying the faulty sensor at any spatial location increases.

One can choose large value of k to improve robustness. But the computational overhead increases with the increase in the value of k.

The reliability index of the algorithm depends on the size of faulty sensor set. The reliability index is 1.0, when there is only one faulty sensor irrespective of the variation of depth parameter k, while it reduces with the increase in the size of the faulty sensor set.

While devising the algorithm it is assumed that the two end sensors in each sensor combination belongs to working sensors set group. However, it may not happen always and can easily be judged from the results obtained in the first level. If all the sensor combinations are found to be faulty it can be concluded that the end sensors are not working sensors. If this happens the algorithm stalls. To overcome this, the two end sensors will be removed from the combinations and tested for faulty groups.

Choosing an optimal value for the depth parameter, k is essential in reducing the number of computations. The depth parameter, k chosen twice the size of specified faulty sensor set in the network appears to be optimal. The reliability index of the

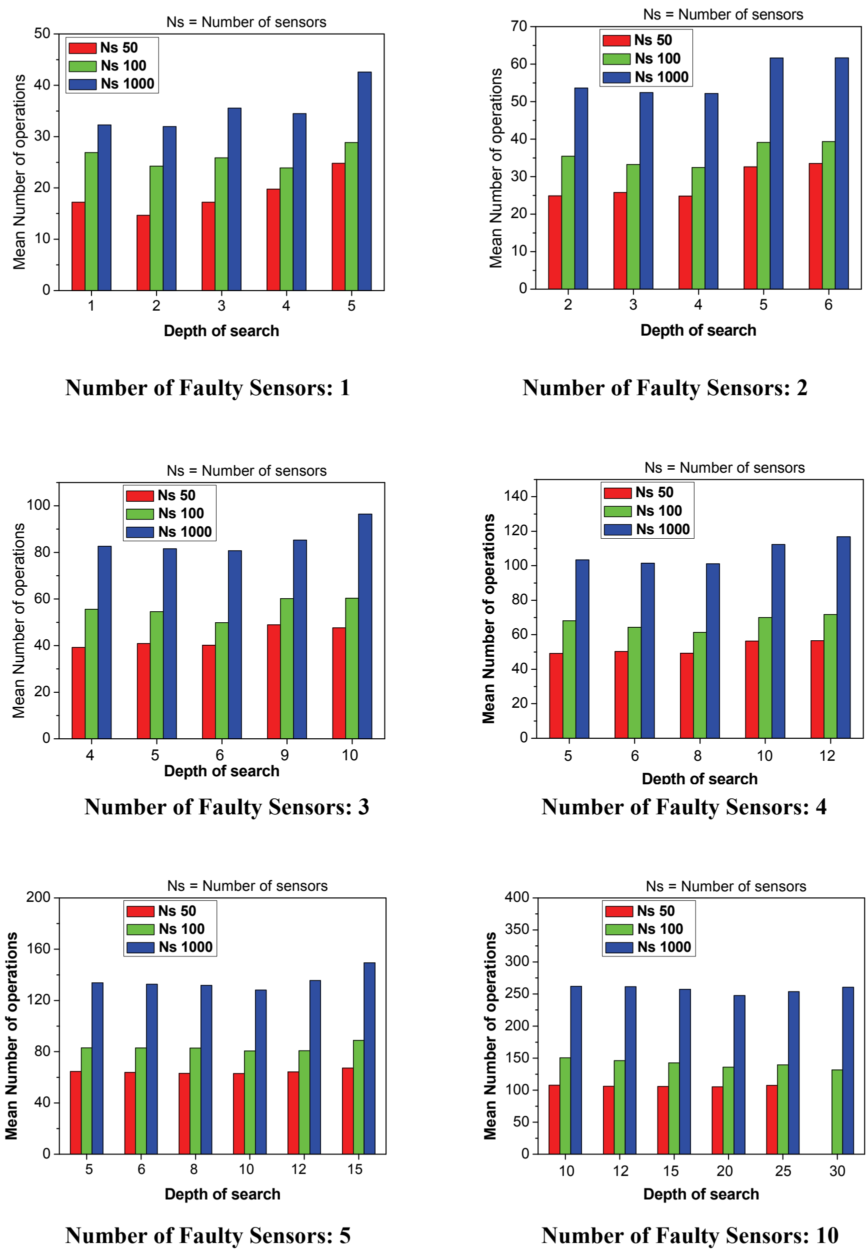


Figure 11: Performance of the proposed search algorithm with different values of k (Depth parameter)

Table 2: Performance of the proposed multi-level search algorithm for sensor fault isolation

Depth of search (k)	Total number of sensors:10				Total number of sensors :50						Total number of sensors :100						Total number of sensors :1000					
	Number of faulty sensors				Number of faulty sensors						Number of faulty sensors						Number of faulty sensors					
	1	2	3	4	1	2	3	4	5	10	1	2	3	4	5	10	1	2	3	4	5	10
1	6 (1.0)	-	-	-	17.21 (1.0)	-	-	-	-	-	26.88 (1.0)	-	-	-	-	-	32.26 (1.0)	-	-	-	-	-
2	10 (1.0)	-	-	-	14.66 (1.0)	24.9 (0.86)	-	-	-	-	23.23 (1.0)	35.48 (0.91)	-	-	-	-	30.95 (1.0)	53.65 (0.92)	-	-	-	-
3	11.4 (1.0)	13.78 (0.98)	34.12 (0.97)	41.1 (0.98)	17.22 (1.0)	25.8 (0.98)	39.2 (0.64)	-	-	-	25.86 (1.0)	33.22 (0.97)	55.6 (0.90)	-	-	-	35.54 (1.0)	52.44 (0.94)	82.65 (0.93)	-	-	-
4	16 (1.0)	18.8 (1.0)	20.88 (0.98)	36.1 (0.96)	19.76 (1.0)	24.8 (1.00)	42.4 (0.91)	49.2 (0.62)	-	-	23.92 (1.0)	32.46 (1.0)	55.2 (0.93)	68.1 (0.94)	-	-	34.48 (1.0)	50.18 (1.00)	82.02 (0.95)	103.45 (0.92)	-	-
5	20.6 (1.0)	22.9 (1.0)	24.3 (1.0)	33.1 (1.0)	24.78 (1.0)	32.65 (0.99)	41.89 (0.99)	49.4 (0.98)	64.5 (0.65)	-	28.82 (1.0)	39.13 (0.99)	54.6 (0.99)	66.2 (0.97)	83.1 (0.91)	-	42.58 (1.0)	61.64 (1.00)	81.54 (0.97)	102.4 (0.96)	133.8 (0.95)	-
6	-	-	23.2 (1.0)	34.2 (1.0)	28.98 (1.0)	33.54 (1.0)	40.92 (1.00)	50.33 (0.99)	63.9 (0.94)	-	-	39.34 (1.0)	49.87 (1.0)	64.31 (0.98)	82.99 (0.93)	-	-	61.70 (1.0)	80.05 (1.00)	101.54 (0.97)	132.7 (0.97)	-
8	-	-	-	26.6 (1.0)	-	-	47.86 (1.0)-	50.29 (1.00)	63.18 (0.99)	-	-	-	54.37 (1.0)	67.36 (1.0)	82.87 (0.99)	-	-	-	82.34 (1.0)	100.94 (1.0)	131.75 (0.99)	-
9	-	-	-	-	-	-	48.92 (1.0)	55.01 (1.0)	62.89 (0.99)	-	-	-	60.18 (1.0)	70.01 (1.0)	82.19 (0.99)	-	-	-	85.29 (1.0)	105.29 (1.0)	130.16 (0.99)	-
10	-	-	-	-	-	-	47.63 (1.0)	55.38 (1.0)	60.97 (1.0)	107.9 (0.74)	-	-	60.39 (1.0)	70.05 (1.0)	80.62 (1.0)	150.53 (0.63)	-	-	96.42 (1.0)	112.39 (1.0)	128.10 (1.0)	262.1 (0.32)
12	-	-	-	-	-	-	-	56.58 (1.0)	64.29 (1.0)	106.3 (0.91)	-	-	-	71.71 (1.0)	80.73 (1.0)	145.89 (0.97)	-	-	-	116.78 (1.00)	135.53 (1.0)	261.5 (0.9)
15	-	-	-	-	-	-	-	-	67.20 (1.00)	105.73 (1.0)	-	-	-	-	88.84 (1.0)	142.42 (0.99)	-	-	-	-	149.422 (1.0)	257.42 (1.0)
20	-	-	-	-	-	-	-	-	-	101.32 (1.0)	-	-	-	-	-	135.82 (1.0)	-	-	-	-	-	247.51 (1.0)
25	-	-	-	-	-	-	-	-	-	107.58 (1.0)	-	-	-	-	-	139.48 (1.0)	-	-	-	-	-	253.87 (1.0)
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	141.70 (1.0)	-	-	-	-	-	260.73 (1.0)
Number of trials	10	50	100	100	50	200	200	200	200	200	500	500	500	500	500	500	500	500	500	500	500	500
Number of iterations with out search algorithm	10	45	120	210	50	1225	20e03	23e04	21e05	1e+010	100	4950	16e04	39e05	75e06	1.7e13	1000	50e04	166e07	4.2e10	8.25e12	2.63e+23

Note:

- The values given in the table indicates the average number of iterations (required to identify faulty sensors) over number of trials made
- The values given in the parenthesis indicates the reliability index of the algorithm computed over the number of trials made

algorithm is also found to be 1.0 when k is chosen two times the Faulty sensor set. It is evident from the bar chart presented in Figure 11 and also Table 2.

4 Conclusions

In this paper the sensor fault identification, isolation and correction scheme for large sensor networks are attempted with principal component analysis using subspace angles. Since the state-of-the-art online continuous structural health monitoring strategies are based on large sensor networks, it is desirable that the sensor correction scheme is robust and fast enough to handle the sensor faults and correct the data online, if necessary. The studies presented in this paper, clearly indicate that the PCA based algorithm is robust and can detect and isolate all the possible sensor faults. However, it requires repeated formation of subspaces and computation of angle of current subspace and reference subspace to isolate the faulty sensor. In view of this, the computational cost is likely to increase prohibitively with the increase in the sensor network size and also size of faulty sensors. In order to overcome this problem, a fast multi-level search algorithm is proposed in this paper. Numerical simulation studies have been carried out with sensor network configurations ranging from small to large size of sensor networks with varying size of faulty sensor set. The reliability of the algorithm is also investigated by conducting several trials on each sensor network choosing the faulty sensors at various sensor locations randomly in each trail. Studies clearly indicate that the proposed multi level search algorithm is robust and also highly reliable. The computational cost of the PCA based sensor fault detection algorithm with the proposed multi-level search strategy, reduces remarkably and the amount of reduction in computational cost increases with the increase in the sensor network size and also the size of faulty sensor set. The proposed multi-level search algorithm is extremely useful for online sensor fault detection, isolation in current SHM schemes.

Acknowledgement: This paper is being published with the permission of the Director, CSIR-Structural Engineering Research Centre (SERC), Chennai. This work is partially supported by the NPMASS programme, Ministry of Defence, Govt. of India(PARC#3:18)

References

Abdelghani, M., Friswell, M.I. (2004): Sensor validation for structural systems with additive sensor faults, *Structural Health Monitoring, 3(3),* 265–275.

Abdelghani, M., Friswell, M.I. (2007): Sensor validation for structural systems with multiplicative sensor faults, *Mechanical Systems and Signal Processing*, 21, 270–279.

Dunia, R., Qin, S.J., Edgar, T.F., McAvoy, T.J. (1996): Identification of faulty sensors using principal component analysis, *Ai Che Journal*, 42(10), 2797–2812.

Dunia, R., Qin, S.J. (1998): Joint diagnosis of process and sensor faults using principal component analysis, *Control Engineering Practice, 6,* 457–469.

Feeny B.F., Kappagantu R. (1998): On the physical interpretation of proper orthogonal modes in vibrations, Journal of Sound Vibration, 211 (4), 607–616.

Friswell, M.I., Inman, D.J. (2000): Sensor validation of smart structures, in: Proceedings of IMAC-XVIII: *A Conference on Structural Dynamics, San Antonio, Texas,* 483–489.

Friswell, M.I., Inman, D.J. (1999): Sensor validation of smart structures, *Journal of Intelligent Material Systems and Structures, 10,* 973–982.

Golub G.H., Van Loan C.F. (1996): Matrix computations (3rd ed.), The Johns Hopkins University Press, Baltimore.

Hernandez-Garcia, M.R., Masri, S.F. (2008): Multivariate statistical analysis for detection and identification of faulty sensors using latent variable methods, *Advances in Science and Technology, 56,* 501–507.

Kerschen, G., DeBoe, P., Golinval, J., Worden, K. (2005): Sensor validation using principal component analysis, *Smart Materials and Structures, 14,* 36–42.

Kerschen, G., DeBoe, P., Golinval, J.C., Worden, K. (2004): Sensor validation for on-line vibration monitoring, in*: Proceedings of the Second European Workshop on Structural Health Monitoring, Munich, Germany*, 819–827.

Kramer, M.A. (1992): Auto associative neural networks, *Comput. Chem. Eng.* 16, 313-328.

Kullaa, J. (2009): Three models for sensor validation, in: Proceedings of the 7^{th} International Workshop on Structural Health Monitoring, Stanford, CA, 529–536.

Kullaa, J. (2007): Sensor fault identification and correction in vibration-based multi channel structural health monitoring, in: *Proceedings of the 6th International Work shop on Structural Health Monitoring, Stanford, CA,* 606–613.

Kullaa, J. (2006): Sensor fault identification and correction in structural health monitoring, in: *Proceedings of ISMA2006, International Conference on Noise and Vibration Engineering, Leuven, Belgium,* 873–884.

Zhiling Li., Koh, B. H., Satish Nagarajaiah. (2007): Detecting Sensor Failure via Decoupled Error Function and Inverse Input–Output Model, *Journal of Engineering Mechanics, ASCE, 133(11), 1222-1228.*