SHM
Structural Health Monitoring

Aggressive Data Reduction for Damage Detection in Structural Health Monitoring

Chiwoo Park,[1] Jiong Tang[2,*] and Yu Ding[1]

[1]*Department of Industrial & Systems Engineering, Texas A&M University College Station, TX 77843, USA*
[2]*Department of Mechanical Engineering, University of Connecticut Storrs, CT, 06269, USA*

While wireless sensors are increasingly adopted in various applications, the need of developing data reduction methods to alleviate data transmission rate issue between the sensors and the data interpretation unit becomes more urgent. This article presents a new data reduction method for sensors used in structural health monitoring application. Our goal is to achieve an effective data reduction capability while maintaining adequate power for damage detection. We propose to establish an explicit measure of damage detection capability for the features in the response signals and use this measure to select the subset of the features that balance between the degree of data reduction and the damage detection capability. We also explore a computationally efficient procedure searching for the best subset of the features. This new method is tested on experimentally obtained Lamb wave signals for beam damage detection. Performance comparisons with respect to the existing methods demonstrate the strength of the proposed method.

Keywords damage detection · data reduction · feature subset selection · energy-efficient sensor network · wavelet shrinkage

1 Introduction

The development of structural health monitoring (SHM) system has received significant attention in recent years due to its broad applications to civil, mechanical, and aerospace structures. The timely detection of damage occurrences in these structures can enhance the safety and security, elongate the product service life, and reduce the operational and maintenance costs.

Generally, a SHM system must include at least two components: the data collection unit and the data interpretation unit. The former is used in the process in which certain data that can reflect the anomaly/damage in the structure are collected. A wide variety of data have been explored for SHM applications, which, when categorized, correspond to various damage detection methods including acoustic emission methods [1,2], magnetic field methods [3], eddy-current techniques [4], thermal field methods [5], vibration/frequency response methods [6], Lamb wave methods [7,8], impedance-based techniques [9,10], etc. In these methods, the data are collected by different types of sensors either through passive approaches (i.e., without actively exciting the structure) or by

*Author to whom correspondence should be addressed.
E-mail: jtang@engr.uconn.edu
Figures 5–7 appear in color online: http://shm.sagepub.com

Vol 9(1): 0059–16
[1475-9217 (201001) 9:1;59–16 10.1177/1475921709341017]

active means (i.e., active interrogation using actuators to excite the structure). With these data, the data interpretation unit then decides whether damage has occurred in the structure. In many cases, the primary concern is to predict the occurrence of damage, and for that, a conformity check between the online measurements and the healthy baseline is used for decision making. In some cases, a direct map between the data anomaly and specific damage location and severity can be established, which can further allow the identification of damage.

Traditional SHM systems employ coaxial wires for the communication and data transmission between the sensors and the decision making (data interpretation) unit. Recently, there has been fueled excitement of developing new SHM systems using the wireless communication technology. It is believed that wireless sensor networks can significantly reduce the installation and operation costs and increase the sensor density as well as the coverage area [11]. While many promising aspects of wireless sensors for SHM applications have been suggested, such idea also leads to a number of challenges. In particular, the issue of data transmission between the sensors and the decision making unit, which may pose challenge even in wired sensors, now becomes extremely serious. One bottleneck in the wireless health monitoring systems is the data transmission rate. The data involved in many SHM systems are time-domain structural responses that have large data sizes. In general, damage detection methods with higher detection sensitivity are often associated with higher sampling rate. For example, it is well known that the sensitivity of vibration/wave propagation based damage detection schemes is closely related to the frequency band, i.e., the wave length of the signal should be smaller than the characteristic length of the damage to be detected [10]. It is not unusual for a wave propagation-based damage detection system to excite and sense wave motions at megahertz range. High-frequency sampling and excitation pose multiple challenges to wireless sensor development, one of which is the timely transmission of large amount of data.

In a wireless sensor network, the handling of large amount of data collected by individual sensors requires the collective considerations from multiple aspects, including local data storage capability, local data processing capability, power consumption, and wireless transmission capability, etc. More specifically, the sensor data, which carry the damage features, are inevitably contaminated by noise. Although directly transmitting the original sensor data to the data interpretation unit can retain the signal fidelity for comprehensive data analysis, it may lead to prohibitive burden to the wireless data transmission, especially for applications that need near real-time decision making. One tempting idea is to perform data pre-processing at the local sensor level, so that the data can be greatly compressed/reduced while the critical features reflecting the damage effects can be preserved. Such data reduction, also known as feature extraction, if effectively established, can alleviate the data transmission rate issue between the sensors and the data interpretation unit. On the other hand, it requires higher computing capability and the associated power consumption at the local sensor level.

It is worth emphasizing that the research and development of wireless sensors for SHM is an on-going, exploratory effort. Both the wireless communication technology and the sensor-related microelectronics are experiencing rapid advancements. Therefore, at the current stage, it is generally difficult to precisely quantify the optimal balance between the local data pre-processing and the transmission reduction. Nevertheless, the current consensus among the wireless communication community is that the power consumption for data transmission is about one magnitude higher than that for local data pre-processing [12]. For this reason, it appears worthwhile to pursue aggressive data reduction (or feature extraction) methods that can significantly reduce the amount of data that need to be transmitted from the local sensor to decision making unit while maintaining adequate degree of damage detection capability. Motivated by this pressing practical need, this article concerns the development of effective data reduction techniques for the purpose of damage detection in SHM.

The rest of the article is organized as follows. In the subsequent section, we review the literature relevant to our research efforts. The detailed description of our proposed data reduction and damage detection method is presented in Section 3. Section 4 includes a case study that demonstrates the performance of our method against the existing methods. The case study is based on an experiment performed using Lamb wave signals to detect structural damages. Finally, we conclude this investigation in Section 5.

2 Relevant Studies

Data de-noising technique is a major set of methodologies used for the purpose of data reduction because measurement signals are inevitably contaminated by noises, and eliminating the noise components apparently reduces the data dimension. There is a rich body of literature focusing on the generic de-noising methodology [13–16], which is to distinguish *purely random noises* from the data components that could potentially be a meaningful signal. Because purely random noises usually have higher frequency and lower energy than a meaningful signal, one can establish a threshold to set apart the noises and the signals. A typical procedure usually starts with a wavelet transformation. One can retain a subset of above-threshold wavelet coefficients while setting the below-threshold ones to zero.

One principal drawback of the generic denoising methods is its ineffectiveness in reducing the data, i.e., that the wavelet coefficients retained are still too many to be handled. Reported by Lada et al. [17], the de-nosing methods could keep 49–67% of the original coefficients; our case study will confirm this understanding. This does not come as a surprise because the de-noising methods intend to be applicable to different applications and thus resort to the most conservative way to define what should be deemed as 'noise.' Without specific application objectives as the guideline, the generic methods simply run a statistical test that identifies the independent, identically distributed components with zero means as the pure noises and treats everything else as signals.

Some recent developments attempt to improve the effectiveness of data. Lada et al. [17] proposed a new criterion, labeled as relative reconstruction error (RRE) as follows:

$$\mathrm{RRE}(p) = \mathrm{MSE} + \lambda \frac{p}{N}$$

where the first term, MSE, stands for the mean square error between the original signal and the signal reconstructed from the chosen coefficients, N is the dimension of the original signal, p is the dimension of the signal after reduction, and λ is a pre-determined weighting coefficient. The second term in RRE is actually the data reduction ratio; the more aggressive a data reduction, the smaller p or p/N is. Lada et al. [17] recommended choosing the subset of wavelet coefficients that minimize the RRE. By including the data reduction ratio as a penalty term in RRE, it helps prevent too many wavelet coefficients being kept for the sake of attaining a very marginal decrease in MSE. In essence, this criterion is the same as the Akaike information criterion (AIC) [18].

Later, Jeong et al. [19] found that the effectiveness in using the RRE for data reduction can be further enhanced. Their idea is to revise the way to establish the threshold for wavelet coefficients. Simply put, both the de-noising methods and the RRE criterion use some sort of threshold when it comes to choose the wavelet coefficients. The RRE's threshold is larger (meaning that it would choose fewer coefficients) than the denoising methods because it includes the data reduction ratio as a penalty. The threshold works as a boundary deciding if the wavelet coefficients keep their original values or set to zero. If the value of a wavelet coefficient is over the threshold, the coefficient keeps its value; otherwise, it is set to zero. This procedure is called hard-thresholding. Jeong et al. [19] pointed out that hard-thresholding causes large variance of the selected coefficients and is very sensitive to small changes in data. In order to make sure that hard-thresholding will not adversely affect the signal quality, certain degree of conservativeness is still needed. Jeong et al. [19] proposed a different RRE based on a soft-thresholding procedure, which first subtracts the threshold from each of the wavelet coefficients and sets the negative values among the results to zero. They derived the new objective function of RRE when soft-thresholding is used. In the new objective, the ratio of the first norm of the remaining values after soft-thresholding and the first norm of all original values is used as the penalty, rather than just a data reduction ratio. They showed that their soft-thresholding method worked better for data reduction purpose. The new RRE is called RRE_s.

Due to their general applicability, the de-nosing methods are expected to be applicable to SHM, e.g., in [20]. On the other hand, both RRE criteria were developed in the context of semiconductor manufacturing, and to our knowledge, they have not yet been applied to SHM applications but in principle they could be used. Since the dimension of the features extracted should be smaller than the original dimension, feature extraction is often used as an *interchangeable* phrase for data reduction. Researchers have explored extracting damage features from structural vibration responses to identify structural damages as well as to transmit sensory data in energy-efficient ways. Classical ways to extract damage features typically involves using vibration modal parameters such as dominant frequencies, mode shapes and displacements. Yam et al. [21] compared the energy spectrums of vibration signals from a normal beam and damaged ones for each time slot. In a wavelet transformation, the wavelet coefficients can be considered features.

Motivated by similar data-reduction needs as mentioned in Section 1, researchers in the SHM field also introduced some simple scalar values to summarize and characterize the original signals and argued that these scalar values can be used for structural damage detection. Wang et al. [22] introduced the distributed damage index detection (DDID) as a measure for the possibility of damage occurrence. DDID is a 1D counting number measuring the difference of raw data between normal structures and the current one, and is calculated at each sensor node and transferred to a base station. Sohn et al. [23] developed a wavelet-based damage index to quantify the attenuation of Lamb waves by structural damages and conducted damage detection by comparing their index with the baseline index deduced from a normal structural. Bukkapatnam et al. [24] used the distortion energies to quantify the damage-induced distortions by examining a wavelet representation of the difference in strain responses between the damaged and the undamaged structures. The distortion energies are calculated for each of several resolution levels of a wavelet representation. The number of the resolution levels is normally less than 10 so that its effectiveness for data reduction is presumably much better than the generic de-noising methods or using the RRE criteria.

In summary, the above-reviewed approaches can be grouped into two categories: (1) The category that tends to keep too many features, not all of which are relevant to damage detection. This category includes the de-noising methods and the RRE criteria. Obviously, if a method is capable of searching for the set of features necessary for damage detection but discarding those irrelevant ones, a further data reduction can be achieved. (2) The second category that uses a summary index or a few indices for damage detection. This group includes DDID, Sohn's damage index and the distortion energy indices. The potential limitation with the second group is that even though those indices are intuitively related to structural integrity, there is no formal linkage between the indices and the detection capability. In other words, they are surrogate measures of damage detection at best. Because of this, as we will show later in the case study, these indices may not be able to capture the optimal set of features needed for damage detection, and thus only lead to suboptimal detection performance in terms of false positive or false negative probabilities. We also believe that using 1D index is too aggressive to capture all the relevant features caused by complex changes in a damaged structure.

The aim of this article is to present a new method to find the minimal subset of the features (in the form of wavelet coefficients) relevant to structural damages. We develop an explicit measure for damage detection power and then link it to the selected features. Subsequently, an iterative algorithm is devised to search for the smallest feature subset, which in the meantime maintains adequate detection capability. Based on the explicit link between the detection power and the damage features, we expect that the proposed method is able to overcome the shortcomings of the above-mentioned data reduction methods in both categories; our case study in Section 4 demonstrates such improvement.

3 Data Reduction Method

3.1 Method Overview

In this section, we present a wavelet-based feature selection method for structural damage detection. It attempts to retain the smallest possible number of features, or equivalently, to achieve the highest possible data reduction while still maintaining adequate damage detection capability. Given its high data reduction ability, the method is suitable for enabling an energy-efficient SHM based on wireless sensor networks.

In our method, we define the feature space as the set of wavelet coefficients resulting from the wavelet transform of the original signals. Please note that the wavelet transform is broadly used in the SHM applications [21,23,24]. We also perform a generic de-noising procedure to remove the purely random noises as part of the data pre-processing. Doing so can help improve the consistency of the subsequent data reduction (or feature extraction) actions. De-noising usually regards as 'noise' the wavelet coefficients of zero mean and lower energy than the given threshold. If we set the threshold too high, we might overlook small local damages of very low energy. If we set it too low, we keep too many wavelet coefficients and the de-noising does not have effects on our method. Therefore, choosing a good threshold is very important. In our method, we used the threshold defined in VisuShrink [13], which is one of the widely used de-noising methods developed by Donoho. The reason we choose this method is because it is the most computationally efficient among the few widely used de-noising methods. In terms of performance, all the de-noising methods are comparable.

After obtaining the pre-processed, 'clean' structural monitoring signals, we present our method by addressing the following two questions: (1) what should be the measure or criterion for feature selection; (2) how to find the best (i.e., the smallest) subset in a computationally efficient way.

The major steps in our method are outlined in Figure 1. In Section 3.2, we derive the formula for computing the damage detection power for a given set of features (i.e., a given subset of wavelet coefficients). The damage detection power is established based on a statistical monitoring and detection formulation. It utilizes the baseline statistics established by using the data collected from intact, healthy structures.

Section 3.3 presents the searching algorithm to find the best subset of coefficients. After a feature selection criterion is set up in Section 3.2, one may use an exhaustive search to pinpoint the best one, which, however, requires excessive computational cost and thus is not suitable for local sensors with limited capacity. We instead devise a forward selection greedy search. The search algorithm starts with an empty feature selection and iteratively selects one additional feature (i.e., a wavelet coefficient) that can improve the selection criterion

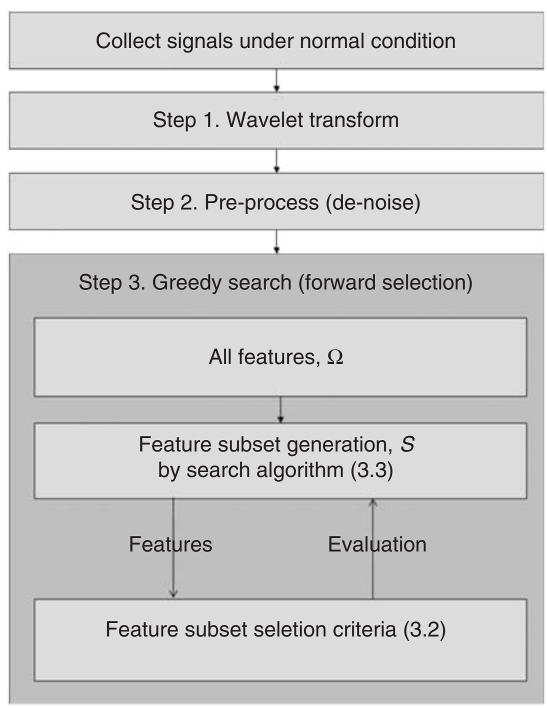

Figure 1 Feature subset selection procedure for damage detection. (3.2) and (3.3) refer to the corresponding subsections where the detailed algorithms or procedures are presented.

(defined in Section 3.2) the most. The iterations terminate when there is no sensible improvement in the selection criterion.

3.2 Feature Selection Criterion

Our primary objective is to detect structural damage using a subset of the features, S, from the set of all features, Ω (recall that here a feature is a wavelet coefficient and we use these two terminologies interchangeably hereafter). Intuitively, the best subset should be the one containing only all the necessary features required to discriminate a damaged structure from the intact structures. As discussed before, a sensible way to do so is to link explicitly the selection of features to the detection capability using that set of features.

In light of this, we want to compute the damage detection capability for a given subset S. Let $Q(S)$ represent the damage detection power for S. According to the statistical detection theory [25], the detection power is typically defined as the complement of the probability of misdetection (or false negative), also known as the β error probability. Suppose that $H(S)$ is a statistical hypothesis testing to decide whether structural damages exist using the feature set S. Then, the damage detection power of S, $Q(S)$ is

$$Q(S) = 1 = \beta(H(S)). \tag{1}$$

In order to perform a statistical detection, a baseline needs to be established; such a baseline is used to characterize what constitutes the normal structural behavior. This can be done by collecting a set of sensory data from healthy, intact structures. Here we assume that the wavelet coefficients from an intact structure follow $N(\underline{\mu}, \sum)$ where $\underline{\mu}$ and $\sum$ are the mean vector and covariance matrix, respectively. With the historical data $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$, from the intact structures, we can estimate $\underline{\mu}$ using the sample average $\hat{\underline{\mu}}$, and estimate $\sum$ using the sample covariance matrix $\hat{\sum}$ as

$$\hat{\underline{\mu}} = \frac{1}{m}\sum_{i=1}^{m}\underline{x}_i \quad \text{and} \quad \hat{\sum} = \frac{1}{m-1}\sum_{i=1}^{m}(\underline{x}_i - \hat{\underline{\mu}})(\underline{x}_i - \hat{\underline{\mu}})^T \tag{2}$$

As such, the multivariate normal distribution with parameters $\hat{\underline{\mu}}$ and $\hat{\sum}$ and serves as the baseline for future damage detection.

When damage occurs in a structure, it presumably causes a different behavior in the structural response and thus a different distribution in the observed sensory measurements. The statistical detection theory [26] tells us that the maximum likelihood statistic for testing the hypothesis $H(S)$, namely whether a new observation deviates from the baseline distribution, is the Hotelling's T^2 statistic. Consider that $\underline{x}$ is the newly measured structural response. Then the Hotelling's T^2 statistic is [27]:

$$T^2 = (\underline{x} - \hat{\underline{\mu}})^T \hat{\sum}^{-1} (\underline{x} - \hat{\underline{\mu}}) \tag{3}$$

It has also been established that this Hotelling's T^2 statistic follows an F distribution. Specifically,

$$\frac{p(m-1)(m+1)}{m(m-p)} T^2 \sim F_{p,m-p} \tag{4}$$

where p is the number of features in the set S and m is the number of the samples used to estimate the parameters $\hat{\underline{\mu}}$, $\hat{\sum}$. Suppose one has chosen the confidence level of detection (namely the probability of α error or false positive) as α_0. Then the decision boundary, also called upper control limit (UCL), is given as:

$$\text{UCL} = \frac{p(m-1)(m+1)}{m(m-p)} F_{p,m-p,\alpha_0} \tag{5}$$

where $F_{p,m-p,\alpha_0}$ is the inverse cumulative distribution function (cdf), deciding the critical value of an F distribution at α_0. Given the T^2 statistic value for any new sensor observations and UCL, the damage detection procedure is straightforward: when $T^2 \geq \text{UCL}$, we can conclude there is damage with (1−α)100% confidence; otherwise, we can conclude there is no damage.

Given the above statistical detection framework, the question is how to compute the β error probability so that the detection power can be quantified. The β error probability can be computed for any given change in the distribution: a shift in the mean component or a change in the covariance matrix. Please note that an advantage of using the Hotelling's T^2 statistic is that it is sensitive to detecting changes in both distribution parameters (mean and variance), i.e., it is unlikely to leave out any damage undetected unless such damage does not cause any distribution change from the baseline.

Generally, a structural damage causes a sustained shift in the mean parameter. Suppose that the mean parameter shifts to a new value, which is different from the baseline by $i\underline{\hat{\sigma}}$, where i is a constant coefficient and $\underline{\hat{\sigma}}$ is the estimate of the baseline variability. Thus, $i\underline{\hat{\sigma}}$ is a multiple of the standard deviation of the baseline. As such, the new distribution mean is $\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}$. Given such change, the β error probability is the conditional probability when a change occurs but the detection method fails to detect (i.e., the T^2 is smaller than the UCL), namely

$$\beta(H(S)) = P(T^2 \leq \mathrm{UCL} \big| \underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}). \quad (6)$$

Utilizing the properties of multivariate normal distribution, the β error probability can be further derived as

$$\begin{aligned}
\beta(\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}) &= P\Big((\underline{x} - \underline{\hat{\mu}})^T \hat{\sum}^{-1} (\underline{x} - \underline{\hat{\mu}}) \\
&\quad \leq \mathrm{UCL} \Big| \underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}\Big) \\
&= P((\underline{x} - \underline{\mu})^T \hat{\sum}^{-1} (\underline{x} - \underline{\mu}) + i^2 \underline{\hat{\sigma}}^T \\
&\quad \hat{\sum}^{-1} \underline{\hat{\sigma}} \leq \mathrm{UCL} \Big| \underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}) \\
&= P\Big(T^2 \leq \frac{p(m-1)(m+1)}{m(m-p)} F_{p,m-p,\alpha_0} - i^2 \underline{\hat{\sigma}}^T \hat{\sum}^{-1} \underline{\hat{\sigma}}\Big) \\
&\quad F_{p,m-p}\left[F_{p,m-p,\alpha_0} - \frac{m(m-p)}{p(m-1)(m+1)} i^2 \underline{\hat{\sigma}}^T \hat{\sum}^{-1} \underline{\hat{\sigma}}\right].
\end{aligned} \quad (7)$$

In the above equation, $F_{p,m-p,\alpha_0}$ is the part of the decision boundary and its value depends on the confidence level α_0 previously specified.

One difficulty of using the above β error formulation is that we usually do not know the exact magnitude of change that actually occurs in the distribution parameter caused by structural damage. It depends on the types and severities of damages. In practice, there could be different ways of handling this problem. For example, we can specify a magnitude of change as the worst case. Typically, the smaller the magnitude, the less likely a change can be detected. So if we choose a small enough threshold magnitude to be detected and make sure that our detection method can successfully detect it with an acceptably low β error, then it is reasonable to conclude that other larger magnitude of change can be detected as well. Another treatment is to calculate a weighted average β error probability for a broad range of changes. For example, the multiple i can take values from 1 to a large value I. Then, a weighted average β error probability

$$\overline{\beta} = \sum_{i=1}^{I} \alpha_i \beta(\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}}), \quad (8)$$

where α_i is the weight associated with the i-th magnitude of change. Usually, we may assign more weights to a small change and less weights to a large change since small changes are inherently difficult to detect. The choice of the value I can be different in different applications. We recommend choosing the value by checking the magnitude of $\beta(\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}})$ As the multiple i increases, the β value decreases. Therefore, we can set the range value I to the smallest i so that $\beta(\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}})$ is below a certain level (e.g., 0.01). Then, for, $i > I$, the magnitude of $\beta(\underline{\mu} = \underline{\hat{\mu}} + i\underline{\hat{\sigma}})$ becomes ignorable, so Equation (8) provides a good approximation.

For our application, we recommend using the weighted average $\overline{\beta}$ because the worst-case approach likely leads to a conservative data reduction performance. Using $\overline{\beta}$, we can obtain the damage detection power as

$$Q(S) = 1 - \overline{\beta} \quad (9)$$

On the other hand, the effectiveness of our data reduction (or feature extraction) method is measured by the data reduction ratio, defined as the ratio of the reduced and the total size of the feature space:

$$R(S) = \frac{|S|}{|\Omega|} \quad (10)$$

where $|\cdot|$ is the operator for cardinality.

Our feature selection is realized through minimizing $R(S)$, subject to that the detection power $Q(S)$ is larger than a required level. This gives us a constrained minimization problem to solve. Here, we adopt the approach used by the RRE criterion, which is to define a unified feature selection criterion combining both $Q(S)$ and $R(S)$ as (in fact we use $1 - R(S)$ in the following criterion because we want to ensure both terms to be maximized subsequently):

$$C(S) = Q(S) + \lambda(1 - R(S)) \quad (11)$$

Similar to RRE, the above criterion is to use the data-reduction ratio as a penalty in the cost function to force an aggressive data reduction. But the difference lies in that the first term $Q(S)$ in the above criterion is not the same as the MSE in RRE: the $Q(S)$ measures the damage detection power of using the features in S, whereas the MSE measures the signal differences, a large portion of which may not be relevant to the damage detection objective.

Remark. In Equation (11), λ is a positive real number and it functions as weighting between data reduction capability and damage detection power. The appropriate values of λ can be different from applications. In applications considering data reduction more important, its value should be larger. If damage detection is more important, a small λ should be chosen. In this article, we use $\lambda = 1$ equally weighting data reduction and damage detection power.

3.3 Subset Generation and Searching Procedure

Given the criterion defined in Equation (11), the feature subset selection problem can be formulated as follows:

$$S^* = \arg\max_{S \subset \Omega}[C(S)] \quad (12)$$

Equation (12) entails a combinatorial optimization problem. Exhausting all possible subsets of Ω to find the proper S^* could be computationally demanding. Suppose that the number of the elements in Ω is n. Then, the total number of the subsets to search through in a complete exhaustive manner is 2^n, which can be a very large number even for a moderate n, say 50.

Considering the possible resource limitation on a sensor node, rather than trying all possible subsets, we devise a greedy search (forward selection) method because of its computational simplicity. The forward selection starts with an empty set and adds one feature from Ω to S, which can result in the largest increase in the objective function in (12). The method will then remove the feature just selected from Ω. In the subsequent iterations, this operation will be repeated and it augments the set S by adding one new feature at a time. In each iteration, the augmentation procedure generates r new subsets if the number of the features remaining in Ω is r, because each new subset contains the features selected from previous iterations plus a new one chosen from the features remaining in Ω. The iterations continue until any new addition to S can no longer improve the objective function. That is, if the improvement of damage detection power by adding a new feature does not outweigh the loss of data reduction rate, the feature selection procedure stops.

The pseudo code of this search algorithm is presented in Figure 2. In this pseudo code, $C(S)$ means the result of evaluation of Equation (11) and MAX is the incumbent maximum value of $C(S)$. The algorithm starts with MAX $= 0$ and $S = \emptyset$, which means that no feature is chosen in the subset initially. In each iteration, the algorithm evaluates $C(S_{+x})$, where the function $C(S_{+x})$ is evaluated using Equation (11) and S_{+x} is the set of the features generated by adding x (in T) to S. Suppose that MF is the newly added feature x that maximizes $C(S_{+x})$ among all the evaluations. Then, the algorithm removes MF from T and adds it to S. The next iteration starts with the updated sets of features, S and T. The iteration continues until we cannot find MF that increases $C(S_{+x})$.

Figure 3 illustrates a simple example of the search algorithm, where the feature space has seven coefficients $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$. It shows the first two iterations of the search algorithm. For the first iteration, the algorithm generates the seven subsets of the features, $S_{+1}, S_{+2}, \ldots, S_{+7}$, each of which can be generated by adding one feature in T to S. Then, the algorithm evaluates $C(S_{+x})$ for each subset. In the example $C(S_{+1})$ is the largest among the evaluations. Thus, the algorithm sets S to S_{+1} and removes feature 1 from T. The second iteration starts with the updated S and T. By the same procedure as the first iteration, the algorithm first evaluates the function $C(S_{+x})$ for each of the possible subsets, $S_{+2}, S_{+3}, \ldots, S_{+7}$, generated by adding one feature in T to $S=\{1\}$. Since $C(S_{+5})$ is the largest, the algorithm adds feature 5 to S and removes it from T. As a result, the algorithm sets $S=\{1,5\}$ and $T=\{2, 3, 4, 6, 7\}$, and the second iteration ends. The third iteration will start with $S=\{1,5\}$ and $T=\{2, 3, 4, 6, 7\}$ These iterations repeat until the algorithm cannot improve $C(S_{+1})$ any more.

```
S = Ø, T = Ω
Do while C (S) is improved
    MAX = 0,  MF = Ø
    Repeat for each x in T
        S+x = S ∪ {x}
        If C(S+x) > MAX , then MAX = C (S+x),  MF = {x}
    End
    S = S ∪ MF, T = T \ MF
END
```

Figure 2 Subset selection algorithm.

The forward-selection greedy search algorithm is computationally efficient. Nevertheless, because of the 'greedy' nature of the algorithm, its performance and the choice of feature subset sometimes could be sensitive to the use of different historical data samples (employed for establishing the baseline). In order to deliver a consistent performance and make our method

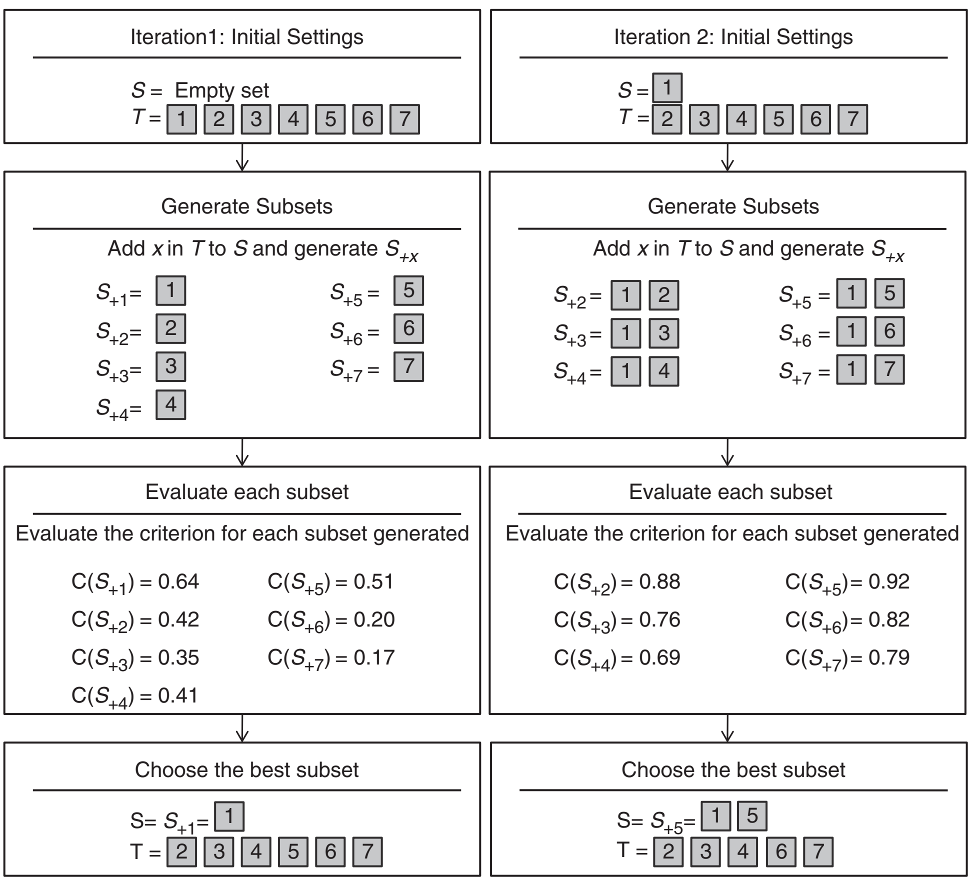

Figure 3 Example of the search algorithm.

robust, here we recommend using a bootstrapping technique [18] to reduce selection variability. Bootstrapping is an effective statistical re-sampling technique to handle variability or inconsistency caused by sampling. Its idea is to divide the baseline data samples into smaller subsets and perform data analysis on each sample subset and then pool the analysis results together in order to cover the difference due to sampling.

Specifically for our application, the first step is to randomly draw K datasets of size M without replacement from the original baseline data samples. The second step is to apply the greedy search algorithm and find the best feature subset for each of the K datasets. Consequently, we will have K resulting subsets, namely $S_1, S_2, \ldots, S_k$ (and these subsets may contain different features). Eventually, we create the final best subset S^* by taking the union of the K best subsets:

$$S^* = \bigcup_{i=1}^{K} S_i \qquad (13)$$

By including this bootstrapping procedure, the number of the features in the best subset will increase but we expect to gain the robustness in performance for damage detection.

4 Case Studies

In this section, we apply the proposed method to the actual signals collected by a laboratory structural health monitoring system in order to verify its effectiveness. We also compare our methods with the existing methods to demonstrate its advantage.

In this experimental system, the Lamb wave approach is used to detect the surface crack damage in a beam structure. The reason we choose the Lamb-wave-based approach for implementation is that the data (time-domain responses) are complicated, large-sized, and typically contaminated by certain level of noise. Meanwhile, the Lamb wave based damage detection is considered as a highly sensitive but 'local' approach with limited coverage area around an actuator-sensor pair, which provides an ideal basis for the future wireless sensor network for structural health monitoring. The proposed reduction method, however, can be applied to other types of data.

4.1 Lamb Wave Approach for SHM and Experimental Setup

Lamb waves are elastic guided waves propagating in a solid plate or layer with free boundaries [28]. It has been recognized that the propagation of Lamb waves in such structures may be sensitive to the structural defects. Therefore, the change of Lamb wave propagation pattern with respect to the pattern under the intact baseline state can be used to indicate the occurrence of damage. Typically, the Lamb waves are excited at relatively high frequency, and thus small-sized damage can be detected.

The aluminum beam structure used in the experiment has the following geometric parameters: 814 mm (length) × 7.62 mm (width) × 3.175 mm (thickness). Figure 4 shows (a) the intact undamaged beam, (b) a damaged beam having a notch with dimensions 0.6 mm (length) × 7.62 mm (width) × 1 mm (depth), and (c) a damaged beam having a notch with 1.6 mm depth. For each beam, a pair of piezoelectric actuator and sensor is attached at 406.5 mm from the left end and 456.5 mm from the left end, respectively.

The hardware of the experimental investigation includes an AGILENT 33220A waveform generator (used to drive the piezoelectric actuator) and an INSTEK GDS-820S digital storage oscilloscope (used to collect the sensor signal). The waveform generator can produce an up to 64k-point arbitrary signal with 50MSa/s (mega samples per second). The oscilloscope can store two signals each up to 125 k data points, and the sampling rate per channel is up to 100 MSa/s. To reduce the dispersion, the excitation signals should have narrow bandwidth. Therefore, sinusoidal waves under Hann window are used as the transient excitation signals, which is the common practice in the Lamb-wave-based damage detection. In this experiment, five-cycle Hann windowed sinusoidal wave with center frequency 90 kHz is used. For this specific beam structure, 90 kHz is the so-called 'sweet spot' frequency [8] that maximizes the peak wave amplitude ratio

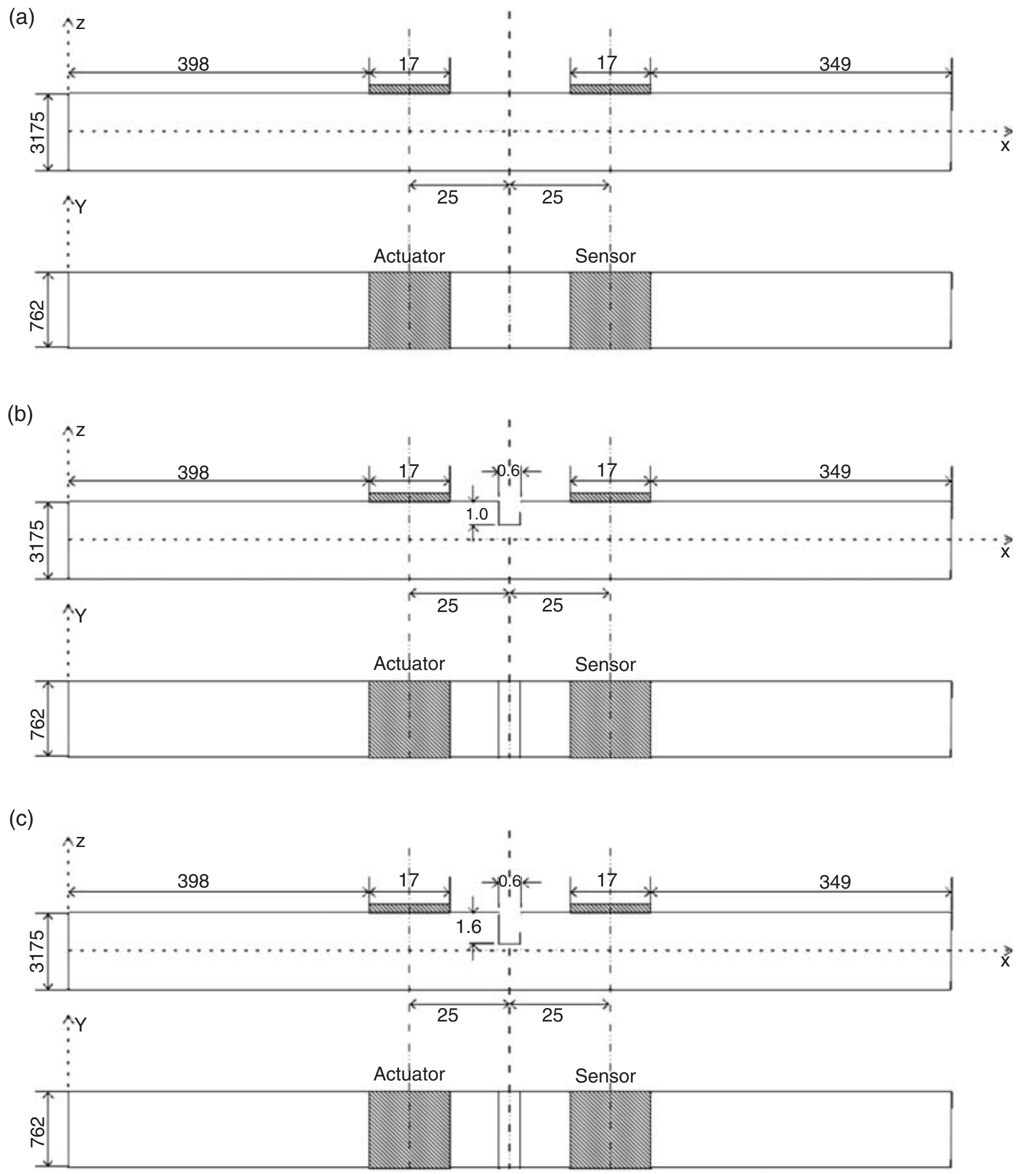

Figure 4 Side and top view of the experimental setups (dimensions in mm): (a) experimental setup I: normal beam, (b) experimental setup III: abnormal beam II, (c) experimental setup III: abnormal beam II.

between the lowest symmetric wave mode and the lowest antisymmetric wave mode, which can highlight the damage effect.

4.2 Detection Results and Discussions

For the purpose of data reduction, we apply our feature extraction algorithm to the Lamb wave signals. The following are the detailed settings used in our method: $I = 3$, chosen by using the guideline stated in Section 3.2, meaning that we take $\overline{\beta}$ in Equation (9) as the average of three distribution changes; the weights are chosen as $\alpha_i = ((I - i + 1)^2 / \sum_{i=1}^{I}(i)^2)$, which assigns greater weights to small changes and less weights to large changes; confidence level $\alpha_0 = 0.0027$; and the bootstrapping constants $K = 5$ and $M = 40$, where the confidence level corresponds to the range of standard six sigma quality level in industrial quality control practices [29]. The bootstrapping constants are decided according to the rule of thumb given in the analysis of RRE criterion [17].

Each sample of Lamb wave signals is digitalized to be a $50{,}001 \times 1$ vector. We collected the following datasets: 300 samples from the intact beam (experimental setup (a)), 150 samples from the beam with 1.0 mm depth crack (experiment setup (b)) and 312 samples from the beam with 1.6 mm depth crack (experiment setup (c)). Among 300 samples from the experimental setup (a), we randomly choose 200 samples (so $m = 200$ in Equation (5)) and use them to establish the baseline. The remaining 100 samples from the intact structure, together with the 462 ($=150 + 312$) samples from damaged structures, will be reserved as testing data, in order to assess the false positive probability (α error) and the false negative probability (β error) of the proposed reduction method. The analysis of the experimental results also includes the dimension of the reduced data, p, and the data reduction rate, R (defined in Equation (10)). In order to ensure that the performance is not drastically affected by the random sampling of the 200 baseline samples, the detection procedure is repeated 50 times on the original dataset of 300 samples, and the average of each of the performance indices (α, β, p, R) is reported. Figure 5 shows how the data samples are collected and subsequently used for data reduction and damage detection.

We also apply the existing methods on the same dataset for data reduction and damage detection. The same set of performance indices (α, β, p, R) is assessed. The existing methods include: (1) SureShrink [13], VisuShrink [14], RiskShrink [14], and AMDL [16] from de-noising methods, (2) the hard thresholding RRE [17] and the soft thresholding RRE_s [19] from data reduction methods, (3) the damage index method [23] and distortion energy method [24] from SHM applications. The performance comparison of all the data reduction methods is presented in Table 1.

From Table 1, we can easily note that our method strikes a good balance in terms of data reduction ability and detection power. It retains

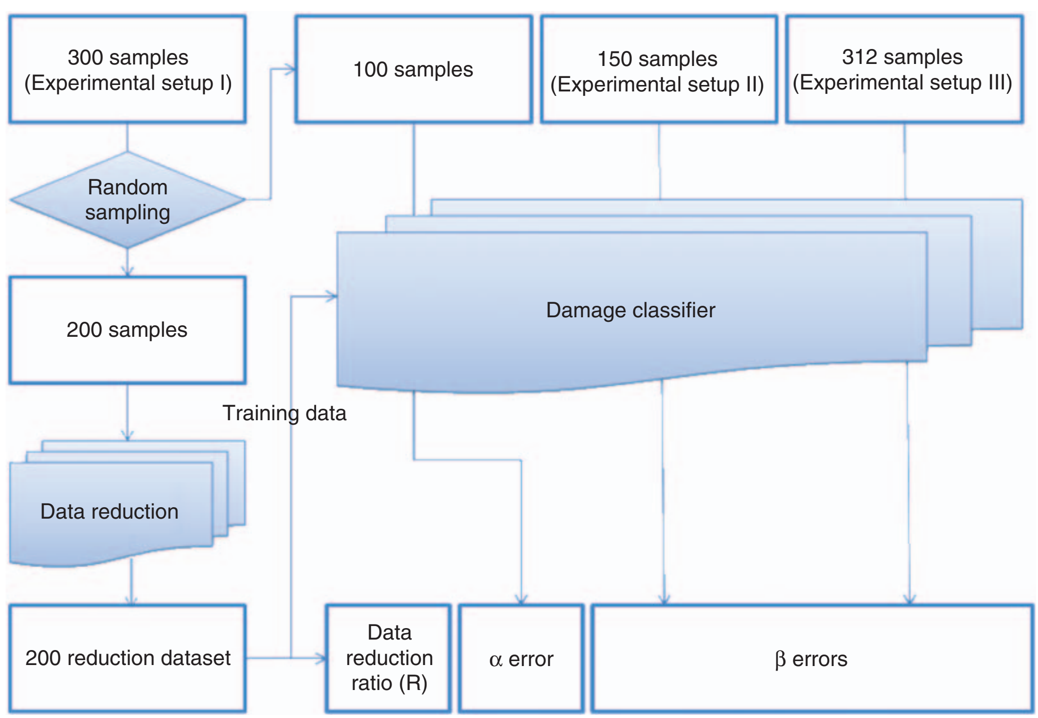

Figure 5 Data reduction and damage detection procedure using the experimental datasets. For the methods that do not specify a damage classifier, the maximum likelihood classifier is used, which is the same as the one used in our method (explained in Section 3.2); otherwise, their own damage classifiers are used.

only six features, which is the second best in terms of data reduction rate, while achieving a detection capability almost as good as using the RRE_s. The damage index method achieves the most aggressive data reduction as we expected, but it suffers from an unusually high false detection probability (a high β value). This confirms the concern that using a single scalar index may be too aggressive to keep all the relevant features for effective damage detection. The method of distortion energy uses eight resolution levels, comparable to our data reduction effectiveness, but its detection capability is not as good as ours.

On the other hand, most of the generic data reduction methods suffer from keeping too many features, not all of which are relevant to the damage detection objective. We also noticed that the method using RRE_s produces competitive results. It keeps more features but helps further reduce the false positive probability (α value). Therefore, we deem it the second best method overall.

Based on the performance comparison, we see that the six features selected by our method could facilitate the damage detection with the similar performance as a larger set of features selected by a general data reduction method such as $RRE_{s.}$ Our conjecture is that there is a high degree of redundancy among the large number of features selected by a general data reduction method. To verify the conjecture, we compare the features from our method with the features chosen by $RRE_{s.}$

Here, a feature is presented by its index in the set of wavelet coefficients after the same wavelet transform. Table 2 shows the indices of the features selected by both methods. In Table 2, the indices of the boldface font are the common set of features chosen by both methods. If our conjecture is correct, the features from RRE_s not belonging to the common set should be highly dependent on the common set of features.

To verify the relevance of these features to the structural damage, we plotted two wavelet coefficient maps in Figure 6: one for the signal from the normal beam and the other for the signal from abnormal beam I. The area inside the black-bordered box corresponds to the area on which six of the seven features from our method are concentrated. When comparing the two black-bordered areas from normal beam and abnormal beam I, we see that the area from

Table 2 Indices of the features selected by our method and RREs.

Data reduction methods	*Indices of the features*
Our method	**1676 1677 1678 1703 1753** 1879
RRE_s	1674 **1676 1677 1678** 1700 1701 **1703** 1726 1727 1729 1730 **1753** 1756 1779 1780 1832

Table 1 Performance of all methods compared, the values marked with asterisks (*) correspond to a case where the resulting data dimension p is greater than the sample size of the baseline data ($m = 200$). When this happens, the original version of the Hotelling's T^2 statistic cannot be used directly. We have to cut each Lamb wave signal into a number of segments, each of which has a reduced dimension less than 200, in order to perform damage detection and assess the two error probabilities. The segmentation causes α error larger than using the RRE_s and our method.

Data reduction techniques	*p*	*R (%)*	*α errors*	*β errors*
Our method	6	99.988	0.0302	0.0000
Damage index	1	99.998	0.0611	0.9972
Distortion energy	8	99.984	0.0900	0.1062
RRE_s	16	99.968	0.0287	0.0000
RRE	626	98.748	0.0351*	0.0000*
AMDL	961	98.073	0.0957*	0.0000*
VisuShrink	2574	94.907	0.0413*	0.0000*
RiskShrink	3733	92.534	0.0513*	0.0000*
SureShrink	12,765	74.471	0.0401*	0.0000*

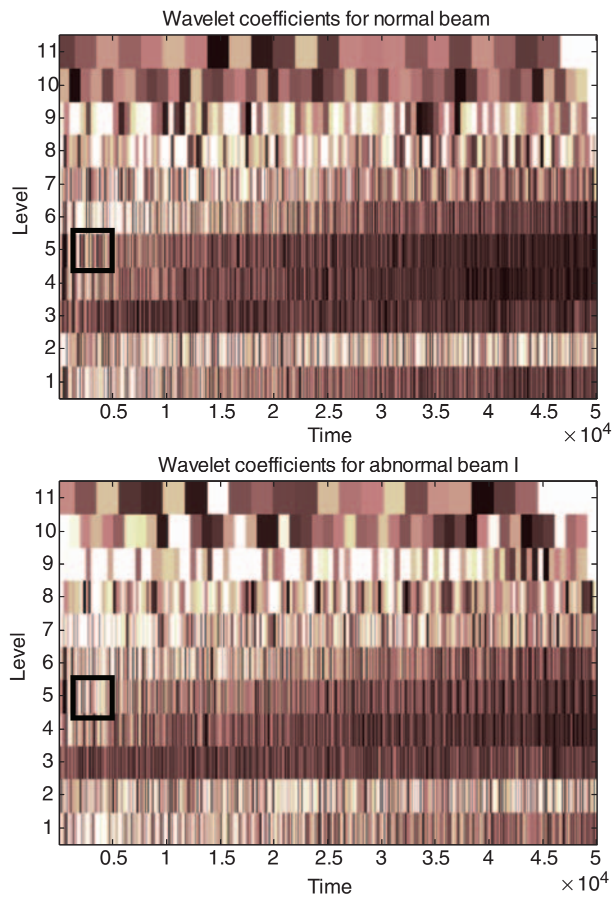

Figure 6 Discrete wavelet coefficient map.

normal beam is darker than the area from abnormal beam I.

To evaluate the dependency among these features, we use the partial correlation Corr $(S|T)$ as a measure of the amount of information, which can be explained by the features in S but not by the features in T. For more details regarding the partial correlation, please see [30]. In a multivariate case where Corr $(S|T)$ is a matrix (similar to a covariance matrix), one often uses the summation of the eigenvalues of Corr $(S|T)$ as the measure of the above-mentioned information amount.

We treat the union of the indices from both methods as T and assign S as an empty set initially. Then, we remove one index from T at a time and add it into S. In Figure 7, we plot the ratio of the summation of eigenvalues of Corr $(S|T)$ and the summation of eigenvalues of the original set T. This ratio represents the amount

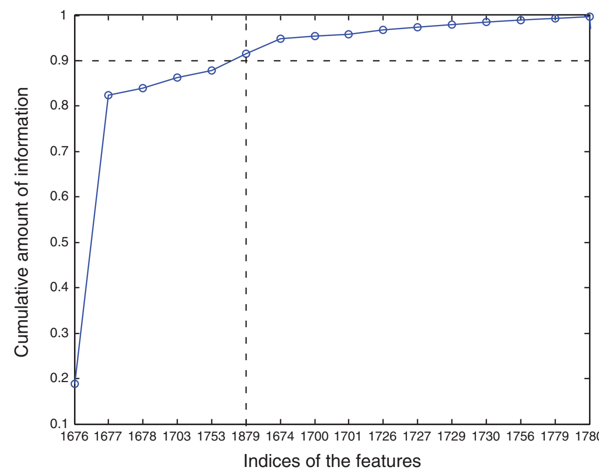

Figure 7 Partial correlation plot.

of information carried by the features only in S. The horizontal axis in Figure 7 represents the indices added to S in a sequential manner. When S contains $\{1676, 1677, 1678, 1703, 1753, 1879\}$, namely the set of the features selected by our method, we can see that more than 90% of the information is carried by S. In other words, the remaining features in set T are indeed highly redundant to those features already in S.

5 Conclusion

In this article, we develop an aggressive data reduction method to choose only the necessary features for the purpose of detecting structural damages. One of the major benefits from a high degree of data reduction is to enable energy-efficient operations for a wireless sensor network used for monitoring structural damages and integrity.

The reason that the proposed method can perform well is because that we have established an explicit measure of damage detection power, while the existing methods use certain type of surrogate measures that are only implicitly related to damage detection. The explicit measure of detection power allows us to find the optimal subset of features that contribute the most to the goal of damage detection. By contrast, the existing methods either retain too many features, not all of which are relevant, or use too few features with some important ones left out.

Through an experimental study based on Lamb wave signals, we have demonstrated that the proposed method outperforms the existing methods. It testifies the benefit of using the explicit measure for damage detection.

Acknowledgment

This research is supported by NSF under grants CMMI 0427878 and CMMI 0428210.

References

1. Green Jr, R.E. and Duke Jr, J.C. (1979). Ultrasonic and acoustic emission detection of fatigue damage. *International Advances in Nondestructive Testing*, *6*, 125–177.
2. Rogers, L.M. and Keen, E.J. (1986). *Detection and Monitoring of Cracks in Offshore Structures by Acoustic Emission*, West Midlands, UK: Engineering Materials Advisory Services Ltd, pp. 205–217.
3. Ghorbanpoor, A. and Shi, S. (1996). Assessment of corrosion of steel in concrete structures by magnetic based NDE techniques. *ASTM Special Technical Publication*, *1276*, 119–131.
4. Dobson, D.C. and Santosa, F. (1998). Nondestructive evaluation of plates using eddy current methods. *International Journal of Engineering Science*, *36*, 395–409.
5. Mirchandani, M.G. and McLaughlin Jr, P.V. (1986). Thermographic NDE of impact-induced damage in fiber composite laminates. *Review of Progress in Quantitative Nondestructive Evaluation*, *5B*, 1245–1252.
6. Doebling, S.W., Farrar, C.R. and Prime, M.B. (1998). A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest*, *30*, 91–105.
7. Kessler, S.S. and Spearing, S.M. (2002). Design of a piezoelectric based structural health monitoring system for damage detection in composite materials. *Proceedings of SPIE, Smart Materials and Structures*, *4701*, 86–96.
8. Giurgiutiu, V. (2005). Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. *Journal of Intelligent Material Systems and Structures*, *16*, 291–305.
9. Sun, F.P., Chaudhry, Z., Liang, C. and Rogers, C.A. (1995). Truss structure integrity identification using PZT sensor-actuator. *Journal of Intelligent Material Systems and Structures*, *6*, 134–139.
10. Park, G., Sohn, H., Farrar, C.R. and Inman, D.J. (2003). Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest*, *35*, 451–463.
11. Lynch, J.P. and Loh, K.J. (2006). A summary review of wireless sensors and sensor networks for structural health monitoring. *The Shock and Vibration Digest*, *38*, 91–128.
12. Crose, S., Marcelloni, F. and Vecchio, M. (2007). Reducing power consumption in wireless sensor networks using a novel approach to data aggregation. *The Computer Journal*, Advance Access Published on July 11.
13. Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association*, *90*(*432*), 1200–1224.
14. Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biomefrika*, *81*, 425–455.
15. Donoho, D.L. and Johnstone, I.M. (1995). De-noising by Soft-thresholding. *IEEE Trans-action on Information Theory*, *41*(*3*), 613–627.
16. Saito, N. (1994). Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion. In: Foufoula-Georgiou, E. and Kumar, P. (eds), *Wavelets in Geophysics*, pp. 299–324, New York: Academic Press.
17. Lada, E.K., Lu, J.-C. and Wilson, J.R. (2002). A wavelet-based procedure for process fault detection. *IEEE Transactions on Semiconductor Manufacturing*, *15*(*1*), 79–90.
18. Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, pp. 200–205, New York: Springer.
19. Jeong, M.K., Lu, J.-C., Huo, X., Vidakovic, B. and Chen, D. (2006). Wavelet-based data reduction technique for process fault detection. *Technometrics*, *48*(*1*), 26–40.
20. Xu, S., Rangwala, S., Chintalapudi, K.K., Ganesan, D., Broad, A., Govindan, R. and Estrin, D. (2004). A wireless sensor network for structural monitoring. In: *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*, Baltimore, MD.
21. Yam, L.H., Yan, Y.J. and Jiang, J.S. (2003). Vibration-based damage detection for composite structures using wavelet transform and neural network identification. *Composite Structures*, *60*(*4*), 403–412.
22. Wang, M., Cao, J., Chen, B., Xu, Y. and Li, J. (2007). Distributed processing in wireless sensor networks for structural health monitoring. *Lecture Notes in Computer Science*, *4611/2007*, 103–112.

23. Sohn, H., Park, G., Wait, J.R. and Limback, N.P. (2003). Wavelet based analysis for detecting delamination in composite plates. In: Chang, F.-K. (ed.), *Proceedings of 4th Int. Workshop on Structural Health Monitoring*, Stanford, CA, USA, pp. 567–574.
24. Bukkapatnam, S.T.S., Nichols, J.M., Seaver, M., Trickey, S.T. and Hunter, M. (2005). A Wavelet-based, Distortion Energy Approach to Structural Health Monitoring. *Structural Health Monitoring*, *4*(*3*), 247–258.
25. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn, Hillsdale, NJ: Lawrence Erlbaum.
26. Graybill, F.A. (2000). *Theory and Application of the Linear Model*, 1st edn, North Scituate, MA: Duxbury Press.
27. Montgomery, D.C. (1997). Introduction to Statistical Quality Control, 3rd edn, New York: Wiley.
28. Rose, J.L. (1999). *Ultrasonic Waves in Solid Media*, Cambridge: Cambridge University Press.
29. Tennant, G. (2001). *Six Sigma: SPC and TQM in Manufacturing and Services*, Aldershot, UK: Gower Publishing, Ltd.
30. Cumming, J.A. and Wooff, D.A. (2007). Dimension reduction via principle variables. *Computational Statistics and Data Analysis*, *52*, 550–565.