Journal of Sound and Vibration (2002) **258**(4), 741–761
doi:10.1006/jsvi.2002.5148, available online at http://www.idealibrary.com on IDEAL®

NOVELTY DETECTION IN A CHANGING ENVIRONMENT: REGRESSION AND INTERPOLATION APPROACHES

K. Worden[†], H. Sohn and C. R. Farrar

Group ESA-EA, Los Alamos National Laboratories, Los Alamos U.S.A.
E-mail: k.worden@sheffield.ac.uk

(*Received* 13 *June* 2001, *and in final form* 4 *February* 2002)

The technique of novelty detection is now established as a means of performing the lowest level of damage identification. Data are accumulated while the system or structure is operating in normal condition and used to construct a reference model. During subsequent operation of the system, data are compared to the reference and any significant deviations are taken to indicate damage. This approach has potential problems if the system or structure is embedded in a changing environment. If the reference data are only characteristic of a limited range of the environmental parameters, measurements from the system in an undamaged condition but from a different environmental state, may cause the diagnostic to register novelty and thus falsely infer damage. This paper demonstrates a potential solution to the problem via the construction of a reference set parametrized by an environmental variable. Two approaches are considered: regression and interpolation.

1. INTRODUCTION

The problem of damage identification has a hierarchical structure [1]. At the lowest level, the idea is to say with confidence, whether damage has occurred or not. At the highest level, it is required to locate and size the damage and also to estimate the residual safe life of the system or structure. One of the more promising approaches to the damage problem is based on *pattern recognition*. Data are measured from the system or structure and converted by a process of *feature extraction*, into a representation where variations due to damage are highlighted. This representation is usually of reduced dimension. Once a feature vector is obtained, it can be processed by a *classification algorithm* which will associate with it a class label or damage state. If only the lowest level of identification is required, the class label will have only two values, *normal* and *damaged*. If a higher-level diagnostic is needed, the class label will encode further information such as: damage type, damage location and damage extent. In this latter case, the classification algorithm will need to have *a priori* knowledge of all the classes that can occur. If the algorithm gains this knowledge by learning from examples, features (and hence data) must be available for each class. This is a formidable problem and requires detailed modelling of the system or extensive experimental programmes in order to collect the requisite data.

In many cases, the most important question will simply be—is the system damaged or not? In this restricted case, a simpler means of training the classifier is available and this is

[†] On study leave from: Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, U.K.

0022-460X/02/$35.00

the basis of the approach termed *novelty detection* [2–4]. The idea is to extract features from the data that characterize only the normal condition and these are used as a template or reference. During monitoring, data are measured, the appropriate features are extracted and compared (in some sense) with the reference. Any significant deviations from the reference are considered to signal novelty or damage. The major advantage of this strategy is that the training data (which establishes the reference) is *only* from the undamaged system, and this vastly reduces the modelling or data collection requirements.

The simple strategy above is only appropriate when the normal condition does not vary with time. Suppose the reference model is constructed from normal condition data measured around a nominal time t_0. If subsequent data are measured at a time t_m, where the operational or environmental conditions have changed but are still considered to be normal, the new data will be flagged as novel. Implying damage from novelty under these circumstances is clearly not correct. The variation with time may be implicit or explicit. The latter case is extremely important for engineering, as the system of interest may be sensitive to the environmental conditions and these conditions may vary with time. A classic example is a bridge whose measured properties vary with temperature. The changes in measured features during a day/night cycle can be large enough to obscure any changes due to damage.

One solution to the problem is to collect the training data over a long enough time period to span all the possible normal conditions. This is the approach followed in references [5,6]. The latter reference considered damage detection in a model offshore platform where the deck mass was changing with time as a result of storing oil. This approach is valid only if the damage produces changes in the features that are orthogonal in some sense to the changes produced by environmental variation. A related problem is that the enlarged feature set may cause a decrease in sensitivity (this will be illustrated later). For ease of reference, such models will be termed *large normal condition models*.

A more satisfactory (and in a sense optimal) solution is to compare new data *only* with reference data from the same environmental conditions. There are two possible situations: (a) the environment is uniquely characterized by a group of *measureable* parameters (temperature, humidity, etc.), (b) the environment cannot be characterized so. The current paper is concerned only with the first (simpler) situation. The idea is to build a set of reference models parametrized by the environmental variables; during structural monitoring, new data are evaluated with respect to a reference model for the appropriate conditions. The training data will be exactly the same as for the large normal condition model; however, during monitoring, only a (potentially small) subset of the training data will effectively be used to assess novelty. Two strategies are considered here, one based on regression and the other based on interpolation.

The process of removing the effect of environmental variation in the observed data is part of the larger issue of *normalization*. The idea is to create a damage indicator that is invariant under changes in the structure or its environment that are not relevant for diagnosing damage. For example if the raw data from a structure are a response time series and the excitation is unobservable and uncontrollable, the effect of the input amplitude on the response can be removed by standardizing the response (removing the mean and scaling by the inverse standard deviation). This is a valid step if, for example, the structure is known to be linear.

In order to illustrate the approach here, data will be generated synthetically from a simple lumped-mass system.

The layout of this paper is as follows. Section 2 describes the particular novelty detection algorithm used in this work—namely outlier analysis—and outlines how it can be adapted to cope with environmental change. Section 3 describes the system of interest;

data are obtained from computer simulation. Section 4 illustrates the regression approach to the normalization process using a low-dimensional feature vector and section 5 shows how the approach is applied in higher dimensions and also highlights some of the problems which can occur in the latter case. Section 6 shows how an interpolation approach can overcome problems with the regression strategy.

2. OUTLIER ANALYSIS

2.1. OUTLIERS IN UNIVARIATE DATA

A *discordant* observation or *outlier* in a data set is an observation that is surprisingly different from the rest of the data in some sense, and therefore is believed to be generated by a different mechanism to the other data. The *discordancy* of a candidate outlier is some measure that can be compared against some corresponding objective criterion, and allows the outlier to be judged as statistically likely or unlikely to have come from the assumed generating model. The application to damage detection is clear; the discordancy should be evaluated with respect to a model constructed from a normal condition of the system of interest. The standard reference for outlier analysis is reference [7].

The case of outlier detection in univariate data is relatively straightforward in the sense that outliers will "stick out" from one end or other of the data set. There are numerous discordancy tests. One of the most common, and the one whose extension to multivariate data will be employed later, is based on deviation statistics and given by

$$z_\zeta = (x_\zeta - \bar{x})/s, \tag{1}$$

where x_ζ is the potential outlier and $\bar{x}$ and s are the mean and standard deviation of the sample respectively. The latter two values may be calculated with or without the potential outlier in the sample depending upon whether *inclusive* or *exclusive* measures are preferred. This discordancy value is then compared to a threshold value and the observation declared, or not, to be an outlier. The value of the threshold is critical; unfortunately it is usually necessary to make some assumptions about the data in order to establish it. Suppose the normal condition data are assumed Gaussian. In this case, there is a 95% probability that a sample x_ζ drawn from the same distribution will lie in the range $[-1{\cdot}96, 1{\cdot}96]$. If a point observed during the monitoring period lies outside this range, there is only a 5% chance that the point is a sample from the same normal condition distribution. In practice, the reference set will not be Gaussian, however, if the deviation is small, the assumption of Gaussianity may yield a sensible threshold.

2.2. OUTLIERS IN MULTIVARIATE DATA

A multivariate data set consisting of n observations in p variables may be represented as n points in a p-dimensional space. It is clear that detection of outliers in multivariate data is more difficult than the univariate case due to the potential outlier having more "room to hide".

The discordancy test that is the multivariate equivalent of equation (1), is the Mahalanobis squared distance measure given by

$$D_\zeta = (\{\mathbf{x}_\zeta\} - \{\bar{\mathbf{x}}\})^{\mathrm{T}}[\mathbf{S}]^{-1}(\{\mathbf{x}_\zeta\} - \{\bar{\mathbf{x}}\}), \tag{2}$$

where $\{\mathbf{x}_\zeta\}$ is the potential outlier, $\{\bar{\mathbf{x}}\}$ is the mean of the sample observations and $[\mathbf{S}]$ the sample covariance matrix.

As with the univariate discordancy test, the mean and covariance may be inclusive or exclusive measures. In many practical situations the candidate outlier is not known beforehand and so the test would necessarily be conducted inclusively. In the case of health monitoring though, the potential outlier is always known beforehand—it is the most recently measured observation—and so it is more sensible to calculate a value for the Mahalanobis squared distance without this observation "contaminating" the statistics of the normal condition data. Whichever method is used, the Mahalanobis squared distance of the potential outlier is checked against a threshold value, as in the univariate case, and its status determined.

2.3. CALCULATION OF CRITICAL VALUES OF DISCORDANCY

In order to label an observation as an outlier or an inlier there needs to be some critical value or threshold against which the discordancy value can be compared. This value is dependent on both the number of observations in the training set, n, and the number of dimensions of the problem being studied p.

For the work presented here, a Monte Carlo method was used to arrive at the threshold value. The procedure for this was to construct a $(p \times n)$ (number of dimensions × number of observations) matrix with each element being a randomly generated number from a zero mean and unity standard deviation Gaussian distribution $N(0, 1)$. The Mahalanobis squared distances were calculated for all the elements, using equation (2) where $\{\bar{\mathbf{x}}\}$ and $[\mathbf{S}]$ are inclusive statistics, and the largest value stored. This process was repeated for at least 10 000 trials whereupon the array containing all the largest Mahalanobis squared distances was ordered in terms of magnitude. The critical values for 5 and 1% tests of discordancy for a p-dimensional sample of n observations are then given by the Mahalanobis squared distances in the array above which 5 and 1% of the trials occur. The inclusive threshold is computed because it is far less expensive computationally, it can be converted into the exclusive threshold by the use of a simple formula [7]. As in the univariate case described above, there is an implicit assumption here that the reference or training set is multivariate Gaussian.

2.4. OUTLIER ANALYSIS IN A CHANGING ENVIRONMENT

The framework above is suitable only for when the normal condition distribution is time-invariant, i.e., the statistics $\{\bar{\mathbf{x}}\}$ and $[\mathbf{S}]$ are constants. In a changing environment, the statistics of the normal conditions will be functions of the environmental parameters. To simplify matters here, it will be assumed that the environment is parametrized by a single measurable variable which will be arbitrarily called temperature T. Thus $\{\bar{\mathbf{x}}\} = \{\bar{\mathbf{x}}(T)\}$ and $[\mathbf{S}] = [\mathbf{S}(T)]$. Suppose that there is variation with time, but it is slow compared to the typical time period of acquisition of a reference set. By measuring data at various points in the environmental cycle, it will be possible to collect a set of statistics characteristic of a set of temperatures $\{\mathbf{T}_i, i = l, \ldots, N_T\}$. In order to build the parametrized reference model, a polynomial regression model in T is fitted for each coefficient of the mean vector and covariance matrix, i.e., for the ith coefficient of the mean,

$$\bar{x}_i \approx \sum_{j=0}^{N_p} a_i^j T^j, \tag{3}$$

where N_p is the polynomial order, and the a_i^j are the regression coefficients. Similarly for the covariance matrix,

$$S_{ij} \approx \sum_{k=0}^{M_p} a_{ij}^k T^k, \tag{4}$$

where M_p need not equal N_p.

A total of p least-squares regressions are required to model the mean, and $p(p+1)/2$ are needed for the covariance as it is symmetric.

The monitoring strategy is now clear. When a new set of observations is tested for novelty, the temperature T at the time of measurement is used to estimate the appropriate statistics $\{\bar{\mathbf{x}}(T)\}$ and $[\mathbf{S}(T)]$ from the regression models, and these are used to compute the Mahalanobis distance.

3. THE DATA FOR THE CASE STUDY

3.1. A LUMPED-MASS SYSTEM

The system selected for generation of the simulated data was a lumped-mass system with equations of motion

$$\begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix}\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} + \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} + \begin{pmatrix} 2k & -k \\ -k & 2k \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}. \tag{5}$$

The raw data for the diagnostic was the response y_1. Because there is a single input x, the required response is also obtained as the solution of

$$m^2\ddddot{y}_1, +2\mathrm{cm}\dddot{y}_1, +(\mathrm{c}^2+4\mathrm{km})\ddot{y}_1 + 4\mathrm{ck}\dot{y}_1 + 3\mathrm{k}^2 y_I = \mathrm{m}\ddot{x} + \mathrm{c}\dot{x} + 2\mathrm{k}x. \tag{6}$$

Since this is a single-input single-output system, the subscript distinguishing the responses will be dropped from this point forward.

The excitation chosen was a white Gaussian sequence with zero mean and unit variance. Simulation was carried out using the function *Isim* from the MATLAB Control System Toolbox [8]. In order to simulate the effects of damage and temperature variation on the system, the following coefficients were prescribed: $m = 1$, $c = 20(1 + D)$ and $k = 10^4(1 - D/2 - T/200)$. The temperature T is allowed to take values in the range [0, 100], and the damage index D is allowed to take values [0,1]. At the reference temperature $T = 0$ with no damage, the system has undamped natural frequencies of 15.9 and 27.6 Hz. A time step of 0·002 s was chosen for the simulation, giving an effective sampling frequency of 500 Hz. For the extreme values of the parameters, $T = 100$, $D = 0$ or $T = 0$, $D = 1$, the system experiences a 50% reduction in stiffness. The difference is that the damage also causes an increase in damping. The effects of temperature and damage have deliberately been chosen to be similar in order to expose the limitations of the large normal condition model.

3.2. THE FEATURES

The raw time data are inappropriate for damage detection as the response is a random variable and the individual values of a measured time record will always be different to any other. In order to construct a reference model, it is necessary to convert the data to a *feature set* containing time-invariant observables of the response. A popular choice is to use Fourier transformation to convert the time data to spectra. This forces the analyst to choose the most significant spectral lines in order to obtain feature vectors of low dimension. It is important that the dimension be low, as the size of the training set needed to properly sample the probability distribution of the features grows explosively with dimension. An alternative strategy that gives a low-dimensional representation of the spectral content is to use an *auto-regressive* (AR) model. This is a model for the process,

$$y_i = \sum_{j=1}^{N_{AR}} \varphi_j y_{i-j} + \varepsilon_i. \tag{7}$$

The response y_i at a given sampling instant t_i is a weighted sum of N_{AR} past values of the response plus an added shock ε_i [9]. It is possible to show that the spectrum of y is estimated by

$$S_{yy}(\omega) = \sigma_\varepsilon^2 \Big/ \left| 1 - \sum\nolimits_{j=1}^{N_{AR}} \varphi_j \mathrm{e}^{-\mathrm{i}j\omega\Delta t} \right|^2, \tag{8}$$

where Δt is the sampling interval and σ_ε^2 is the variance of the shock sequence ε_i. The accuracy of the prediction and the accuracy of the spectral estimate increases with the AR model order until the appropriate order is reached. For feature extraction purposes, it is useful to know the correct model order. One means of estimating the order is based on the following observation [9]. If the correct model order for an AR process is p, any coefficients beyond the pth in a higher order model are distributed as a Gaussian with zero-mean and standard deviation $\sigma_p = 1/\sqrt{N_w}$, where N_w is the number of points in the estimation window or record. In practice, the magnitude of the last AR coefficient $\varphi_{N_{AR}}$ is plotted against the model order; beyond the correct order, the coefficients will fall below the threshold σ_p. The method is sometimes referred to as *partial auto-correlation* (PAC). To illustrate the method, data from the system described above will be analyzed.

A response sequence of length 10 000 samples was generated for the system with $T = 0$ and $D = 0$, and the PAC plot was constructed as shown in Figure 1. Figure 1 also shows

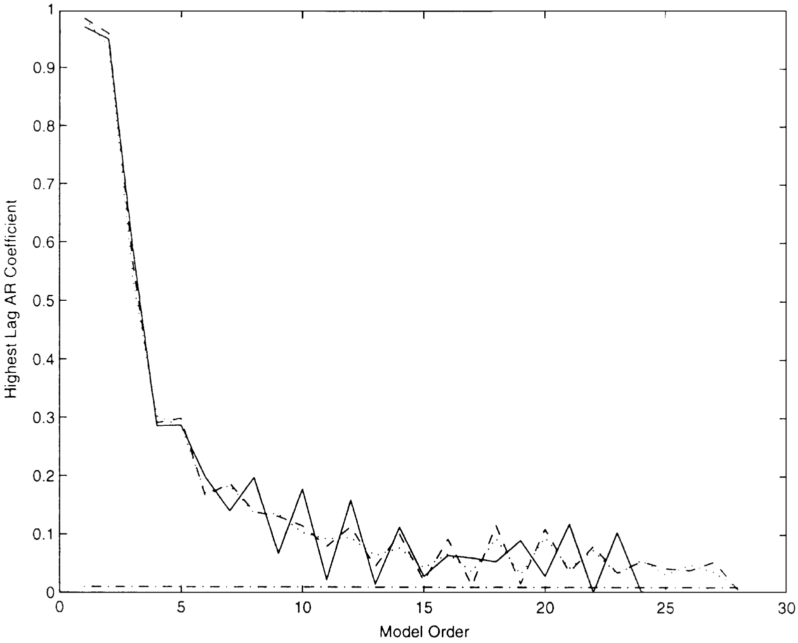

Figure 1. Partial auto-correlation plot for data from system of interest. ——, $T = 0$: $D = 0$; – – –, $T = 100$: $D = 0$; . . ., $T = 0$: $D = 1$.

the corresponding plots for similar data records with $T = 100$, $D = 0$ and $T = 0$, $D = 1$, i.e., for the two extreme conditions of the system. The analysis was carried out using MATLAB [10] and made use of routines from the System Identification Toolbox [11].

The PAC plot shows that the appropriate AR model order for the $T = 0$, $D = 0$ data is 23, while for the high temperature and damaged data the right model order is 27. (Clearly, this is an *ad hoc* procedure and may be subject to variation between data samples.) Figure 1 also shows that the first two coefficients are dominant, and it may be that two dimensions suffice to illustrate some aspects of the regression technique. If the equation of motion (6) is converted to a discrete-time representation using centred differences to approximate the derivatives, an ARX (in terms of past y and x values) model is obtained with order (4,2); however, this is only a guide to the necessary model order. If an AR or MA (only lags in x present) model is required, the order will be higher in order to compensate for the missing variables. The order will also depend on how accurate the central difference is at the given sampling frequency.

4. ANALYSIS USING A LOW-DIMENSIONAL FEATURE SPACE

The analysis in this section assumes that the appropriate AR model order for novelty detection is 2. The first objective here is to demonstrate the fundamental limitation of the large normal condition model.

A large training set was generated as follows. For values of T between 0 and 100 inclusive, at intervals of $\Delta T = 10$, a response series containing 10 000 points was generated from the system described in section 3 with $D = 0$. A feature set at each temperature was created by moving a 1000-point window through the record with an overlap of 950 points. In each window, an $AR(2)$ model was fitted by a simple least-squares method. This procedure gave a set of 181 two-dimensional vectors for each of the 11 temperatures. A total of 1991 vectors were obtained spanning the whole environmental range of the system. Note that because the windows overlap substantially, there will be a high degree of correlation between the feature vectors. The true measure of the size of the training set should probably be the number of feature vectors that can be obtained from the response when there is *no* overlap. Although the data here only allows 10 independent measurements, the additional feature vectors obtained from the overlap procedure are useful for visualisation purposes.

A testing set was created by the same procedure, except that the responses were generated with $T = 0$ and the damage index D running from 0 to 1 with a step size of 0·1.

Figure 2 shows a plot of the training and testing data. The training data at $T = 0$, which are the proper reference data for the testing set is highlighted. The figure shows that the large training set cluster (marked by crosses) overlaps substantially with the damage cluster (marked by points). This means that a diagnostic trained on the large set will be insensitive to the lower levels of damage. In contrast, the $T = 0$ component of the reference (marked by circles) overlaps less. More importantly, because of the high variance of the large training set, the Mahalanobis distance will grow much more slowly as the Euclidean distance from the set increases.

To illustrate the effect further, Figure 3 shows the result of carrying out an outlier analysis using the large training set and compares with the results of using the $T = 0$ subset which is appropriate for this particular damage data.

The testing data in Figure 3 is for damage indices $D = 0$–1 in steps of 0·2. The vertical dotted lines in the figure separate the different damage regimes (each segment is of 181 points). It is immediately clear that the diagnostic trained on the full range of normal

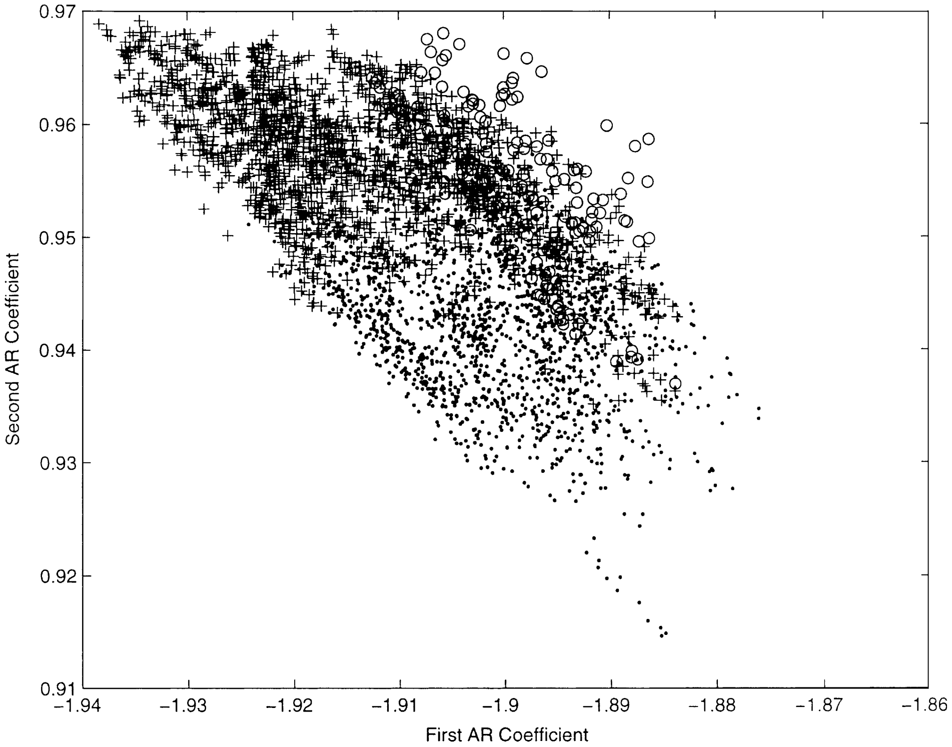

Figure 2. Scatter plot showing distribution of data for AR(2) features. ○, $T = 0$. $D = 0$; +, $T > 0$. $D = 0$; •, $T = 0$. $D > 0$.

conditions is less sensitive than that trained on the $T = 0$ subset. The threshold shown as a horizontal line is the 99·9% confidence limit for the $T = 0$ subset. Because the subset contains only a fraction of the points used for the large training set, the 99·9% limit for the large set would be higher, so the situation is actually worse than the figure suggests. Even when the appropriate training set is used for comparison, it can be seen that the sensitivity of the diagnostic is low. This is due to the minimal nature of the feature vector. It will be shown later that higher order AR models give higher sensitivity to damage.

The analysis above shows the clear advantage of using a reference appropriate to the environmental conditions. The next results will illustrate the use of the regression approach. The training data used for this example is identical to the large training set used previously; however, this time it is partitioned according to the temperature at the time of measurement. In all, 11 data sets are available spanning the temperature range. These data are used to construct 11 mean vectors and 11 covariance matrices, each labelled by the measurement temperature. The final regression model is obtained by fitting a parametric model to each vector and matrix coefficient in turn as described in section 3. Because 11 points are available for each curve-fit, the polynomial order was chosen as 3 to avoid overfitting.

Figure 4 shows the variation in the mean of the first AR coefficient as the temperature varies. It is clear that a low order polynomial should be used for fitting, any higher than cubic and the curve-fit would start to reproduce the statistical fluctuations in the data.

The next illustration shows how the regression model performs on data from the damaged system at different temperatures. As before, response data were simulated and

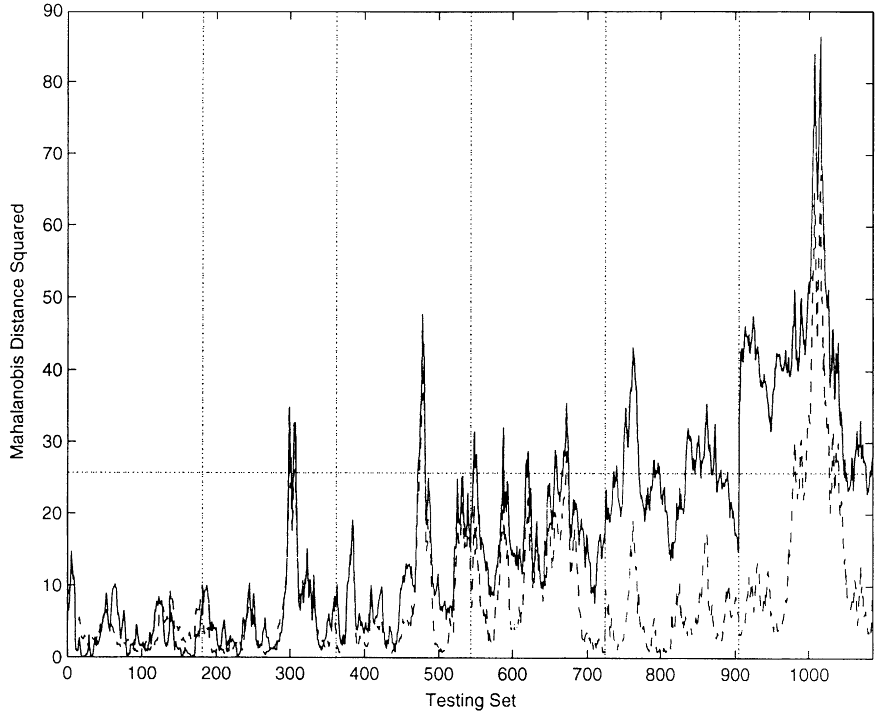

Figure 3. Comparison of outlier statistics for a large training set and an appropriate reference. ——, $T = 0$ training set; – –, Large training set.

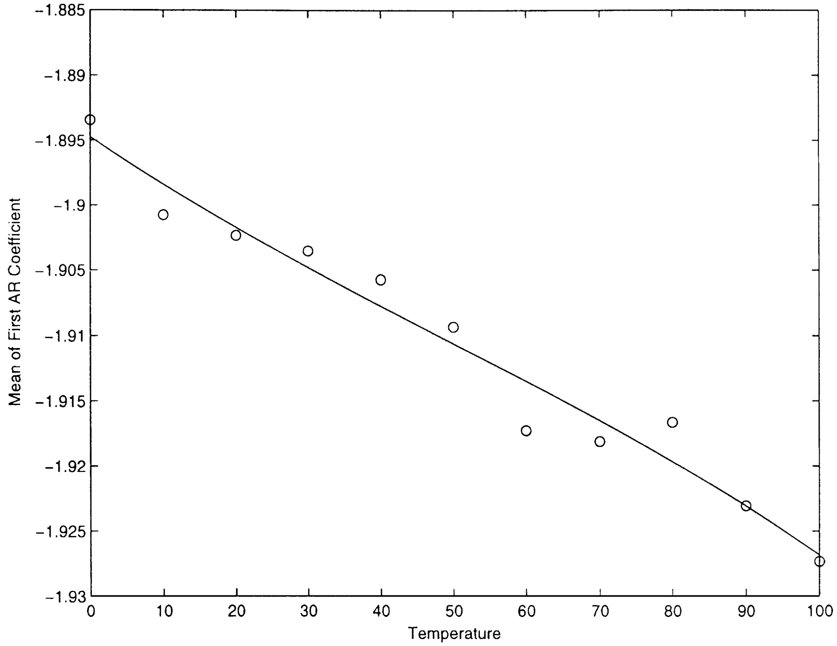

Figure 4. The mean of the first AR coefficient in the AR(2) feature set as a function of temperature together with the least-squares curve-fit. ○, measured data; ——, cubic curve fit.

converted into sets of feature data. Four feature sets were obtained: (a) $T = 20$, $D = 0$, (b) $T = 20$, $D = l$, (c) $T = 75$, $D = 0$, (d) $T = 75$, $D = 1$. The significance of this simulation is that for the first two data sets the temperature occurred in the training set, for the latter two the regression model is interpolating between measured points. The results of the outlier analysis using the regression model on the features above are shown in Figure 5.

The results from the regression model are very satisfactory. Both of the sets of features from normal condition are below the threshold; both of the damage states are clearly flagged.

The regression approach has been validated here on a low-dimensional feature set. This leaves something to be desired as the sensitivity of the regression model to the damage is rather low—Figure 5 shows results for extreme damage states. It is expected that using a higher order AR model and consequently a higher-dimensional feature set is likely to increase sensitivity. In physical terms, using a higher order model will give a better resolved estimate of the response spectrum which is likely to be more sensitive to damage. This is investigated in the next section.

5. ANALYSIS USING A HIGH-DIMENSIONAL FEATURE SPACE

The PAC analysis in section 3 indicated that the appropriate model order for the response data was between 23 and 27. This means that all coefficients in a model up to this order will contain information about the system, and will therefore contain information about discordancy if the system response changes. The object of this section is to repeat the regression analysis in a higher-dimensional feature space. The model order for the following analysis will be set at 24.

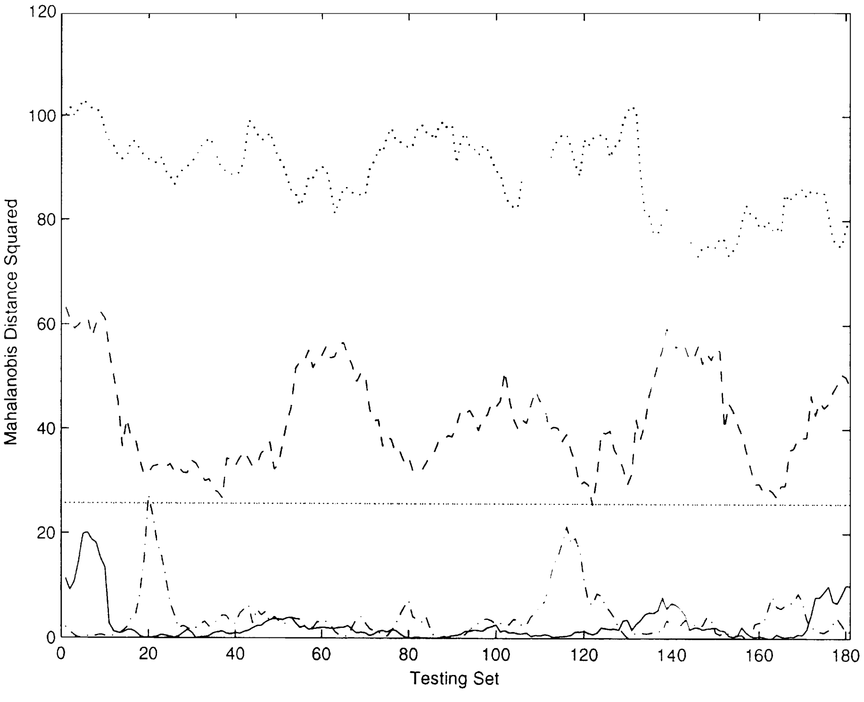

Figure 5. Results of outlier analysis using regression model on data from two different temperatures: low-dimensional features. ——, $T = 20$, $D = 0$; – –, $T = 20$, $D = 1$; –, $T = 75$, $D = 0$; . . ., $T = 75$, $D = 1$.

Now, as the object of the exercise here is to investigate if increased sensitivity results from higher order AR models, the first simulation will be to consider sensitivity. As in the previous section, a large training set was generated by simulating the response at 11 temperatures over the expected range. The same moving window strategy was used to generate features, the main difference being that an AR(24) model was fitted in each window instead of an AR(2). Another important difference concerns the *size* of the training sets. In the case of the two-dimensional feature set, 181 samples constituted an adequate training set. This is not the case for 24 dimensions, particularly as the samples are correlated (recall that the estimation windows for the AR coefficients overlap substantially). For the present exercise, time records containing 100 000 points were used. With a window length of 1000 points and an overlap of 800 points, this gave a training set with 496 samples per temperature. Although the time histories may appear long, because the sampling frequency is 500 Hz, the acquisition time is 200 s and this is certainly short compared to the expected time scales for environmental change. For the testing set, the same damage states as those illustrated in Figure 3 were used.

Figure 6 shows the results of two outlier analyses. The first uses the large normal condition model and uses training data over the whole temperature range. The second uses only the appropriate subset corresponding to the appropriate $T = 0$ subset.

The results are much more consistent than those from the two-dimensional features. There are a few excursions above threshold on the data from the lowest damage state ($D = 0.2$) and many excursions for the next level of damage ($D = 0.4$). Note that for the higher-dimensional features, there is substantial overlap between the damage states and the extended normal condition of the large training set. This renders the diagnostic insensitive to damage.

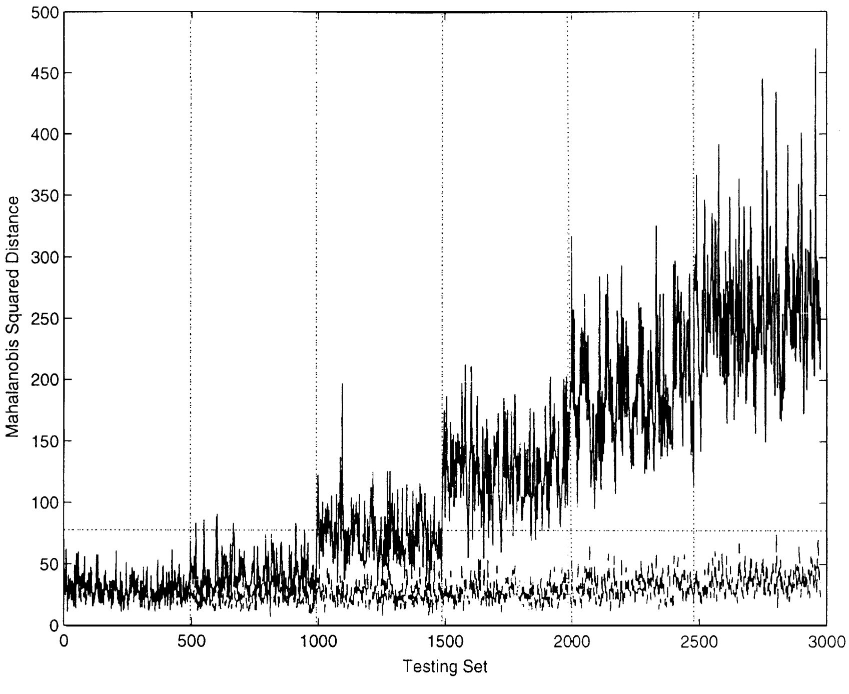

Figure 6. Comparison of outlier statistics for a large training set and an appropriate reference. ——, $T = 0$ training set; – –, large training set.

Just to emphasize that the increase in sensitivity of the diagnostic is due to the increased dimensionality of the feature vectors and not to the increased size of the training set, Figure 7 shows the Mahalanobis distances for the two-dimensional feature vectors when the training and testing sets are of the increased size. By comparison with Figure 6, it is shown that the higher-dimensional features are superior.

Moving on to the regression model, a curve-fit was performed as before to the means and covariances as a function of temperature. Figure 8 shows the outlier statistics for the same testing sets as in Figure 5, namely: (a) $T = 20$, $D = 0$, (b) $T = 20$, $D = 1$, (c) $T = 75$, $D = 0$, (d) $T = 75$, $D = 1$.

The results are excellent, the two damage states are well above threshold, while the two normal condition sets are well below. In comparison with Figure 5, the separation between the two normal conditions and the corresponding two damage conditions is more consistent. While the $T = 75$ results were further above threshold in Figure 5, some of the $T = 20$ were barely above and one point was actually flagged as normal.

It seems that the regression approach to normalization works very well. However, there is a caveat. A serious problem can occur if the normal condition sets are inadequate for proper training. In order to illustrate this, a training set was constructed using the prescription for the low-dimensional feature set. For each of the 11 temperatures, 10 000 points of time data were generated and a 1000 point window was stepped through the data with an overlap of 950 points. This gave 181 training patterns per temperature that were highly correlated.

In this case, two testing sets were used: the first with $T = 20$, $D = 0$ and the second with $T = 20$, $D = 1$. Figure 9 shows the result of computing the outlier statistic on the basis of the inadequate training data.

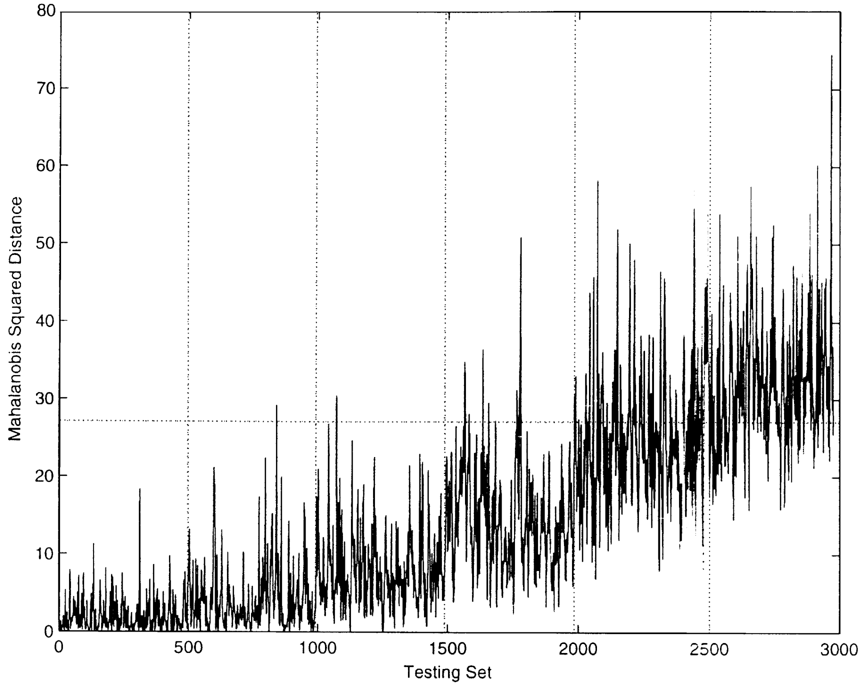

Figure 7. Outlier statistics for two-dimensional feature set but with training set of increased size.

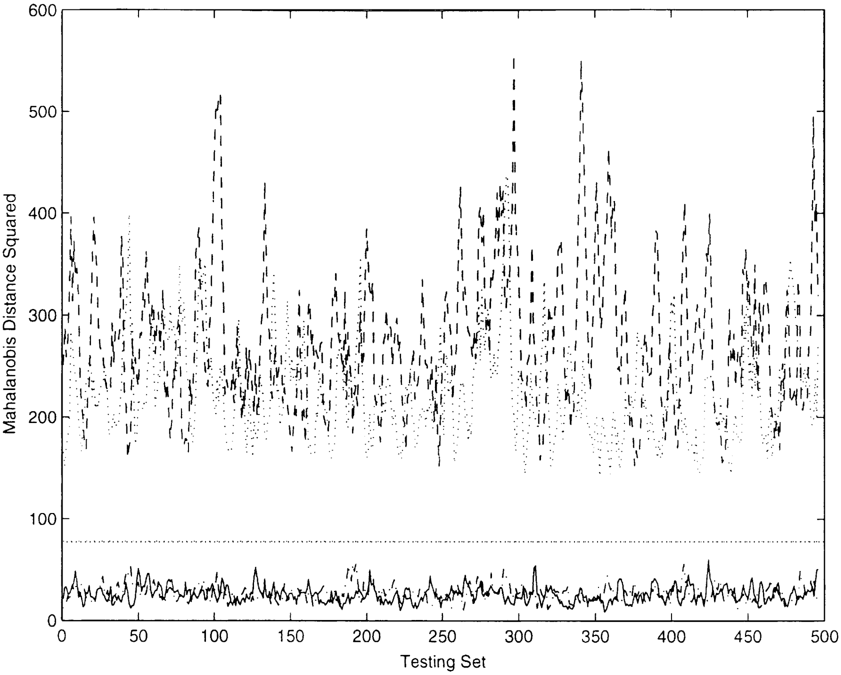

Figure 8. Results of outlier analysis using regression model on data from two different temperatures: high-dimensional features. ——, $T = 20$, $D = 0$; – –, $T = 20$, $D = 1$; –, $T = 75$, $D = 0$; . . ., $T = 75$, $D = 1$.

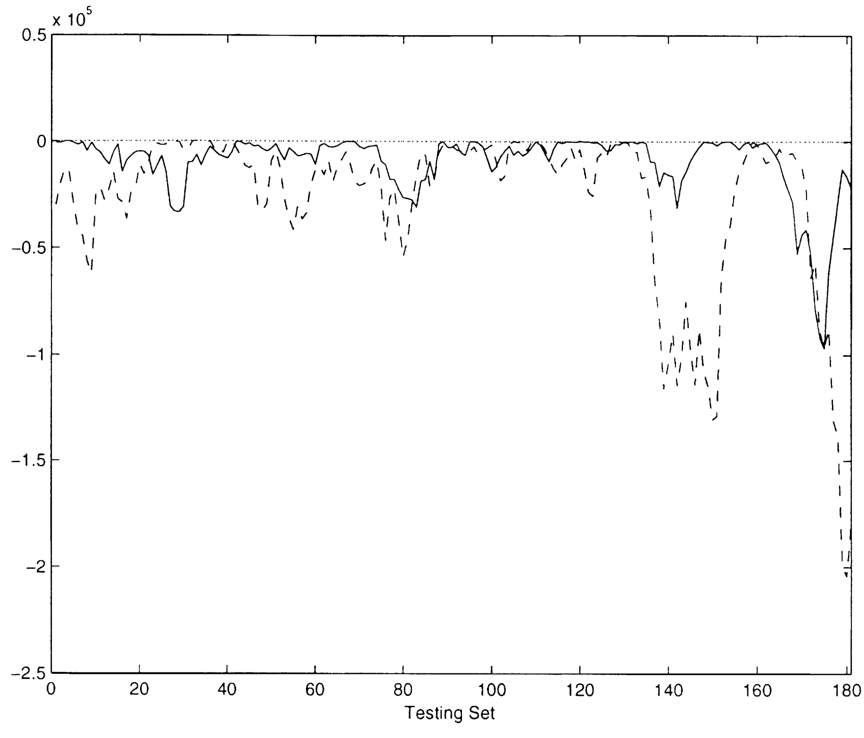

Figure 9. Results of outlier analysis using regression model with inadequate training data. ——, $T = 20$, $D = 0$; – –, $T = 20$, $D = 1$.

For this example, the Mahalanobis squared distance is *negative*. At first this seems rather unlikely. However, recall that the covariance matrix for the test above is built from component-wise curve-fits. While the covariance matrices in the training set were all symmetric and positive semi-definite by construction, the matrices recovered from the regression models are only constrained to be symmetric. In fact, an eigenvalue analysis of the matrix $[\mathbf{S}(T=20)]$ showed that the smallest eigenvalue was in fact negative. When $[\mathbf{S}(T)]$ is inverted to form the distance measure in equation (2), the negative eigenvalue becomes the largest.

In this case, the negative distances are assumed to be a result of an inadequate training set. In a Monte Carlo simulation that generated 100 training sets of an appropriate size for the high-dimensional feature set (496 points per set), none gave negative eigenvalues for the covariance matrices. However, this is no guarantee that there will never be problems. To compensate, it is possible to modify the distance measure in order to make it positive semi-definite as follows.

First, one generates $[\mathbf{S}(T)]$ from the regression model as usual. Singular value decomposition is then used to give the breakdown,

$$[\mathbf{S}] = [\mathbf{U}][\mathbf{s}][\mathbf{U}]^{\mathrm{T}}, \tag{9}$$

where $[\mathbf{s}] = \mathrm{diag}(s_1, s_2, \ldots, s_p)$ is a diagonal matrix containing the eigenvalues in descending order of magnitude. If the first negative eigenvalue occurs in the nth element, then define $[\mathbf{s}^d]^{-1} = \mathrm{diag}(s_1^{-1}, s_2^{-1}, \ldots, s_{n-1}^{-1}, 0, \ldots, 0)$. The approximate, *but positive semi-definite by construction*, inverse to the original $[\mathbf{S}]$ is given by

$$[\mathbf{S}^d]^{-1} = [\mathbf{U}][\mathbf{s}^d]^{-1}[\mathbf{U}]^{\mathrm{T}}. \tag{10}$$

When this approximate inverse is used to compute the discordancy for the testing data shown in Figure 9, the results are as shown in Figure 10.

The result of using the approximate inverse is excellent. Note that the testing set containing normal data is above threshold, this is because the training set did not adequately sample the full normal condition at $T = 0$, this cannot be compensated for. However, the modified distance measure is only to be used in isolated cases where an adequate training set still generates a covariance matrix with a negative eigenvalue. Note that the procedure effectively eliminates the smallest eigenvalues of $[\mathbf{S}]$, and thus the largest eigenvalues of $[\mathbf{S}]^{-1}$. This means that the corrected Mahalanobis distance will always be smaller than one which uses an unbiased estimate of the covariance matrix. The more eigenvalues are deleted, the smaller will be the Mahalanobis distance. For the example shown above, deletion of one or two eigenvalues still gave a sensitive diagnostic; however, if eight eigenvalues were deleted, the Mahalanobis distance squared for the most severely damaged case ($D = 1$) fell below the threshold.

In order to assess the applicability of the training set, one can plot the number of eigenvalues deleted as a function of temperature over the range of interest as in Figure 11.

The plot shows that the main problems occur at the ends of the temperature range, and this is to be expected. In the range $T = 20$–28, only one eigenvalue is deleted and this would be adequately–compensated by using the approximate inverse for the covariance matrix. In the range $T = 62$–73, using the approximation would result in an almost complete loss of sensitivity to the damage. The figure confirms that the training set is inadequate.

When an eigenvalue deletion plot was constructed for the appropriately populated training set described at the beginning of this section, there were *no* deletions at any temperature.

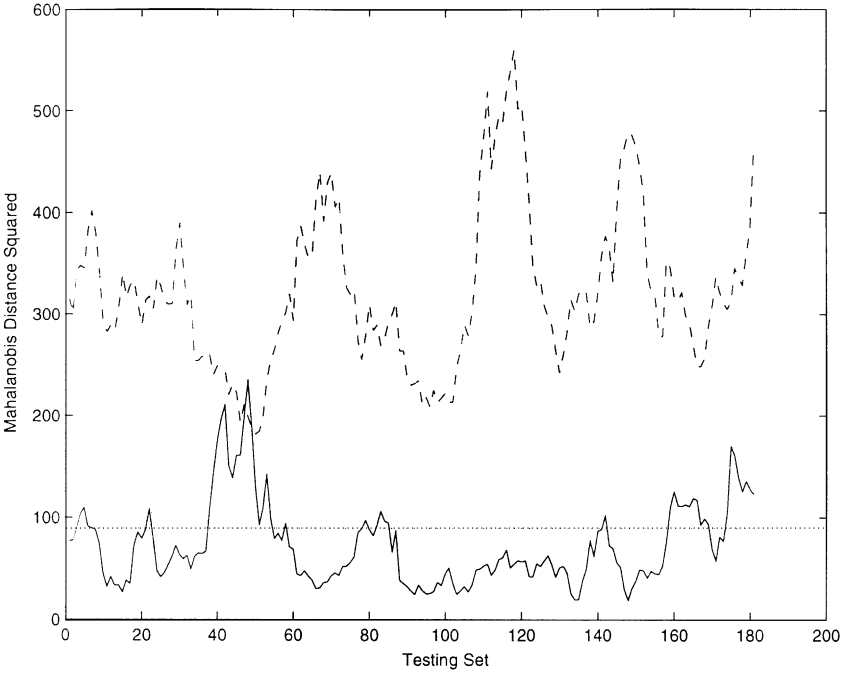

Figure 10. Results of outlier analysis using regression model with inadequate training data: compensating distance measure used. ——, $T = 20$, $D = 0$; – –, $T = 20$, $D = 1$.

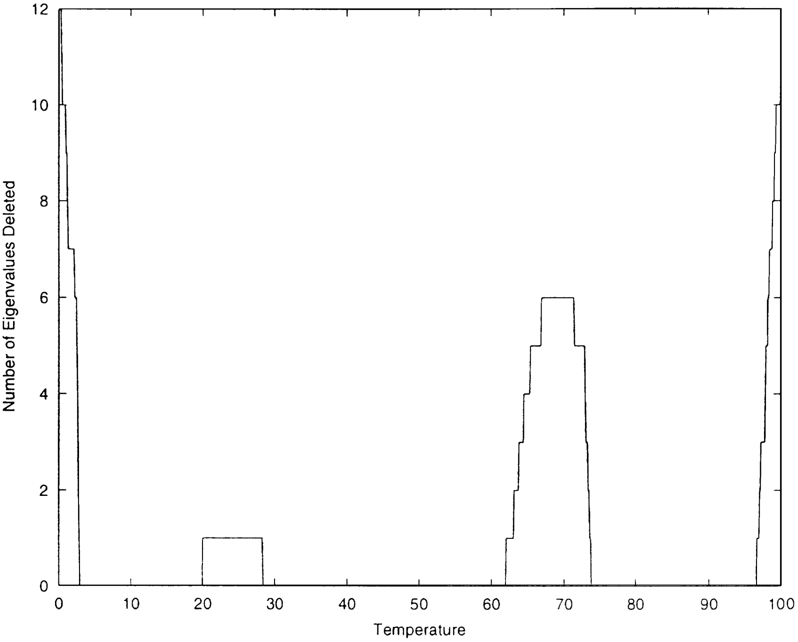

Figure 11. Number of eigenvalue deletions as a function of temperature.

6. AN INTERPOLATION APPROACH TO NORMALIZATION

The last section concluded that there are potential problems with the regression approach to normalization.

The problem is: given a training set, i.e., a set of mean vectors $\{\bar{\mathbf{x}}\}_i$ and covariance matrices $[\mathbf{S}]_i$, estimated for a set of measured temperatures T_i, $(i = 1, \ldots, N_T)$, estimate the appropriate mean and covariance matrix for a temperature T not in the measured set. The estimate for the covariance matrix should be symmetric and positive semi-definite.

There is a simple solution to this based on *interpolation* as opposed to regression.

Suppose $T_i < T < T_{i+1}$, define the interpolant $[\mathbf{S}(T)]$ by

$$[\mathbf{S}(T)] = (1 - \alpha)[\mathbf{S}]_i + \alpha[\mathbf{S}]_{i+1} \tag{11}$$

(with a similar formula for the mean), where

$$\alpha = (T - T_i)/(T_{i+1} - T_i). \tag{12}$$

This has the desired properties that if $T = T_i$, then $[\mathbf{S}(T)] = [\mathbf{S}]_i$ and similarly for $T = T_{i+1}$. Also, most importantly, the estimate is guaranteed to be positive semi-definite. Suppose $\{\mathbf{v}\}$ is an arbitrary vector, consider the scalar

$$\{\mathbf{v}\}^{\mathrm{T}}[\mathbf{S}(T)]\{\mathbf{v}\} = (1 - \alpha)\{\mathbf{v}\}^{\mathrm{T}}[\mathbf{S}]_i\{\mathbf{v}\} + \alpha\{\mathbf{v}\}^{\mathrm{T}}[\mathbf{S}]_{i+1}\{\mathbf{v}\}. \tag{13}$$

Because $[\mathbf{S}]_i$ and $[\mathbf{S}]_{i+1}$ are positive semi-definite, and α is between zero and one, the RHS is always greater than or equal to zero. Thus $\{\mathbf{v}\}^{\mathrm{T}}[\mathbf{S}(T)]\{\mathbf{v}\} \geqslant 0$ for any $\{\mathbf{v}\}$ and so $[\mathbf{S}(T)]$ is positive semi-definite.[‡]

In order to illustrate the approach, the same training set as in section 5 was used. The features were the AR(24) parameters estimated from the 100 000 point time records. Thus, each covariance matrix was estimated from 496 patterns. The temperatures ranged from $T = 0$ to 100 with a step of $\Delta T = 10$.

The testing set was made up of four components as before: (a) $T = 20$, $D = 0$, (b) $T = 20$, $D = 1$, (c) $T = 75$, $D = 0$, (d) $T = 75$, $D = 1$. Recall that $T = 75$ was *not* represented in the training set. Figure 12 shows the results of a discordancy test for each condition. The results are as good as those from the regression model as shown in Figure 8.

7. GENERALIZATION TO SYSTEMS WITH MORE THAN ONE ENVIRONMENTAL PARAMETER

7.1. REGRESSION

In the regression approach, the generalization to the case with more than one environmental parameter is simple to *formulate*. Suppose, for simplicity, that instead of temperature, there are two parameters θ_1, and θ_2, the idea is to compute mean vectors and covariance matrices in conditions spanning the range of environmental change and then fit *multinomial* regression models, i.e.,

$$\bar{x}_i \approx \sum_{j=0}^{N_p}\sum_{k=0}^{N_p} a_i^{jk}\theta^j\theta^k \quad \text{and} \quad S_{ij} \approx \sum_{k=0}^{M_p}\sum_{l=0}^{M_p} a_{ij}^{kl}\theta^k\theta^l. \tag{14, 15}$$

[‡] Thanks to Mike Fugate of the Pattern Recognition and Machine Learning group at LANL for pointing this out.

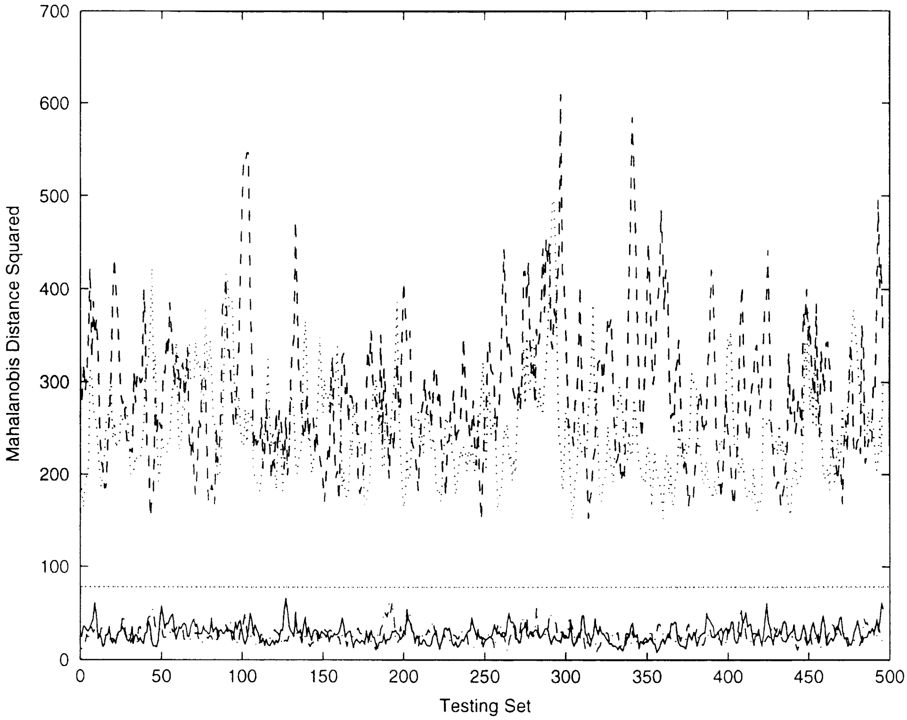

Figure 12. Results of outlier analysis using interpolation model on data from two different temperatures: high-dimensional features. ——, $T = 20$, $D = 0$; – –, $T = 20$, $D = 1$; –, $T = 75$, $D = 0$; . . ., $T = 75$, $D = 1$.

The problem with this approach is that the number of coefficients to fix grows explosively with the polynomial order of the model and the number of environmental parameters. A new curse of dimensionality comes into play. Not only should there be enough training patterns for each environmental condition to fix the normal condition distribution there, there should be enough normal conditions to adequately sample the space spanned by the environmental parameters.

7.2. INTERPOLATION

The generalization to more than one environmental parameter for the interpolation approach can be a little more demanding. However, the case of two parameters can be addressed using currently available software, so the strategy will be outlined here.

The simplest situation to deal with is if the parameters are sampled on a two-dimensional grid i.e., the points in the set can be labelled (i,j); the values of the parameters at the grid points will be denoted $\underline{\theta}^j = (\theta^j_{1i}, \theta^j_{2i})$. The spacings between points at i and $(i+1)$, for example, need not be constant; however, it is assumed that the sampling is organized so that the cells of the mesh are rectangular. Associated with each point is a mean vector $\{\bar{\mathbf{x}}\}^j_i$ and a covariance matrix $[\mathbf{S}]^j_i$. Interpolation in this case is fairly straightforward, suppose the values of the statistics are required at a new point with environmental co-ordinates $\underline{\theta}$. First, it is necessary to identify which cell in the mesh includes the new point as in Figure 13.

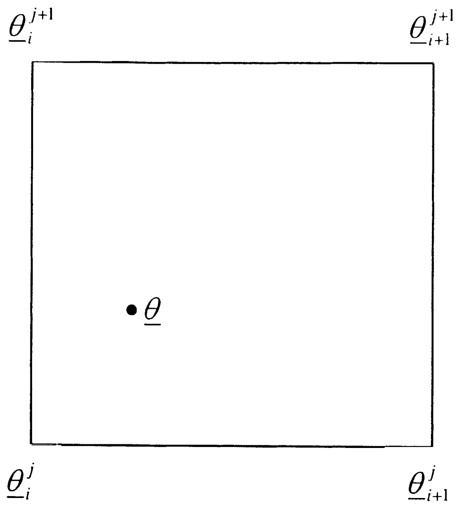
$\underline{\theta}_i^{j+1}$
$\underline{\theta}_{i+1}^{j+1}$
$\underline{\theta}$
$\underline{\theta}_i^{j}$
$\underline{\theta}_{i+1}^{j}$

Figure 13. Interpolation over a rectangular cell.

The interpolated value for $[\mathbf{S}(\underline{\theta})]$ is then given by Press *et al.* [12],

$$[\mathbf{S}(\underline{\theta})] = (1-\alpha)(1-\beta)[\mathbf{S}]_i^j + \alpha(1-\beta)[\mathbf{S}]_{i+1}^j + (1-\alpha)\beta[\mathbf{S}]_i^{j+1} + \alpha\beta[\mathbf{S}]_{i+1}^{j+1}, \qquad (16)$$

where

$$\alpha = (\theta_1 - \theta_{1i}^j)/(\theta_{1i+1}^j - \theta_{1i}^j) \quad \text{and} \quad \beta = (\theta_2 - \theta_{2i}^j)/(\theta_{2i}^{j+1} - \theta_{2i}^j) \qquad (17,18)$$

By a similar argument to that in the last section, the interpolant $[\mathbf{S}(\underline{\theta})]$ is guaranteed symmetric and positive semi-definite. This particular strategy generalizes straightforwardly to higher dimensions as long as the statistics are sampled on a regular grid, where in this case regular simply means that the cells of the mesh are rectilinear.

In general, it may be impossible to arrange that the samples of the statistics are obtained on a regular grid. In two dimensions, the interpolation is still tractable; the idea is that of Sibson's natural neighbour interpolation method [13].

The first stage in the process is to construct the Delauny triangulation defined by the sample points. The plane region of interest is decomposed into a contiguous set of triangles by the algorithm described in reference [13]. Each of the triangles will have three of the sample points as its vertices. The second part of the interpolation process is to find which triangle the new point $\theta = (\theta_1, \theta_2)$ falls inside. Once this is known, the situation is as shown in Figure 14; without loss of generality, the vertices are labelled $\underline{\theta_1}, \underline{\theta_2}$ and $\underline{\theta_3}$. Associated with the vertices are covariance matrix estimates: $[\mathbf{S}]_1$, $[\mathbf{S}]_2$ and $[\mathbf{S}]_3$.

The new point θ divides the triangle into three subtriangles. Each subtriangle is associated with the vertex opposite. If the areas of the subtriangles are A_i, $i = 1, \ldots, 3$, and the total area of the triangle is A, the normalized areas are defined as: $\lambda_i = A_i/A$. It is obvious that each $\lambda_i > 0$; it also follows that $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The values λ_i uniquely fix the position of the point θ within the triangle; because they also satisfy the properties given above, they are called a *barycentric co-ordinate* system within the triangle. Most importantly, they define a linear interpolant on the triangle defined by

$$[\mathbf{S}(\underline{\theta})] = \sum_{i=1}^{3} \lambda_i [\mathbf{S}]_i. \qquad (19)$$

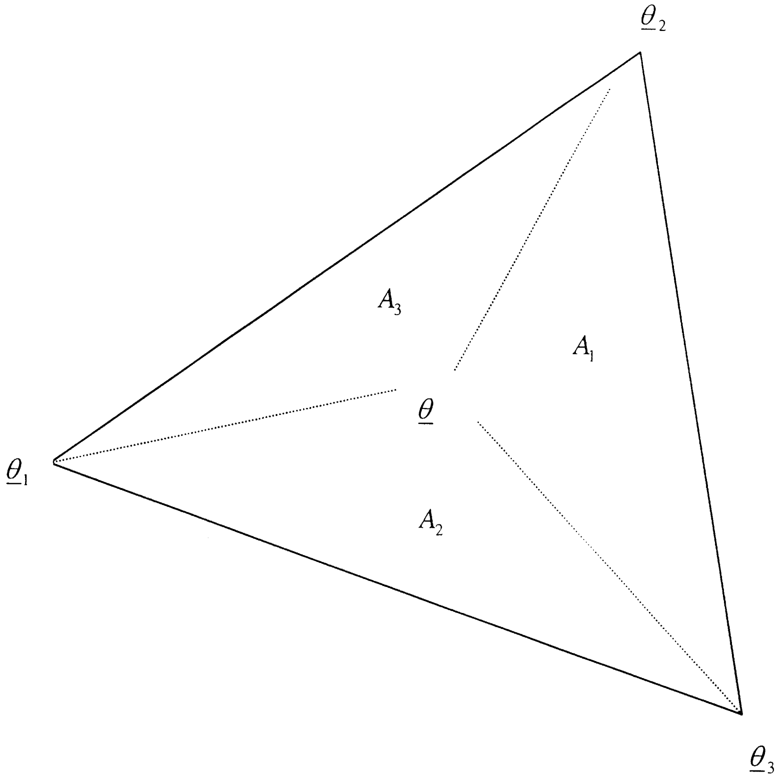

Figure 14. Interpolation over a triangular cell.

This interpolant is guaranteed to be symmetric and positive semi-definite by the same argument as used previously.

Software to construct the Delauny triangulation and the linear interpolant is available for the two-dimensional case [14]. For higher dimensions, commercial software is, to the author's knowledge, not available.

8. DISCUSSION AND CONCLUSIONS

This report discusses two possible approaches to normalization in novelty detection. Namely, the problem of removing environmental variations as a factor in deciding if new data are anomalous with respect to a previously measured normal condition set. The techniques discussed here are limited to the case where the environmental parameters of interest are measurable.

The first approach is via *regression*. The reference set is parametrized using the measured environmental parameters and a polynomial regression model is fitted for each coefficient of the relevant statistics. In this case, as basic outlier analysis is used as the novelty detector, regression models are fitted for the components of the mean vector and covariance matrix of the relevant features. The procedure was illustrated using features extracted by AR modelling of a group of time-histories, and the environment variable was a fictitious temperature parameter. It was shown that the regression approach was more (potentially much more) sensitive to damage than the previous normalization approach investigated by the author, where a large training set was constructed spanning all environmental conditions of interest. A potential problem associated with the regression approach when used with outlier analysis is that the model estimate of the covariance matrix at a given temperature is not guaranteed positive definite. This problem only appears to occur if the training data do not adequately sample the normal condition distribution of interest. A means of circumventing the problem is demonstrated together with a graphical method of assessing the suitability of the training set. The approach does not store the statistics from the training data, but stores the coefficients of the regression model.

Next, an interpolation method is demonstrated which does not suffer from the positive definiteness problem associated with regression. In contrast to the regression approach, the method stores the database of statistics spanning the environmental range of interest and interpolates within it as required. The interpolation shown here is linear.

For data with a sparse set of normal conditions, the regression approach may be attractive as the least-squares curve-fitting has a smoothing effect on the estimates. If adequate training data is available there is not expected to be too much difference between the two approaches. Both methods are shown to be effective on data from a system characterized by a single environmental parameter. For more than one parameter, both methods will suffer from the “curse of dimensionality”. The number of normal condition states required in order to adequately sample the range of environmental conditions will grow quickly with dimension. This requirement is distinct from the need to adequately sample the probability distribution of the features for a given environmental state. In procedural terms, the regression model is more easily generalized to higher numbers of environmental parameters, particularly if the training data is based on a set of environmentally varying samples that are irregularly distributed in the parameter space.

One issue which has been ignored here concerns measurement noise. In the case discussed here, the features are derived quantities—the AR parameters are estimated from measured excitation and response data. It is known that the presence of Gaussian noise on the measurements will induce a scatter on the AR parameter estimates, the extent of the scatter increasing with the r.m.s. of the original noise process. This may cause problems with the regression approach akin to those from inadequate training sets. If the noise is not Gaussian, the AR parameter estimates may be biased. If the noise is stationary, this may not be a problem. However, if the noise is non-stationary, the features will move in the feature space in a manner which might be interpreted as the result of damage. However, this caveat is equally valid applied to novelty detection problems with no environmental variation. The question of noise will be addressed in the next phase of this work; it is the intention of the authors to carry out an experimental study.

ACKNOWLEDGMENTS

The first author would like to thank Los Alamos National Laboratory for financial support while this study was carried out.

REFERENCES

1. A. Rytter 1993 *Ph.D. Thesis, Department of Building Technology and Structural Engineering, University of Aalborg, Denmark*. Vibration based inspection of civil engineering structures.
2. C. M. Bishop 1994 *IEEE Proceedings on Vision and Image Signal Processing* **141** 217–222. Novelty detection and neural network validation.

3. L. Tarassenko, P. Hayton, Z. Cerneaz and M. Brady 1995 *Proceedings of the 4th IEE International Conference on Artificial Neural Networks, Cambridge, U.K.* IEE Conference Publication, Vol. 409, 442–447. Novelty detection for the identification of masses in mammograms.
4. K. Worden 1997 *Journal of Sound and Vibration* **201** 85–101. Structural fault detection using a novelty measure.
5. V. Barnett and T. Lewis 1994 *Outliers in Statistical Data*. New York: John Wiley and Sons; third edition.
6. C. Surace and K. Worden 1997 *Proceedings of the 3rd International Conference on Modern Practice in Stress and Vibration, Dublin,* 89–94. Some aspects of novelty detection methods.
7. C. Surace and K. Worden 1998 *Proceedings of the 8th ISOPE, Canada,* 64–70 A novelty detection approach to diagnose damage in structures: an application to an offshore platform.
8. *Control System Toolbox for use with MATLAB* 1998. Natick, MA: The Math Works Inc.
9. G. E. P. Box, G. M. Jenkins and G. C. Reinsel 1994 *Time Series Analysis, Forecasting and Control*. Englewood Cliffs, NJ: Prentice-Hall.
10. *MATLAB—The Language of Technical Computing* 1998. Natick, MA: The Math Works Inc.
11. L. Ljung 1995 *System Identification Toolbox for use with MATLAB*. Natick, MA: The Math Works Inc.
12. W. H. Press, B. P. Filannery, S. A. Teukolsky and W. T. Vetterling 1986 *Numerical Recipes—The Art of Scientific Computing*. Cambridge: Cambridge University Press.
13. R. Sibson 1981a in *Interpreting Multivariate Data* V. Barnett, (editor) Chichester: John Wiley and Sons. A brief description of natural neighbour interpolation.
14. R. Sibson 1981b *TILE4—A Users Manual*. Department of Mathematics and Statistics, University of Bath.