Statistical Damage Classification using Sequential Probability Ratio Tests

Hoon Sohn† and David W. Allen

Engineering Sciences & Applications Division, Weapon Response Group, M/S P946
Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

Keith Worden

Department of Mechanical Engineering
University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.

Charles R. Farrar

Engineering Sciences & Applications Division, Weapon Response Group, M/S C946
Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

Abstract

The primary objective of damage detection is to ascertain with confidence if damage is present or not within a structure of interest. In this study, a damage classification problem is cast in the context of the statistical pattern recognition paradigm. First, a time prediction model, called an Auto Regressive-Auto Regressive model with Exogenous inputs (AR-ARX), model is fit to a vibration signal measured during a normal operating condition of the structure. When a new time signal is recorded from an unknown state of the system, the prediction errors are computed for the new data set using the time prediction model. When the structure undergoes structural degradation, it is expected that the prediction errors will increase for the damage case.

† E-mail: sohn@lanl.gov (corresponding author)

Based on this premise, a damage classifier is constructed using a sequential hypothesis testing technique called a sequential probability ratio test (SPRT). The SPRT is one form of parametric statistical inference tests and the adoption of the SPRT to damage detection problems can improve the early identification of conditions that could lead to performance degradation and safety concerns. The sequential test assumes the probability distribution of the sample data sets, and a Gaussian distribution of the sample data sets is often assumed. This assumption, however, might impose potentially misleading behavior on the extreme values of the data i.e. those points in the tails of the distribution. As the problem of damage detection specifically focuses attention on the tails, the assumption of normality is likely to lead the analysis astray. To overcome this difficulty, the performance of the sequential hypothesis test is improved by integrating extreme values statistics, which specifically model behavior in the tails of the distribution of interest, into the sequential probability ratio test.

KEYWORDS: damage detection, time series analysis, sequential probability ratio test, extreme value statistics, statistical pattern recognition, vibration test.

1. Introduction

The most primary goal of structural health monitoring and damage detection is simply to identify from measured data if a structure of engineering interest has deviated from a normal operational condition. Particularly, vibration-based damage detection techniques assume that changes of the structure's integrity affect characteristics of the measured vibration signals enabling one to detect damage. The area of the SHM that receives the most attention in the technical literature is feature extraction (Doebling et al., 1998). Feature extraction is the process of the identifying damage-sensitive properties, derived from the measured vibration response, which allows one to distinguish between the undamaged and damaged structures. On the other hand, the least attention is paid to the development of statistical inference tools to enhance the actual damage classification process. A statistical inference is concerned with the implementation of the algorithms that operate on the extracted features to quantify the damage state of the structure.

In this paper, a unique combination of time series analysis, statistical pattern recognition techniques, and extreme value statistics is presented to automate the damage identification problem with a special attention to the statistical modeling for decision-making. The structure of the report is as follows. Section 2 briefly reviews the time series analysis of vibration signals using the AR-ARX model. Section 3 outlines the main theory of sequential probability ratio test, and Section 4 extends the SPRT to extreme value statistics. The SPRT is applied to numerical and experimental data in Sections 5 and 6, respectively. Section 7 concludes and summarizes the findings of this study.

2. Time Series Analysis

The time series analysis begins with the assumption that a "pool" of signals is acquired from a known structural condition of the system. In the experimental example reported later on, multiple time series are recorded from the undamaged structure. The collection of these time series is called the "reference database" in this study. The construction of this reference database is shown to be useful for normalizing data with respect to varying operational and environmental conditions. The applications of this time series analysis to data normalization are presented in Sohn and Farrar (2001) and Sohn et al. (2001).

A linear prediction model combining AR and ARX models is employed to compute the damage-sensitive feature. In this case, the damage-sensitive feature is the residual error between the prediction model and measured time series.

First, all time signals are standardized prior to fitting an AR model such that:

$$\hat{x} = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

where $\hat{x}$ is the standardized signal, μ_x and σ_x are the mean and standard deviation of *x*, respectively. This standardization procedure is applied to all signals employed in this study. (However, for simplicity, *x* is used to denote $\hat{x}$ hereafter.)

For each time series *x(t)* in the reference database, an AR model with *r* auto-regressive terms is constructed. An AR(*r*) model can be written as (Box et al., 1994):

$$x(t) = \sum_{j=1}^{r} \phi_{xj}\ x(t-j) + e_x(t) \tag{2}$$

This step is repeated for all signals in the reference database.

Employing a new segment *y(t)* obtained from unknown structural condition of the system, repeat the previous step. Here the new segment $y(t)$ has the same length as the signal $x(t)$:

$$y(t) = \sum_{j=1}^{r} \phi_{yj}\ y(t-j) + e_y(t) \tag{3}$$

Then, the signal segment $x(t)$ "closest" to the new signal block $y(t)$ is defined as the one that minimizes the following difference of AR coefficients:

$$\text{Difference} = \sum_{j=1}^{r} (\phi_{xj} - \phi_{yj})^2 \tag{4}$$

This "data normalization" is a procedure to select the previously recorded time signal from the reference database, which is recorded under operation and/or environmental conditions closest to that of the newly obtained signal. If the new signal block is obtained from an operational condition close to one of the reference signal segments and there has been no structural deterioration or damage to the system, the dynamic characteristics (in this case, the AR coefficients) of the new signal should be similar or "closest" to those of the reference signal based on the Euclidean distance measure in Equation (4).

When a time prediction model is constructed from the selected reference signal, this prediction model should be able to appropriately predict the new signal if the new signal is "close" to the reference signal. On the other hand, if the new signal were recorded under a structural condition different from the conditions where reference signals were obtained, the prediction model estimated from even the "closest" signal in the reference database would not reproduce the new signal well.

For the construction of a two-stage prediction model proposed in this study, it is assumed that the error between the measurement and the prediction obtained by the AR model ($e_x(t)$ in Equation (2)) is mainly caused by the unknown external input. Based on this assumption, an ARX model is employed to reconstruct the input/output relationship between $e_x(t)$ and $x(t)$.

$$x(t) = \sum_{i=1}^{p} \alpha_i \; x(t-i) + \sum_{j=1}^{q} \beta_j \; e_x(t-j) + \varepsilon_x(t) \tag{5}$$

where $\varepsilon_x(t)$ is the residual error after fitting the ARX(*p,q*) model to $e_x(t)$ and $x(t)$ pair. The feature for damage diagnosis will later be related to this quantity, $\varepsilon_x(t)$. Note that this AR-ARX modeling is similar to a linear approximation method of an Auto-Regressive Moving-Average (ARMA) model presented in Ljung 1987 and references therein. Ljung (1987) suggested keeping the sum of p and q smaller than r ($p+q \le r$). Although the p and q values of the ARX model are set rather arbitrarily, similar results are obtained for different combinations of *p* and *q* values as long as the sum of *p* and *q* is kept smaller than *r*.

Next, it is investigated how well this ARX(*p,q*) model estimated in Equation (5) reproduces the input/output relationship of $e_y(t)$ and $y(t)$:

$$\varepsilon_y(t) = y(t) - \sum_{i=1}^{p} \alpha_i \; y(t-i) - \sum_{j=0}^{q} \beta_j \; e_y(t-j) \tag{6}$$

where $e_y(t)$ is considered to be an approximation of the system input estimated from Equation (3). Again, note that the α_i and β_j coefficients are associated with $x(t)$ and obtained from Equation (5). If the ARX model obtained from the reference signal block $x(t)$ and $e_x(t)$ pair were not a good representative of the newly obtained signal segment $y(t)$ and $e_y(t)$ pair, there would be a significant change in the standard deviation of the residual error, $\varepsilon_y(t)$, compared to that of $\varepsilon_x(t)$. Therefore, the standard deviation of the residual error is defined as the damage-sensitive feature and the increase of this standard deviation is monitored to using the following sequential probability ratio test.

3. Damage Classification using Sequential Probability Ratio Tests

The SPRT procedure is particularly relevant if the data is collected sequentially (Wald, 1947). Examples of such sequential collection include failures on a production line, patient throughput in a hospital or relapses in behavioral interventions. Sequential Analysis is different from classical hypothesis testing where the number of cases tested or collected is fixed at the beginning of the experiment. In classical hypothesis testing the data collection is executed without analysis and consideration of the data. After all data are collected, the analysis is done and conclusions are drawn. However, in sequential analysis every case is analyzed directly after being collected, the data collected up to that moment is then compared with threshold values, incorporating the new information obtained from the freshly collected case. This approach allows one to draw conclusions during the data collection, and a final conclusion can possibly be reached at a much earlier stage as is the case in classical hypothesis testing. The advantages of sequential analysis are easy to see. As data collection can be terminated after fewer cases and decisions drawn earlier, the savings in terms of human life and misery, and financial savings, might be considerable. Particularly, the framework of this sequential analysis suits the paradigm of continuous structural health monitoring very well.

3.1 Sequential Test

A sequential statistical inference starts with observing a sequence of random variables $\{x_i\}$ $(i = 1, 2, \ldots)$. This accumulated data set at stage n is denoted as:

$$X_n = (x_1, \ldots x_n) \tag{7}$$

The goal of a statistical inference is to reveal the probability model of X_n, which is assumed to be at least partially unknown. When the statistical inference is cast as a parametric problem, the functional form of X_n is assumed known and the statistical inference poses some questions regarding the parameters of the probability model. For instance, if $\{x_i\}$ are independent and identically distributed (i.i.d.) normal variables, one may pose some statistical test about the mean and/or the variance of this normal distribution.

A sequential test is one of the simplest tests for such a statistical inference where the number of samples required before reaching a decision is not determined in advance. An advantage of the sequential test is that on average a smaller number of observations are needed to make a decision compared to the conventional fixed-sample size test. First, a simple hypothsis test containing only two distinct distributions is considered. Here, the interest is in discriminating two simple hypotheses:

$$H_o : \theta = \theta_o, \quad H_1 : \theta = \theta_1, \quad \theta_o \neq \theta_1 \tag{8}$$

where θ is a particular parameter value in question[1], and it is assumed that θ can take either θ_o or θ_1 only. When a sequence of observations $\{x_i\}$ are available, the purposes of any sequential test for the above hypotheses are (1) to reach the correct decision about H_o with the least probability of type I and II errors[2], and (2) to minimize the sampling number before the correct decision is made, and (3) to eventually terminate with either the acceptance or rejection of H_o as the sampling size n increases. When a sequential test satisfies the last condition, the test is defined *closed*. Otherwise, an *open* test may continue infinitely observing data without reaching any terminal decision about H_o.

It turns out that the simultaneous achievment of all three goals is impossible by any test. Therefore, a reasonable compromise among the these conflicting goals needs to be achieved. For the well-established fixed-sampling tests, the sample size n is fixed, and an upper bound on the type I error is pre-specified. Then, an optimal fixed-sample test is selected by minimizing the probability of type II error. On the other hand, a sequential test specifies upper bounds on the probabilities of type I and II errors and minimizes the following average sample number, $\mathrm{E}(n \mid \theta)$:

---

[1] In general, θ can be a vector of multiple parameters. However, θ is assumed to be a single parameter for simplicity in this paper.

$$E(n\,|\,\theta) = \sum_{n=1}^{\infty} n\, p(n\,|\,\theta) \tag{9}$$

where $p(n\,|\,\theta)$ is the probability mass function of n when θ is the true value of the parameter. Note that for a closed test $p(n<\infty\,|\,\theta)=1$ for $\theta = \theta_o$ or θ_1.

There exist a class of sequential tests, and sequential tests, which satisfy the following criteria are called *valid* (Ghosh, 1970):

$$\begin{aligned}&(1)\ \text{The test is closed.}\\ &(2)\ 1-Q(\theta)\le\alpha \ \text{ for } \theta=\theta_o\\ &(3)\ Q(\theta)\le\beta \ \text{ for } \theta=\theta_1\end{aligned} \tag{10}$$

where α and β are the preassined type I and II errors, respectively. $Q(\theta)$ is the probability that any sequential test accepts H_o as $n\to\infty$. In other words,

$$Q(\theta) = \sum_{n=1}^{\infty} \int_{X_n\in R_n^o} f(X_n\,|\,\theta)\,dX_n \tag{11}$$

where $f(X_n\,|\,\theta)$ (change the equation size to 12 pt later) is the conditional probabilty of observing the accumuated data set X_n given the assumption of θ. The integral in Equation (11) is evaluated over the acceptance region of H_o ($X_n \in R_n^o$). The second criterion in Equation (10) states that for all values of n, the true type I error, $1-Q(\theta_o)$, should be less than the pre-assigned risk α. In a similar fashion, the third criterion indicates that the true type II error $Q(\theta_1)$ should be less than β. Among various valid sequential tests, it can be proven that the SPRT minimizes on average the sample size required to make a correction making it an optimal sequential test.

---

[2] Type I error arises if H_o is rejected when in fact it is true. Type II error arises if H_o is accepted when it is false.

Because of this extreme sensitivity of the SPRT to signal disturbance, the SPRT has been applied for the surveillance of nuclear power plant components (Gross and Humenik, 1991).

When implementing the SPRT, a trade-off must be considered before assigning values for α and β. When there is a large penalty associated with false positive alarms (for example, alarms that shut down traffic over a bridge), it is desirable to keep α smaller than β. On the other hand, for safety critical systems such as nuclear power plants, one might be more willing to tolerate a false positive alarm to have a higher degree of safety assurance. In this case, it is not uncommon to specify β larger than α.

3.2 Sequential Probability Ratio Test

A SPRT, S(*b,a*), for the hypothesis test in Equation (8) is defined as follows (Ghosh, 1970): Observe a sequence of observations $\{x_i\}$ $(i=1,2,\ldots)$ successively, and at stage n;

$$\begin{aligned} &(1)\ \text{Accept } H_o \text{ if } Z_n \le b \\ &(2)\ \text{Reject } H_o \text{ if } Z_n \ge a \\ &(3)\ \text{Continue observing data if } b \le Z_n \le a \end{aligned} \tag{12}$$

where the transformed random variable Z_n is the natural logarithm of the probability ratio at stage n (It should be clear by now why this test is called a sequential probability ratio test):

$$Z_n = \ln \frac{f(X_n \,|\, \theta_1)}{f(X_n \,|\, \theta_o)} \text{ for } n \ge 1 \tag{13}$$

Without any loss of generality, Z_n is defined zero when $f(X_n \,|\, \theta_1) = f(X_n \,|\, \theta_o) = 0$. b and a are the two stopping bounds for accepting and rejecting H_o, respectively, and they can be estimated by the following Wald approximations (Wald, 1947):

$$b \cong \ln\frac{\beta}{1-\alpha} \text{ and } a \cong \ln\frac{1-\beta}{\alpha} \tag{14}$$

Although closed form solutions of a and b are available for several probability models, it has been a standard practice to employ Equation (14) to approximate the stopping bounds in all practical applications. The continuation region $b \le Z_n \le a$ is called the *critical inequality* of S(*b,a*) at stage n.

In many practical problems, it is often more realistic to formulate the hypothesis test as discrimination between two one-sided hypotheses:

$$H_o : \theta \le \theta_o, \quad H_1 : \theta \ge \theta_1, \quad \theta_o < \theta_1 \tag{15}$$

The criteria in Equation (10) are now equivalent to

(1) The test is closed.

(2) $1 - Q(\theta) \le \alpha$ for $\theta \le \theta_o$ (16)

(3) $Q(\theta) \le \beta$ for $\theta \ge \theta_1$

Ghosh (1970) shows that the previous SPRT shown in Equation (12) also provides an optimal solution to this hypothesis test defined in Equation (15).

3.3 Applications to Normal Distribution

In the damage detection problem presented, the main interest is t examine how the probability distribution function of the residual errors broadens as data are recorded under a damaged condition of a system. Therefore, the following hypothesis test is constructed using the standard deviation of the residual errors as the parameter in question:

$$H_o : \sigma \le \sigma_o, \quad H_1 : \sigma \ge \sigma_1, \quad 0 < \sigma_o < \sigma_1 < \infty \tag{17}$$

Here, when the standard deviation of the residual errors, σ, is less than a user specified standard deviation value σ_o, the system in question is considered undamaged. On the other hand, when σ becomes equal to or larger than the other user specified standard deviation σ_1, the system is suspected to be damaged. It should be noted that the selection of σ_o and σ_1 is structure dependent, and it might be necessary to use signals from a few damage cases in order to establish these two decision boundaries.

If modified observations $\{z_i\}$ $(i = 1, 2, \ldots)$ are defined as follows;

$$z_1 = \ln \frac{f(X_1 \mid \sigma_1)}{f(X_1 \mid \sigma_o)} \text{ and } z_i = \ln \frac{f(X_i \mid \sigma_1) f(X_{i-1} \mid \sigma_o)}{f(X_i \mid \sigma_o) f(X_{i-1} \mid \sigma_1)} \tag{18}$$

then, Z_n becomes:

$$Z_n = \sum_{i=1}^{n} z_i \tag{19}$$

Assuming that X_n has a normal distribution with mean μ and standard deviation σ, z_i can be related to x_i:

$$z_i = \frac{1}{2}(\sigma_o^{-2} - \sigma_1^{-2})(x_i - \mu)^2 - \ln \frac{\sigma_1}{\sigma_o} \tag{20}$$

In a graphical representation of a SPRT S(*b*,*a*), Z_n, which is the cumulative sum of the transformed variable z_i, is continuously plotted against the two stopping bounds *b* and *a*. It should be noted that the mean μ of the distribution is assumed to be known. Even when μ is unkown, the aforementioned procedure is still valid if x_i is replaced by y_i:

$$y_i = \left(\sum_{j=1}^{i} x_j - i\, x_{i+1} \right) \Big/ \sqrt{i(i+1)} \text{ for } i = 1, 2, \ldots \tag{21}$$

It can be shown that now $\{y_i\}$ has i.i.d. normal distribution with zero mean and the same standard deviation as $\varepsilon_y(i)$.

4. Extreme Value Statistics

In the previous section, the SPRT procedure is formulated assuming that the sampled data have a normal distribution. However, the assumption of normality might impose potentially misleading behavior on the extreme values of the data, namely, those points in the tails of the distribution. An alternative approach can be based on extreme value statistics. This branch of statistics was developed to specifically model behavior in the tails of the distribution of interest.

In fact, there is a large body of statistical theory that is explicitly concerned with modeling the tails of distributions, and these statistical procedures are applied to the current problem of damage classification. The relevant field is referred to as *extreme value statistics*, a branch of *order statistics*. There are many excellent textbooks and monographs in this field. Some are considered classics (Gumbel, 1958; Galambos, 1978), and others are more recent (Embrechts et al., 1997; Kotz and Nadarajah, 2000). Castillo (1988) is notable in its concern with engineering problems in fields like meteorology, hydrology, ocean engineering, pollution studies, strength of materials, etc. Although extreme value statistics has been widely applied, there has been little application of these techniques to damage detection.

The major problems with modeling the normal condition of a system are that the functional form of the distribution is unknown and that there are an infinite number of candidate distributions that may be appropriate for the prediction applications. The researcher must choose among various distributions and then estimate parameters based on training data. This process is largely subjective. If instead of working with the central statistics of a distribution, extreme value statistics are applied to the tails, there are only three candidate distributions for the tails and the problem of model selection and parameter estimation becomes more objective.

Suppose that one is given a vector of samples $(x_1,\ldots,x_n)$ from an arbitrary *parent distribution*. The most relevant statistic for studying the tails of the parent distribution is the maximum operator, max($x_1,\ldots,x_n$), which selects the point of maximum value from the sample vector. Note that this statistic is relevant for the right tail of a univariate distribution only. For the left tail, the minimum should be used. The pivotal theorem of extreme value statistics (Fisher and Tippett, 1928) states that in the limit as the number of vector samples tends to infinity, the induced distribution on the maxima of the samples can only take one of three forms: Gumbel, Weibull, or Frechet. The rest of this section will be concerned with elaborating on this fact.

If the values of the sequence $(x_1,\ldots,x_n)$ are arranged in ascending order, the r^{th} element of this sequence x_r is called the *r*th *order statistic*. The basic question, which now arises is, what are the distributions of the order statistics, in particular, the minimum, x_1, and the maximum, x_n.

Following Castillo (1988), let $m_n(x)$ be the number of samples for which $x_j \leq x.$ Each time one chooses a value x_j from the sample, one is conducting a Bernoulli experiment, an experiment that has one of two outcomes, with a probability $F(x)$, the cumulative distribution function (CDF), that $x_j \leq x$, and the complementary probability, $1-F(x)$, that $x_j > x$. The CDF of $m_n(x)$ is, therefore, a binomial distribution with $F^k(x)$ denoting the probability of success

$$F_{m_n(x)}(r) = \text{Prob}[m_n(x) \le r] = \sum_{k=0}^{r} \binom{n}{k} F^k(x)[1 - F(x)]^{n-k} \tag{22}$$

Now, because the event $(x_r \le x)$ is basically the same as the event $(m_n(x) \ge r)$, $\text{P}(x_r \le x)$ is identical to $p(m_n(x) \ge r)$ = $1 - P(m_n(x) < r)$. In addition, it follows that $F_{x_r}(x)$ = $1 - F_{m_n(x)}(r-1)$ or

$$F_{x_r}(x) = \text{P}(x_r \le x) = \sum_{k=r}^{n} \binom{n}{k} F^k(x)[1 - F(x)]^{n-k} \tag{23}$$

If one is concerned with the maximum of the sample, the relevant order statistic is X_n and the relevant distribution is

$$F_{x_n}(x) = F^n(x) \tag{24}$$

Concentrating now on the maximum, let $n \to \infty$, then the limit distribution for the maximum will satisfy

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 1 & \text{If } F(x) = 1 \\ 0 & \text{If } F(x) < 1 \end{cases} \tag{25}$$

This distribution doesn't make sense because a CDF is developed on the assumption that it is continuous, but here the limit is discontinuous. The way around this discontinuity is to normalize the independent variable with a sequence of constants ($x \to a_n + b_n x$) in such a way that

$$\lim_{n\to\infty} F^n(a_n + b_n x) = F_M(x) \tag{26}$$

where $F_M(x)$ is a non-degenerate limit function. In fact, it is required that $F_M(x)$ be continuous.

Fisher and Tippett (1928) state that, in the limit as the number of vector samples tends to infinity, the induced distribution $F_M(x)$ in Equation (26) can only take one of the following three forms: Gumbel, Weibull, or Frechet:

Frechet:
$$F_M(x) = \begin{cases} \exp\left[-\left(\dfrac{\delta}{x-\lambda}\right)^{\beta}\right] & \text{if} \quad x \geq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

Weibull
$$F_M(x) = \begin{cases} 1 & \text{if} \quad x \geq \lambda \\ \exp\left[-\left(\dfrac{\lambda - x}{\delta}\right)^{\beta}\right] & \text{otherwise} \end{cases} \tag{28}$$

Gumbel
$$F_M(x) = \exp\left[-\exp\left(-\frac{x-\lambda}{\delta}\right)\right] \qquad \begin{matrix} -\infty < x < \infty \\ \delta > 0 \end{matrix} \tag{29}$$

where λ, α, and β are the model parameters, which should be estimated from the data. $F_x(x)$ is, in fact, a cumulative density function of maxima and the subscript "M" is used to denote that the distribution is for the maxima. Note that these distributions are relevant for the right tail of a univariate distribution only. For the left tail, similar distributions for the minimum can be obtained.

Now given samples of maximum data from a parent population, it is possible to select an appropriate limit distribution and fit a parametric model to the data. It is also possible to fit models to portions of the parent distribution's tails as these models are equivalent in the tail to the appropriate extreme value distribution. Once the appropriate model is obtained, the SPRT can be reformulated using the know distribution type of the extreme values. In this paper, the discussion is limited to the Gumbel distribution for maxima but similar derivation can be obtained for the other extreme distributions.

4.1 A Sequential Probability Test using a Gumbel Distribution for Maxima

Now, the SPRT is extended to the extreme values of the parent distribution, the distribution of the residual errors. In the previous section, the SPRT is formulated assuming that the residual errors have a normal distribution. However, slight errors in the normality assumption of the parent distribution can lead to larger errors for the extremes resulting in errneous false positive/negative indications of damage. To avoid this problem, the SPRT is reformualted using the probability distributions of extreme values. It should be reminded that there are only three possible choices for the distributions of the extremes regardless the parent distrubution types. Particulary, because the maxima of a normal distribution are known to have a Gumbel distribution and the resiual errors of the experimental study presented later are close to a normal distribution, the derivation presented here focuses on incorprating Gumbel distribution for maxima values into the SPRT. Similar formualtion can be easily derived for other types of extreme value distribution and for minima values.

Similar to Equation (17), the following hypothesis test is constructed using the standard deviation of the maxima as the parameter in question:

$$H_o : \sigma_M \leq \sigma_o, \quad H_1 : \sigma_M \geq \sigma_1, \quad 0 < \sigma_o < \sigma_1 < \infty \tag{30}$$

Now, σ_M is the standard deviation of the residual error maxima, and the subscript "M" denotes a qunatity related to the maxima. σ_o is a user specified lower limit of the standard deviation for the undamage contion, and σ_1 is the other user specified upper limit for the damage condition. It is observed that the change of the maxima distribution's standard deviation is monotomically related to the change of the parent distribution's standard deviation. Here, an indirect statitstical inference on the standard deviation of the parent distribution (the distribution of the residaul errors) is conducted by examining the standard deviation of the maximum values.

It can be shown that the model parameters, λ and σ, of the Gumble distribution are related to its mean μ_M and standard deviation σ_M (Castillo, 1987):

$$\delta = \frac{\sqrt{6}}{\pi}\sigma_M \text{ and } \lambda = \mu_M - 0.57772\delta \tag{31}$$

If the distribution of the maxima is preprocessed such that the mean value is zero, Equation (18) can be rewritten in terms of λ and σ:

$$z_1 = \ln\frac{f(X_1 \mid \lambda_1, \delta_1)}{f(X_1 \mid \lambda_o, \delta_o)} \text{ and } z_i = \ln\frac{f(X_i \mid \lambda_1, \delta_1) f(X_{i-1} \mid \lambda_o, \delta_o)}{f(X_i \mid \lambda_o, \delta_o) f(X_{i-1} \mid \lambda_1, \delta_1)} \tag{32}$$

If $\{x_i\}$ are independent and identically distributed (i.i.d.), $f(X_i \mid \lambda_1, \delta_1)$ becomes $f(x_1 \mid \lambda_1, \delta_1) \times f(x_2 \mid \lambda_1, \delta_1) \times \cdots \times f(x_i \mid \lambda_1, \delta_1)$ and Equation (32) can be further simplified as follows:

$$z_i = \ln\frac{f(x_i \mid \lambda_1, \delta_1)}{f(x_i \mid \lambda_o, \delta_o)} \text{ for } i = 1, 2, \ldots, n \tag{33}$$

Next, the probability density function of the Gumbel distributionfor maxiam is obtained by differentiating the cumulative density function presented in Equation (29).

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\delta}\exp\left(-\frac{x-\lambda}{\delta}\right)\exp\left[-\exp\left(-\frac{x-\lambda}{\delta}\right)\right] \tag{34}$$

By substituting Eqaution (34) into Equation (33), z_i can be related to x_i:

$$z_i = -\ln\frac{\delta_1}{\delta_o} + \left(\frac{x_i - \lambda_o}{\delta_o}\right) - \left(\frac{x_i - \lambda_1}{\delta_1}\right) + \exp\left(-\frac{x_i - \lambda_o}{\delta_o}\right) - \exp\left(-\frac{x_i - \lambda_1}{\delta_1}\right) \tag{35}$$

By relating λ and σ to σ_M as shown in Equation (31), Equation (35) can be furhter simplied as follows:

$$z_i = -\ln\frac{\sigma_1}{\sigma_o} + \frac{\pi}{\sqrt{6}}\left(\sigma_o^{-1} - \sigma_1^{-1}\right)x_i + \exp\left(-\frac{x_i + 0.4504\,\sigma_o}{\sqrt{6}\;\sigma_o/\pi}\right) - \exp\left(-\frac{x_i + 0.4504\,\sigma_1}{\sqrt{6}\;\sigma_1/\pi}\right) \tag{36}$$

Finally, the cumulative sum of the transformed variable, Z_i, is monitored against the two stopping bounds, *a* and *b*.

4.2 A SPRT using Extreme Value Statistics with a Known Gaussian Parent Distribution

In this section, the SPRT is modified assuming that the parent distribution of the maxima has a known Gaussian distribution. Equaiton (24) shows that when the parent distribution has a cumulative densition function $F(x)$, the cumulative density function for the maxima extracted from a sample size *n* becomes $F^n(x)$. Then, the cumulative density function and the associated probability density function of maxima are obtained:

$$F_M(x) = F^{n-1}(x) \text{ and } f_M(x) = n\,F^{n-1}(x)\,f(x) \tag{37}$$

where $F_M(x)$ and $f_M(x)$ are the CDF and PDF of the maxima values, and $F(x)$ and $f(x)$ are the CDF and PDF of a normal distribution, respectively. By substituting Equation (37) into Equation (18), the following z_i statistics are obtained:

$$\begin{aligned} z_i &= \ln\frac{f_M(x_i\,|\,\sigma_1)}{f_M(x_i\,|\,\sigma_o)} = \ln\frac{n\,F^{n-1}(x_i\,|\,\sigma_1)\,f(x_i\,|\,\sigma_1)}{n\,F^{n-1}(x_i\,|\,\sigma_o)\,f(x_i\,|\,\sigma_o)} \\ &= \frac{1}{2}(\sigma_o^{-2} - \sigma_1^{-2})(x_i - \mu)^2 - \ln\frac{\sigma_1}{\sigma_o} + (n-1)\,\ln\frac{F(x_i\,|\,\sigma_1)}{F(x_i\,|\,\sigma_o)} \end{aligned} \tag{38}$$

Note that, when the sampling size for the maxima becomes one $(n=1)$, Equation (38) reduces back to Equation (20).

5. Numerical Examples

In this section, the performances of three variations of the SPRT are compared for different types of parent distributions. The three variations of the SPRT include (1) the conventional SPRT with the normality assumption of data sets [Equation (20)], (2) a SPRT using a Gumbel distribution for maxima [Equation (36)], and (3) a SPRT using extreme value statistics with a known Gaussian parent distribution [Equation (38)]. Hereafter, these techniques are referred to as SPRT-1, SPRT-2, and SPRT-3, respectively. These three SPRT techniques are applied to simulated data sets with Gaussian, lognormal, and Gamma distributions.

From a given distribution type of population, two data sets are randomly generated. The first set of data consists of 8192 data points and has a known standard deviation of σ_x. The second data set also consists of 8192 data points and has an increased standard deviation of $\sigma_y = F\sigma_x$. Here, F is an amplifying constant varying from 0.90 to 1.00, 1.10, 1.15, 1.45, 1.50, 1.60 and 1.70. The first data set simulates the residual errors from the initial intact condition of the structure, and the second data set represents the residual errors from a new structural condition of the structure.

The damage classification problem is cast in such a way that, if the standard deviation of the new signal, σ_y, becomes above a predetermined upper limit, $1.4\sigma_x$, then the new signal is considered from a damaged state of the system. On the other hand, if σ_y is less than the other predetermined lower limit, $1.2\sigma_x$, the new signal is assumed to be from the undamaged condition. Otherwise (when $1.2\sigma_x$ < σ_y < $1.4\sigma_x$), the damage classifier cannot make a confident decision regarding the current state of the structure and needs to continue collecting additional data. This sequential hypothesis test can be stated in a simplified format:

$$H_o : \sigma_y \leq 1.2\sigma_x \quad \text{and} \quad H_1 : \sigma_y \geq 1.4\sigma_x \tag{39}$$

Because the statistical inference in Equation (39) is cast only for the unknown standard deviation σ_y, it is assumed that the mean of the signals is known. Therefore, the mean of each signal is subtracted from the raw signal.

When the SPRT is combined with maximum value statistics (SPRT-2 and SPRT-3), a moving window of width 16 samples is stepped through the 8192 points of each data set to generate 512 maxima for each condition. For all numerical examples, the upper bounds of type I and II errors are set to 0.001. The corresponding two bounds are b = –6.9 and a = 6.9, respectively It should be noted that because the parent distribution is assumed unknown for SPRT-2, the hypothesis test in Equation (30) cannot be performed and an alternative hypothesis test is conducted on the standard deviation of the maximum values.

$$H_o : \sigma_{y,M} \leq 1.2\sigma_{x,M} \quad \text{and} \quad H_1 : \sigma_{y,M} \geq 1.4\sigma_{x,M} \tag{40}$$

where the subscript denotes the quantifies for maximum values.

Three different parent distribution types are investigated in this section: normal, lognormal, gamma distributions. It should be noted that the maxima of all three distributions have a Gumbel distribution.

5.1 Gaussian Parent Distribution

In the first example, the parent distribution is assumed normal. Then, the three SPRTs are applied to the simulation data. Table 1 summarizes the results of the sequential hypothesis testing. Each entry in Table 1 has three numbers. The first number denotes the number of tests accepting the right hypothesis, and the second number denotes the number of tests rejecting the right decision. The last one is the number of cases where SPRT cannot make either decisions based on the given data sets. For example, when F=1.10, Table 1 reports that SPRT-2 accepts the right null hypothesis 63 times out 100 simulations, and rejects the null hypothesis 16 times and no decision is made for the remaining 21 cases.

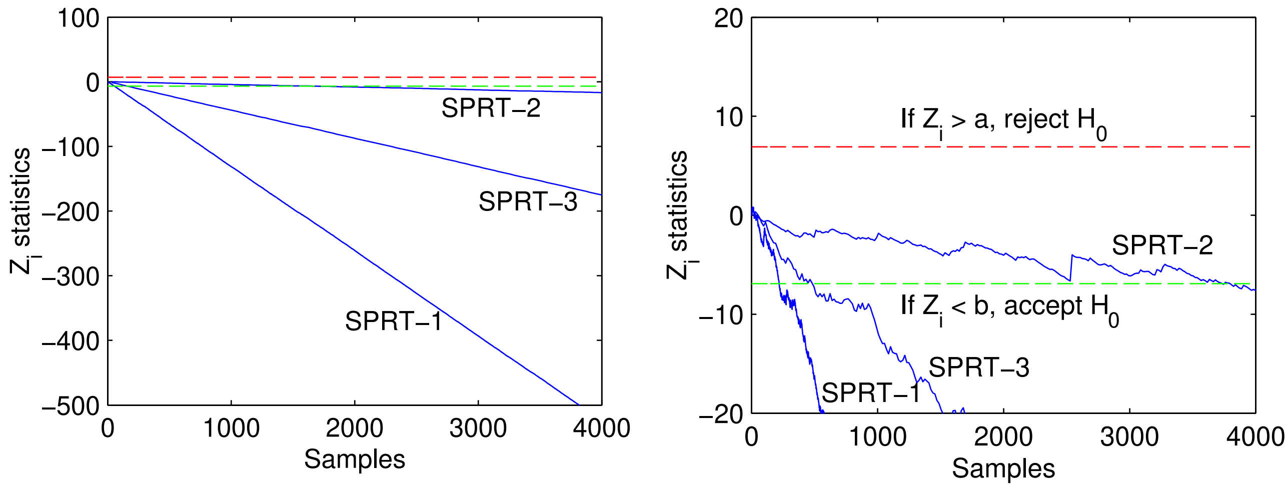
As expected, SPRT-1 and SPRT-2, which are based on the normality assumption of the parent distribution, have accepted the correct hypothesis 100%. However, SPRT-2 with the Gumbel distribution of the maxima has several misclassifications near the lower decision boundary (when F=1.10 and 1.15). These misclassifications are mainly caused by the discrepancy between the stated hypothesis test and the actual hypothesis test conducted for SPRT-2. The original hypothesis test is supposed to be performed on the standard deviation of the “parent distribution”. However, because the parent distribution type is unknown for SPRT-2, the actual hypothesis test is conducted on the standard deviation of the “extreme distribution”. Therefore, cautions should be paid when the classification results in Table 1 are compared for the three different SPRTs.

Figure 1 shows the typical results of the sequential tests. When the Z statistics goes below the lower bound at b= –6.9, the null hypothesis is accepted. On the other hand, when the Z statistics becomes larger than the upper bound a= 6.9, the null hypothesis is rejected and the alternative hypothesis is accepted. In this particular case shown in Figure 1, the accepting the null hypothesis is the correct answer, and all sequential tests make the right classification. It is shown that SPRT-1 generally comes to a decision earlier than the other two sequential tests. Because the extreme values for SPRT-2 and SPRT-3 are sampled at every 16 points of the parent data, it is naturally expected that the statistical inference using SPRT-1 will be faster than those using SPRT-2 and SPRT-3 with extreme statistics.

Table 1: Damage classification results for normal distribution data

Hypo	H_o				H_1			
F	0.90	1.00	1.10	1.15	1.45	1.50	1.60	1.70
SPRT-1	100/0/0*	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
SPRT-2	100/0/0	100/0/0	63/16/21	31/47/22	100/0/0	100/0/0	100/0/0	100/0/0
SPRT-3	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0

*The first number denotes the number of accepting the right hypothesis, and the second number denotes the number of rejecting the right decision. The last one is the number of cases where SPRT cannot make either decisions based on the given data sets. For example, 100/0/0 means that, out of 100 simulations, 100 times are correctly categorized and there were no misclassification or undecided cases.



(a) Shown from 1 to 4000 points (b) Shown from 1 to 600 points

Figure 1: A typical damage classification result for data sets with a normal distribution (Correct decision: accepting H_0)

5.2 Lognormal Parent Distribution

In the second numerical example, the parent distribution is assumed lognormal instead of normal. A random variable, x, has a lognormal distribution if the natural logarithm of x is normal. In this case, the density function of x becomes

$$f(x) = \frac{1}{\sqrt{2\pi}\, s\, x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \nu}{s} \right)^2 \right] \tag{41}$$

where ν and s are the mean and standard deviation of $\ln x$, respectively. For this simulation, $\nu = 1.0$ and $s = 0.5$ are assumed. The associated lognormal density function is displayed in Figure

2. The skewness and kurtosis of this distribution are 1.74 and 8.45, respectively. Note that, for a normal distribution, the values of the skewness and kurtosis are 0.0 and 3.0, respectively.

The analysis results are summarized in Table 2. Although the formulation of SPRT-1 is based on the normality assumption, SPRT-1 surprisingly performs well even for a lognormal distribution. The performance of SPRT-2 is compatible with the previous result for the normal case. Again, the several misclassifications in Table 2 are mainly attributed to the difference between the state and actual hypothesis tests. SPRT-3 completely misses the true hypothesis when F=1.10 and 1.15. It seems that the false assumption of the parent distribution produces accumulated errors in the extreme statistics leading the results of SPRT-3 astray. As shown in Figure 3, SPRT-1 again makes the fastest decision among all three SPRTs, and SPRT-2 takes the longest time to elect a hypothesis.

Table 2: Damage classification results for lognormal distribution data

Hypo	H_o				H_1			
F	0.90	1.00	1.10	1.15	1.45	1.50	1.60	1.70
SPRT-1	100/0/0*	100/0/0	100/0/0	99/1/0	100/0/0	100/0/0	100/0/0	100/0/0
SPRT-2	100/0/0	100/0/0	93/1/6	66/15/19	100/0/0	100/0/0	100/0/0	100/0/0
SPRT-3	100/0/0	11/89/0	0/100/0	0/100/0	100/0/0	100/0/0	100/0/0	100/0/0

*The first number denotes the number of accepting the right hypothesis, and the second number denotes the number of rejecting the right decision. The last one is the number of cases where SPRT cannot make either decisions based on the given data sets. For example, 100/0/0 means that, out of 100 simulations, 100 times are correctly categorized and there were no misclassification or undecided cases.

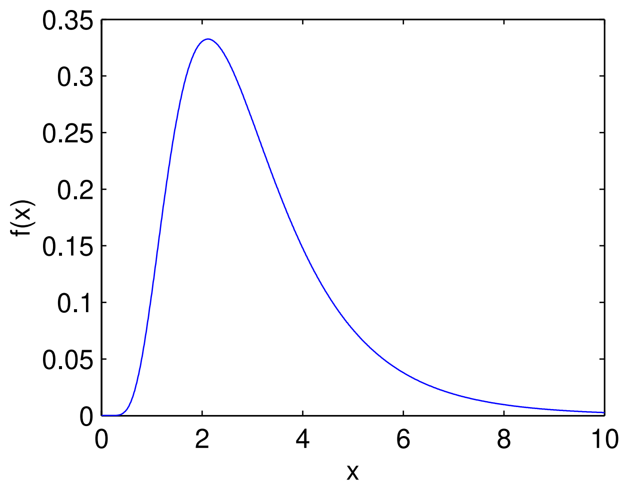


Figure 2: A lognormal density function with $\nu = 1.0$ and $s = 0.5$

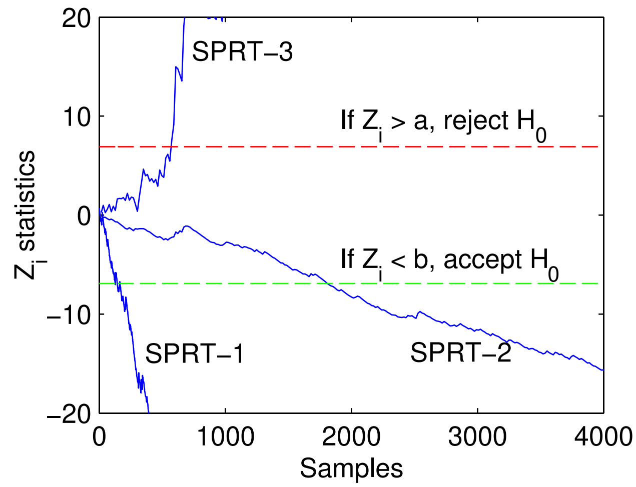


Figure 3: A typical damage classification result for data sets with a lognormal distribution (Correct decision: accepting H_0)

5.3 Gamma Parent Distribution

Finally, the sequential tests are applied to data sets simulated from a gamma parent distribution. A gamma distribution is often used to describe the *k*th occurrence of an event, which constitutes a Poisson process with a mean rate of occurrence, ν (Ang and Tang, 1975). The corresponding density function, therefore, is

$$f(x) = \frac{\nu\,(\nu\,x)^{k-1}}{\Gamma(k)} \exp[-\nu x] \quad x \geq 0 \tag{42}$$

where $\Gamma(k)$ is the gamma function. Note that the exponential and chi-square distributions are special cases of the gamma distribution, and obtained by setting k=1.0 and ν=0.5 in Equation (42), respectively. The gamma distribution is skewed to the right especially for smaller value k. As the degrees of freedom, k, increases the gamma distribution converges to the normal distribution. In this example, the sample data are generated from a gamma distribution with k=3 and ν=0.2. This gamma distribution has the skewness value of 1.15 and kurtosis of 5.00, respectively. The associated density function is plotted in Figure 4.

Hypothesis results similar to the case of the lognormal distribution are obtained in Table 3 and Figure 5. For all three distribution types considered in the examples, SPRT-1 outperforms SPRT-2 and SPRT-3. Humenic and Gross (1990) report a smilar condition that the SPRT is robust in the sense that the SPRT works well even if the underlying distribution is not exactly Gaussian. Again, the results of SPRT-3 seem unreliable especially near the lower decision bound. Therefore, the application of SPRTs to the subsequent experimental data is limited to SPRT-1 and SPRT-2.

Table 3: Damage classification results for Gamma distribution data

Hypo	H_o				H_1			
F	0.90	1.00	1.10	1.15	1.45	1.50	1.60	1.70
SPRT-1	100/0/0*	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
SPRT-2	100/0/0	100/0/0	99/1/0	83/1/16	98/0/2	100/0/0	100/0/0	100/0/0
SPRT-3	100/0/0	93/7/0	0/100/0	0/100/0	100/0/0	100/0/0	100/0/0	100/0/0

*The first number denotes the number of accepting the right hypothesis, and the second number denotes the number of rejecting the right decision. The last one is the number of cases where SPRT cannot make either decisions based on the given data sets. For example, 100/0/0 means that, out of 100 simulations, 100 times are correctly categorized and there were no misclassification or undecided cases.

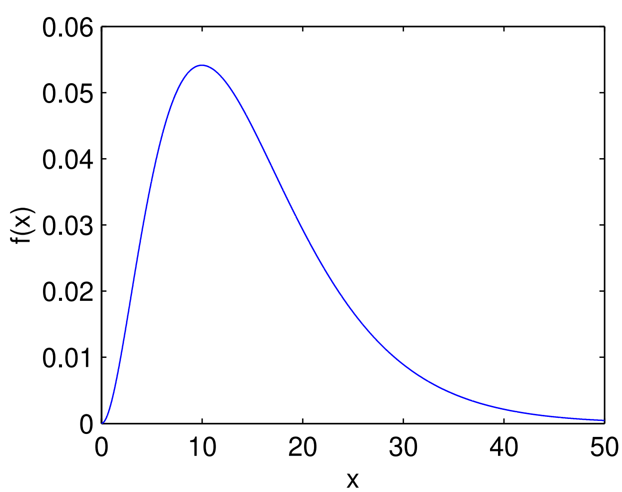


Figure 4: A lognormal density function with k =3.0 and $\nu = 0.2$

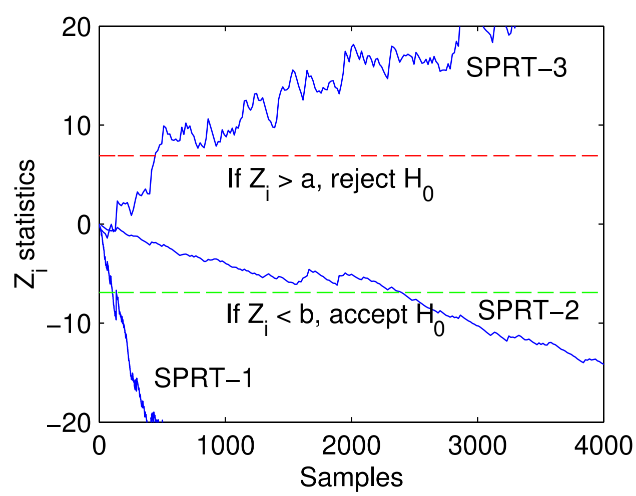


Figure 5: A typical damage classification result for data sets with a gamma distribution (Correct decision: accepting H_0)

6. Experimental Test

6.1 Description of a Test Structure

The structure tested is a three-story frame structure model as shown in Figure 6. The structure is constructed of Unistrut columns and aluminum floor plates. The floors are 1.3cm-thick (0.5in) aluminum plates with two-bolt connections to brackets on the Unistrut. The base is a 3.8cm-thick (1.5in) aluminum plate. Support brackets for the columns are bolted to this plate and hold the Unistrut columns. The details of these joints are shown in Figure 7 and Figure 8. The floor layout from the top of the structure is shown in Figure 9. All bolted connections are tightened to a torque of 0.7 Nm (60 inch-pounds) in the undamaged state. Four Firestone air mount isolators, which allow the structure to move freely in horizontal directions, are bolted to the bottom of the base plate. The isolators are inflated to 140 kPa gauge (20 psig) and then adjusted to allow the structure to sit level with the shaker.

The structure is instrumented with 24 piezoelectric single-axis accelerometers, two per joint as shown in Figure 9. The accelerometers are numbered from the corner A to B, C, and D counterclockwise and from the top floor to the first floor. Accelerometers are mounted on the aluminum blocks that are attached by hot glue to the plate and column. This configuration allows relative motion between the column and the floor to be detected. The nominal sensitivity of each accelerometer is 1 V/g. The shaker is coupled to the structure by a 15cm-long (6in), 9.5mm-diameter (0.375in) stinger connected to a tapped hole at the mid-height of the base plate. The shaker is attached at corner D of the base floor (below floor 1), as shown in Figure 6, so that both translational and torsional motions can be excited. The RMS voltage of the shake was fixed at 2 volts, and random signals were generated from the shaker. A 10-mV/lb-force transducer is also mounted between the stinger and the base plate. This force transducer is used to measure the input to the base of the structure. A commercial data acquisition system controlled from a laptop PC is used to digitize the accelerometer and force transducer analog signals. The data sets that were analyzed in the feature extraction and statistical modeling portion of the study were the acceleration time histories. Each time signal gathered consisted of 8192 points and were sampled at 1600Hz.

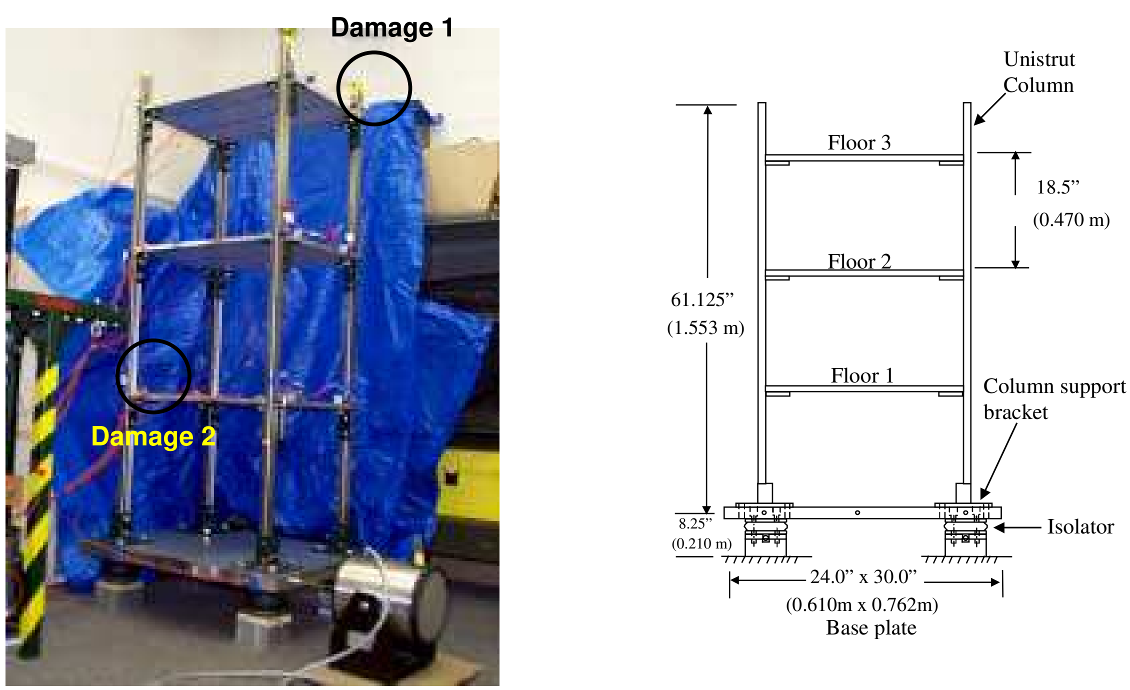


Figure 6: a three-story frame structure with dimension and damage locations

Figure 7: a bolted joint of the test structure

Figure 8: the connection to the base plate

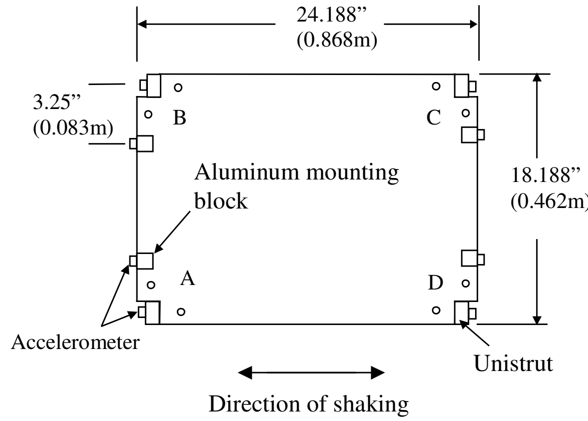


Figure 9: floor layout as viewed from above

Two damage cases are investigated in this experiment. The first damage is introduced to the corner A of the first floor (Damage 1) and the second damage is placed at the corner C of the third floor (Damage 2). These two damage locations are shown in Figure 6. For each damage case, the bolts are loosened until hand tight, allowing relative movement between floor plate and column. After the damage cases, all the bolts were tightened again to the initial torque of 0.7 Nm (60 in-pounds). Five time series are measured from the initial undamaged case, and these time series are used for training, constructing the reference database. Five time series are recorded under each damage case, and additional five time series are obtained after tightening all bolts to the initial torque values. These time series are used for testing the proposed SPRT procedure. That is, a total of 20 time series are used for this experiment.

6.2 Damage Classification Results

Instead of independently analyzing 24 time histories from each accelerometer, the point-by-point difference between time series from the two adjacent accemeroters at a joint is first computed. Then, the resulting 12 time series corresponding to each joint are used for the AR-ARX modeling. The order r in the AR model [see Equation (2)] is set to 25, and the p and q orders for the ARX model [see Equation (5)] are set to 20 and 5, respectively. Satisfactory prediction errors mostly less than 10% error are achieved for all the reference signals indicating that the selected AR-ARX model aprropriately characterizes the underlying dynamic system of each signal readings.

Next, SPRT-1 and SPRT-2 are applied to the damage sensitive feature obtained from the AR-ARX modeling, the residual errors. The type I & II errors are set to 0.001 as before. The formulation of the sequential probability test here is based on the premise that, when a system being monitored undergoes a structural change such as damage, a signal measured under the new structural condition will be significantly different from the signal obtained from the initial undamage case. Therefore, when a time prediction model is constructed using the baseline undamaged time signal, the prediciton error of the newly obtained signal, which is again from the dacasmage case, will departure from that of the baseline signal. Particularly, the prediction error of the new singal is expected to increase. Based on this observation, the sequential hypothesis test is cast as follows for SPRT-1:

$$H_o : \sigma \le \sigma_o, \quad H_1 : \sigma \ge \sigma_1, \quad 0 < \sigma_o < \sigma_1 < \infty \tag{43}$$

In this particular example, σ_0 and σ_1 are set to 0.40 and 0.42, respectively. Note that the establishment of the σ_0 and σ_1 values are based on the observation of actually damage cases. That is, changes of the standard deviation should be first monitored for the corresponding damage cases to select the appropriate σ_0 and σ_1 values. This selection of the σ_0 and σ_1 values categorizes the proposed method as a supervised learning method. In a similar fasion, the sequential hypothesis test for SPRT-2 is cast as follows:

$$H_o : \sigma_M \leq \sigma_o, \quad H_1 : \sigma_M \geq \sigma_1, \quad 0 < \sigma_o < \sigma_1 < \infty \tag{44}$$

σ_0 and σ_1 are set to 0.24 and 0.26 based on the similar obervations as before.

The results of damage classification using SPRT-1 and SPRT-2 are reported in

Table **4** and Table 5. To briefly summarize the results, SPRT-1 and SPRT-2 illustrate comparable performance. Both SPRT-1 and SPRT-2 do not show any false-positive indications of damage for all five undamaged cases. For the first damage case (Damage 1), the damaged joint is located at the corner A on the first floor, and this joint is associated with sensor readings from channels 17 and 18. Using SPRT-1 and SPRT-2, the correct damage location is correctly revealed for all five cases. For the second damage case (Damage 2), where the bolt at the corner C on the third floor is loosened corresponding to channel 5 and 6 readings, SPRT-1 indicates that the adjacent joint at the corner D on the same floor is most likely damaged. SPRT-2 also suggests the existance of damage at the same adjacent joint but correclty identifies the actually damaged joint 3 times out of the five exampled time series.

7. Conclusion

A unique integration of time series analysis, statistical inference, and extreme value theory is provided to address the issue of damage identification. Time series analysis techniques, which solely based on the measured vibration signals, are first employed to extract damage sensitive features for damage classification. While there had been increasing interest in the filed of structural health monitoring, decision as to whether a structure is damaged or not tend to be made on the basis of exceeding some heuristic threshold. In this study, the sequential probability ratio test (SPRT) is employed to provide a more principled statistical tool for this decision-making procedure, excluding unnecessary interpretation of the measured data by users. Finally, the performance and robustness of damage classification is improved by incorporating extreme values statistics of the extracted features into the SPRT. The applicability of the SPRT to structural health monitoring is demonstrated using measured time signals from a three-story frame structure tested at a laboratory environment. The framework of the proposed SPRT is well suited for developing a continuous monitoring system, and can be easily implemented on digital signal processing (DSP) chips automating the damage classification process.

Table 4: Damage classification results using SPRT-1

Test Case	Ch1-Ch2	Ch3-Ch4	Ch5-Ch6	Ch7-Ch8	Ch9-Ch10	Ch11-Ch12	Ch13-Ch14	Ch15-Ch6	Ch17-Ch18	Ch19-Ch20	Ch21-Ch22	Ch23-Ch24
Undamaged	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
Damage 1	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
Damage 2	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0

*The zero '0' denotes that the null hypothesis is accepted indicationg no damage is present at that joint, and the unity '1' denotes that the null hypothesis is rejected and the corresponding joint is damaged. The shaded areas represent the locations of the acutally damaged joints, and the hypothesis results in these shadded ares should ideally correspond to 1. The hypothesis results should be zero otherwise.

** For each undamaged and damage cases, five time series are recorded, and the corresponding damage classification results are shown.

Table 5: Damage classification results using SPRT-2

Test Case	Ch1-Ch2	Ch3-Ch4	Ch5-Ch6	Ch7-Ch8	Ch9-Ch10	Ch11-Ch12	Ch13-Ch14	Ch15-Ch6	Ch17-Ch18	Ch19-Ch20	Ch21-Ch22	Ch23-Ch24
Undamaged	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
Damage 1	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0
Damage 2	0	0	1	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	1	1	0	0	0	0	0	0	0	0
	0	0	1	1	0	0	0	0	0	0	0	0

*The zero '0' denotes that the null hypothesis is accepted indicationg no damage is present at that joint, and the unity '1' denotes that the null hypothesis is rejected and the corresponding joint is damaged. The shaded areas represent the locations of the acutally damaged joints, and the hypothesis results in these shadded ares should ideally correspond to 1. The hypothesis results should be zero otherwise.

** For each undamaged and damage cases, five time series are recorded, and the corresponding damage classification results are shown.

References

[1] Ang, A.H-S. and Tang, W.H. (1975) *Probability Concepts in Engineering Planning and Design*, John Wiley & Sons, Inc., New York, NY.

[2] Box, G. E., Jenkins, G. M., and Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control*, Third Edition, Prentice-Hall, Inc., New Jersey.

[3] Castillo, E. (1987) *Extreme Value Theory in Engineering*, Academic Press, Inc., San Diego, CA.

[4] Doebling, S. W., C. R. Farrar, M. B. Prime and D. W. Shevitz, (1998) "A Review of Damage Identification Methods that Examine Changes in Dynamic Properties," *Shock and Vibration Digest* 30 (2), pp. 91-105.

[5] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) *Modeling Extremal Events*, Springer-Verlag, New York, NY,.

[6] Fisher R.A. & Tippett L.H.C. (1928) *Limiting forms of the frequency distributions of the largest or smallest members of a sample*, Proceedings of the Cambridge Philosophical Society 24 pp.180-190.

[7] Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York, NY.

[8] Ghosh, B.K. (1970) *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Menlo Park, CA.

[9] Gross, K.C. and Humenik, K.E. (1991) "Sequential Probability Ratio Tests for Nuclear Plant Component Surveillance," *Nuclear Technology*, Vol. 93, pp. 131-137.

[10] Gumbel, E.J. (1958) *Statistics of Extremes*, Columbia University Press, New York.

[11] Humenik, K.E. and Gross, K.C. (1990) "Sequential Probability Ratio Tests for Reactor Signal Validation and Sensor Surveillance Applications," *Nuclear Science and Engineering*, Vol. 105, pp. 383-390.

[12] Kotz, S. and Nadarajah, S. (200) *Extreme Value Distributions. Theory and Applications,* Imperial College Press, London, UK.

[13] Ljung, L. (1987) *System Identifications: Theory for the User*, Prentice Hall, New Jersey.

[14] Sohn, H. and Farrar, C.R. (2001) "Damage Diagnosis Using Time Series Analysis of Vibration Signals," *Journal of Smart Materials and Structures*, Vol. 10, pp. 446-451.

[15] Sohn, H., Farrar, C.R., Hunter, N.F., and Worden, K. (2001) "Structural Health Monitoring Using Statistical Pattern Recognition Techniques," *ASME Journal of Dynamic Systems,*

Measurement and Control: Special Issue on Identification of Mechanical Systems, Vol. 123, No. 4, pp. 706-711.

[16] Wald, A. (1947) *Sequential Analysis*, John Wiley and Sons, New York, NY.