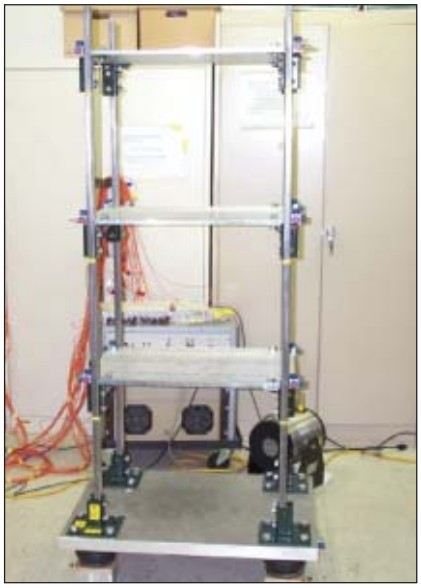
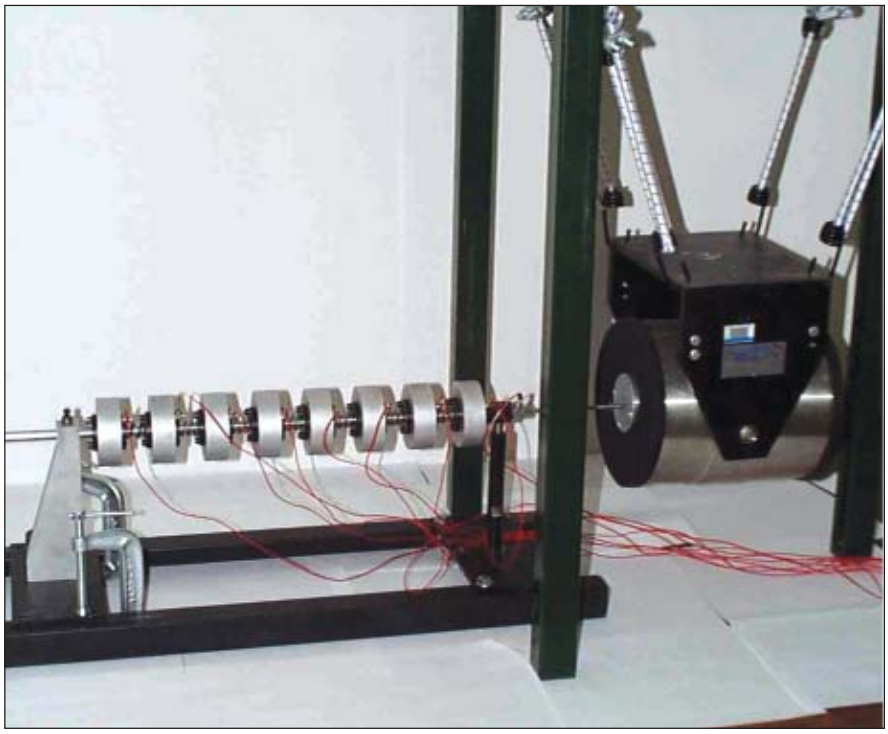
LA-13903-MS
Approved for public release;
distribution is unlimited.



Extreme Value Statistics for Damage Detection in Mechanical Structures



Los Alamos
NATIONAL LABORATORY

Funding for this study was primarily provided by Los Alamos National Laboratory's Damage Prognosis LDRD program.

Edited by Ed Lorusso, Group IM-1

Cover: The three-story frame structure, which is shown in the upper photograph, demonstrates the applicability of the extreme value statistics to structural health monitoring. Damage is introduced to the structure by loosening bolts in a joint. The eight-degree of freedom spring-mass system, shown in the lower photograph, is also used for a similar damage identification study. Nonlinear damage is simulated by placing a bumper between two masses. Acceleration time series measured from the two mechanical systems are used for damage diagnosis.

LA-13903-MS
Issued: August 2002

Extreme Value Statistics for Damage Detection in Mechanical Structures

Keith Worden*
David W. Allen
Hoon Sohn
Daniel W. Stinemates
Charles R. Farrar

*University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

Los Alamos
NATIONAL LABORATORY

CONTENTS

EXTREME VALUE STATISTICS FOR DAMAGE DETECTION IN MECHANICAL STRUCTURES

Keith Worden*, David W. Allen, Hoon Sohn, Daniel W. Stinemates and Charles R. Farrar

Abstract: The first and most important objective of any damage identification algorithm is to ascertain with confidence if damage is present. Considering most real world applications of damage detection, this detection must be accomplished in an unsupervised learning mode. Here, the term "unsupervised learning" implies that data from a damaged state are not used to aid in the damage detection process. Methods have been proposed for unsupervised damage detection based on ideas of statistical process control, which monitors such parameters as the sample mean and standard deviation. Statistical process control is currently based on the assumption that the underlying distribution of data is Gaussian. However, the assumption of normality imposes potentially misleading behavior on the extreme values of the data, namely, those points in the tails of the distribution. As the problem of damage identification specifically focuses attention on these tails, the assumption of normality is likely to lead any analyses astray. 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. This report shows a number of approaches to extreme value analysis contrasted with the standard approach where it is assumed that the damage-sensitive features are normally distributed. Both approaches are applied to numerical and experimental data to illustrate the difference between the two methods.

---

[1]*On study leave from Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK. E-mail: k.worden@sheffield.ac.uk

1. Introduction

This report is concerned with *Statistical Process Control* (SPC) in an unsupervised learning mode, which is the first level of damage identification. When applied to structural health monitoring (SHM), unsupervised learning means that data from the damaged condition are not available to aid in the damage detection process.

The objective of unsupervised SPC is to establish a model of the system's normal condition and thereafter to signal statistically significant departures from this condition. A significant change in the system is an indication of damage. This objective can be accomplished in several ways. Some methods include the use of the control chart (Fugate et al., 2001) based on the mean and standard deviation of a feature, or the sequential probability ratio test (Gosh, 1970) that monitors a probability ratio between undamaged and damaged cases. The main limitation of all of these methods is that they make unwarranted assumptions about the nature of the feature distribution tails. These assumptions are potentially hazardous, because extreme events that reside in the tails of the normal condition are likely to be mistaken for damage. More specifically, SPC relies on a model based on central statistics (the sample mean and standard deviation) and the analysis is largely insensitive to the structure of the tails. Another way of regarding this problem is as a question of setting appropriate control limits or thresholds.

The major problems with modeling the undamaged condition of a system are that the functional form of the distribution is often unknown and there are an infinite number of candidate distributions that may be appropriate for the prediction applications. Furthermore, in some cases, only extreme values of events may be recorded due to sensor or storage limitations. For example, seismic stations are primarily interested in recording *strong ground motion*, motion beyond certain magnitude with sufficient strength to affect people and their environment (Kramer, 1996). The measurements of peak strains or accelerations are enough to monitor the base isolation systems of buildings and bridges (Takahira and Mita, 2002). In the current procedures, a knowledgeable operator makes a choice among the infinite distributions and then estimates parameters based on training data. This process is largely subjective. Any choice of the distribution and parameters will constrain the behavior of the tails to that prescribed distribution.

In fact, there is a large body of statistical theory that is explicitly concerned with modeling the tails of distributions, and these statistical procedures can be applied to the problem of SPC. The relevant field is referred to as *extreme value statistics* (EVS), 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; Reiss and Thomas, 2001). Castillo (1988) is notable in its concern with engineering problems in fields like meteorology, hydrology, ocean engineering, pollution studies, strength of materials, etc. Roberts (1998 and 2000) introduced the ideas of EVS into novelty detection and applied them in the biosignal processing context. Although EVS has been widely used, there has been little application of these techniques to SPC and damage identification. This report illustrates the use of EVS in their own right and not as another way of looking at Gaussian distributions in an effort to avoid such assumptions.

The layout of this report is as follows: Section 2 provides an introduction to EVS, followed by section 3, which describes parameter estimation techniques for fitting EVS distributions to extreme value data using simulated examples. Section 4 shows a comparison between SPC thresholds calculated using Gaussian assumptions with those calculated using EVS for three different distributions. Sections 5 and 6 explore the integration of EVS into damage detection for two physical structures, a three-story frame structure and an eight degree-of-freedom (DOF) spring-mass system. Sections 7 and 8 finish the report with the summary and conclusions of the work. Two appendices are included to present various computer programs used to estimate parameters for EVS distributions and to give a more in depth view of statistical pattern recognition techniques used in the analysis.

2. Extreme Value Statistics

The Gaussian distribution occupies its central place in statistics for a number of reasons; not least is the central limit theorem (Benjamin and Cornell, 1970). The central limit theorem states that if $\{X_1, X_2, ..., X_n\}$ is a set of random variables with an arbitrary distribution, the sum variable $X_\Sigma = X_1 + X_2 + \ldots + X_n$ will have a Gaussian distribution as $n \to \infty$. Although this theory is arguably the most important limiting theorem in statistics, it is not the only one. If the problem at

hand is concerned with the tails of distributions, there is another theorem that is more appropriate.

Suppose that one is given a vector of samples $\{X_1, X_2, \ldots, X_n\}$ from an arbitrary *parent distribution*. The most relevant statistic for studying the tails of the parent distribution is the maximum operator, $\max(\{X_1, X_2, \ldots, X_n\})$, which selects the 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 EVS 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 (Fisher, 1928). The rest of this section will be concerned with elaborating on this fact.

If the values of the sequence $X_1, X_2, \ldots, X_n$ are arranged in ascending order $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, the r^{th} element of this sequence $X_{r:n}$ is called the r^{th} *order statistic*. In order statistics it is customary to include the total sample size, *n*, in the notation. The basic question that now arises is what are the distributions of the order statistics, in particular, the minimum, $X_{1:n}$, and the maximum, $X_{n:n}$.

Following Castillo (1988), let $m_n(x)$ be the number of samples for which $X_{r:n} \le x$. Each time one chooses a value $X_{r:n}$ 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 $X_{r:n} \le x$, and the complementary probability, $(1 - F(x))$, that $X_{r:n} > x$. The CDF of $m_n(x)$ is therefore a binomial distribution with $F^k(k)$ denoting the probability of success,

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

Now, because the event $\{X_{r:n} \le x\}$ is basically the same as the event $\{m_n(x) \ge r\}$ so that, $\text{Prob}[X_{r:n} \le x] = \text{Prob}[m_n(x) \ge r] = 1 - \text{Prob}[m_n(x) < r]$, and $F_{X_{r:n}}(x) = 1 - F_{m_n(x)}(r-1)$, or

$$F_{X_{r:n}}(x) = \text{Prob}[X_{r:n} \le x] = \sum_{k=r}^{n} \binom{n}{k} F^k(x)[1-F(x)]^{n-k} \tag{2}$$

If one is concerned with the maximum of the sample, the relevant order statistic is $X_{n:n}$ and the relevant distribution is,

$$F_{X_{n:n}}(x) = F^n(x) \tag{3}$$

If one is concerned with the minimum of the sample, the relevant order statistic is $X_{1:n}$ and the appropriate distribution is,

$$F_{X_{1:n}}(x) = 1-[1-F(x)]^n \tag{4}$$

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

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 1 & \text{If } F(x) = 1 \\ 0 & \text{If } F(x) < 1 \end{cases} \tag{5}$$

This distribution does not 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,

$$\lim_{n\to\infty} F^n(a_n + b_n x) = H(x) \tag{6}$$

where *H(x)* is a non-degenerate limit function. In fact, it is required that *H(x)* be continuous. The situation for minima is similar: A sequence of normalizations is required such that,

$$\lim_{n\to\infty} 1-[1-F(c_n + d_n x)]^n = L(x) \tag{7}$$

and *L(x)* is a non-degenerate continuous limit function.

The fundamental theorem of EVS states (Fisher and Tippett, 1928):

Theorem: Feasible limit distributions for maxima

The only three types of non-degenerate distributions $H(x)$ *satisfying Equation (6) are*

$$\text{FRECHET}: \qquad H_{1,\beta}(x) = \begin{cases} \exp\left[-\left(\dfrac{\delta}{x-\lambda}\right)^{\beta}\right] & \text{if} \quad x \geq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$\text{WEIBULL}: \qquad H_{2,\beta}(x) = \begin{cases} 1 & \text{if} \quad x \geq \lambda \\ \exp\left[-\left(\dfrac{\lambda - x}{\delta}\right)^{\beta}\right] & \text{otherwise} \end{cases} \tag{9}$$

$$\text{GUMBEL}: \qquad H_{3,0}(x) = \exp\left[-\exp\left(-\frac{x-\lambda}{\delta}\right)\right] \qquad -\infty < x < \infty \quad \text{and} \quad \delta > 0 \tag{10}$$

Or, in the appropriate form for minima,

Theorem: Feasible limit distributions for minima

The only three types of non-degenerate distributions $L(x)$ *satisfying Equation (7) are*

$$\text{FRECHET}: \qquad L_{1,\beta}(x) = \begin{cases} 1-\exp\left[-\left(\dfrac{\delta}{\lambda-x}\right)^{\beta}\right] & \text{if } x \leq \lambda \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

$$\text{WEIBU LL}: \qquad L_{2,\beta}(x) = \begin{cases} 0 & x \leq \lambda \\ 1 - \exp\left[-\left(\dfrac{x - \lambda}{\delta}\right)^{\beta}\right] & x > \lambda \end{cases} \tag{12}$$

$$\text{GUMBEL}: \qquad L_{3,0}(x) = 1-\exp\left[-\exp\left(\frac{x-\lambda}{\delta}\right)\right] \qquad -\infty < x < \infty \text{ and } \delta > 0 \tag{13}$$

where λ, α, and β are the model parameters that are estimated from the data.

Now, given samples of maximum data from a number of *n*-point populations, 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 parametric model is obtained, it can be used

to compute effective thresholds for SPC based on the true statistics of the data as opposed to a blanket assumption of a Gaussian distribution.

The next section illustrates the types of analysis that are possible on a number of synthetic data sets.

3. Selection of Extreme Value Distributions and Parameter Estimation Techniques

Based on the EVS theories presented in the previous section, this section presents techniques for selecting a proper extreme value distribution for given data sets and estimating the parameters associated with the chosen distribution. This section refers to a number of MATLAB (MathWorks, 1998) programs that are included in Appendix A. These programs are translations of the original BASIC routines in Castillo (1988).

3.1. Generation of Data

Data from various extreme value distributions can be simulated using a standard approach. If $F(x)$ is the CDF of the distribution of interest and y is a uniform deviate, random variable with a unit probability density function over the interval [0,1], then $F^{-1}(y)$ has the required distribution. This property is coded in the routine *simulationPD* (Appendix A). The routine takes advantage of a useful property of order statistics conditional distributions. It can be shown for two consecutive order statistics that (Castillo, 1988)

$$F_{X_{i:n}|X_{i+1:n}}(x_1|x_2) = \left(\frac{F(x_1)}{F(x_2)}\right)^i \tag{14}$$

and this fact can be used to generate the samples *in decreasing order*.

3.2. Probability Paper

Once a population of samples is obtained, it is a simple matter to plot the *empirical CDF*. The data are first placed in increasing order. The data are now the order statistic $X_{r:n}$. One associates with each order statistic a *plotting position* or assignment of probability. A naïve approach

assigns the value r/n to $X_{r:n}$. However, this assignment does not behave well under certain nonlinear transformations of the data that will be described later. A more robust choice is to assign the value $(r-0.5)/n$ to $X_{r:n}$. There are numerous different formulae for plotting positions; several are represented here and are coded in the function *draw_samplesPD* (Appendix A).

The purpose of the function *draw_samplesPD* is to show the empirical CDF in a number of different coordinate systems, each appropriate for a given extreme value distribution. To illustrate, consider the Gumbel CDF for maxima in Equation (10). Let $y = H_{3,0}(x)$ be the formula for the CDF. If one makes the nonlinear transformation $g(x)$ and $h(y)$ of the x and y coordinates,

$$\begin{aligned} \xi &= g(x) = x \\ \eta &= h(y) = -\ln[-\ln(y)] \end{aligned} \tag{15}$$

where $\ln(y)$ represents the natural logarithm, then the new coordinates ζ and η satisfy,

$$\eta = \frac{\xi - \lambda}{\delta} \tag{16}$$

The Gumbel CDF will appear as a straight line in this coordinate system. Such a plot will be referred to as "on Gumbel probability paper" or "in Gumbel coordinates". Figure 1 shows the empirical CDF for 1000 data points generated from the Gumbel maximum distribution with $\lambda = 50$ and $\delta = 10$. In Gumbel maximum coordinates one obtains a straight line as required. The final point in Figure 1 and some subsequent figures seem to stray greatly from the rest of the data. This aberration may be an artifact of slow convergence to the extreme value distribution for the most extreme points.

The function *draw_samplesPD* also allows the user to superimpose a linear regression line together with the 95% confidence interval on the empirical CDF. If a straight line adequately models the data from an unknown distribution in Gumbel maximum coordinates, then this modeling provides support for the hypothesis that the unknown distribution is Gumbel maximum. A similar transformation carries data from the Gumbel minimum distribution into a coordinate system where the empirical CDF is a straight line.

Figure 2 shows data from a Weibull maximum distribution ($\lambda = 50,\ \delta = 100,\ \beta = 2$) plotted in Gumbel maximum coordinates. As one might expect, the result is not a straight line. In fact, there is definite curvature (concavity). Figure 3 shows data from a Frechet maximum distribution ($\lambda = 0,\ \delta = 30,\ \beta = 5$) plotted in Gumbel maximum coordinates. In this case, the curvature is not as marked as Figure 2 but is clearly in the opposite sense (convexity).

This curvature is one of the tests for the limiting distribution for maxima. First, the empirical CDF is plotted in Gumbel coordinates for maxima. The user then makes an assessment as to whether the curvature deviates significantly from unity, and the limit distribution is assigned accordingly. The routine *selectionEV* (Appendix A) computes the curvatures in this fashion. However, its use is somewhat limited without appropriate confidence intervals for the Gumbel case.

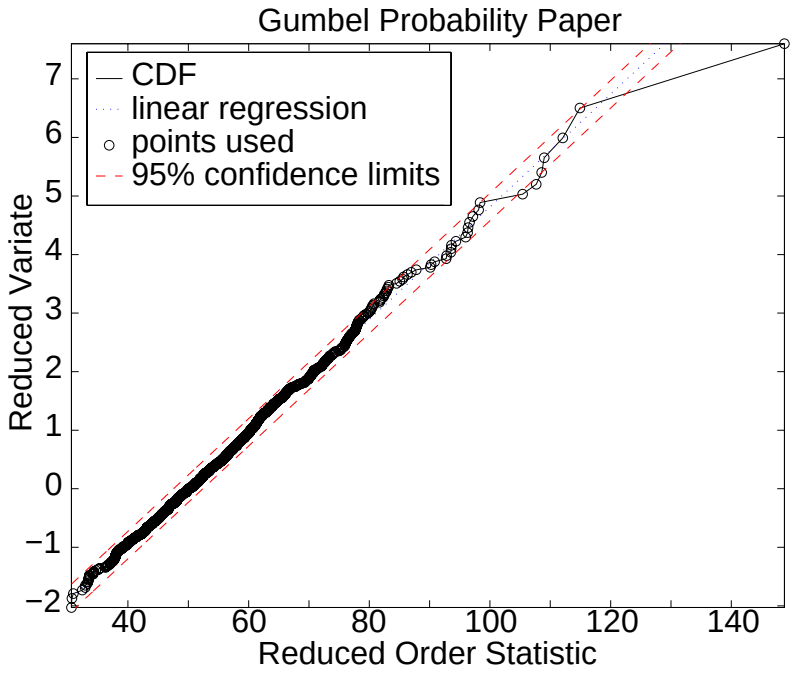


Figure 1: The empirical CDF for Gumbel maximum distributed data in Gumbel maximum coordinates.

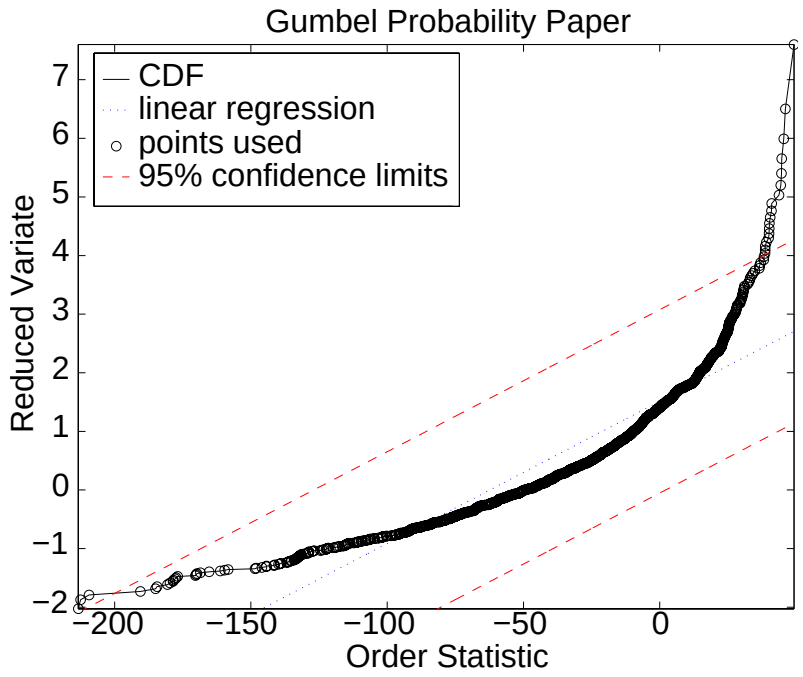


Figure 2: The empirical CDF for Weibull maximum distributed data in Gumbel maximum coordinates.

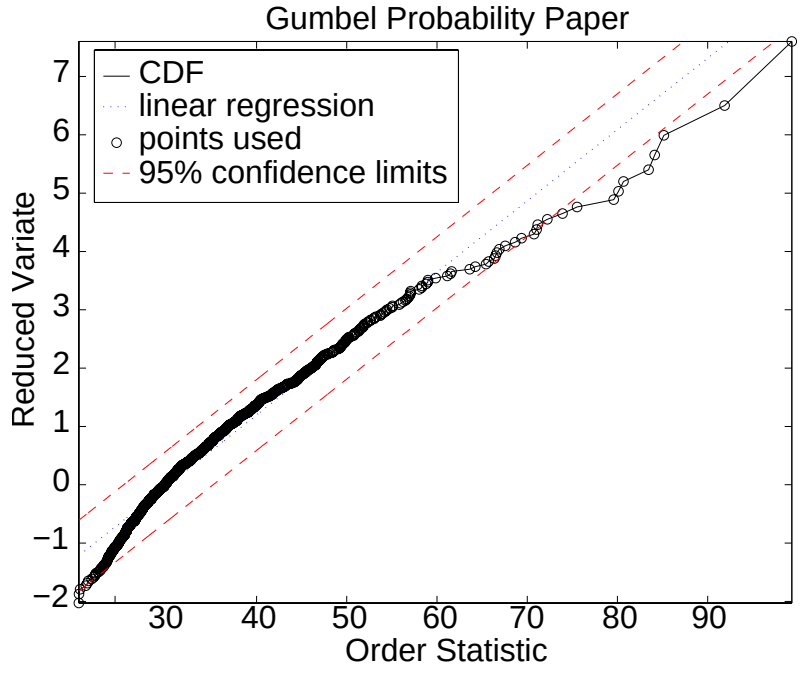


Figure 3: The empirical CDF for Frechet maximum distributed data in Gumbel maximum coordinates.

If the data are known to come from a Weibull distribution for maxima as in Equation (9) with the empirical CDF $y = H_{2,\beta}(x)$, then the transformation,

$$\begin{aligned} \xi &= g(x) = -\ln(\lambda - x) \\ \eta &= h(y) = -\ln[-\ln(y)] \end{aligned} \tag{17}$$

carries the empirical CDF into the straight line,

$$\eta = \beta(\xi - \ln\delta) \tag{18}$$

The difference in this situation is that the transformation requires an *a priori* estimate of the location parameter λ.

Figure 4 shows the data from the Weibull distribution for maxima shown in Figure 2, but this time plotted in Weibull coordinates with the correct location parameter used in the transformation (λ = 50). The required straight line is obtained. If an incorrect λ had been used, the plot would not have been linear. It would be possible to design an optimization procedure to assign the λ value that makes the plot in Weibull coordinates maximally linear. However, it proved convenient here to simply vary λ by trial-and-error until a satisfactory line was obtained.

If the data are known to come from a Frechet distribution for maxima as in Equation (8) with the empirical CDF $y = H_{1,\beta}(x)$, then the transformation,

$$\begin{aligned} \xi &= g(x) = \ln(x - \lambda) \\ \eta &= h(y) = -\ln[-\ln(y)] \end{aligned} \tag{19}$$

carries the empirical CDF into the straight line,

$$\eta = \beta(\xi - \ln\delta) \tag{20}$$

As before, an *a priori* estimate of λ is required. Figure 5 shows the data from the Frechet distribution for maxima shown in Figure 3 but this time plotted in Frechet coordinates with the correct location parameter used in the transformation $(\lambda = 0)$. Once again, the required straight line is obtained.

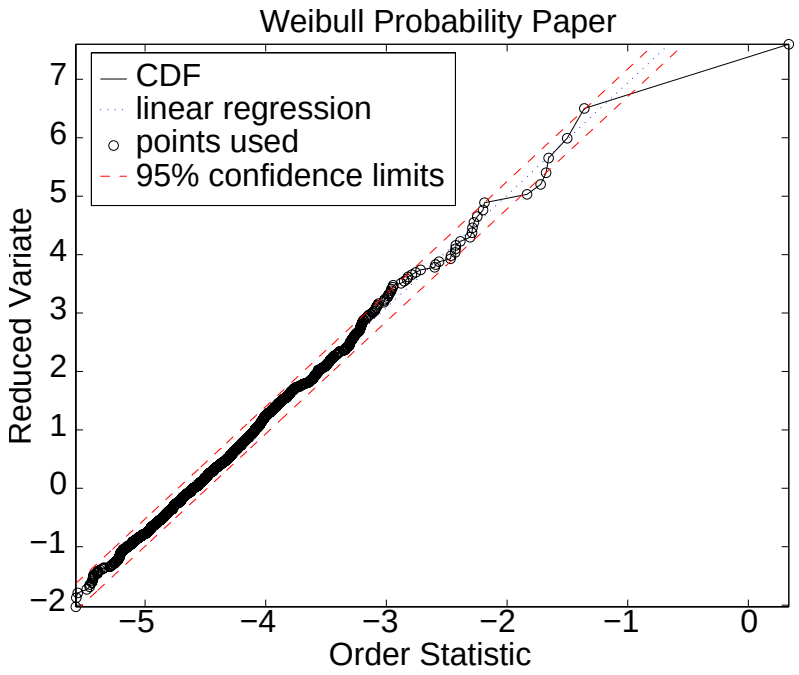


Figure 4: The empirical CDF for Weibull maximum distributed data in Weibull maximum coordinates (with the appropriate λ).

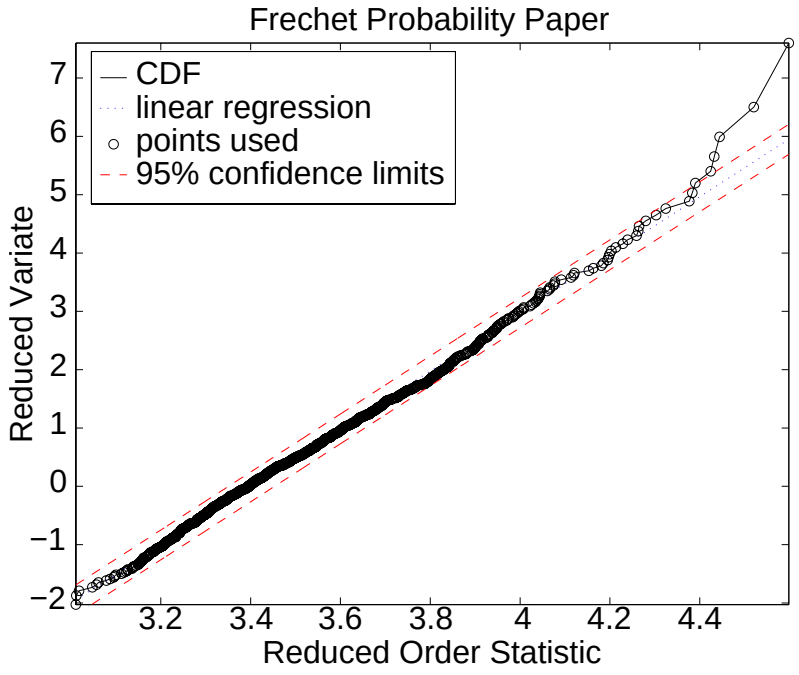


Figure 5: The empirical CDF for Frechet maximum distributed data in Frechet maximum coordinates (with the appropriate λ).

3.3. Parameter Estimation

Having established the appropriate limit distribution, the next stage in the analysis is to estimate parameters of the chosen distribution. The function *estimate* (Appendix A) can be used to fit the best parameters for both the least-squares and maximum likelihood cases. Because the function *estimate* only fits parameters to one canonical model form, the Gumbel distribution for minima, it should be noted that the following pre-processing is required before the actual curve fitting.

First, if the data are distributed as maxima, the transformations $x \to -x$ and $\lambda \to -\lambda$ carry each maximum CDF into the corresponding minimum CDF at least as far as optimization is concerned.

Next suppose the data have the Weibull distribution for minima. Then, the transformation $Y = \ln(X - \lambda)$ carries the Weibull distribution X into the Gumbel distribution Y with the following relations between the parameters,

$$\lambda_G = \ln(\delta_W) \text{ and } \delta_G = \frac{1}{\beta_W} \tag{21}$$

where the subscripts *G* and *W* denote Gumbel and Weibull distributions, respectively. As in the plotting problem, this transformation requires an *a priori* estimate of λ, but this transformation can be obtained by optimizing the linearity of the empirical CDF plot in Weibull coordinates.

If the data have the Frechet distribution for minima, the transformation $Y = -\ln(\lambda - X)$ carries the Frechet distribution X into the Gumbel distribution Y, with the following relations between the parameters,

$$\lambda_G = -\ln(\delta_F) \text{ and } \delta_G = \frac{1}{\beta_F} \tag{22}$$

where the subscript *F* denotes a Frechet distribution. This relation means that the parameter estimation problem is reduced to fitting the data to the limit distribution of the form in Equation (13).

The optimization estimates the parameters λ and δ, by minimizing some error criterion. The most straightforward error criterion is the weighted least-squares method, which seeks to minimize the following objective function *G*,

$$G = \sum_{i=1}^{n} w_i [p_i - L_{3,0}(x_i; \lambda, \delta)]^2 \tag{23}$$

where the training data are the points on the empirical CDF $\{(x_i, p_i), i = 1, \ldots, n\}$, $p_i's$ are an appropriate choice of plotting positions, and w_i's are a set of weights. Although there are various possibilities, Castillo (1988) recommends,

$$w_i = \frac{1}{p_i} \tag{24}$$

With these weights, the method is referred to as *least-squares probability relative error* (LSPRE). The other approach to optimization is Maximum Likelihood (ML). The reader is referred to Castillo (1988) for details, although the method is available as part of the function *estimate*.

For the Gumbel distribution, a very simple but often inaccurate approach called *the method of moments* is available. It is possible to show that the mean, $\bar{x}$, and the variance, σ^2, of the Gumbel distribution maxima and minima are related to λ and δ,

$$\bar{x} = \lambda - \gamma\delta \text{ and } \sigma^2 = \frac{\pi^2\delta^2}{6} \tag{25}$$

where γ is Euler's constant (≈ 0.57772). From the above mean and sample variance, the moment estimates of the parameters can be calculated as:

$$\delta = \frac{\sigma\sqrt{6}}{\pi} \text{ and } \lambda = \bar{x} + \gamma\delta \tag{26}$$

Figure 6 shows the LSPRE curve-fit to the Gumbel maximum data shown in Figure 1. The estimated parameters $\lambda = 50.12$ and $\delta = 10.16$ compare favorably with the exact values of $\lambda = 50$ and $\delta = 10$, respectively. The ML estimator gave estimates of 50.03 and 10.38, respectively.

Figure 7 shows the LSPRE curve-fit to the Weibull data in Figure 4. The correct value of $\lambda = 50$ was assumed. The parameter estimates were $\delta = 99.42$ and $\beta = 1.97$, compared to the true values of 100 and 2. The corresponding ML estimates were 99.85 and 1.93, respectively.

Figure 8 shows the LSPRE curve-fit to the Frechet data in Figure 5. The correct value of $\lambda = 0$ was assumed. The parameter estimates were $\delta = 30.53$ and $\beta = 5.38$, compared to the true values of 30 and 5. The corresponding ML estimates were 29.92 and 4.82, respectively. A summary of the estimation results can be found in Table 1.

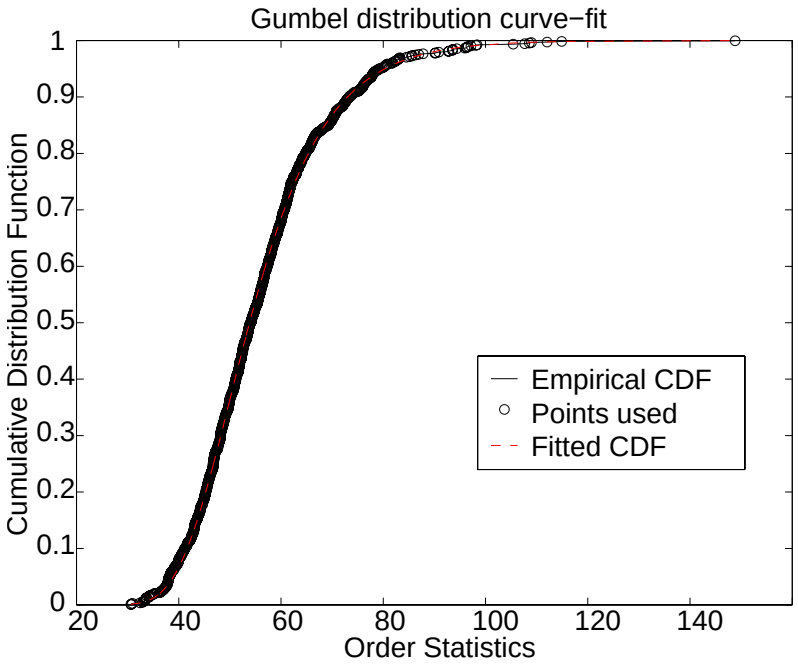


Figure 6: LSPRE curve-fit to Gumbel maxima data in Figure 4.

Table 1: Summary of Parameter Estimation Exercise for Gumbel, Weibull and Frechet Distributions.

	λ			δ			β		
	Exact	LSPRE*	ML**	Exact	LSPRE	ML	Exact	LSPRE	ML
Gumbel	10.00	10.16	10.38	10.00	10.16	10.38	N/A	N/A	N/A
Weibull	50.00	Exact value used	Exact value used	100.0	99.42	99.85	2.00	1.99	1.93
Frechet	0.00			30.00	30.35	29.95	5.00	5.38	4.82

*Least-Squares Probability Relative Error (LSPRE) method: This method requires an estimate of the λ parameter for Weibull and Frechet distributions. In this case the known exact value is used.

**Maximum Likelihood (ML) method. This method requires an estimate of the λ parameter for Weibull and Frechet distributions. In this case the known exact value is used.

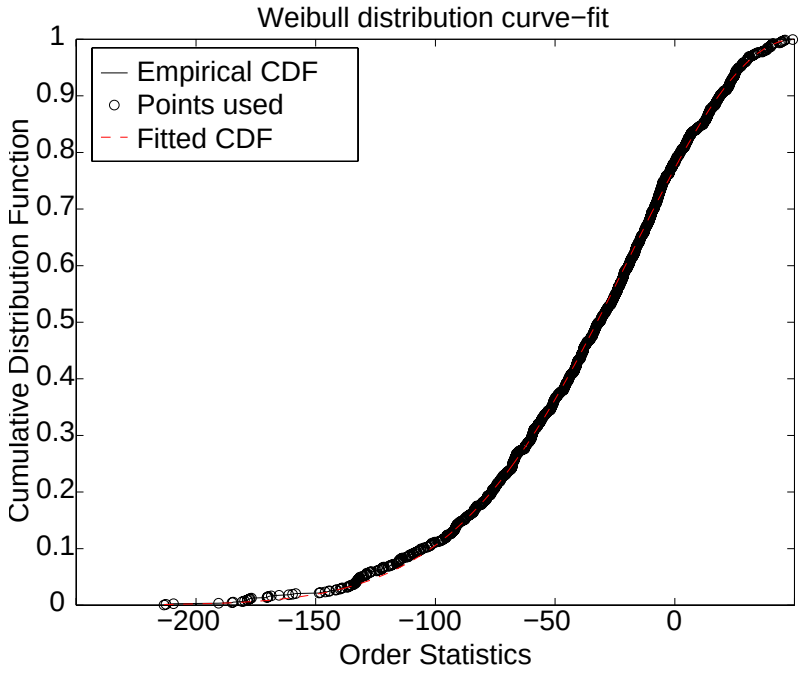


Figure 7: LSPRE curve-fit to Weibull maxima data in Figure 4.

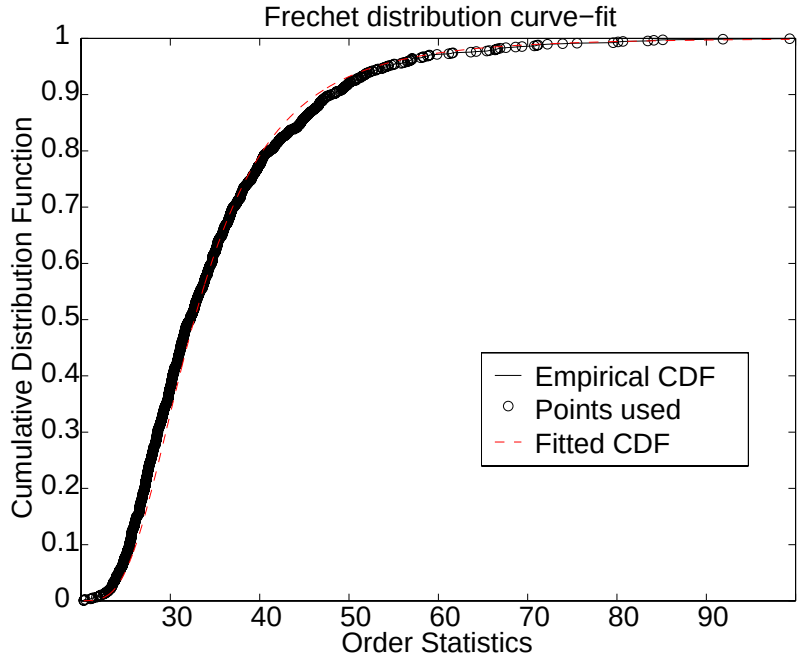


Figure 8: LSPRE curve-fit to Frechet maxima data in Figure 5.

4. Numerical Examples

Simulated random signals from three different distributions are used to demonstrate the usefulness of EVS in accurately modeling the tails without any assumptions of the parent distribution. In each example, the 99% confidence interval for SPC analysis is computed based on the following three methods:

1. The assumed true parent distribution
2. A best-fit normal distribution where the sample mean and standard deviation are estimated from the random data generated from the assumed parent distribution.
3. An extreme value distribution, the parameters of which are estimated from either the top or bottom fraction of the simulated random data.

Hereafter, the confidence interval estimation methods based on the above three distributions are referred to as Method 1, Method 2, and Method 3, respectively.

Setting a confidence interval on the parent distribution using either Method 1 or 2 is fairly trivial. The lower and upper limits of the confidence interval are constructed based on the probability of a type I error that one intends to tolerate. When the probability of the type I error is specified to be α $(0 \le \alpha \le 1)$, $100 \times (1-\alpha)\%$ of data from a normal condition should be encompassed within the confidence interval. In other words, $100 \times \alpha\,\%$ of data will be outliers. Accordingly, the lower and upper limits of the confidence interval can be set at $F^{-1}(\alpha/2)$ and $F^{-1}(1-\alpha/2)$, respectively. Here $F^{-1}(x)$ is the inverse CDF of the known parent distribution. These threshold limits correspond to a $100 \times (1-\alpha)\%$ confidence interval. For instance, when the type I error is set 0.1, this type I error corresponds to a 90% confidence interval. In addition, the lower and upper limits are set so that 5 % and 95% of the normal data are below each of these threshold values (90% are within the two bounds). Because the true CDF of the parent distribution is unknown in Method 2, the CDF of the best-fit normal distribution is used instead of the true CDF to compute the lower and upper limits.

When Method 3 is applied to compute the threshold values, cautions much be taken in selecting the probability of type I error for the distribution of the extreme values. For instance, let's assume that 10,000 sample points are generated from a parent distribution and the type I error is set to 1% $(\alpha = 0.01)$. Then, by the definition of the type I error, it is expected that there will be about 1 % or 100 outliers out of 10,000 samples. If either the maximum or minimum value is extracted from a moving window of size 10 $(n = 10)$, 1,000 extreme values will be obtained from the original 10,000 samples. In other words, 10% of the original data will be used to fit the extreme value distribution. In the next step, the type I error of the extreme value distribution should be set so that this type I error produces the same number of outliers as the type I error of the parent distribution does. To do this, the type I error of the extreme value distribution should be set to 10% (or $\alpha \times n$) in order to produce 100 outliers out of 1,000 extreme values (or out of 10,000 original samples). That is, the lower and upper limits of the confidence interval can be set at $F^{-1}(n \times \alpha/2)$ and $F^{-1}(1 - n \times \alpha/2)$, respectively.

For the computation of the lower limit at $F^{-1}(n \times \alpha/2)$, the Gumbel distribution for minima is used to approximate the CDF function. For the given cumulative probability value at $n \times \alpha/2$, Equation (13) becomes,

$$n \times \frac{\alpha}{2} = 1 - \exp\left[-\exp\left(\frac{x - \lambda}{\delta} \right) \right] \tag{27}$$

By solving Equation (27) with respect to x, the lower limit x_m at $F^{-1}(n \times \alpha/2)$ is obtained,

$$\text{Lower limit: } x_m = \lambda + \delta \ln\left(-\ln\left(1 - n \times \frac{\alpha}{2} \right) \right) \tag{28}$$

The upper limit x_M at $F^{-1}(1 - n \times \alpha/2)$ is obtained from Equation (10) in a similar fashion,

$$\text{Upper limit: } x_M = \lambda - \delta \ln\left(-\ln\left(1 - n \times \frac{\alpha}{2} \right) \right) \tag{29}$$

Note that the λ and δ values in Equations (27) and (28) are obtained by fitting the maxima values to the Gumbel distribution for minima, and the λ and δ values in Equation (29) correspond to the Gumbel distribution for maxima.

Three distributions are chosen to investigate the number of false-positives, or Type I errors, produced by each of the three methods discussed previously. The normal, lognormal and gamma distributions are modeled using the three methods and the number of outliers is compared for a 99% confidence interval. The normal distribution will provide a sanity check to make sure that the establishment of the confidence intervals based on EVS and best-fit normal distribution produce similar thresholds. The lognormal and the gamma distributions are both skewed and will provide an opportunity to dramatically illustrate the shortcomings of the confidence interval estimation based on a normal assumption of the data. The probability density function (PDF) for each of the three distributions are as follows:

$$\text{Gaussian:} \quad f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{30}$$

$$\text{Lognormal :} \quad f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \geq 0 \tag{31}$$

$$\text{Gamma :} \quad f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}} \quad \text{for} \quad x \geq 0 \tag{32}$$

where $\Gamma(a)$ is the gamma function.

Castillo (1988) shows that both the minimum and the maximum for the normal and lognormal distributions can be modeled with a Gumbel distribution, thereby reducing the effort of finding the best-fit distribution in this example. On the other hand, the gamma distribution has Gumbel distributed maxima and Weibull distributed minima. Distributions of varying sample size from N = 1000 to N = 10^6 were created and analyzed. The typical analysis results only for the sample size of N = 10,000 are presented in this study. Similar results are, however, observed for the other examined sample sizes. Tables 2, 3 and 4 summarize the results of the parameter

estimation and number of outliers for the N = 10,000 sets of data from each of the three distributions. Only the first 1,000 data points are plotted for illustrative purposes in Figures 8, 9 and 10.

Looking at the normally distributed data in Figure 9, it seems that the thresholds obtained from Methods 1, 2 and 3 are comparable. For Method 3, the *least-squares return period relative error* (LSRPRE) estimation technique (Castillo, 1988) is used to compute parameters of the Gumbel distributions for the maxima and minima of the normally distributed data. Initially several techniques of parameter estimation suggested by Castillo (1998) were investigated and the LSRPRE turned out to produce the best-it result for the given data sets.

Table 2 shows the upper and lower confidence limits computed from Methods 1, 2 and 3, and the associated numbers of outliers. As can be seen in Figure 9 and Table 2, even though Method 3 returns thresholds that are slightly different from the known PDF, the number of outliers is closer to the expected 1% than Method 2.

In the second numerical example, the parent distribution is lognormal instead of normal. For this simulation, μ = 1.0 and σ = 0.5 are assumed for the parameter values in Equation (31). The associated lognormal density function is displayed on the left side of Figure 10. The skewness and kurtosis of this distribution are 1.74 and 8.45, respectively. Note that, for all normal distributions, the values of the skewness and kurtosis should be 0.0 and 3.0, respectively (Wirsching et al., 1995). Therefore, the departure of the skewness and kurtosis values from 0.0 and 3.0 indicates the non-Gaussian nature of the data. Figure 10 and Table 3 display similar analysis results for the lognormal parent distribution. Again the LSRPRE estimation technique is employed for the maxima of the lognormal data. The minima, however, are fitted using the *least-squares probability absolute error* method. For the lognormal example, Method 3 only returns 3 more false-positive indications than the expected 100 outliers as calculated from Method 1. Method 2, however, shows over double the number of false-positive indications due to the upper threshold being far too low. On the other hand, the lower limit based on normality completely misses all of the minimum values because the lognormal distribution contains only positive data points.

Finally, the sequential tests are applied to data sets simulated from a gamma parent distribution. In this example, the sample data are generated from a gamma distribution with $k = 3$ and $\nu = 0.2$ for the parameter values in Equation (32). This gamma distribution has the skewness value of 1.15 and kurtosis of 5.00, respectively. The associated density function is plotted on the left side of Figure 11. The gamma distribution is skewed to the right for a small value of k. As the degrees of freedom, k, increases the gamma distribution converges to the normal distribution. The maxima of the gamma parent distribution are fit using the LSRPRE method, while the minimum values are fit using the *standard weighted least-squares* method with a weighting factor of 1. The extreme value method again shows a distinct advantage over the normal assumption. For the gamma distribution, Method 3 returns four lower numbers of false-positives than expected from Method 1, while Method 2 again returns almost twice as many false-positives as Method 1. The number of false positive indications returned by Method 2 might lead to incorrect damage diagnosis of the system.

A drawback of EVS is that different methods of parameter estimation are optimal for fitting different distributions. Once the parameter values of the extreme value distribution are estimated, there is, however, a noticeable advantage of EVS over normality assumption in properly setting the threshold values. The next two sections apply the techniques demonstrated in Section 4 to real world test structures for damage detection.

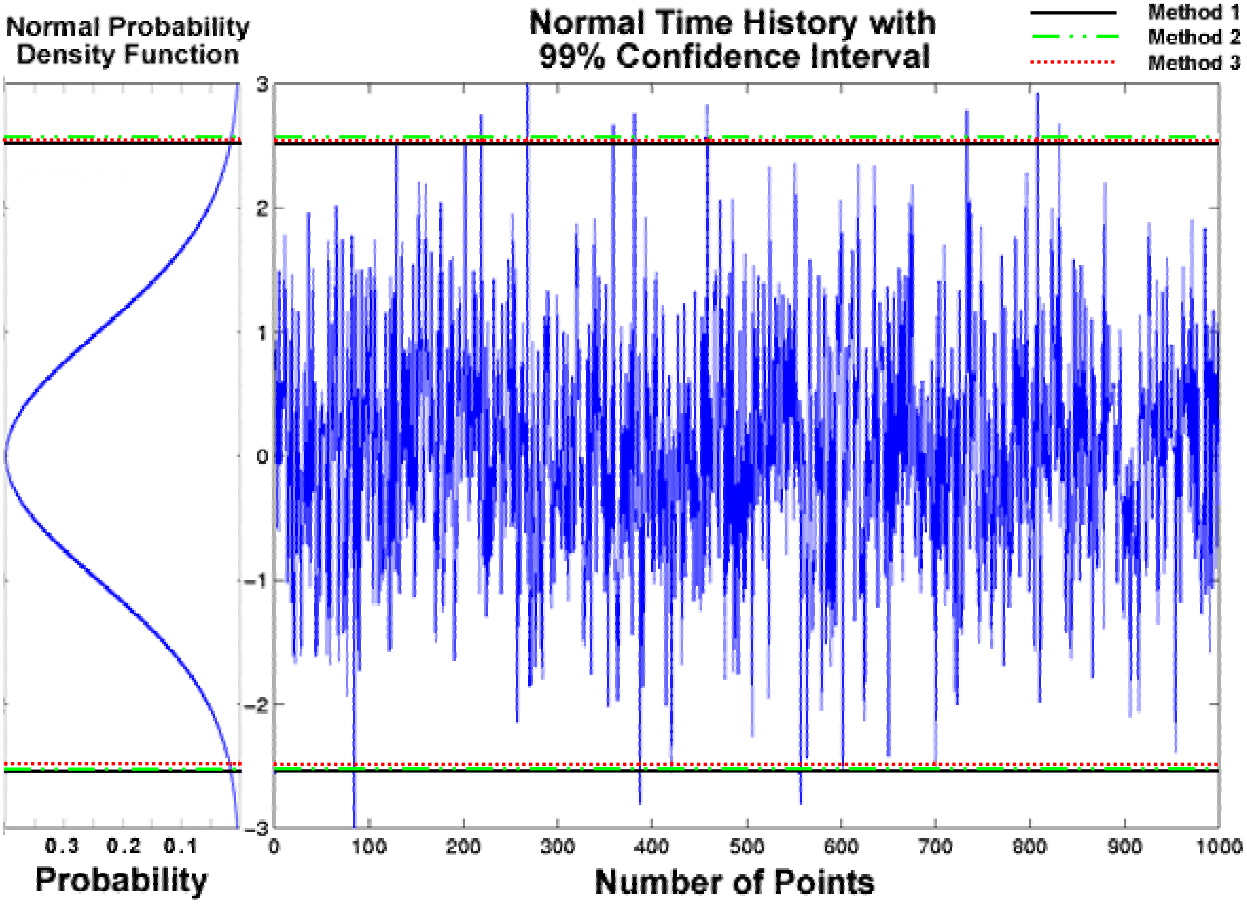


Figure 9: The exact 99% confidence interval of a normal parent distribution compared with that from extreme values statistic. This figure shows the first 1000 data points from a 10,000 data point set.

Table 2: Estimation of 99% confidence intervals for the 10,000 data points generated from a Gaussian parent distribution.

Estimation method	Upper confidence Limit	Lower confidence Limit	# of outliers out of 10,000 samples. ($\alpha = 0.01$)
Method 1 (Exact)	2.548	-2.548	100
Method 2 (Normal)	2.551	-2.545	91
Method 3 (Gumbel)	2.549	-2.482	99

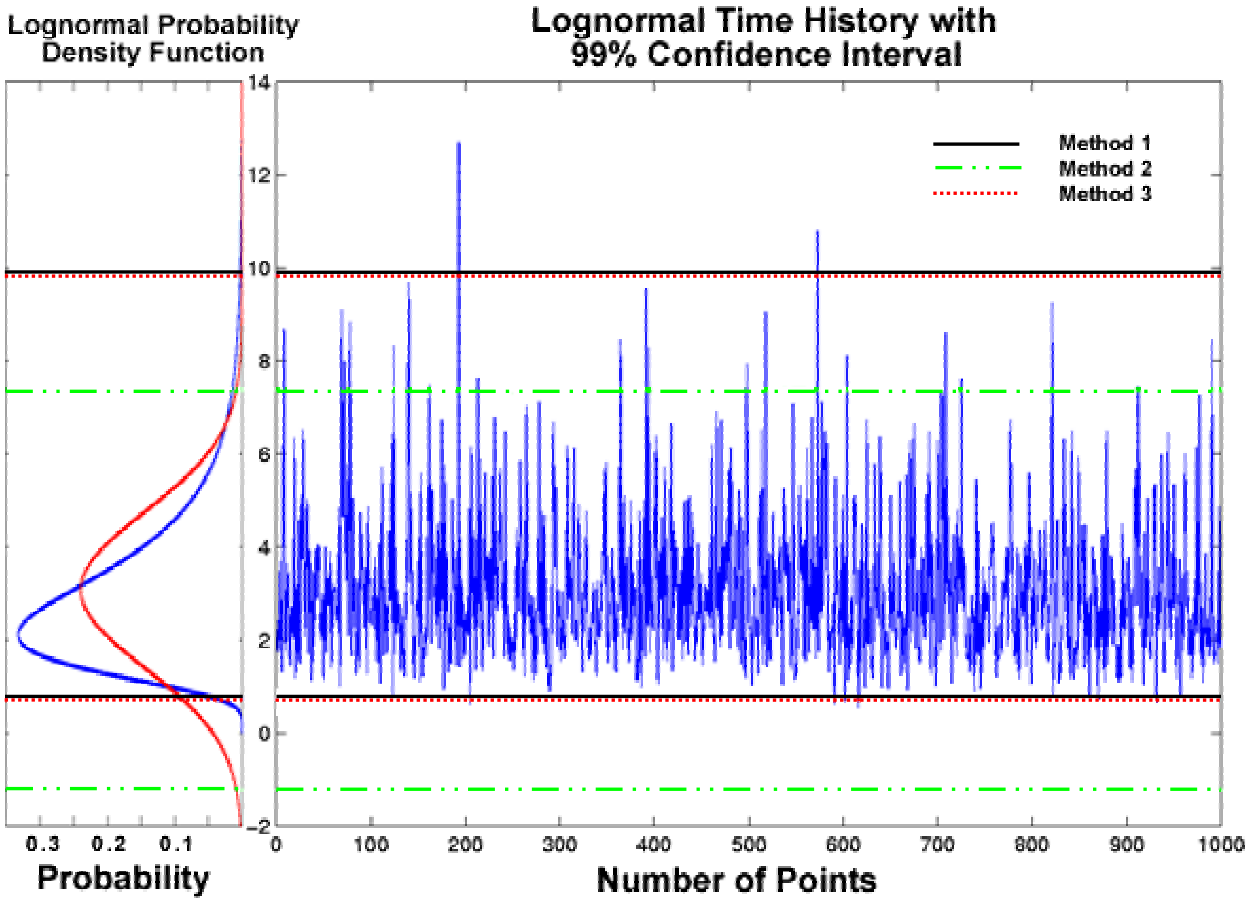


Figure 10: The exact 99% confidence interval of a lognormal parent distribution compared with those computed from either extreme values statistic or the normality assumption.

Table 3: Estimation of 99% confidence intervals for the 10,000 data points generated from a lognormal parent distribution.

Estimation method	Upper confidence Limit	Lower confidence Limit	# of outliers out of 10,000 samples. (α = 0.01)
Method 1 (Exact)	9.854	0.750	100
Method 2 (Normal)	7.378	-1.206	230
Method 3 (Gumbel)	9.827	0.715	103

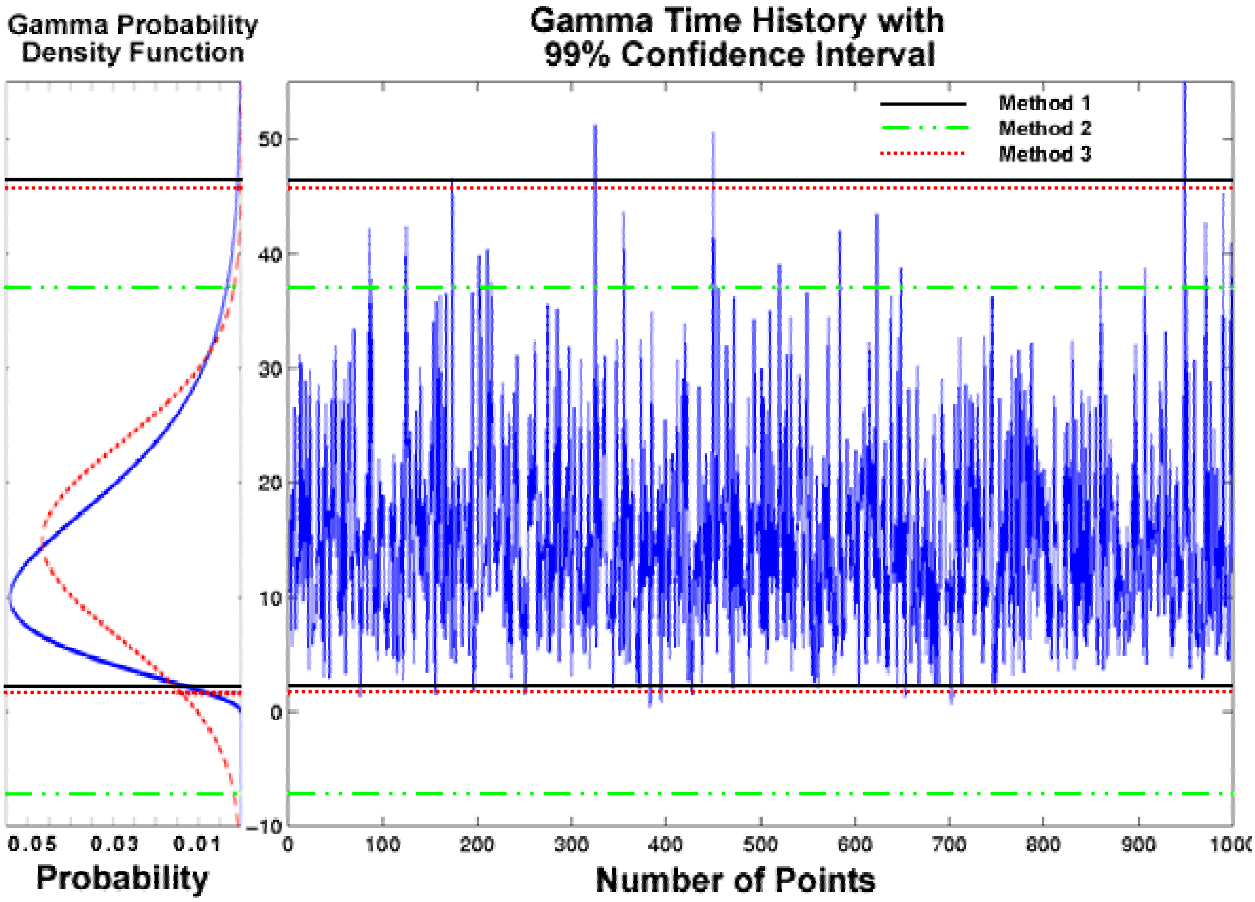


Figure 11: The exact 99% confidence interval of a gamma parent distribution compared with those computed from either extreme values statistic or the normality assumption.

Table 4: Estimation of 99% confidence intervals for the 10,000 data points generated from a gamma parent distribution.

Estimation method	Upper confidence Limit	Lower confidence Limit	# of outliers out of 10,000 samples. (α = 0.01)
Method 1 (Exact)	46.369	1.689	100
Method 2 (Normal)	37.016	-7.142	191
Method 3 (Gumbel)	45.693	1.600	96

5. Three-Story Frame Structure

5.1. Test Structure Specifications

One of the structures tested is a three-story frame structure model shown in Figure 12. The structure is constructed of Unistrut columns and aluminium floor plates. The floors are 1.3-cm-thick (0.5 in) aluminium plates with two-bolt connections to brackets on the Unistrut. The base is a 3.8-cm-thick (1.5 in) aluminium 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 12. Dimensions of the test structure are displayed in Figures 15 and 16. 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 shaker is coupled to the structure by a 15-cm-long (6 in), 9.5-mm-diameter (0.375-in) stinger connected to a tapped hole at the mid-height of the base plate. The shaker is attached at corner D, as shown in Figure 14, so that both translational and torsional motions can be excited.

5.2. Test Setup and Data Acquisition

The structure is instrumented with 24 piezoelectric single-axis accelerometers, two per joint as shown in Figures 13 and 16. 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 accelerometers are numbered from the corner A to B, C, and D counter-clockwise and from the top floor to the first floor. The accelerometers on the plate have odd channel numbers, while the accelerometers on the column have even numbers. The nominal sensitivity of each accelerometer is 1 V/g. 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 analogue signals. The data sets that were analyzed in the

feature extraction and statistical modeling portion of the study were the acceleration time histories. Each set of data gathered consisted of 4096 points.

In each test case, three separate data sets were collected with the shaker input level at 3, 5, or 7 volts. The bandwidth of the shaker and the sampling rate of the response were also varied from 800, 1600 to 3200 Hz in each test case to determine in which frequency bandwidth the feature extracted from the date would be most sensitive to the induced damage. All input from the shaker to the base was random.

Two damage cases were investigated in this experiment. The first damage were introduced to the corner C of the third floor (3C in Table 5) and the second damage were placed at the corner A of the first floor (1A in Table 5) along with the damage at 3C for the second damage case. These two damage locations are shown in Figure 12. For each damage location, four bolts at each joint were loosened until hand tight, allowing relative movement between the floor plate and column. After each damage case, all the bolts were tightened again to the initial torque of 0.7 Nm (60 in-pounds).

Figure 12: Photo of the full test structure.

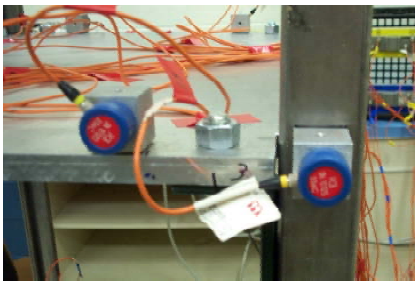


Figure 13: Photo of sensor location on a joint.

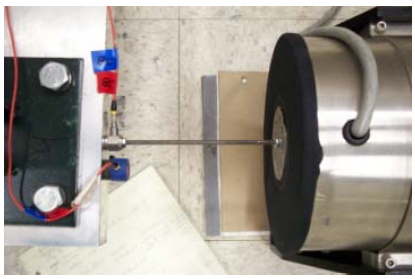


Figure 14: Photo of the shaker and force transducer.

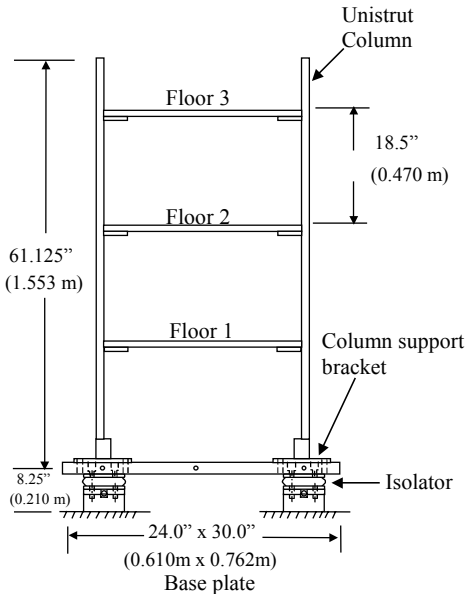


Figure 15: Basic dimensions of the three-story frame structure.

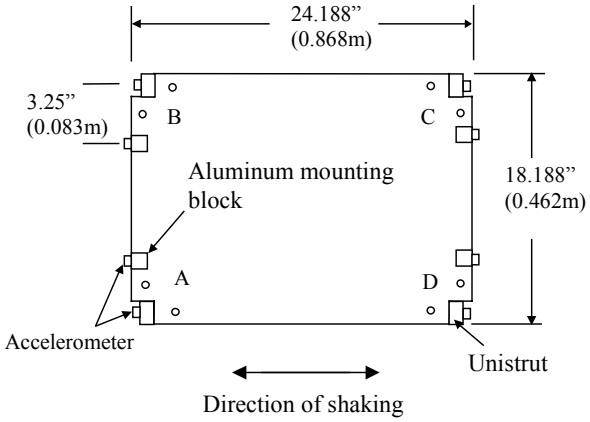


Figure 16. Floor layout as viewed from above.

Table 5 shows the operational variability in data sets from which the EV data were drawn.

Table 5: Test Matrix Showing Operational Variability Considered in the Data Sets.

Case	Sampling rate	Input Voltage Level
Undamaged	800 Hz	3,5,7 V
Undamaged	1600 Hz	3,5,7 V
Undamaged	3200 Hz	3,5,7 V
Corner 3C Hand Tight	800 Hz	3,5,7 V
Corner 3C Hand Tight	1600 Hz	3,5,7 V
Corner 3C Hand Tight	3200 Hz	3,5,7 V
Corners 3C and 1A Hand Tight	800 Hz	3,5,7 V
Corners 3C and 1A Hand Tight	1600 Hz	3,5,7 V

5.3. Finite Element Analysis of the Three-Story Frame Structure

For most real world applications, information from the damaged system will not be easily available. Therefore, numerical simulations can be utilized to define SHM system properties

prior to deploying a monitoring system on real world structures. Examples of such properties include, but are not limited to, bandwidth, sensitivity, dynamic range of sensors, optimal location of sensors, sensitivity of extracted features to damage, and possible excitation source waveforms. The ultimate goal of this section is to develop a finite element model of the three-story frame structure that can be "damaged" or "repaired" as needed, and that will eventually facilitate parameter specification of the SHM system prior to its actual implementation. Figure 17 provides a view of the FE model, which was constructed to match the geometry and physical properties of the three-story frame structure. The geometry for the model was developed using the CAD software IDEAS (www.sdrc.com) and analyzed in a general finite element analysis program ABAQUS (www.abaqus.com). Modeling of each component is detailed in the following.

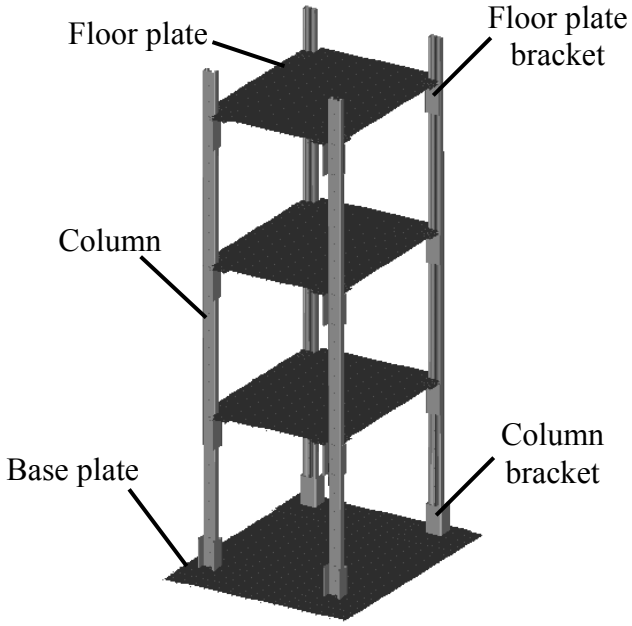


Figure 17: Finite element model of experimental structure.

Floor plates

Each of the three floor plates is a rectangular aluminum plate with squares cut out of each corner for the columns. The floor plates were modeled with shell elements using aluminum material properties and the same thickness as the physical floor plates. The base plate was also modeled with shell elements and aluminum material properties.

Air-bearings

To model the air-bearings, springs were attached from nodes on the base plate to the ground. The nodes were placed at locations corresponding to the physical connections between the air-bearing and the floor plate. Values for the air-bearing spring constants were approximated using specifications given by the air-bearing manufacturer. These spring constants were then updated to match the natural frequencies computed from the finite element model with those obtained from an experimental modal test of the structure. This model updating properly reproduced rigid body motions associated with the structure rocking on its base.

Columns

The columns are 152.4 cm-long (60-in.) B-line brand stainless steel channels. Figure 18 shows the cross-section shape and major dimensions of the columns. Beam elements with the same geometric cross-section as the physical columns were used in the model.

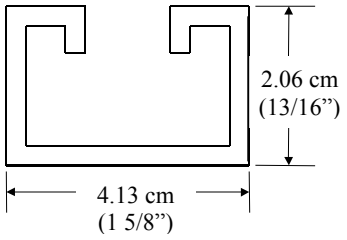


Figure 18: Cross-section of column.

Column brackets

To attach the columns to the base plate, B-line brackets were used. These brackets are referred to as column brackets throughout this report. Each bracket is comprised of a flat plate that is bolted to the base plate and a U-shaped channel that extends perpendicular to the base plate and encompasses the bottom 8.89 cm (3.5 in) of the column. The column is bolted to the U-shaped channel of the bracket. The plate portion of each bracket was modeled with shell elements. To model the U-shaped channel, the cross-section of the column was increased to a length that

included the geometry of the bracket. The cross-section of the column including the bracket is shown in Figure 19.

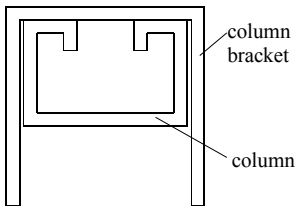


Figure 19: Cross-section of column with column bracket.

Floor plate brackets

Brackets are used to attach each floor plate to the four columns. Each bracket, referred to as a floor plate bracket, has two flat plates attached to the floor plate and an L-shaped channel attached to the column. Each of the flat plates is rectangular with dimensions of 3.81 cm (1.5") x 4.7625 cm (1.875"). The L-shaped channel is 9.21 cm (3.625") tall, and fits with the column as shown in Figure 20. As was done with the column brackets, the cross-section of the column is adjusted in the regions where the floor plate brackets are attached to include the cross-section geometry of the brackets. The plates of each bracket are rigidly connected to the columns with beam multiple point constraints (MPCs).

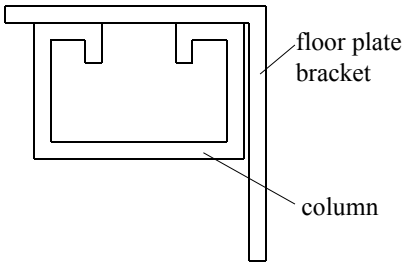


Figure 20: Cross-section of column with floor plate bracket.

Contact Surfaces

The original intent of the FE modeling was to create a linear model of the structure. Therefore, contact between the floor plates and the floor plate brackets was simulated by introducing linear

spring elements from nodes on the floors to nodes on the brackets. For the floor plates, a normal spring constant K_u was calculated by,

$$K_u = \frac{(AE)}{L} \tag{33}$$

where A is the contact area, E is the Young's Modulus, and L is half of the shell thickness because the nodes are located on the mid-surface of the shell elements and represent movement of that surface. The same method applies for the lower normal contact spring K_l (Figure 21).

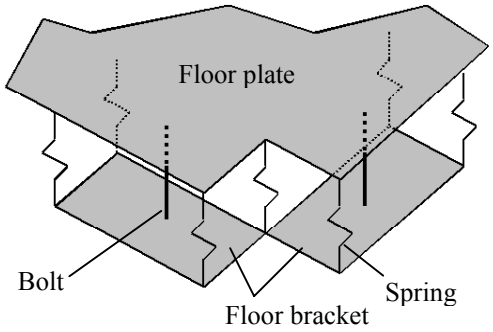


Figure 21: Isometric view of a corner joint (with column removed) showing locations of contact springs.

The acceleration time histories from the linear model were not desirable because the data from the damaged and undamaged cases were exactly the same, so the springs were changed to bilinear springs. The nonlinear springs resist motion with linearly increasing force up to a certain magnitude, after which the resisting force remains constant. In this way, the nonlinear springs model friction and slippage in the joint. The spring constants for the nonlinear springs were estimated based on the motion in the joint in the linear model. These spring constants should be updated and validated in future work with experimental data from the test structure.

Damage Simulation

The bolts in the joints of the structure were modeled as beam elements connecting nodes on the floor plates to nodes on the floor brackets. The tightness of the bolts was introduced by applying a stress initial condition in the axial direction of the bolts. The magnitude of the preload was first calculated from the following equation,

$$F_i = \frac{T}{0.20d} \tag{34}$$

where F_i is the preload in the bolt, T is the torque to which the bolt is tightened, and d is the fastener size. The stress can then be calculated as,

$$\sigma = \frac{F_i}{\frac{\pi}{4}d^2} \tag{35}$$

For damaged cases, the preload in the bolts at one of the joints was reduced to model loosening of the bolts. The damage can be simulated in three different ways in this model. The first case is to reduce the preload of the bolts at the beginning of the analysis and leave them constant throughout the procedure. This method models a joint that has been damaged, but is not being further damaged. The second case is to reduce the stress in the bolt linearly throughout the analysis. This method models a joint that is progressively loosening or being damaged. The third case of damage introduction is a step reduction of stress at one time point during the analysis. Many other damage cases could be introduced to the model in further studies. The ability to easily apply different types and magnitudes of damage to the structure is one of the main advantages this analytical method has over experimental data collection.

Model Validation

Before being used to generate acceleration response data, the model was validated by comparing it to the physical structure. The two areas compared were weights of the components for the model and the physical structure and modal analysis results for each.

The weights of the components of the finite element model and the physical test structure are shown in Table 6. As seen in the table, the weights of individual components are not exactly equal, but they are close, and the total weights of the two cases are almost exactly equal. One discrepancy that needs clarification is the weight of the base plate and air-bearings. When the physical structure was weighed, the air-bearings were left attached to the base plate. Therefore, the weight of the air-bearings is included in the weight of the base plate listed in the table.

Experimental and analytical modal analyses were performed on the physical structure and with the finite element model, respectively. The frequencies for the first several modes of each are shown in Table 7.

Table 6: Weights of structure components.

	Mass (kg)	
Component	Model	Physical
Floor plates	28.58	29.48
Base plate	47.99	54.43
Air-bearings	4.581	
Base brackets	4.627	6.360
Floor brackets	6.940	5.911
Columns	17.15	11.68
Base bolts	0.6804	1.270
Floor bolts	0.2268	1.678
Total mass	110.8	110.8

Table 7: Natural frequencies (Hz) from experimental data and finite element model.

Mode number	Experimental	FE Model
1	2.290	3.030
2	3.040	3.870
3	12.57	6.760
4	13.90	7.270
5	14.46	11.00
6	24.87	20.12
7	32.04	34.95
8	40.08	39.89
9	49.82	50.84
10	69.10	57.59

5.4. Damage Diagnosis Results

The signal chosen for analysis was the raw time-series data from channel 21, the sensor on the plate for damage location 3A of the structure. Data from the highest level of excitation, 7V, were selected. If the data were plotted in series, data corresponding to the damaged state would be observed to have a lower variance, and this feature could be picked up immediately by an

appropriate hypothesis test. Figure 22 shows the data from the undamaged condition followed by the data from the damaged condition. The two lines denote a 99% confidence interval computed from the normality assumption of the undamaged data. However, if one is interested in a simple threshold-crossing criterion, one could not deduce damage from the figure because the damaged data with a low variance clearly falls between the 99% confidence interval of the undamaged data.

Raw data would not usually be used in this manner and there is no reason to suppose that damage would generally produce an increase in the amplitude of the signal. The object of this exercise is to show that one can construct a threshold-crossing diagnostic by using EVS.

A moving window of width 64 samples was stepped through the 4096 points of each data set to generate 64 maxima for each condition. When the empirical CDF was plotted in coordinates appropriate for Gumbel maximum probability paper, the results in Figure 23 were obtained. A straight line fit to the 64 highest order statistics gave a satisfactory degree of agreement with the data. This fit was interpreted as evidence that the maxima were Gumbel distributed. The bilinear nature of the plot in Figure 23 might be regarded as some cause for concern. However, an analysis of the parent distribution indicated that the raw accelerations were very close to Gaussian, and it is known that the maxima from Gaussian distributions converge slowly to the required Gumbel distribution (Castillo, 1988). In fact, a plot of the original 4096 data points from channel 21 in Gumbel maximum coordinates also showed a satisfactory straight line fit in the right tail (Figure 24).

The next stage in the analysis was to estimate parameters to fit a Gumbel maximum distribution to the empirical CDF. The results of this process are shown in Figure 25. The ML method was used, and the parameters obtained were $\lambda = 0.2637$ and $\delta = 0.0373$.

Having obtained estimates for the parameters, it is a trivial matter to use the CDF to generate values for the $100(1-\alpha)$ % confidence interval of the distribution, as in Equations (28) and (29), to give the associated confidence limits for the data.

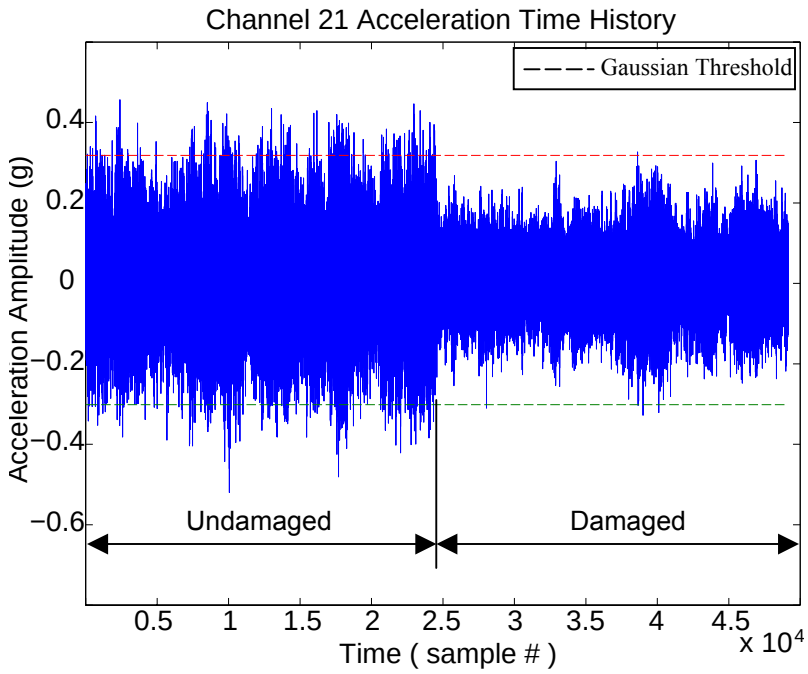


Figure 22: The raw time-series data from channel 21 with a 99% confidence interval; undamaged system data followed by damaged system data.

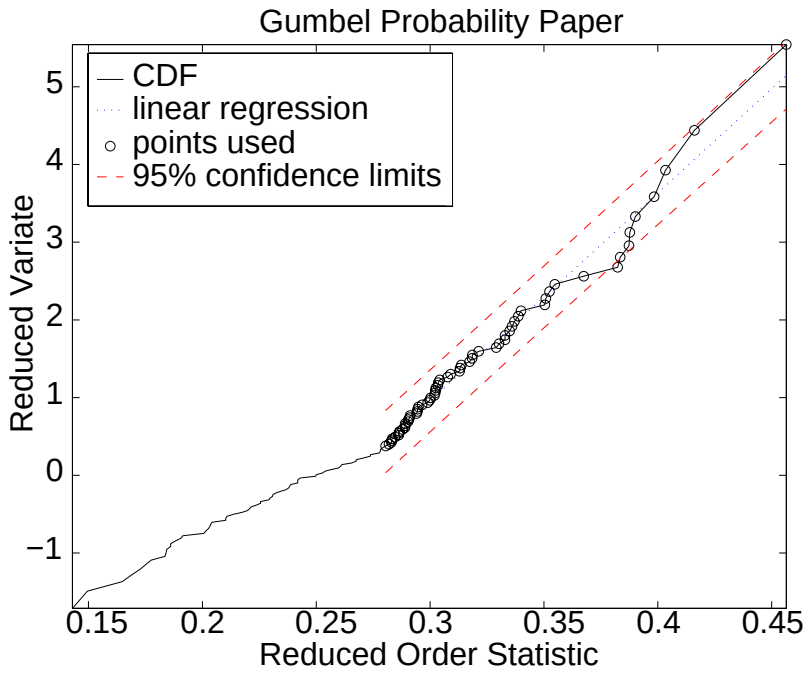


Figure 23: Plot of windowed maxima in Gumbel coordinates.

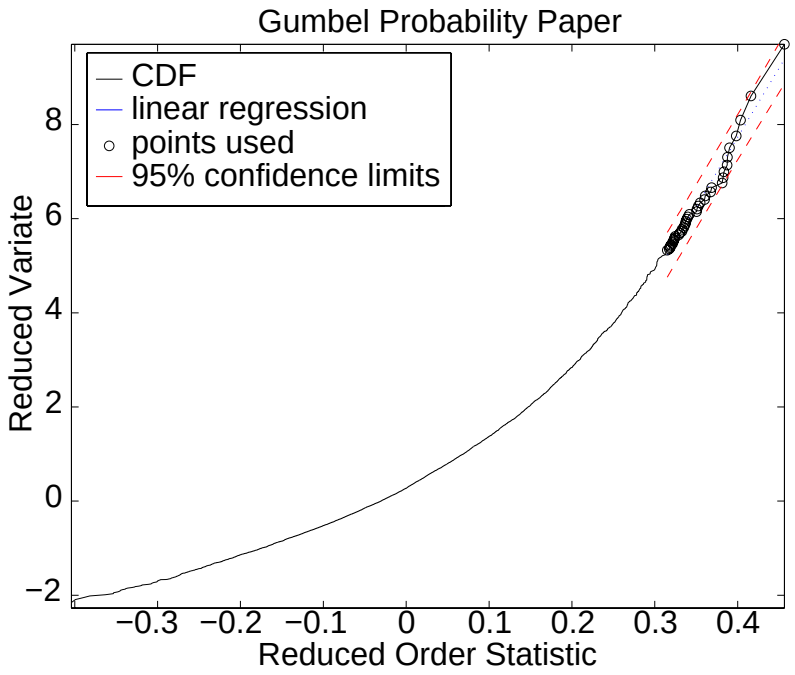


Figure 24: Plot of channel 21 raw acceleration data in Gumbel maximum coordinates.

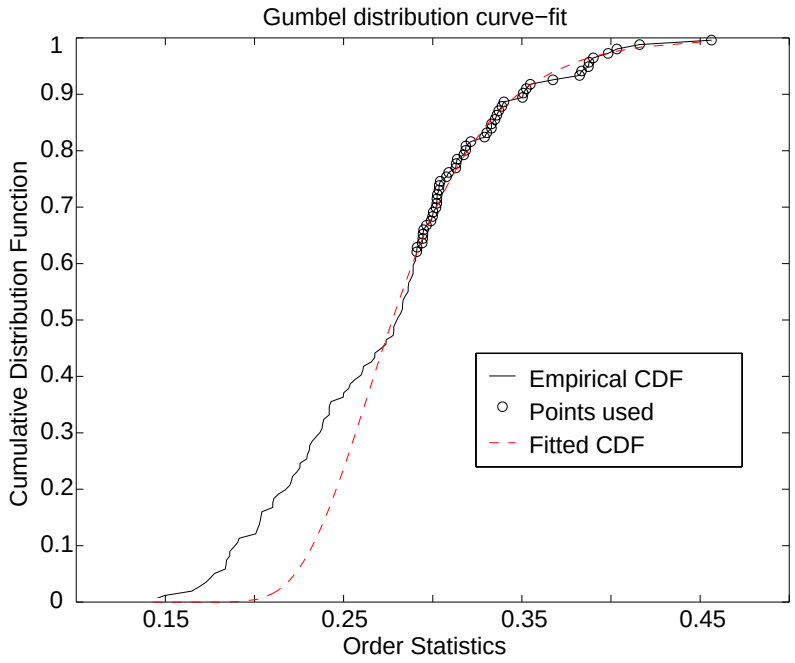


Figure 25: Fit of Gumbel maximum distribution to empirical CDF from channel 21 maxima.

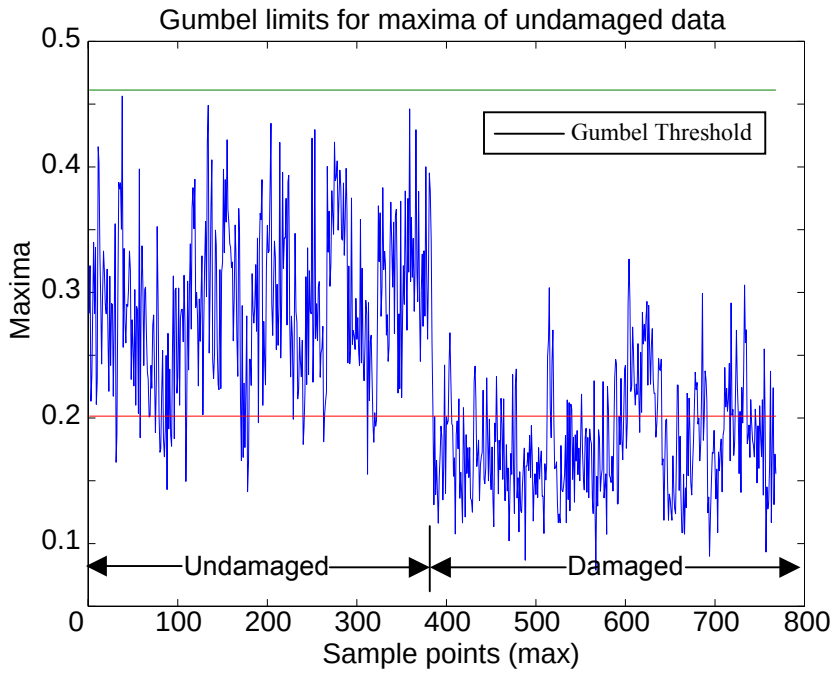


Figure 26. Windowed maxima from channel 21 accelerations. Upper and lower thresholds are 99.5 and 0.5 percentiles from the Gumbel distribution for maxima shown here as solid lines.

The results in Figure 26 show that a good upper threshold is obtained for maximum values, as opposed to Figure 22 that used the raw data, i.e., the data from the parent distribution. There is also a lower threshold here that serves to partially separate the undamaged and damaged conditions. In fact, this threshold is *incorrect.* The problem is that the lower threshold comes from the Gumbel distribution for maxima (here the concern is with *maxima of maxima*). The lower threshold should actually be computed from a distribution for minima (the *minima of maxima*). This fact is clear from the number of threshold crossings on the normal condition data set.

To find the appropriate distribution, the empirical CDF was first plotted in coordinates appropriate for Gumbel minimum probability paper resulting in a figure that had marked curvature in both tails. The next attempt used Weibull minimum coordinates. The location parameter λ was adjusted by trial and error until reaching a value of $\lambda = 0.12$, yielding the plot shown in Figure 27. This plot shows that the *whole population* of channel 21 maxima was well described by a Weibull distribution for minima. This description was confirmed by the maximum likelihood curve fit shown in Figure 28. Although only points in the left tail were used to fit

parameters, the curve describes the empirical CDF over the entire range. The remaining parameters for the distribution were estimated as $\delta = 0.1696$ and $\beta = 2.6976$. Using the whole data for the parameter estimation yielded estimates of $\delta = 0.1720$ and $\beta = 2.7056$.

Once the parameters for the distribution were obtained, the lower and upper thresholds could once again be computed using the inverse CDF of Equation (12).

$$x_m = \lambda + \delta[-\ln(1-\frac{\alpha}{2})]^{1/\beta} \qquad x_M = \lambda + \delta[-\ln(\frac{\alpha}{2})]^{1/\beta} \tag{36}$$

This parameter estimation gave the upper and lower threshold values of 0.4347 and 0.1438, respectively, as shown in Figure 29.

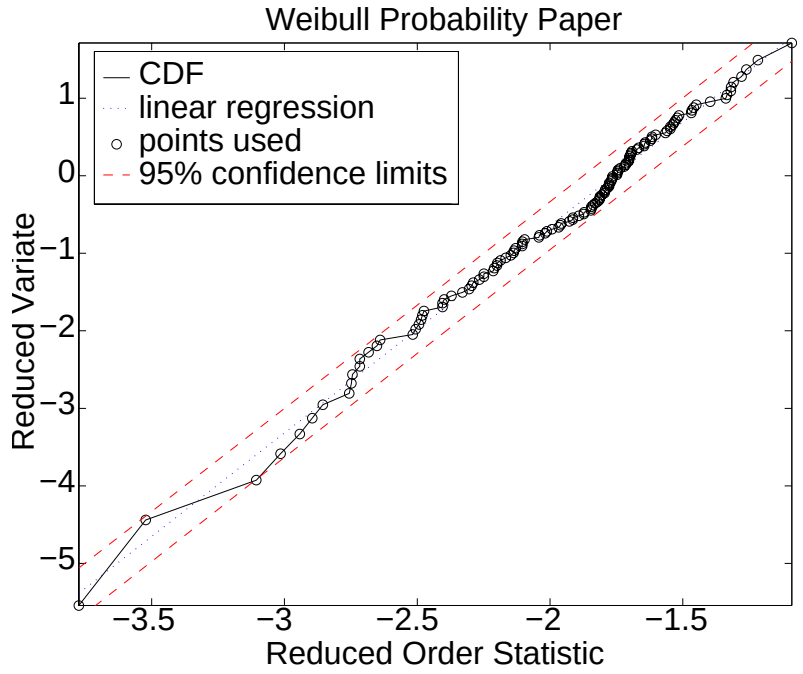


Figure 27: Plot of channel 21 windowed maxima in coordinates appropriate for a Weibull minimum distribution.

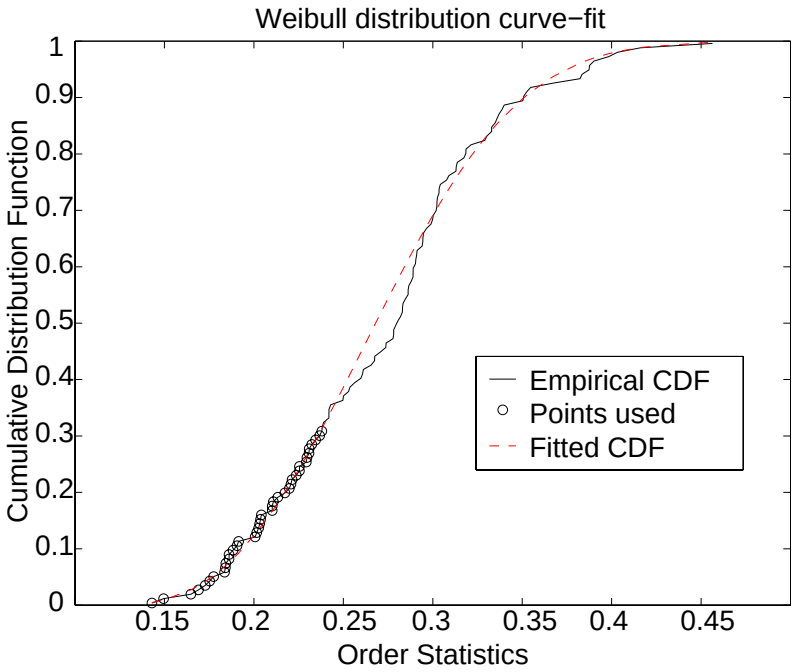


Figure 28: Weibull minimum distribution fit to channel 21 windowed maxima.

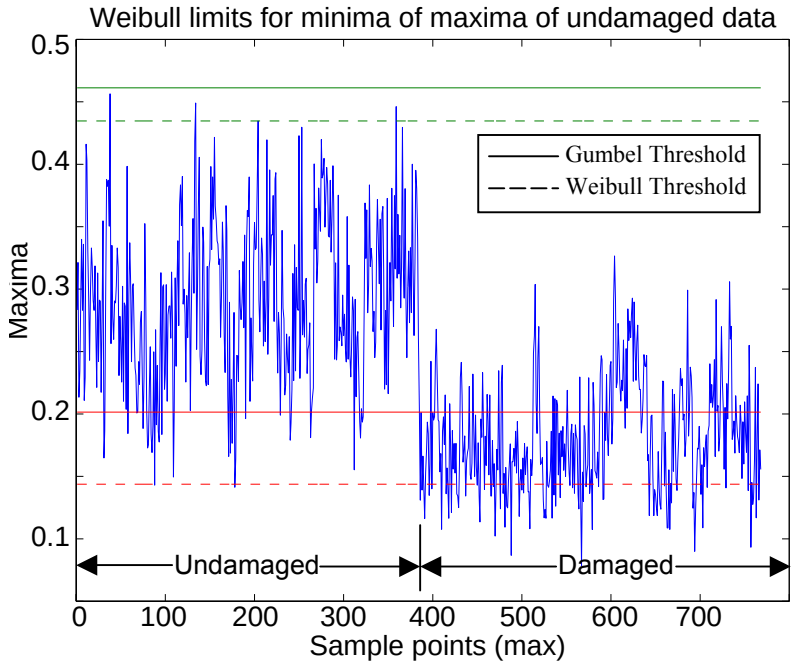


Figure 29: Windowed maxima from channel 21 accelerations. Upper and lower thresholds are 99.5 and 0.5 percentiles from the Weibull distribution for minima, shown here as dashed lines.

The lower threshold is much better than the previous threshold based on the Gumbel distribution for maxima (solid line in Figure 29). The improved threshold estimation appears to be the major factor in favor of using EVS for damage identification. The EVS allows for better control of the thresholds that are used to signify statistical deviance from the undamaged condition. To drive the point home, Figure 30 shows the data from Figure 26 with thresholds computed on the assumption that the maxima data are Gaussian (dotted thresholds); the 0.5 and 99.5 percentiles are clearly very conservative. Many more of the damaged condition maxima would be judged normal.

The final analysis for these data employs the Auto Regressive–Auto Regressive with Exogenous Inputs (AR-ARX) process for damage detection (Sohn and Farrar, 2001). The AR-ARX model is a linear, time predictive model from which residual errors are calculated. These residual errors are then used as the damage sensitive feature that is monitored. The details can be found in Appendix B.

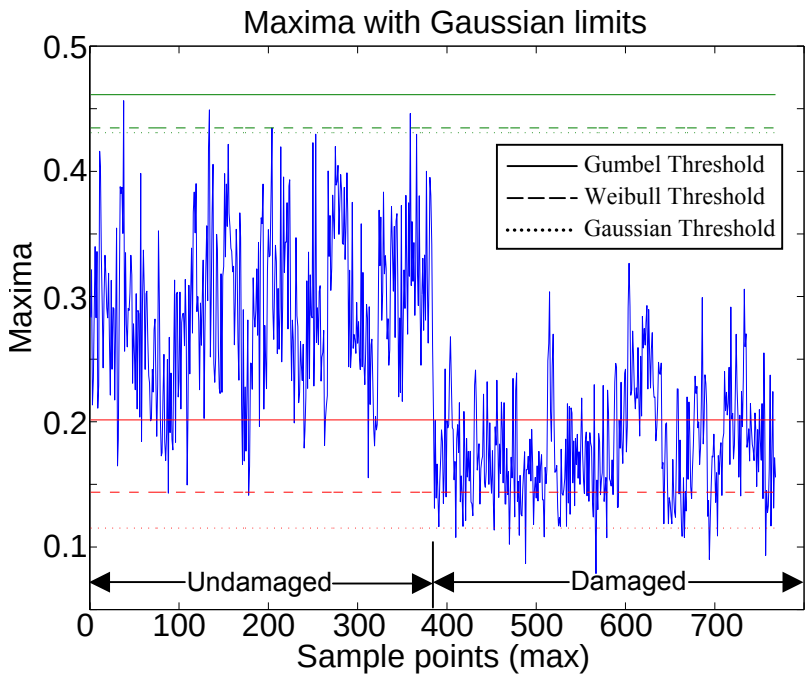


Figure 30: Windowed maxima from channel 21 accelerations. Upper and lower thresholds are 99.5 and 0.5 percentiles from the Gaussian distribution, shown here as dotted lines.

First, the difference between acceleration time histories between channels 21 and 22, which span the damaged joint 3A, was computed for the damaged and undamaged cases. The AR-ARX model was then fit to the undamaged signal and used to construct a vector of prediction errors. The same model was then used to find a sequence of prediction errors on the damaged data. The underlying assumption is that the errors on the damaged sequence will be significantly higher than those on the undamaged sequence because the AR-ARX model is developed from fits to the undamaged data.

The EVS computes an accurate threshold for the undamaged data. To apply the method, a window of 64 points was again stepped through the data, and the maxima from each of the windows were extracted. As before, this procedure resulted in 64 maxima from the original 4096 points.

Figure 31 shows this sequence of maxima plotted in coordinates appropriate for the Gumbel maximum distribution. The transient nature of the AR-ARX time prediction model at a few initial time points engendered anomalously high initial residual errors. This transient resulted in two spurious points out of the 64-point record of maxima, which were duly ignored during the initial analysis. Figure 31 shows that the hypothesis of a Gumbel distribution for the right tail is tenable, so the parameters of the Gumbel distribution were estimated from the maxima data. The comparison of the empirical Gumbel CDF and the analytical one with fitted parameters are shown in Figure 32.

The parameters obtained from the curve fit were $\lambda = 0.0506$ and $\alpha = 0.0071$. Substituting these values into Equations (28) and (29) gave a lower threshold of 0.0388 and an upper threshold of 0.0882. However, note that the lower threshold is inappropriate because it is based on maxima values as previously discussed.

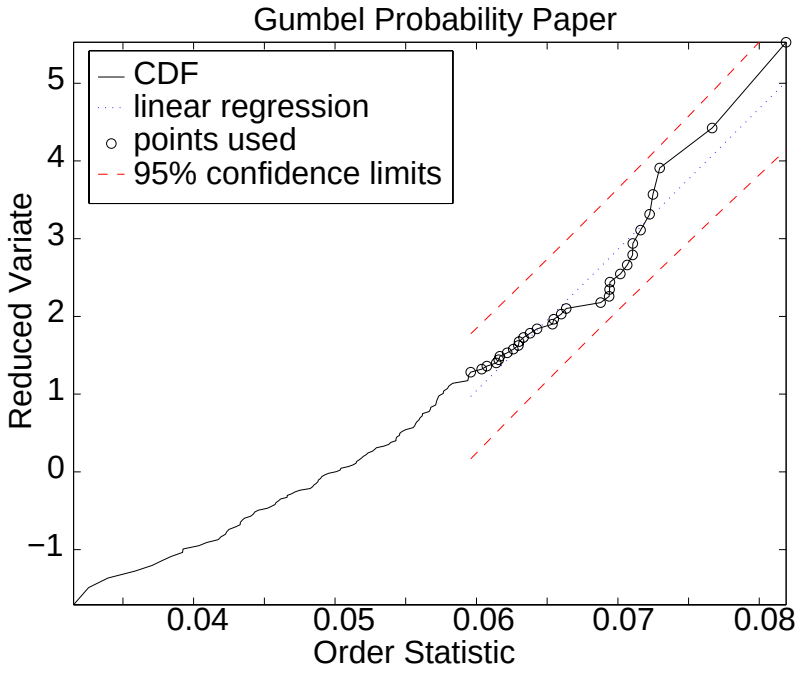


Figure 31: Plot of windowed maxima of AR-ARX errors in a Gumbel maximum coordinates.

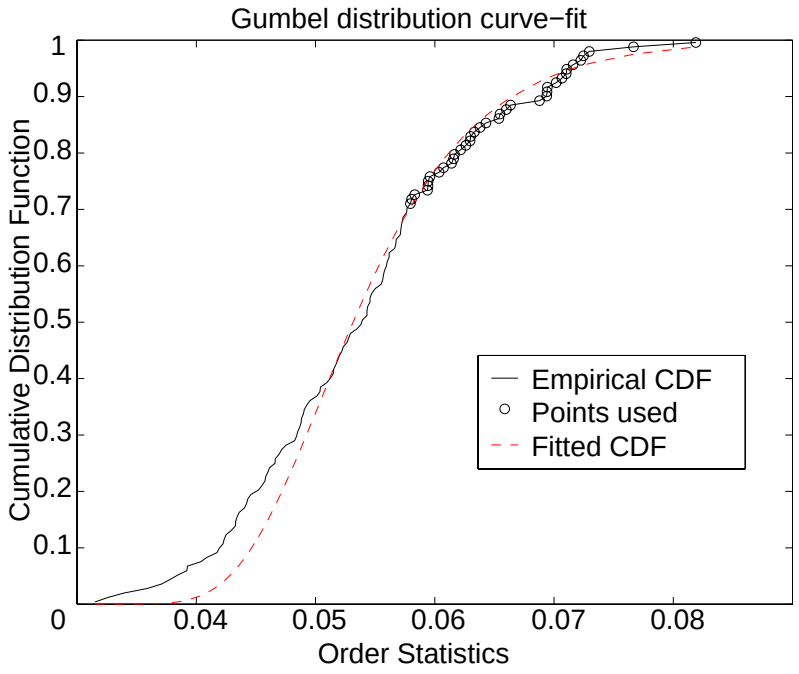


Figure 32: Gumbel maximum distribution fit to windowed maxima of AR-ARX errors.

Figure 33 shows the maxima of the prediction errors obtained from the AR-ARX analysis for the undamaged and damaged cases. With a type I error probability of $\alpha = 0.005$, a one-sided confidence interval is computed on the right tail of the prediction errors. Note that because the occurrence of damage will increase the prediction errors, the detection of a prediction error that goes beyond a certain threshold value is the primary goal in this example. In other words, the monitoring of smaller prediction errors are not main concerns. Therefore, only the upper control limit is computed from Equation (29) using α instead of $\alpha/2$. As expected, the upper threshold provides a good upper bound on the normal condition data and separates it successfully from the damaged data. In this case, because the undamaged and damaged data sets are so distinctive, Gaussian limits would probably have worked as well. However, as the earlier analysis showed, there are situations where the Gaussian limit will break down.

The next analysis involves more data from channel 21, but coming from a separate experiment in which the level of excitation was 5 V and the frequency bandwidth of the excitation was 1.6 kHz. The features are once again the prediction errors from the AR-ARX analysis. In this case, three data sets were considered: the first two corresponding to the undamaged condition and the final one corresponding to the damaged case. As before, 4096 points per set were available and 64-point sets of maxima were extracted. The empirical CDF for the maxima from the first undamaged case is shown in Figure 34 on the Gumbel probability paper. The excellent fit implies that the maxima have a Gumbel distribution. Figure 35 shows the maximum likelihood curve-fit to the empirical CDF. An estimate of the parameters yielded $\lambda = 0.2804$ and $\delta = 0.0497$. The upper threshold value was obtained from Equation (29) again using α instead of $\alpha/2$ and gave a threshold value of 0.5439.

The windowed maxima of the errors are shown in Figure 36 with the EV and Gaussian thresholds applied. As one can see, the Gaussian threshold encompasses the entire undamaged feature, but also encompasses a significant portion of the damaged data feature. The EVS threshold does show a few false-positive indications in the undamaged data as should be expected. The EVS threshold, however, shows a marked improvement in distinguishing the damaged feature set from the undamaged data set. Next, these same EVS techniques will be applied to a spring-mass system in which the feature set has been extracted and normalized in a slightly different manner than the three-story frame structure.

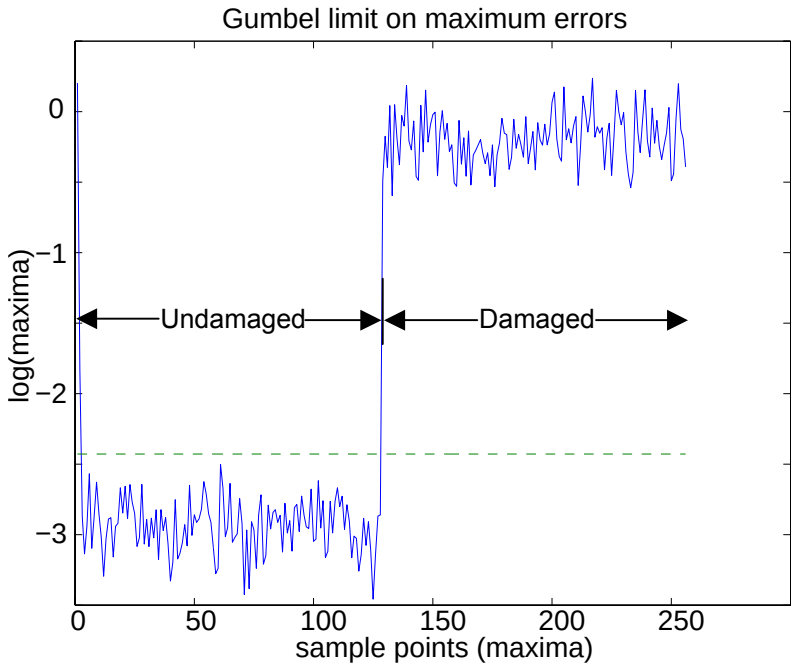


Figure 33: Windowed maxima from the AR-ARX prediction errors. The upper threshold is 99.5 percentile from the Gumbel distribution for the maxima.

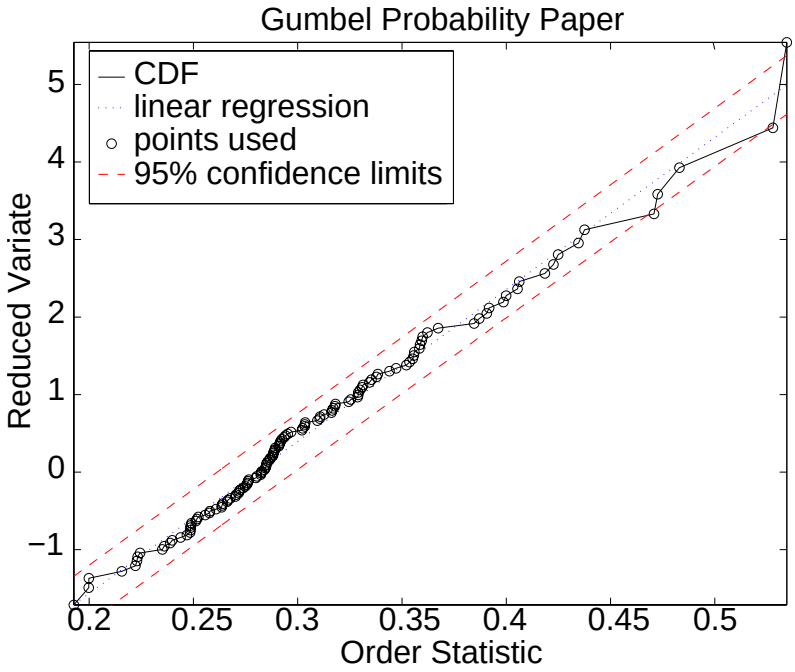


Figure 34: Undamaged experimental data from channel 21 are plotted in Gumbel coordinates to show a satisfactory Gumbel fit to the data.

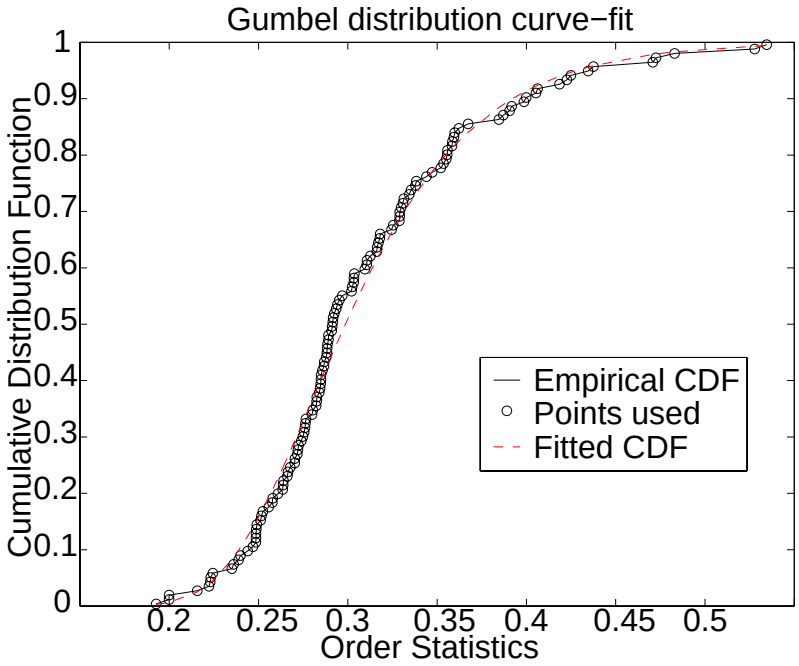


Figure 35: Gumbel distribution fit to channel 21 data.

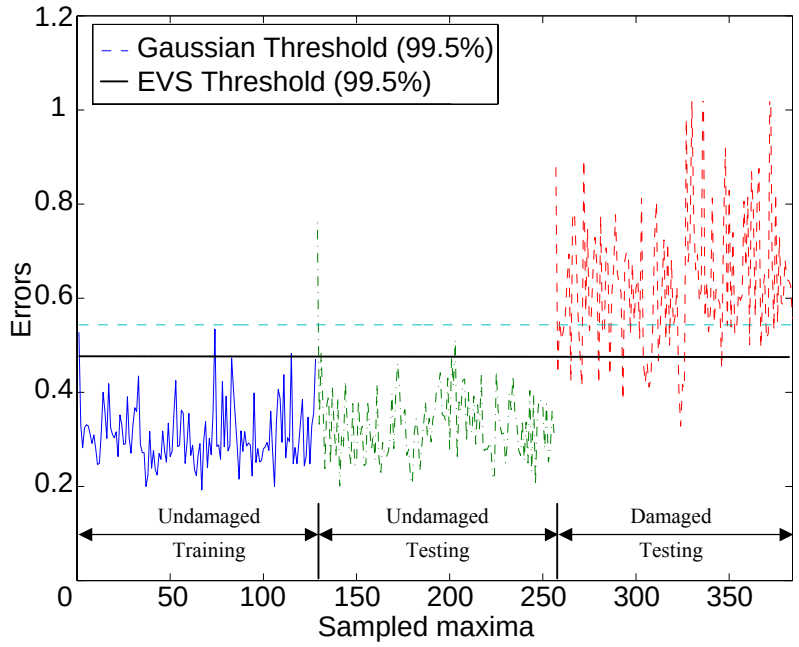


Figure 36: Undamaged and damaged system data with threshold values determined from EV analysis and Gaussian analysis.

6. Eight degree-of-freedom spring-mass system

6.1. Description of the Test Structure

The effectiveness of EVS is further demonstrated using acceleration time series recorded from an 8 DOF spring-mass system shown in Figure 37. The system is formed with eight translating masses connected by springs. Each mass is an aluminum disc that is 25.4 mm thick and 76.2 mm in diameter with a center hole. The hole is lined with a Teflon bushing. There are small steel collars on each end of the discs (Figure 38). The masses all slide on a highly polished steel rod that supports the masses and constrains them to translate only along the rod. The masses are fastened together with coil springs epoxied to the collars that are, in turn, bolted to the masses.

The DOFs, springs and masses are numbered from the right end of the system, where the excitation is applied, to the left end, as shown in Figure 37. The nominal value of mass 1 (m1) is 559.3 grams. Again, this mass is located at the right end where the shaker is attached. m1 is greater than the others because of the hardware needed to attach the shaker. All the other masses (m2 through m8) are 419.4 grams. The spring constant for all the springs is 56.7 kN/m for the initial condition. Damping in the system is caused primarily by Coulomb friction. Every effort is made to minimize the friction through careful alignment of the masses and springs. A common commercial lubricant is applied between the Teflon bushings and the support rod. The undamaged configuration of the system is the state for which all springs are identical and have a linear spring constant. Placing a bumper between two adjacent masses simulates nonlinear damage by limiting the movement of one mass relative to the other mass.

Figure 38 shows the hardware used to simulate nonlinear damage. When one end of a bumper, which is placed on one mass, hits the other mass, impact occurs. This impact simulates damage caused by the impact from the closing of a crack during vibration. Changing the amount of relative motion permitted before contact and changing the hardness of the bumpers on the impactors can control the degree of damage. For all damage cases presented, the initial clearance is set to zero. Table 8 summarizes each of the four damage cases. In damage case 3, five of the twenty-five data sets were ignored because the excitation level was low enough that the bumpers did not close the 2 mm gap to the other mass, resulting in effectively undamaged cases.

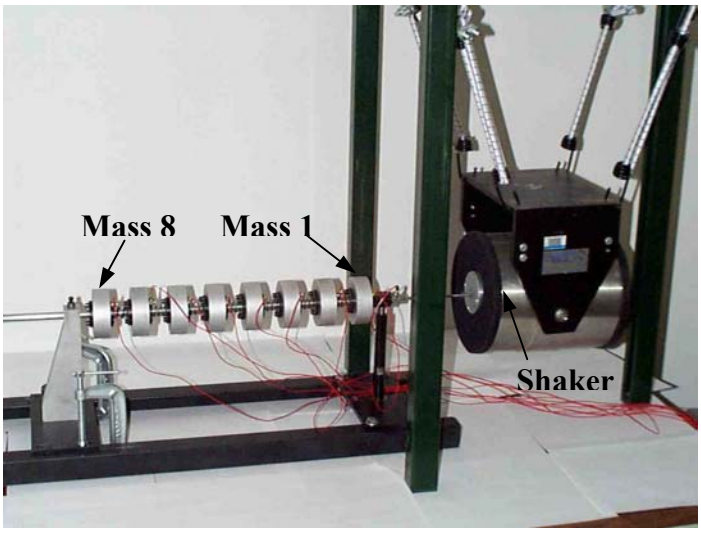


Figure 37: The 8 DOF spring-mass system attached to a shaker with accelerometers mounted on each mass.

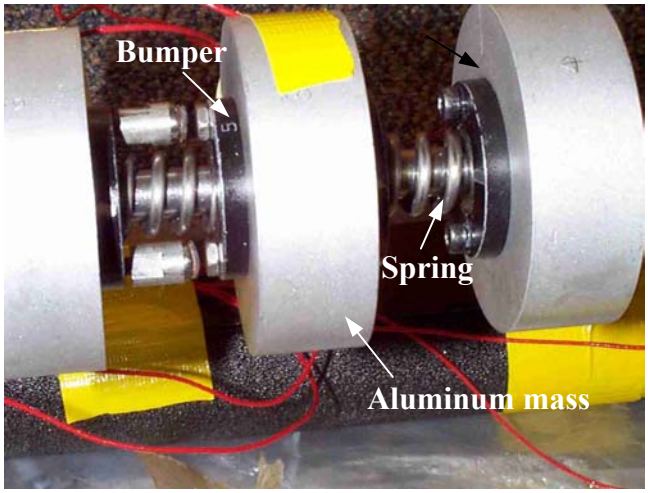


Figure 38: A typical bumper used to simulate nonlinear damage.

Table 8: List of time series employed in this study.

Case	Description	Input level	Data # per input	Total data #
0	No bumper	3, 4, 5, 6, 7 Volts	15 sets	75 sets
1	Bumper between m1-m2	3, 4, 5, 6, 7 Volts	5 sets	25 sets
2	Bumper between m5-m6	3, 4, 5, 6, 7 Volts	5 sets	25 sets
3	Bumper between m7-m8	4, 5, 6, 7 Volts	5 sets	20 sets

6.2. Damage Diagnosis Results

In this example, like the frame structure, the AR-ARX model (Appendix B; Sohn and Farrar, 2001) is first fit to an acceleration time history measured from the baseline condition of a system. If a time prediction model obtained from the baseline system is used to predict a new time signal measured under a damaged condition, the prediction errors will increase. Based on this premise, SPC is performed using the prediction errors as features. However, because the 8 DOF system is also subject to changing excitation levels, the varying input levels might result in unwanted false outliers. To overcome this difficulty, an *auto-associative neural network* is employed for data normalization. Here, *data normalization* is a procedure of "normalizing" data sets such that signal changes caused by operational and environmental variations of the system can be separated from structural changes of interest, such as structural deterioration or degradation. Detailed discussion of data normalization using the auto-associative neural network can be found in Sohn, et al. (2002).

Because there are 4096 points in each case and a 99% confidence interval is being used, one would expect that for an undamaged case there would be 21 statistically deviant points, or outliers, on each side of the distribution; 42 outliers in total. Table 9 summarizes the diagnosis results of the 8 DOF experiment.

The outliers in the undamaged data were higher than was expected, but both the normality assumption and the extreme value method yielded similar results. Several normality assessment techniques revealed that the prediction errors used as features were fairly close to normal, therefore there is no surprise that the normality assumption and EVS returned similar results in this case. Looking at m6 in the third damage case, the number of outliers is definitely above the undamaged case and would most likely show up as a false-positive indication. m1 has consistently the lowest number of outliers. This is likely because it is connected to the shaker and has less variability than the other masses in the system. In most of the damage cases the two masses between which the bumper is placed show a large increase in outliers, as expected.

Table 9: Summary of the 8 DOF system test results showing the predicted number of outliers contrasted with the normal assumption and the extreme value statistics.

Case	m1	m2	m3	m4	m5	m6	m7	m8
Undamaged	54 (42)	70 (47)	63 (62)	60 (63)	60 (62)	65 (72)	67 (75)	68 (82)
Damage 1	60 (47)	369 (381)	118 (123)	90 (93)	78 (89)	88 (63)	68 (73)	73 (93)
Damage 2	53 (43)	55 (47)	68 (67)	68 (71)	383 (444)	155 (197)	82 (99)	72 (100)
Damage 3	54 (41)	58 (50)	77 (78)	64 (71)	88 (103)	155 (181)	526 (586)	290 (331)

* Highlighted cells show locations of damage where the bumper is placed between two masses and the number of outliers are expected to increase.

- Entries in the table represent the number of outliers for 99% confidence thresholds. The first number is obtained using the EVS method. The second number in parenthesis is obtained using the normality assumption.

7. Summary

Data that lie in the tails of distributions have traditionally been modeled based on a Gaussian distribution. This inherent assumption of many statistical processes can be dangerous for applications such as statistical process control (SPC), which deal mostly with those extreme data points. Extreme value statistics (EVS) takes a closer look at modeling those extreme points independent of any measure of the Gaussian assumption.

Modeling the tails simplifies the statistics to some extent. These extreme points conform to one of three types of distributions: Gumbel, Weibull, or Frechet. There are techniques required in determining which of these distributions the data fall into and in parameter estimation for fitting the data. The SPC can then be reworked to take advantage of the proper distribution and fitted parameters.

The numeric examples demonstrate the ability of EVS when applied to simple SPC. Thresholds obtained from the actual distribution, the best-fit normal distribution and specifically modeling the extreme values are contrasted. In all of the examined cases, EVS produces results that only slightly deviate from those of the true distributions.

The new SPC, extended by incorporating EVS, was applied to a three-story frame structure with bolted connections and an 8 degree-of-freedom (DOF) spring-mass system. The frame structure was created to simulate the failure of a moment resisting connection in an earthquake. A positive indication of damage represents the identification of bolt loosening at a joint on the structure. Analysis of the frame structure data used simple minima and maxima control chart

methods to illustrate differences in threshold values calculated using Gaussian assumption or EVS. In the first pass, a single channel of acceleration data was analyzed from the undamaged and damaged cases. A window was stepped through the data, and the maximum values from each window were extracted as the features. Modeling the features using EVS yielded improved threshold values when compared to the thresholds based on the normality assumption of the data.

The 8 DOF system was created to demonstrate various algorithm robustness in detecting a nonlinear damage introduced into an otherwise linear system. The nonlinear damage is introduced into the system in the form of bumpers placed between the translating masses. Looking at the 8 DOF system, the results were much less drastic than the numeric examples. In the 8 DOF analysis, both the Gaussian and the EVS methods yielded similar results. After looking at several normality tests, it was found that the features were nearly normally distributed. This would indicate that the Gaussian and the EVS methods would return comparable results.

8. Conclusions

The results of this report show that there are advantages of using EVS to help analyze the Structural Health Monitoring (SHM) problem. Many of the SHM techniques implemented at Los Alamos National Laboratory (LANL) involve statistical decision-making based on the normality assumption of the test data. However, damage identification techniques often work with features in the extremities of a distribution that may not be accurately modeled by this Gaussian assumption. Despite the limited scope, this report shows promising results by reworking simple SPC techniques to compute confidence intervals from extreme value statistics instead of Gaussian statistics. By incorporating EVS into damage detection and location algorithms, potentially limiting and erroneous assumptions in these algorithms can be avoided. Although the results obtained in this investigation indicate that EVS only slightly changes the thresholds in the two experimental cases because of the Gaussian nature of the data, EVS seems the correct way to model these features. When structures, data recording or normalization procedures that yield non-Gaussian features need to be monitored in the future, the EVS framework will already be in place.

Appendix A: MATLAB Programs

This Appendix contains the source code for a series of programs that are useful for extreme value statistics. The programs are MATLAB translations of the original Basic source code from the book by Castillo (1988). The following programs may also be found online at www.lanl.gov/damage_id. There are minor variations in the structure to eliminate GOTO functions for example, but in general they are fair copies. The one major exception is in the program *estimate* that makes a call to a optimization function in the MATLAB optimization toolbox rather than using Castillo's original code. Examples of the use of these functions can be found in the main body of this report.

A.1. simulatePD

This function is used to generate samples of data from a given parent distribution. The basic structure is simple. A population of the required size is generated from a uniform distribution. Then, if $F(x)$ is the cumulative distribution for the required parent and y is the relevant sample from the uniform distribution, $F^{-1}(y)$ is a sample from the parent distribution. The function uses a clever trick based on conditional distributions to generate the population of specified sample size (Castillo, 1988). All of the Extreme Value (EV) distributions can be generated from a parent distribution, together with a number of other useful distributions. The main omission is the Gaussian distribution, but this omission is not an issue as the MATLAB function *randn* supplies Gaussian deviates.

```
function samples =
simulatePD(npts,distribution,parameter1,parameter2,parameter3)
% SIMULATE : generates an ordered vector of samples from a given
distribution.
%
% Function SIMULATEPD(NPTS,DISTRIBUTION,PARAMETER1,PARAMETER2,PARAMETER3).
Generates an
% ordered vector of samples, i.e. order statistics, from a specified parent
distribution
% given the appropriate parameters for the said distribution. Routine works
by
% applying a nonlinear transformation to an ordered vector of uniform
deviates.
%
% SAMPLES : is the ordered vector of samples from the population.
%
% NPTS : The number of points in the vector SAMPLE.
```

```
% DISTRIBUTION : An integer indicating the distribution type as follows:
%   (1) Gumbel (maxima)
%   (2) Gumbel (minima)
%   (3) Weibull (maxima)
%   (4) Weibull (minima)
%   (5) Frechet (maxima)
%   (6) Frechet (minima)
%   (7) Uniform
%   (8) Exponential
%   (9) Cauchy
%   (10) Raleigh
%   (11) Pareto
%   (12) p(x) = exp(-1/(x*x))
% PARAMETER1 : If only one parameter is required to specify the distribution,
this is
%    it, otherwise this is the first parameter.
%    If the distribution is Pareto, parameter1 = beta
%    " "  "  " Rayleigh, parameter1 = A
%    " "  "  " Exponential, parameter1 = lambda
%    " "  "  " Uniform, parameter1 = A
%    " "  "  " Gumbel, Weibull or Frechet, parameter1 = lambda
% PARAMETER2 : If two or more parameters are needed to specify the
distribution, this
%    is the second.
%    If the distribution is Uniform, parameter2 = B
%    " "  "  " Gumbel, Weibull or Frechet, parameter2 = delta
%    % PARAMETER3 : If distribution is Weibull or Frechet, parameter3 = beta
%
% (Note that no parameters are needed for distributions 8 or 12.)
%
% This function is a simple variation on the BASIC subroutine SIMUL, provided
in:
%
% Extreme Value Theory in Engineering
% Enrique Castillo
% Academic Press
% ISBN : 0-12-163475-2
%
% Los Alamos : 31/07/01
%

% Generate vector of ordered uniform deviates.

uni=zeros(1,npts);  % Initialise UNI with zeroed entries.

h=1;      % H and V are temporary variables in the notation of the
for i=npts:-1:1   % reference above, UNI is the vector of ordered uniform
 v=rand;    % deviates.
 h=h*v^(1/i);
 uni(i)=h;
end

% Next step is the transformation. If F(x) is the CDF of the desired
distribution,
% then F^(-1)(UNI) is a vector of samples from the desired distribution.

switch distribution
```

```
case 1
 samples = parameter1 - parameter2*log(-log(uni));
case 2
 samples = parameter1 + parameter2*log(-log(1 - uni));
case 3
 samples = parameter1 - parameter2*(-log(uni)).^(1/parameter3);
case 4
 samples = parameter1 + parameter2*(-log(1 - uni)).^(1/parameter3);
case 5
 samples = parameter1 + parameter2*(-log(uni)).^(-1/parameter3);
case 6
 samples = parameter1 - parameter2*(-log(1 - uni)).^(-1/parameter3);
case 7
 samples = parameter1 + (parameter2 - parameter1)*uni;
case 8
 samples = -log(1 - uni);
case 9
 samples = tan( PI*(uni - (1/2)) );
case 10
 samples = sqrt(-parameter1*parameter1*log(1 - uni));
case 11
 samples = (1 - uni).^(-1/parameter1);
case 12
 samples = sqrt(-1/log(uni));
end

% End of function SIMULATEPD
```

A.2. draw_samplesPD

This routine constructs a plot of the empirical CDF for a population of samples on appropriately transformed coordinates. For example, if "Gumbel maximum paper" is specified, the data is appropriately transformed so that data from a Gumbel maximum parent distribution will have a linear empirical CDF. To use the function to assign a distribution type to a population, the user can fit a linear regression to the tail of interest. The implication is that if the points of the data fall within the confidence interval for a straight line on say, Gumbel maximum paper, then the parent distribution is asymptotically Gumbel maximum. To obtain a plot for the Weibull or Frechet distributions, the user must supply an estimate of the location parameter λ.

```
function dummy = draw_samplesPD(samples,tail,paper,position,lambda,fraction)
% DRAW_SAMPLESPD : Generates an empirical CDF on appropriate probability
paper.
%
% Function DRAW_SAMPLESPD(SAMPLES,TAIL,PAPER,POSITION,LAMBDA). Plots the
empirical CDF
% of a sample of data on either Gumbel, Weibull or Frechet probability paper
i.e.
% on appropriately transformed axes.
%
% SAMPLES : is the ordered vector of samples from the population.
% TAIL : Specifies the tail of interest: (1) left tail (minima)
%          (2) right tail (maxima)
% PAPER : Specifies the type of probability paper: (1) Gumbel
%             (2) Weibull
%             (3) Frechet
% POSITION : Specifies the formula for plotting positions: (1) P(I) = I/(N+1)
%               (2) P(I) = (I-0.5)/N
%               (3) P(I) = (I-3/8)/(N+1/4)
%               (4) P(I) = (I-0.44)/(N+0.12)
% LAMBDA : threshold parameter (see book below).
% FRACTION : Fraction of data set used to fit straight line.
%
% This function is a simple variation on the BASIC subroutine DRAW, provided
in:
%
% Extreme Value Theory in Engineering
% Enrique Castillo
% Academic Press
% ISBN : 0-12-163475-2
%
% Los Alamos : 31/07/01
%

% Generate vector of plotting positions.
npts = length(samples);
y = zeros(size(samples));
```

```
for i=1:npts
 switch position
 case 1
  y(i) = i/(npts + 1);
 case 2
  y(i) = (i - 0.5)/npts;
 case 3
  y(i) = (i - 0.375)/(npts + 0.25);
 case 4
  y(i) = (i - 0.44)/(npts + 0.12);
 end
end

% Sort sample into ascending order and read into X array.
x = sort(samples);

% Establish ranges of plot.
xmin = x(1);
xmax = x(npts);
ymin = y(1);
ymax = y(npts);
if( ymax < 0.995 )
 ymax = 0.995;
end
if( ymin > 0.005 )
 ymin = 0.005;
end

% Set threshold parameter.
tiny = 1.0e-10;
if(nargin < 5)
 switch tail
 case 1
  switch paper
  case 2
   lambda = xmin - tiny; % LAMBDA must be smaller than XMIN.
  case 3
   lambda = xmax + tiny; % LAMBDA must be greater than XMAX.
  end
 case 2
  switch paper
  case 2
   lambda = xmax + tiny;
  case 3
   lambda = xmin - tiny;
  end
 end
else
 switch tail
 case 1
  switch paper
  case 2
   if( lambda > xmin )
   lambda = xmin - tiny;
   end
  case 3
   if( lambda < xmax )
```

```
   lambda = xmax + tiny;
   end
  end
 case 2
  switch paper
  case 2
   if( lambda < xmax )
   lambda = xmax + tiny;
   end
  case 3
   if( lambda > xmin )
   lambda = xmin - tiny;
   end
  end
 end
end

% Change coordinates.
for i=1:npts
 [r,q] = change_coordinates(x(i),y(i),tail,paper,lambda);
 x(i) = r;
 y(i) = q;
end

[r,q] = change_coordinates(xmax,ymax,tail,paper,lambda);
xmax = r;
ymax = q;

[r,q] = change_coordinates(xmin,ymin,tail,paper,lambda);
xmin = r;
ymin = q;

% If we have specified FRACTION, then we want a straight line fit.
if( nargin < 6 )
 plot(x,y);
else

% Get range for curve-fit.
 nfit = fix(fraction*npts);
 switch tail
 case 1
  nl = 1;
  nh = nfit;
 case 2
  nl = npts - nfit + 1;
  nh = npts;
 end

 % LS Polynomial (linear here) fit.
 [coeff,error_struct] = polyfit(x(nl:nh),y(nl:nh),1);

 % Reconstruct data from line.
 [predict_y, delta] = polyval(coeff, x(nl:nh), error_struct);

 % DELTA has the 50% confidence interval, convert to 95% and plot bounds.
 upper_y = predict_y + (1.96/0.6745)*delta;
 lower_y = predict_y - (1.96/0.6745)*delta;
```

```
plot(x,y,'k',x(nl:nh),predict_y,'b',x(nl:nh),y(nl:nh),'ko',x(nl:nh),lower_y,'
r',x(nl:nh),upper_y,'r');
 legend('CFD','linear regression','points used','50% confidence limits',0)
end

% Embellish graph
axis([xmin,xmax,ymin,ymax]);
ylabel('Reduced Variate');
switch paper
case 1
 xlabel('Order Statistic');
 title('Gumbel Probability Paper');
case 2
 xlabel('Reduced Order Statistic');
 title('Weibull Probability Paper');
case 3
 xlabel('Reduced Order Statistic');
 title('Frechet Probability Paper');
end

% End of function DRAW_SAMPLESPD
function [r,q] = change_coordinates(x,y,tail,paper,lambda)

switch tail
case 1
 q = log(-log(1.0 - y));
 switch paper
 case 1
  r = x;
 case 2
  r = log(x - lambda);
 case 3
  r = -log(lambda - x);
 end
case 2
 q = -log(-log(y));
 switch paper
 case 1
  r = x;
 case 2
  r = -log(lambda - x);
 case 3
  r = log(x - lambda);
 end
end

% End of function CHANGE_COORDINATES.
```

A.3. selectionEV

This function is used to determine the appropriate extreme value distribution for a population from a given parent. The reader is referred to Castillo (1988) for the details, but there are two methods allowed here. First is the curvature method based on plotting the empirical CDF on Gumbel paper. If the population distribution is Gumbel, the CDF will appear linear. Maxima (minima) data from a distribution asymptotic to the Weibull distribution will show convexity (concavity) in the right (left) tail. This will be reflected in an estimated curvature s greater than unity. For maxima (minima) data from a Frechet asymptotic distribution, the plot will show concavity (convexity) in the right (left) tail. This concavity or convexity will be reflected in a curvature less than unity. The second method here is based on the approach described in Pickands (1975) and Castillo (1988). The reader is referred to these references for details. This function is actually of limited use at the moment, as it requires the estimation of confidence intervals to establish if the curvature is *significantly* greater than unity. A function for this will be provided later.

```
function [s,a,c] = selectionEV(samples,tail,method)
% selectionEV : Determines the limit (EV) distribution for a given parent
%
% Function SELECTIONEV(SAMPLES,TAIL,METHOD).b Determines the domain of
attraction of a
% parent distribution from a sample. Two methods are used (see below).
%
% S : Curvature estimate from curvature method.
% A, C: Parameters for distribution G from Pickands III method (see book
below).
%
% SAMPLES : is the ordered vector of samples from the population.
% TAIL : Specifies the tail of interest: (1) left tail (minima)
%          (2) right tail (maxima)
% METHOD : Specifies the algorithm: (1) Curvature
%         (2) Pickands III
%             %
% This function is a simple variation on the BASIC subroutine SELEC, provided
in:
%
% Extreme Value Theory in Engineering
% Enrique Castillo
% Academic Press
% ISBN : 0-12-163475-2
%
% Los Alamos : 31/07/01
%

% Generate required storage and order the sample vector.
```

```
npts = length(samples);
x = zeros(npts,1);
sort(samples);
x = samples;
y = zeros(npts,1);

switch method
case 1 % Curvature method.

 % Convert to Gumbel probability paper coordinates.

 for i=1:npts
  y(i) = (i - 0.5)/npts; % Basic plotting positions.
  y(i) = -log(-log(y(i))); % Gumbel transformation.
 end

 % if the left tail is of interest, invert the X array and change sign.

 if( tail < 2 )
  x = invert_X(x);
 end

 % Define the two ranges for fitting straight lines. Then fit then using a
simple
 % least-squares procedure.

 n11 = fix(2.0*sqrt(npts));
 n22 = fix(n11/2);

 n1 = npts - n11 + 1;
 n2 = npts - n22 + 1;
 s1 = straight_line(x,y,n1,n2); % gradient of lower segment.

 n1 = n2;
 n2 = npts;
 s2 = straight_line(x,y,n1,n2); % gradient of upper segment.

 s = s1/s2; % curvature.
 a = 0.0; % zero the parameters for Pickands III.
 c = 0.0;

case 2 % Pickands III method.

 % Invert the order of the X array. Keep the original in Y.

 for i=1:npts
  y(i) = x(i);
 end
 for i=1:npts
  x(i) = y(npts - i + 1);
 end

 % if the left tail is of interest, invert the X array and change sign.

 if( tail < 2 )
  x = invert_X(x);
```

```
 end

 % Main body of algorithm.

 tiny = 1.0e-30;

 dm = 1.0e30; % Set minimum distance high.

 n4 = fix(npts/4);

 for l=1:n4
  nt = 4*l;
  dl = 0.0; % Original basic had a ! here, not sure what it means, assuming
it
    % means 'integer'. Anyway, set maximum distance low.

  z24 = x(2*l) - x(4*l); % Find initial estimates of a and c.
  z12 = x(l) - x(2*l);
  c = log10(z12/z24 + tiny)/log10(2.0); % Offset traps underflows of log.
  a = c*z24/(2.0^c - 1);

  for i=1:nt % Loop over the values less than NT to find maximum differences
     % between empirical Fl and Gl.
   aux = 1 + c*(x(i) - x(nt))/a;

   if( aux >= 0 ) % This is an allowed value of AUX, so carry on.

   gl = 1.0 - aux^(-1.0/c); % Empirical distributions.
   fl = (nt - i + 1.0)/nt;

   e = abs(fl - gl); % Error between distributions.

   if( e > dl )
    dl = e; % Store if error is largest so far.
   end

   fl = (nt - i)/nt; % Try different plotting position for Fs.
   e = abs(fl -gl);
   if( e > dl )
    dl = e; % Store if error is largest so far.
   end

   end % End of if( aux > 0 ).
  end % End of loop over I.

  if( dl < dm )
   dm = dl;
   m = l;
   am = a;
   cm = c;
  end
 end % End of loop over L.

 s = 0.0; % Zero the curvature.
 a = am; % Parameters for Pickands III.
 c = cm;
```

```
end % Of case statement.

% End of function SELECTIONEV.

function s = straight_line(x,y,n1,n2) % Bog standard LS gradient estimate.

a1 = 0.0;
a2 = 0.0;
a01 = 0.0;
a11 = 0.0;

for i=n1:n2
 a1 = a1 + y(i);
 a2 = a2 + y(i)*y(i);
 a01 = a01 + x(i);
 a11 = a11 + x(i)*y(i);
end

n8 = n2 - n1 + 1; % Don't ask me - notation of the book above.
s = (n8*a11 - a1*a01)/(n8*a2-a1*a1);

return;

% End of function STRAIGHT_LINE.

function y = invert_X(x)

npts = length(x);
y = zeros(npts,1);
z = zeros(npts,1);

for i=1:npts
 z(i) = -x(i);
end
for i=1:npts
 x(i) = z(npts - i + 1);
end

y = x;

return;

% End of function INVERT_X.
```

A.4. estimate

This function fits a parametric model to the empirical CDF from a sample population. The models allowed are the standard extreme value limit distributions for maxima and minima. For the Gumbel distributions, the function returns the location λ and scale δ parameters. For the Weibull and Frechet distributions, the function returns the scale δ and shape β parameters. However, an *a priori* estimate of the location parameter λ is required for the Weibull and Frechet distributions. Various methods of curve fitting are allowed. However, the recommended option is Maximum Likelihood (ML) (option 5 for the variable *method*). Note that the ML method appears to be more susceptible to local minima. A safe option is Basic Least Squares (LS) (option 1 for *method*). This routine requires the presence of the MATLAB function *fminunc* from the optimization toolbox.

```
function [ lambda, delta, beta] =
estimate(samples,tail,distribution,method,position,nl,nh,lambda,weights,plot_
flag)
% ESTIMATE : Estimates parameters from sample data for a given EV
distribution.
%
% Function ESTIMATE(SAMPLES,TAIL,DISTRIBUTION,METHOD,POSITION,NL,NH,LAMBDA).
Given
% samples of data from a parent distribution, fits parameters for one of the
three
% Extreme Value distributions. If the distribution is Weibull or Frechet, an
estimate
% of the location parameter is needed.
%
% SAMPLES : is the ordered vector of samples from the population.
% TAIL : Specifies the tail of interest: (1) left tail (minima)
%          (2) right tail (maxima)
% DISTRIBUTION : Specifies the type of EV distribution: (1) Gumbel
%              (2) Weibull
%              (3) Frechet
% METHOD : Specifies curve-fitting method: (1) Least-squares probability
absolute error
%           (2) Least-squares return period relative error
%           (3) Standard weighted least-squares
%           (4) Least-squares error with given weights
%           (5) Maximum likelihood
%           (6) Percentile method
%           (7) Method of moments (Gumbel)
% POSITION : Specifies the formula for plotting positions: (1) P(I) = I/(N+1)
%               (2) P(I) = (I-0.5)/N
%               (3) P(I) = (I-3/8)/(N+1/4)
```

```
%               (4) P(I) = (I-0.44)/(N+0.12)
% NL : Lowest order statistic to be used for fitting.
% NH : Highest order statistic to be used for fitting.
% LAMBDA : Estimate of location parameter.
% WEIGHTS : Weights for the weighted least-squares methods.
% PLOT_FLAG : Decides if we plot CDF.
%
% This function is a variation on the BASIC subroutine ESTIM, provided in:
%
% Extreme Value Theory in Engineering
% Enrique Castillo
% Academic Press
% ISBN : 0-12-163475-2
%
% Los Alamos : 09/08/01
%

% Save range parameters for plot in case they get changed (right tail).

nlp = nl;
nhp = nh;

if( nargin < 10 )
 plot_flag = 0;
end

% Sort data into increasing order, place in array X.

npts = length(samples);
sort(samples);
x = zeros(npts,1);
x = samples;

% If right tail is of interest, data and threshold parameters change sign.

if( tail == 2)
 temp = zeros(size(x));
 for i = 1:npts
  temp(i) = -x(i);
 end
 for i = 1:npts
  x(i) = temp(npts - i + 1);
 end
 temp = nl;
 nl = npts - nh + 1;
 nh = npts - temp + 1;
 lambda = -lambda;
end

% If necessary transform Weibull and Frechet data to Gumbel data.

switch distribution
case 2
 for i=nl:nh
  if( lambda > x(i) )
   error(' ! Location parameter (Weibull) incorrect ');
  end
```

```
  x(i) = log( x(i) - lambda );
 end
case 3
 for i=nl:nh
  if( lambda <= x(i) )
   error(' ! Location parameter (Frechet) incorrect ');
  end
  x(i) = - log( lambda - x(i) );
 end
end

% Now the data is Gumbel, normalise (standardise).

xbar = mean(x(nl:nh));
sigma = std(x(nl:nh));
x = (x - xbar)/sigma;

% At this point, we appear to need to fix P0, which are - for Gumbel, the
location
% and scale parameters, and for the other two distributions, the scale and
shape
% parameters.

if( method < 7 )      % For any method other than moments, get
 for i=1:2       % initial estimates of parameters, transform
  switch i       % to Gumbel paper and we get a straight
  case 1       % line. Split the range in half and
   ns1 = nl;       % sum over the data points in each half
   ns2 = fix( (nl + nh)/2 );  % This gives two equations, which can be
  case 2       % solved for the gradient and intercept
   ns1 = fix( (nl + nh)/2 + 1 ); % of the line. Why not use least-squares?
   ns2 = nh;
  end
  s1 = 0.0;
  s2 = 0.0;
  for k=ns1:ns2
   temp = 1 - (k - 0.5)/npts; % Plotting position 2 is used.
   s1 = s1 + log( -log(temp) );
   s2 = s2 + x(k);
  end
  switch i
  case 1
   a11 = ns2 - ns1 + 1;
   a12 = s1;
   c1 = s2;
  case 2
   a21 = ns2 - ns1 + 1;
   a22 = s1;
   c2 = s2;
  end
 end
 de = a11*a22 - a21*a12;
 p0(1) = (c1*a22 - c2*a12)/de;
 p0(2) = (c2*a11 - c1*a21)/de;
else
 p0(2) = sqrt(6.0)/pi; % For method of moments, just use the formulae from
 p0(1) = 0.5772*p0(2); % the book - Table 5.1.
```

```
 % All we need to do now is put the scale back in the data and transform
back.

 [lambda, delta, beta] =
transform_back(p0,xbar,sigma,tail,distribution,lambda);
end

% If we are interested in the percentile method, we have enough now.

if( method == 6 )
 [lambda, delta, beta] =
transform_back(p0,xbar,sigma,tail,distribution,lambda);
end

% Keep on with methods 1 to 5. Generate vector of plotting positions.

if( method < 6 )
 p = zeros(npts,1);

 for i=1:npts
  switch position
  case 1
   p(i) = i/(npts + 1);
  case 2
   p(i) = (i - 0.5)/npts;
  case 3
   p(i) = (i - 0.375)/(npts + 0.25);
  case 4
   p(i) = (i - 0.44)/(npts + 0.12);
  end
 end

 % If weights are needed, generate them.

 if( method < 4 )
  weights = zeros(npts,1);
  for i=nl:nh
   switch method
   case 1
   weights(i) = 1.0;
   case 2
   weights(i) = 1/(p(i)*p(i));
   case 3
   weights(i) = 1.0/(p(i)*(1.0 - p(i)));
   end
  end
 end

 % Optimisation step.

 options = optimset('LargeScale','off');

 popt = zeros(size(p0));

 popt = fminunc('evaluate',p0,options,npts,x,p,method,nl,nh,weights);
```

```
 [lambda, delta, beta] =
transform_back(popt,xbar,sigma,tail,distribution,lambda);
end

% If plotting is required, do it.

if( plot_flag ~= 0 )

 if( method > 5 )
  for i=1:npts
   switch position
   case 1
   p(i) = i/(npts + 1);
   case 2
   p(i) = (i - 0.5)/npts;
   case 3
   p(i) = (i - 0.375)/(npts + 0.25);
   case 4
   p(i) = (i - 0.44)/(npts + 0.12);
   end
  end
 end

 predict_y = zeros(size(x));
 switch tail
 case 1
  switch distribution
  case 1
   predict_y = 1.0 - exp( - exp( (samples - lambda)/delta ));
  case 2
   for i=1:npts
   if( samples(i) >= lambda )
    predict_y(i) = 1.0 - exp( - ((samples(i) - lambda)/delta)^beta );
   else
    predict_y(i) = 0.0;
   end
   end
  case 3
   for i=1:npts
   if( samples(i) <= lambda )
    predict_y(i) = 1.0 - exp( - (delta/(lambda - samples(i)))^beta );
   else
    predict_y(i) = 1.0;
   end
   end
  end
 case 2
  switch distribution
  case 1
   predict_y = exp(-exp( -(samples - lambda)/delta) );
  case 2
   for i=1:npts
   if( samples(i) <= lambda )
    predict_y(i) = exp( -((lambda - samples(i))/delta)^beta );
   else
    predict_y(i) = 1.0;
   end
```

```
   end
  case 3
   for i=1:npts
   if( samples(i) >= lambda )
    predict_y(i) = exp( -(delta/(samples(i) - lambda))^beta );
   else
    predict_y(i) = 0.0;
   end
   end
  end
 end

 plot(samples,p,'k',samples(nlp:nhp),p(nlp:nhp),'ko',samples,predict_y,'r');
 xlabel('Order Statistics');
 ylabel('Cumulative Distribution Function');
 switch distribution
 case 1
  title('Gumbel distribution curve-fit');
 case 2
  title('Weibull distribution curve-fit');
 case 3
  title('Frechet distribution curve-fit');
 end
 legend('Empirical CDF','Points used','Fitted CDF',0);
end

return;

% End of function ESTIMATE.


function [lambda, delta, beta] =
transform_back(p0,xbar,sigma,tail,distribution,lambda)

p0(1) = xbar + sigma*p0(1);
if( tail == 2 )
 lambda = -lambda;
 p0(1) = -p0(1);
end
p0(2) = sigma*p0(2);

switch distribution
 case 1
  lambda = p0(1);
  delta = p0(2);
  beta = 0.0;
 case 2
  if( tail == 1 )    % This step is necessary or appears to necessary
   p0(1) = exp( p0(1) );  % because of the transformation to Gumbel
  else       % coordinates. The effect appears to be different
   p0(1) = exp( - p0(1) ); % for left and right tails. Later on, I'll get my
  end       % head round this and write a better comment.
  p0(2) = 1/p0(2);

  delta = p0(1);
  beta = p0(2);
 case 3
```

```
  if( tail == 1 )    % Likewise here.
   p0(1) = exp( - p0(1) );
  else
   p0(1) = exp( p0(1) );
  end
  p0(2) = 1/p0(2);

  delta = p0(1);
  beta = p0(2);
end

return;
% End of function TRANSFORM BACK.
```

A.5. evaluate

This function is called by *estimate*. Its only purpose is to provide values for the objective function for the optimization in *estimate*. It is not intended for independent use.

```
function fp = evaluate(par,npts,x,p,method,nl,nh,weights)

if( par(2) <= 0 )
 fp = -1.7e30;
 return;
end

fp = 0.0;

if( method < 5 )
 for i=nl:nh
  p1 = 1.0 - exp( - exp( (x(i) - par(1))/par(2) ));
  fp = fp + ( p(i) - p1 )*( p(i) - p1)*weights(i);
 end
else
 xl = x(nl);
 xh = x(nh);
 for i=nl:nh
  p1 = (par(1) - x(i))/par(2);
  fp = fp + p1 + exp(-p1);
 end
 p1 = 1.0 - exp(- exp( (xl - par(1))/par(2) ));
 if( p1 == 0.0 )
  fp = -1.7e30;
  return;
 end
 p2 = log( par(2) );
 p3 = exp( (xh - par(1))/par(2) );
 fp = - fp + (nl - nh - 1)*p2 - (npts - nh)*p3;
 fp = fp + (nl - 1)*log(p1);
 fp = - fp; % We want to minimise - the likelihood.
end

return;

% End of function EVALUATE.
```

A.6. extract_max

This is a simple function that moves a window of a given size over a signal from the parent distribution and extracts the maximum for each window. A trivial modification would allow extraction of minima.

```
function max_array = extract_max(data,nwin)

npts=length(data);

nstep=npts/nwin;

max_array=zeros([1,nstep]);

for i=1:nstep
 max_array(i) = max(data( (i-1)*nwin+1 : i*nwin ));
end

return;
```

Appendix B: AR-ARX Time History Analysis

For Structural Health Monitoring (SHM) strategies that rely on vibration response measurements, the ability to normalize the measured data with respect to varying operational and environmental conditions is essential if one is to avoid false-positive indication of damage. Examples of common normalization procedure include normalizing the response measurements by the measured inputs as is commonly done when extracting modal parameters. When environmental cycles influence the measured data, a temporal normalization scheme may be employed. These strategies for SHM data normalization fall into two general classes: (1). Those employed when measures of the varying environmental and operational parameters are available. (2). Those employed when measures of the varying environmental and operational conditions are not available. This Appendix B focus on addressing the latter case.

The data normalization procedure begins by assuming that a "pool" of signals is acquired from various unknown operational and environmental conditions, but from a known structural condition of the system. The ability of this procedure to normalize the data will be directly dependent on this pool being representative of data measured in as many varying environmental and operational conditions as possible. In this report, multiple time series are recorded from the undamaged test structure at different input force levels and input frequencies. The collection of these time series is called "the reference database."

A two-stage prediction model, combining Auto-Regressive (AR) model and Auto-Regressive model with exogenous inputs (ARX), 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:

$$\hat{x} = \frac{x - \mu_x}{\sigma_x} \qquad \text{B-1}$$

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 *p* auto-regressive terms is constructed. An AR(*p*) model can be written as:

$$x(t) = \sum_{j=1}^{p} \phi_{xj} \; x(t-j) \; + e_x(t) \qquad \text{B-2}$$

This step is repeated for all signals in the reference database. The AR order is set based on a partial auto-correlation analysis.

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)*:

$$y(t) = \sum_{j=1}^{p} \phi_{yj} \; y(t-j) \; + e_y(t) \qquad \text{B-3}$$

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:

$$\text{Difference} = \sum_{j=1}^{p} (\phi_{xj} - \phi_{yj})^2 \qquad \text{B-4}$$

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 B-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 B-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 $x(t)$

$$x(t) = \sum_{i=1}^{a} \alpha_i \; x(t-i) \; + \sum_{j=1}^{b} \beta_j \; e_x(t-j) + \varepsilon_x(t) \qquad \text{B-5}$$

where $\varepsilon_x(t)$ is the residual error after fitting the ARX(a,b) 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. It is suggested to keep the sum of *a* and *b* smaller than p ($a + b \le p$). Although the *a* and *b* values of the ARX model are set rather arbitrarily, similar results are obtained for different combinations of *a* and *b* values as long as the sum of *a* and *b* is kept smaller than *p*.

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

$$\varepsilon_y(t) = y(t) \; - \sum_{i=1}^{a} \alpha_i \; y(t-i) \; - \sum_{j=0}^{b} \beta_j \; e_y(t-j) \qquad \text{B-6}$$

where $e_y(t)$ is considered to be an approximation of the system input estimated from B-3. Again, note that the α_i and β_j coefficients are associated with $x(t)$ and obtained from Equation B-5. Therefore, 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 $\varepsilon_x(t)$.

References

1. Benjamin, J.R. and Cornell, C.A., *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, Inc., New York, NY, 1970.

2. Castillo, E., *Extreme Value Theory in Engineering*, Academic Press Series in Statistical Modeling and Decision Science, San Diego, CA, 1988.

3. Embrechts, P., Kluppelberg, C., and Mikosch, T., *Modeling Extremal Events*, Springer-Verlag, New York, NY, 1997.

4. Fisher, R.A. and Tippett, L.H.C., "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, 1928.

5. Galambos, J., *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York, NY, 1978.

6. Ghosh, B.K., *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Menlo Park, CA, 1970.

7. Gumbel, E.J., *Statistics of Extremes*, Columbia University Press, New York, NY, 1958.

8. Kotz, S. and Nadarajah, S., *Extreme Value Distributions*. Theory and Applications. Imperial College Press, England, 2000.

9. Kramer, S.L., *Geotechnical Earthquake Engineering*, Prentice Hall, Upper Saddle River, NJ, 1996.

10. MATLAB, *The Language of Technical Computing*, The Math Works Inc., Natick, MA, 1998.

11. Pickands, III J., *Statistical inference using extreme order statistics*, Annals of Statistics, **3,** pp.119-131, 1975.

12. Reiss, R.D. and Thomas, M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag AG, Switzerland, 2001.

13. Roberts, S., "Novelty detection using extreme value statistics," *IEE Proceedings in Vision, Image and Signal Processing*, **146,** pp.124-129, 1998.

14. Roberts, S., "Extreme value statistics for novelty detection in biomedical signal processing," *IEE Proceedings in Science, Technology and Measurement*, **147,** pp.363-367, 2000.

15. Sohn, H. and Farrar, C.R., "Damage diagnosis using time series analysis of vibration signals," *Smart Materials and Structures*, **10,** pp.446-451, 2001.

16. Sohn, H., Worden K., Farrar C.R., "Statistical Damage Classification under Changing Environmental and Operational Conditions," *accepted for publication by the Journal of Intelligent Materials Systems and Structures*, 2002.

17. Takahira, S. and Mita, A., "Damage Index Sensors for Structural Health Monitoring," *The Second International Conference on Advances in Structural Engineering and Mechanics*, Busan, South Korea, August 21-13, 2002.

18. Wirsching, H., Paez, T. L., and Ortiz, K., *Random Vibrations Theory and Practice*, John Wiley, New York, NY, 1995.

19. Worden, K., Manson, G. and Fieller, N.J., "Damage detection using outlier analysis," *Journal of Sound and Vibration*, **229,** pp.647-667, 2000.

Distribution List

Masato Abe
University of Tokyo
Dept. of Civil Engineering
Hongo 7-3-1, Bunkyo-ku
Tokyo 113-8656, Japan

Emin Aktan
Drexel University
3201 Arch Street, Suite 240
Philadelphia, PA 19104

Sreenivas Alampalli
New York Dept. of Transportation
1220 Washington Ave.
Albany, NY 12232

Ken Alvin
Sandia National Laboratories
M/S 0439
Albuquerque, NM 87185-5800

Grahem Archer
School of Civil Engineering
1284 Civil Engineering Building
Purdue University
West Lafayette, IN 47907-1284

Alejandro Asfura
EQE, Inc.
44 Montgomery St., Suite 3200
San Francisco, CA 94104
Bill Baker
Los Alamos National Laboratory
ESA-EA, MS P946
P.O. Box 1663
Los Alamos, NM 87545

Doo Byong Bae
Dept. of Civil and Environmental Eng.
Kookmin University
861-1 Chongnung-dong Songbuk-gu
Seoul,136-702, Korea

Luciana Barroso
CE/TTI Building, Room 705-L
Department of Civil Engineering
Texas A&M University
College Station, TX 77845

Janice Barton
University of South Hampton
Dept. of Ship Science
South Hampton SO17 1BJ
UK

Jim Beck
Applied Mechanics and Civil
Engineering
Caltech 104-44
Pasadena, CA 91125

Mrinmay Biswas
Duke University
Dept. of Civil Engineering
Durham, NC 27706

Christian Boller
Daimler Chrysler Aerospace
Munich, D-81663
Germany

James Brownjohn
Nanyang Technological University
School of Civil and Structural
Engineering
Nanyang Avenue
Singapore 639798.

Rune Brinker
Department of Building Technology and
Structural Engineering
Aalborg University
Sohngaardsholmsvej 57, DK-9000
Aalborg, Denmark

Bob Burick
Granite Construction Co.
Heavy Construction Division
Box 50024
Watsonville, CA 95077-5024

Thomas Burton
Dept. of Mech. Engineering
Texas Tech University
Lubbock, TX, 79409-1021

R. Cantieni
Uberlandstrasse 129
Dubendorf, CH-8600 Switzerland

Eric Canuteson
Pacific Microinstruments
509 W. Foothill Blvd.
Monrovia, CA 91016

Tom Carne
Dept. 2741
Sandia National Laboratory
Albuquerque, NM 87185-5800

Peter Cawley
Dept. of Mechanical Engineering
Imperial College
Exhibition Rd
London, SW7 2BX
UK

Sung-Pil Chang
Dept. of Civil Engineering
Seoul National University
Shilimdong, Kwanaku,
Seoul, Korea 151-742

Chih Chen Chang
Department of Civil Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon
Hong Kong

Steve Chase
Federal Highway Administration
6300 Georgetown Pike
McLean, VA 22101-2296

Eu Kyeung Cho
Hyundai Engineering & Construction Co.
San1-1, Mabuk-Ri, Goosung-Myun,Yongin-Si
Kyunggi-Do, Korea, 449-910

Fu-Kuo Chang
Dept. of Aeronautics and Astronautics
Stanford University
Stanford, CA 94305

Tse-Yung Chang
Hong Kong University of Science and Technology
Civil Engineering
Clear Water Bay
Kowloon, Hong Kong

Weiling Chiang
President's Office
National Central University
Chungli, Taiwan

Franklin Cheng
University of Missouri Rolla
Dept. of Civil Eng.
Rolla, MO 65401

Chang Keun Choi
Dep. of Civil Engineering
Korean Institute of Advanced Science and Technology
373-1, Kusong-dong, Yusong-gu,
Taejon ,KOREA, 305-701

Ken Chong
National Science Foundation
4201 Wilson Blvd., Rm. 545
Arlington, VA 22230

Anil Chopra
Earthquake Engineering Research Center
University of California
1301 South 46th St.
Richmond, CA 94804

Lowell Cogburn
Association of American Railroads
P.O. Box 11130
Pueblo, CO 81001

Jerry Conner
Room 1-290
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Joel Conte
Department of Structural Engineering
JSOE School of Engineering
University of California, San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0085

Roy Craig, Jr.
University of Texas at Austin
Aerospace Engineering and Engineering Mechanics Dept., Mail Code C0600
Austin, TX 78712Ð1085

Barry Davidson
Compusoft Engineering Ltd.
PO Box 9493
Newmarket, Auckland
New Zealand

Shirley Dyke
Washington University
One Brookings Hall 1130
St. Louis, MO 63130

Dave Ewins
Imperial College
Mechanical Engineering Dept.
Exhibition Road
London SW7 2BX, UK

Gregory Fenves
Department of Civil and Environmental Engineering, MC 1710
University of California
Berkeley,CA 94720-1710

M. Ferner
Anlauf Ingenicur - Consulting GMBH
Posttach 101259
D-69002 Heidelberg, Germany

Mike Friswell
Dept. of Mechanical Engineering
University of Wales, Swansea
Singleton Park,
Swansea, SA2 8PP
UK

Claus-Peter Fritzen
Institute of Mechanics and Automatic Control
University of Siegen
Paul-Bonatz-Str. 9-11
D-57068 Siegen, Germany

Gongkang F
New York State Dept. of Transportation
1220 Washington Ave.
Albany, NY 12232

Chris Gannon
Penguin Engineering Ltd
PO Box 33 093
Petone, New Zealand

Ephrahim Garcia
DARPA
Dense Science Office
3701 N. Fairfax Dr.
Arlington, VA 22203-1714

Luigi Garibaldi
Dipartimento di Meccanica
Politecnico di Torino
Corso Duca degli Abruzzi, 24
10129 Torino, Italy

Lothar Gaul
Institut A fur Mechanik
Universitat Stuttgart
Pfaffenwaldring 9
70550 Stuttgart, Germany

Michael Grygier
NASA Johnson Space Center
ES43
Houston, TX 77058

Alfredo Guemes
UPM
ETSI Aeronautics
Madrid, 28016
Spain

Joe Hammond
Inst. of Sound and Vibration Research
University of South Hampton
SO17 1BJ
South Hampton, UK

Kurt Hansen
Dept. of Energy Engineering
Technical University of Denmark
Building 404, DTU
DK 2800 Lyngby
Denmark

Nicholas Haritos
Dept. of Civil and Environment Eng.
University of Melbourne
Parkville, Victoria 3052
Australia

Phil Hashimoto
300 Commerce Drive
Suite 200
Irvine, CA 92602

Dan Inman
Virginia Polytechnic Institute of State
Univ.310 New Engineering Building
Mail code 0261
Blacksburg, VA 24061-0219

Jeong Hwan Jang
EJTECH CO. LTD.
5th Floor, Seoweon Building, 2-44
YangJae-Dong, Seocho-Gu, Seoul,
Korea

Jim Johnson
EQE, Inc.
44 Montgomery St., Suite 3200
San Francisco, CA 94104

Tarsem Jutla
Caterpillar, Inc.
Technical Center, Bldg. K
P.O. Box 1875
Peoria, Illinois 61656-1875

Daniel Kammer
University of Wisconsin-Madison
Dept. of Engr. Mech. and Astronautics
3352 Engr. Hall, 1415 Johnson Drive
Madison, WI 53706

Tom Kashanganki
University of Maryland
SMART Materials and Structures Res.
Center
College Park, MD 20742
Robert Kennedy
18971 Villa Terrace
Yorba Linda, CA 92686

Klaus Kerkhoff
Staatliche Materialprufungsanstalt
Universitat Stuttgart
D-70569 Stuttgart (Vailhingen)
Germany

Magdi Khalifa
Civil Engineering Department
University of Nebraska Lincoln
W348 Nebraska Hall
Lincoln, NE 68588-0531

Chul Young Kim
Dept. of Civil & Environmental Engineering
Myong Ji University
San 38-2, Nam-dong, Yongin-si
Kyunggi-do, 449-728, Korea

Hyoung-Man Kim
McDonnell Douglas Aerospace
M/S: MDC-2-3353
13100 Space Center Blvd
Houston, TX 77059

Jae Kwan Kim
Dept. of Civil Engineering
Seoul National University
Shilimdong, Kwanaku,
Seoul, Korea 151-742

Nam Sik Kim
Hyundai Eng. & Construction Co. Ltd.
San1-1, Mabuk-Ri, Goosung-Myun,Yongin-Si
Kyunggi-Do, Korea, 449-910

Anne Kiremidjian
Department of Engineering
Stanford University
Terman Engineering Center 238
Stanford, CA 94305-4020

Poul Kirkegaard
Department of Building Technology and Structural Engineering
Aalborg University
Sohngaardsholmsvej 57, DK-9000
Aalborg, Denmark

Hyun Moo Koh
Dept. of Civil Engineering
Seoul National University
Shilimdong, Kwanaku,
Seoul, Korea 151-742

Richard Kohoutek
Dept. of Civil and Mining Engineering
University of Wollongong
Northfields Avenue
Wollongong 2522 Australia

John Kosmatka
University of California San Diego
Department of AMES
San Diego, CA 92093-0085

Christian Kot
Argonne National Laboratory
9700 South Cass Ave. RE/331
Argonne, IL 60439-4817

Everett Kuo
Ford Research Laboratory
P.O. Box 2053/MD2122
Dearborn, MI 48121

Charles Larson
Boeing
MS.H013-C326
5301 Bolsa Avenue
Huntington Beach, CA 92649

Kincho Law
Dept. of Civil Engineering
Stanford University
Stanford, CA 94305-4020

George Lee
Multidisciplinary Center for Earthquake Engineering Research
SUNY at Buffalo
Red Jacket Quadrangle
Buffalo, NY 14261-0025

H. S. Lew
NIST
Bldg. 226, Rm B168
Gaithersburg, MD 20899

Brett Lewis
APTEK, Inc.
1257 Lake Plaza Dr.
Colorado Springs, CO 80906

Zhong Liang
University of Buffalo
Dept. of Mechanical and Aerospace Eng.
141 Ketter Hall, Buffalo, NY 14260

Nick Lieven
Dept. of Aerospace Engineering
University of Bristol, Queen's Bldg.
Bristol, BS8 1TR, UK

Tae W. Lim
University of Kansas
Dept. of Aerospace Engineering
2004 Learned Hall
Lawrence, KS 66045

Michael Link
Universitat Gesamthochschule Kassal
Fachbereich 14
Bauingenienurwesen, Fachgebiet Leichtbau
Monchebergstr. 7, D-34109 Kassal
Germany

Shih-Chi Liu
National Science Foundation
4201 Wilson Blvd.
Arlington, VA 22230

Richard Livingston
Federal Highway Administration
6300 Georgretown Pike, HRD1-12
McLean, Va 22101

Nuno Maia
IDME/IST
Av. Rovisco Pais
1096 Lisboa Codex
Portugal

Dave Martinez
Sandia National Laboratories
M/S 0439
Albuquerque, NM, 87185-5800

Sami Masri
University of Southern California
Department of Civil Engineering
MC 2531
Los Angeles, CA 90089-2531

Randy Mayes
Sandia National Laboratories
MS0557
PO Box 5800
Albuquerque, NM 87185

David McCallen
Center for Complex Distributed Systems
Lawrence Livermore National Laboratory
Livermore, CA 94550

Ken McConnell
Iowa State University
3017 Black Eng. Bldg.
Ames, Iowa 50011

Akira Mita
Graduate School of Science and Technology, Keio University
8-14-1 Hiyoshi, Kohoku-ku
Yokohama 223-8522
Japan

Jack Moehle
University of California
Department of Civil Engineering
775 Davis Hall
Berkeley, CA 94720

Julio M. Montalvao e Silva
IDME/IST
Av. Rovisco Pais
1096 Lisboa Codex
Portugal

Denby Morrison
Shell E&P Technology Company
Ocean R&D, Bellarie Technology Center
P.O. Box 481
Houston, TX 7700

Peter Moss
University of Canterbury
Department of Civil Engineering
Private Bag 4800
Christchurch, New Zealand

John Mottershead
Dept. of Engineering
Brownlow Hill
University of Liverpool
Liverpool, L69 3GH
UK

Robert Murray
Lawrence Livermore National Laboratory
P.O. Box 808, L-197
Livermore, CA 94550

H.G. Natke
Universtat Hannover
Applestrabe 9A
D-30167 Hannover
Germany

Yi-qing Ni
Department of Civil and Structural Engineering
Hung Hom, Kowloon, Hong Kong

Robert Nigbor
Department of Civil Engineering
University of Southern California
Los Angeles, CA 90089-2531

Ozden Ochoa
Offshore Technology Research Center
1200 Mariner Dr.
Texas A&M University
College Station, TX 77845

Wally Orisamolu
Manager, Structural Integrity & Reliability Group
Components Department
United Technologies Research Center
411 Silver Lane, MS 129-73
East Hartford, CT

Roberto Osegueda
The University of Texas at El Paso
Department of Civil Engineering
El Paso, TX 79912

Richard Pappa
NASA Langley Research Center
MS 230, Hampton, VA 23681

Gerard Pardoen
University of California-Irvine
101 ICEF-Civil Eng.
Irvine, CA 92717

K. C. Park
Center for Aerospace Structures
University of Colorado, Boulder
Campus Box 429
Boulder, CO 80309-0429

Lee Peterson
Center for Aerospace Structures
University of Colorado, Boulder
Campus Box 429
Boulder, CO 80309-0429

Darryll J. Pines
Dept. of Aerospace Engineering
Room 3154 Engineering Classroom Bldg
University of Maryland
College Park, MD 20742

Bruno Piombo
Departimento di Meccanica
Politecnico di Torino
Corso Duca degli Abruzzi, 24
I - 10129 Torino, Italy

Bob Randall
School of Mechanical/Manufacturing Engineering
University of New South Wales,
Sydney 2052 Australia

John Reed
922 Parma
Los Altos, CA 94024

Mark Richardson
Vibrant Technology, Inc.
18141 Main Street
Jamestown, CA 95327

Jim Ricles
Lehigh University
Department of Civil Engineering
117 ATLSS Drive, H Building
Bethlehem, PA 18015-4729

John Ruminer
Los Alamos National Laboratory
DDESA, MS P945
P.O. Box 1663
Los Alamos, NM 87545

Romualdo Ruotolo
Dip. Ingegneria Aeronautica e Spaziale
Politecnico di Torino
10100 Torino
Italy

Anders Rytter
RAMBOLL
Kjaerulfsgade 2
DK-9400 Norresundby
Denmark

Erdal Safak
U.S. Geological Survey
DFC, Box 25046, MS.966
Denver, CO 80225

M. Saiidi
College of Engineering
Department of Civil Eng./258
Reno, NV 89557-0152

Masoud Sanayei
Tufts University
Dept. of Civil and Env. Engineering
Medford, MA 02155

Jose Maria Campos dos Santos
UNICAMP
Caixa Postal 6122
13083-970 Campinas, SPBrazil

Paul Sas
Katholieke Universiteit Leuven
Mechanical Engineering Dept.
Celestijnenlaan 300B
B-3001 Herverlee
Belgium

Bob Shumway
Division of Statistics
Univ. of California, Davis
Davis, CA 95616

Michael Simmons
Caterpillar, Inc.
Technical Center, Bldg. A
P.O. Box 1875
Peoria, Illinois 61656-1875

Suzanne Smith
University of Kentucky
Department of Engineering Mechanics
467 Anderson Hall
Lexington, KY 40506-0046

Cecily Sobey
Earthquake Engineering Research
Center Library
Gift & Exchange Dept.
University of California/RFS 453
1306 South 46th Street
Richmond, CA 94804-4698

Ian Stanley
Kinemetrics, Inc.
222 Vista Ave.
Pasadena, CA91107

Norris Stubbs
Texas A&M University
Department of Civil Engineering
Mechanics & Materials Center
College Station, TX 77843-3136

Fred Tasker
Dept. of Mechanical Engineering
University of Maryland Baltimore County
Baltimore, MD 21228-5398

Mike Todd
Naval Research Laboratory
Optical Sciences Division
Code 5673
4555 Overlook Ave. SW
Washington D.C. 20375

Geoff Tomlinson
The University of Sheffield
Dept. of Mechanical and Process
Engineering
PO Box 600
Mappin St, Sheffield S1 4DU UK

Pavel Trivailo
RMIT University
226 Lorimer St.
Fishermen's Bend 3207 Victoria
GPO Box 2476V
Melbourne 3001 Victoria Australia

Ward Turner
Exxon Production Research Company
P.O. Box 2189
Houston, TX 77252

Herman Van Der Auweraer
LMS International
Interleuvenlaan 68
B-3001 Leuven, Heverlee
Belgium

C. E. Ventura
The University of British Columbia
Dept. of Civil Engineering
2324 Main Mall
Vancouver, B.C.
Canada, V6T 1Z4

Sara Wadia-Fascetti
Northeastern University
Dept. of Civil Engineering
443 Shell Engineering Center
Boston, MA 02115

Gunnar Wang
Norwegian Defense Research Est.
P.O. Box 25
N-2007 Kjeller, Norway

Ming Wang
Dept. of Civil and Material Engineering
(M/C 246)
College of Engineering
842 West Taylor St.
Chicago, Illinois 60607-7023

Semyung Wang
Visiting Research Scholar
Aerospace System Design Lab
School of Aerospace Engineering
Technology
Georgia Institute of Technology
Atlanta, GA 30332-0150

Lloyd Welker, Jr.
Ohio Department of Transportation
25 South Front Street
Columbus, OH 43216-0899

Robert West, Jr.
Structural Imaging and Modal Analysis Lab.
Virginia Polytechnic Inst. of State Univ.
Mechanical Engineering Department
Blacksburg, VA 24061-0238

Ed White
Boeing
P.O. Box 516
St. Louis, MO 63166

Ken White
Dept. of Civil, Agricultural, and Geological Eng.
Box 30001/Dept. 3CE
Las Cruces, NM 88003-0001

Al Wicks
Mechanical Engineering Department
Virginia Tech. Univ.
Blacksburg, VA 24061-0238

P. Winney
P&P Engineering
Consultant Engineers
P.O. Box 36
Billingshurst, West Sussex RH14 OYG

Felix S. Wong
Weidlinger Associates
4410 El Camino Real, Suite 110
Los Altos, CA 94022-1049

Shi-Chang Wooh
Room 1-272
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Keith Worden
The University of Sheffield
Department of Mechanical and Process Engineering
PO Box 600
Mappin St
Sheffield, S1 3JD UK

Fan Wu
Risk Management & Solutions, Inc.
149 Commonwealth Drive
Menlo Park, CA 94025

J. T. P. Yao
Dept. of Civil Engineering
Texas A&M University
College Station, TX 77843-3136

Chung-Bang Yun
Dept. of Civil Engineering
Korean Institute of Advanced Science and Technology
373-1, Kusong-dong, Yusong-gu,
Taejon, KOREA, 305-701

Dave Zimmerman
University of Houston
Department of Mechanical Engineering
Houston, TX 77204-4792

Los Alamos NM 87545