

Extreme Value Modeling and Parametric Investigations of Gust and Maneuver Loads for General Aviation

Gulshan Singh*

General Electric Global Research, Niskayuna, NY, 12309, USA

Juan Ocampo[†] and Harry Millwater[‡]

University of Texas at San Antonio, TX, 78249, USA

Abstract

An extreme value distribution (EVD) of the maximum load per flight of a load spectrum is critical for a probabilistic damage tolerance analysis of a General Aviation aircraft. The EVD parameters are important because the structural integrity of the aircraft depends upon the maximum load seen by the structure during a specified number of flights. It is well known that the load spectrum that an aircraft experiences depends upon a large number of variables including number of flights, type of usages, number of usages, percentage of each usage, maneuver and gust load limit factors, aircraft velocity, flight duration, ground stress, one-g-stress, exceedance curve, and randomness in these variables. This research investigates the effect of three selected variables (type of usage, exceedance curve, and flight length-velocity and flight length-weight matrices) on the maximum load per flight EVD. A computer code (load module) capable of generating a realistic load spectrum for a given set of loading parameters was developed. A generalized extreme value approach was developed to estimate the EVD of the maximum load per flight. A number of parametric investigations were performed to determine the effect of load spectrum variables on the EVD parameters. The preliminary results indicate that exceedance curves and type usage have the largest effect on the EVD parameters.

1 Introduction

A risk assessment of the continued operational safety of a general aviation (GA) fleet can provide important insight into the criticality/severity of a potentially serious structural issue. A probabilistic damage tolerance analysis (PDTA) [1, 2] is necessary to assess the risk and provide a mechanism to include inspection and maintenance operations. A typical goal of a PDTA analysis is to estimate the probability of occurrence of an adverse event such as probability-of-failure during a flight (T) or cumulative probability-of-failure during the first T number of flights or flight

*Corresponding author: (singhg@ge.com) Mechanical Engineer and AIAA Member.

[†]Graduate Research Assistant and AIAA Member.

[‡]Professor, Department Chair, and AIAA Member.

hours. The probability-of-failure during a flight is called *single flight probability-of-failure* and is denoted by $SFPOF(T)$ for the flight number T . The cumulative probability-of-failure during the first T flights is called *cumulative total probability-of-failure* and is denoted by $CTPOF(T)$. The FAA and US Air Force have their standards [3] which define the acceptable level of $SFPOF$ and/or $CTPOF$. There are many methods and definitions available in the literature [2, 4–11] to estimate these probabilities.

It is important to emphasize here that the PDTA methodology and probabilities ($SFPOF$ and $CTPOF$) are shown here to demonstrate that the EVD of a load spectrum is critical in a PDTA analysis. The PDTA methodology is not the focus of this work. The goal of this work is to develop an approach to estimate an EVD of a load spectrum and study the effects of the load spectrum parameters on the EVD parameters. The developed approach and the results of this research are not only applicable to the PDTA methodology shown below but also to the other methodologies that employ a conditional probability approach to compute the probabilities.

During a flight, failure occurs when the stress intensity factor exceeds the fracture toughness of the material

$$\begin{aligned} K_{SIF}(T) &> K_C \\ \sigma\beta(T)\sqrt{\pi a(T)} &> K_C \end{aligned}$$

where K_{SIF} , K_C , σ , $\beta(T)$, and $a(T)$ are the stress intensity factor at flight T , fracture toughness, applied stress, geometry correction factor at flight T , and crack size at flight T .

For given values initial crack size (a_0) and fracture toughness (K_C), the probability-of-failure during a flight can be estimated as following:

$$\begin{aligned} P \left[\left(\sigma\beta(T)\sqrt{\pi a(T)} \right) > K_C \right] \\ P \left[\sigma > \left(\frac{K_C}{\beta(T)\sqrt{\pi a(T)}} \right) \right] \end{aligned} \quad (1)$$

The left hand side in Eq. 1 is applied stress and the right hand side is the residual strength. The equation indicates the probability of an applied stress (during the flight T) exceeding the residual strength. A conservative approximation of this probability can be given by:

$$1 - F_{EVD} \left(\frac{K_C}{\beta(T)\sqrt{\pi a(T)}} \right) \quad (2)$$

where F_{EVD} is the extreme value distribution of the applied load. For given values fracture toughness and initial crack size, this expression can be used to define $SFPOF$:

$$\begin{aligned} SFPOF(T) &= \int_{-\infty}^{\infty} \int_0^{\infty} \left[\prod_{t=1}^{T-1} F_{EVD} \left(\frac{K_C}{\beta(t)\sqrt{\pi a(t)}} \right) \right] \\ &\left[1 - F_{EVD} \left(\frac{K_C}{\beta(T)\sqrt{\pi a(T)}} \right) \right] f_{a_0}(a_0) f_{K_C}(K_C) da_0 dK_C \end{aligned} \quad (3)$$

The first term in the equation for $SFPOF(T)$ is the probability-of-survival for the proceeding $(T - 1)$ flights and the second term is the probability of failure during the flight T . The first term is necessary to ensure that the component survives for $(T - 1)$ flights. The following expression for $CTPOF(T)$ calculates the probability of survival for T flights using the term $\left[\prod_{t=1}^T F_{EVD} \left(\frac{K_C}{\beta(t)\sqrt{\pi a(t)}} \right) \right]$ and then computes the probability of failure.

$$CTPOF(T) = \int_{-\infty}^{\infty} \int_0^{\infty} \left[1 - \prod_{t=1}^T F_{EVD} \left(\frac{K_C}{\beta(t)\sqrt{\pi a(t)}} \right) \right] f_{a_0}(a_0) f_{K_C}(K_C) da_0 dK_C \quad (4)$$

where, T , F_{EVD} , K_C , a_0 , $\beta(T)$, $a(T)$, $f_{a_0}(a_0)$, $f_{K_C}(K_C)$ indicate flight number, EVD of maximum load per flight, fracture toughness, initial crack size, geometric factor at flight T , crack size at flight T , PDF at initial crack size, and PDF of fracture toughness, respectively. The expressions in Equations 3 and 4 can be estimated using a numerical integration or sampling based approach.

As shown in these equations and in the literature, an EVD of the maximum load per flight is a critical part of a PDTA analysis for a given set of aircraft loading parameters. Therefore, a better understanding of the relationship between the load spectrum parameters and EVD parameters is vital. The importance of this research is highlighted by the fact that a load spectrum and its EVD are related. The load spectrum and its EVD must be used together. A PDTA analysis in which the load spectrum is taken from one source and EVD is taken from another source can produce unreliable results.

The generalized extreme value theory [12–14] (EVT) can be used to estimate the extreme value distribution for a given parent probability distribution or observed data. Many researchers in the literature [15–23] have employed the EVT and other statistical techniques in aerospace and general engineering applications. Smith and Adelfeng [16] used extreme value statistics to analyze solar activity for the Space Station. Cober and Isaac [22] used EVT to estimate maximum aircraft icing environment using historic data.

This research integrates: (i) generalized extreme value theory, (ii) maximum likelihood function, and (iii) particle swarm optimization to estimate EVD distribution for a given set of aircraft loading parameters. This developed approach is validated by solving two numerical examples with known solutions. The validated approach is then used to estimate EVD parameters for multiple sets of aircraft loading conditions. The results of the approach are used to investigate the effect of three parameters (type of usage, exceedance curve, and flight length-velocity and flight length-weight matrices).

The remainder of the paper is organized as follows. Section 2 provides a detailed description of load spectrum variables and the load spectrum generation process. This Section also contains a brief introduction to extreme value theory, maximum likelihood approach, and particle swarm optimization. The verification of the developed approach is performed using two numerical examples and the results are compared in Section 3. The results of the EVD parametric investigations are also presented in the results section (Section 3). A summary and the conclusions of the research are in Section 4.

2 EVD Estimation

2.1 EVD Estimation Steps

The EVD estimation approach developed in this research is called Maximum Likelihood Estimation (MLE) approach. A brief summary of the approach is as follows:

- (i) Collect load spectrum parameters (number of flights, type of usage, number of usages, percentage of each usage, maneuver and gust load limit factors (LLF), aircraft velocity, ground stress, one-g-stress, flight length-velocity matrix, flight length-weight matrix, and exceedance curve) for a usage of interest.
- (ii) Generate load spectrum using the load module developed in this research and the load spectrum parameters.
- (iii) Extract the maximum load per flight from every flight of the flight spectrum.
- (iv) Employ generalized extreme value theory, maximum likelihood function, and an optimization approach to estimate the EVD of the load spectrum.

A detailed description of the computer code that generate load spectrum and mathematical techniques (extreme value theory, maximum likelihood function, and particle swarm optimization) used in the MLE approach is given below.

2.2 Load Spectrum Generation

A Fortran code was developed to generate a realistic load spectrum, the code accounts for five different flight stages (Maneuver, Gust, Taxi, Landing and Rebound, and Ground Air Ground) and incomplete cycles at any current flight are saved to be added in future flights. The input parameters for the computer code to generate a load spectrum are given in Table 1. Most of the parameters in the table are either self-explanatory or are explained in the second column of the table. Two parameters: flight matrices and exceedance curves are explained in this section.

2.2.1 Flight Matrices

Two flight matrices: flight length-velocity matrix and flight length-weight matrix are used to capture the flight length, the flight velocity, and the weight (one-g-stress and ground stress) difference between various flights and the correlation between these variables. An example of flight length-velocity matrix and flight length-weight matrix is shown in Tables 2 and 3, respectively.

Flight length-velocity matrix The first two columns of the matrix (Table 2) are related to flight length, where the first column contains the flight time in hours and the second column contains the percentage of the flights that the aircraft is flow for that period of time. For example, 0.45 in row 3 and columns 2 of the table shows that 45% of the total flights flew for 0.5 hours. The second row has the average speed in terms of percentage of the design velocity. For example, 1.0 in row 2 and column 3 is the 100% velocity. Each of the columns below the percentage velocity contains the information about the percentage of the time that the airplane flies at that velocity conditioned to the flight time. The values are the probability density data related to the flight

hours and flight speed. For example, value 0.30 in column 4 and row 3 indicates that 30% of the flights that flew 0.5 hours flew at 95% of the design speed. These properties are for 45% (column 2 row 3) of the total flights. The data in the table was used to create a cumulative distribution function for flight time and flight speed. Figure 1 shows the joint probability distribution function from Table 2.

Flight length-weight matrix The format of the flight length-weight matrix is similar to flight length-velocity matrix and the description of a flight length-velocity matrix can be used for the flight length-weight matrix. The first row in the flight length-weight matrix is weight percentages instead of velocity percentages. The data is used to create a cumulative distribution function for flight time and weight condition. Weight condition data is used to generate realizations of one-g stress and ground stress assuming that these two variables are fully correlated.

2.2.2 Exceedance Curve

During 1962, at the request of the FAA and upon recommendation of the National Aeronautics and Space Administration (NASA) Committee on Aircraft operating problems, the NASA V-G (velocity, normal acceleration)/VGH (velocity, normal acceleration, pressure altitude) General Aviation Program was established. This program recorded more than 105 airplanes with more than 42,155 hours of VGH data. Tabulated data in exceedance form can be found in the literature [24], FAA reports AFS-120-73-2 [19], and AC23-13A [20]. The data for maneuver and gust load are presented as cumulative number of occurrences per nautical miles versus the acceleration fraction (incremental normal acceleration divided by the incremental limit factor, an airplane characteristic).

The gust and maneuver (positive and negative) exceedance curves for a Single Engine Unpressurised Operations usage are shown in Figure 2. The data/plots have been taken from the literature [24] and tables A1-1 and A1-2 of FAA reports AC23-13A [20]. The x-axis is the acceleration fraction and y-axis is the cumulative frequency of exceedance per nautical mile. The exceedance curves can be used to generate the CDF and PDF of acceleration fraction and load spectrum for desired number of flights of the selected usage.

2.2.3 Load Spectrum Generation Procedure

The steps to generate the spectrum are presented as follows:

- (i) Provide input parameters; a summary of the input parameters is presented in Table 1
- (ii) Generate random realization of the parameters: maneuver and gust exceedance curves, flight-length and aircraft velocity as per flight length-velocity and maximum aircraft velocity, and one-g-stress as per flight length-weight and maximum one-g-stress
- (iii) Calculate number of occurrences for each of the flight stages using the methodology in the literature [19, 20, 25], the maximum stress, minimum stress, and exceedance curves
- (iv) After each of the stresses and occurrences are calculated for the current flight, incomplete cycles from previous flights are added to the current flight stresses, then the complete stresses are extracted, and the incomplete stresses from the current flight are saved for the next flight
- (v) Randomize the load pairs generated in the previous steps

(vi) Repeat steps (i) through (v) for the given number of flights

2.3 Generalized Extreme Value Theory

Suppose X_1, X_2, \dots, X_p is a sequence of independent random variables having a common distribution function $F(x)$. If M_p represent the maximum of the process over n observations, then as per extreme value theory, the distribution of M_p can be derived exactly for all the values of p [14]:

$$\begin{aligned} Pr\{M_p \leq z\} &= Pr\{X_1 \leq z, X_2 \leq z, \dots, X_p \leq z\} \\ &= Pr\{X_1 \leq z\} \times Pr\{X_2 \leq z\} \times \dots \times Pr\{X_p \leq z\} \\ &= \{F(z)\}^p \end{aligned} \tag{5}$$

Therefore, if the probability density function (PDF) or the distribution function of a random variable is given then an EVD of the variable over p samples can be estimated using Eq. 5. This may not be immediately helpful in practice because the PDF of aircraft loading is not available in a closed-form equation. However this principle provides the exact solution for a standard distribution such as uniform, normal, or Weibull distribution. When the PDF of the parent distribution is not available and the above approach cannot be used, the following approach can be employed.

From the extreme value theory, it is known that the asymptotic form of extreme value data as $p \rightarrow \infty$ can take one of three forms: Gumbel, Frechet, Weibull (Types *I*, *II*, and *III*). The three possible models for the maximum can be encapsulated in the generalized extreme value model as [21]:

$$G(z) = exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \tag{6}$$

The distribution in Eq. 6 $G(z)$ is known as the generalized extreme value distribution. Here μ , σ , and ξ indicate the location, scale, and shape parameters of the generalized extreme value distribution, respectively. The value of the shape parameter decides the type of the distribution. The extreme value distribution converges to Weibull, Gumbel, or Frechet if the shape parameter (ξ) value is less than zero, equal to zero, or greater than zero, respectively. A numerical approach, such as maximum likelihood function integrated with an optimization approach, can be used to determine the EVD distribution (Gumbel, Frechet, Weibull) and its parameters. It is possible that due to numerical reasons, the shape parameter (ξ) may not converge to the exact zero or to a Gumbell distribution. To address this issues, a suitable range of the shape parameter is selected (such as $-0.05 \leq \xi \leq 0.05$) for Gumbel distribution.

There are two parameters that correspond to number of samples and number of data points. The parameter p in this section corresponds to number of samples. For example, p realizations are generated and the maximum value of the realizations is saved. If this process is repeated a large number of times and the distribution of the collected data (maximum of p realization) is estimated using the EVD approach then the number of samples is p . The total number of data points collected or observed data during this process is called *number of data points* and is denoted using N .

2.4 Maximum Likelihood Function

It is possible, as in the case of General Aviation, that the distribution of realizations/data is not available in a closed-form. It is also possible that the data may not fall into a standard distribution (Uniform, Normal, or Weibull etc.) that can be characterized using a closed-form equation. The maximum likelihood function [14,21] is a common approach to determine distribution parameters for such data.

In this approach, it is not necessary to know the distribution of the original realizations or its EVD, based on the extreme value theory, since the extreme value distribution of the realizations is supposed to be one of the three possible extreme value distributions: Gumbel, Frechet, and Weibull. The maximum likelihood function can be used to determine the extreme value distribution (EVD) and the distribution parameters. If x_1, x_2, \dots, x_N are independent realization of a random number having probability density function $f(x, \theta)$, the likelihood function is

$$L(x, \theta) = \prod_{i=1}^N f(x_i, \theta) \quad (7)$$

where θ is an unknown parameter or a set of unknown parameters of $f(x, \theta)$. The statistical model of the realizations $L(\theta)$ is a joint PDF; thus the normalizing condition holds:

$$\int_X L dx = 1$$

For the problem considered in this paper, the realization x_1, x_2, \dots, x_N is a given quantity. This means that the Eq. 7 is a function of the parameter θ . In other words, the parameter θ is the unknown parameter in Eq. 7. When $L(x; \theta)$ is considered to be a function of θ , $L(x; \theta)$ is called the *likelihood function* of the data samples.

Since the logarithm function is monotonic, the log-likelihood takes its maximum at the same point as the likelihood function, so that maximum likelihood estimator also maximizes the corresponding log-likelihood function. The log-likelihood function of the likelihood function $L(\theta)$ is:

$$\ell(\theta) = \ln(L(\theta)) = \sum_{i=1}^N \ln(f(x_i, \theta)) \quad (8)$$

A simple approach to obtain the maximum likelihood estimator is to differentiate the log-likelihood with respect to θ and equate the results to zero. For complicated problems a numerical approach such as an optimization can be used to maximize the log-likelihood function with respect to the parameter θ . This parameter can be a set of parameters such as $\theta = \{\mu, \sigma\}$ for a normal distribution or $\theta = \{\mu, \sigma, \xi\}$ for a general extreme value distribution.

2.5 Particle Swarm Optimization

A maximization of log-likelihood is an unconstrained optimization problem. There are many optimization algorithms available to solve such problems. The selection of an algorithm is made easy by a low computational cost associated with the computation of a likelihood or log-likelihood function. An optimization algorithm should be able to find an optimum after using a large

number of function evaluations. In this work, the particle swarm optimization (PSO) algorithm is used because of a lower number of user input parameters, use of an elitist strategy, and ease of implementation.

PSO, as proposed by [26], modifies the population from step to step based on a set of rules. In this set of rules, each individual particle uses its current fitness, its best fitness so far, the best fitness of all the individuals, and a communication structure to determine the movement parameters. These parameters change over the course of iterations, and the particle population tends to converge and provide an optimum solution. The movement/velocity of the i^{th} particle at iteration $k + 1$ is calculated as:

$$v_i^{k+1} = w_i v_i^k + c_1 r_{i1} (pbest_i - s_i^k) + c_2 r_{i2} (gbest - s_i^k) \quad (9)$$

$$s_i^{k+1} = s_i^k + v_i^{k+1} \quad (10)$$

where v_i^k is the i^{th} particle velocity at iteration k , w_i is the i^{th} particle weight function, c_j is c_1 or c_2 and is the weight coefficient of each term (j^{th}) of the PSO equation, r_{ij} is a uniform random number between 0 and 1, s_i^k is the i^{th} particle position at iteration k , $pbest_i$ is the particle with the best objective function values of the i^{th} particle over the iteration history (1 through k), and $gbest$ is the particle with the best objective function value of all the particles. Eq. 9 was used to calculate velocity of the i^{th} particle in each dimension. The velocity value was substituted in Eq. 10 to calculate the particle position at iteration $k + 1$. The dimension index was not shown in order to keep the equation simple. The constants values w_1 , c_1 , and c_2 were taken from the literature [26, 27] and are 0.529, 1.494, and 1.494, respectively.

An optimization process begins by randomly generating a number of possible designs representing different EVD configurations. Each set of parameters configuration is called a particle in the PSO terminology. The total number of alternatives being evaluated at any one time is equal to the number of particles in the PSO. The objective function values (log-likelihood) are calculated for each particle. Based on the objective function values, the parameters $pbest_i$ and $gbest$ are determined. Then Equations 9 and 10 were used to calculate the new particle positions. This process was repeated until the convergence criterion was met. Due to the low computational cost of the log-likelihood function calculation, the maximum number of iterations (between 40 and 100) was used as the convergence criterion.

3 Results and Discussion

Two numerical examples (uniform and normal distributions) were solved to test and verify the developed maximum likelihood function and optimization approaches. These cases were selected because the exact extreme value distribution for both cases is known analytically. Upon verification, the approaches were employed to determine EVD for many cases and parametric studies were performed to compute the effect of three variables: exceedance curve, type usage, and velocity-duration and weight duration matrices.

3.1 Numerical Example

3.1.1 Uniform Distribution

A uniform distribution between 0 and 1 was selected for comparing the results from the maximum likelihood function estimate and the exact solution. The PDF of a uniform distribution between 0 and 1 is:

$$f_X(x) = \begin{cases} 1; & \text{if } 0 \leq x \leq 1 \\ 0; & \text{if elsewhere} \end{cases} \quad (11)$$

and the CDF is:

$$F_X(x) = x; \quad 0 \leq x \leq 1$$

According to extreme value theory, the EVD of a uniform distribution of a sample size p is:

$$F_{M_p}(z) = z^p; \quad 0 \leq z \leq 1 \quad (12)$$

and the corresponding PDF:

$$f_{M_p}(z) = \begin{cases} pz^{p-1}; & \text{if } 0 \leq z \leq 1 \\ 0; & \text{if elsewhere} \end{cases} \quad (13)$$

Eq. 13 was used to compute the exact EVD of the extreme value of a sample size p of the uniform distribution where p was 5, 10, 15, 20, and 25. The number of data points (N) used for the MLE approach was 100,000.

The results obtained from maximizing the log-likelihood function are shown in Table 4. The table shows that the shape parameter (Column 4) of the EVD is negative in all cases. A negative value of the shape parameter indicates that the EVD of the standard uniform distribution is a Weibull distribution. Table 4 shows the three parameters of Weibull EVD for all five cases selected for the comparison. The PDFs from the exact solution and the maximum likelihood are plotted in Figure 3.

The results in Figure 3 and Table 4 show that the maximum likelihood approach is in agreement with the exact solutions. A slight disagreement between the exact and MLE solution at the right tail of the distribution can be ignored for this research purpose because it is in a very small part of the total range and within acceptable limits (2%). Probability and quantile plots were also used to verify the distribution provided by the maximum likelihood approach. Probability and quantile plots showed that the maximum likelihood function approach is able to provide accurate results for the uniform distribution.

3.1.2 Standard Normal Distribution

The CDF of a standard normal distribution is:

$$F(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (14)$$

Exact Solution: CDF of a standard normal distribution is given in Eq. 14. Since the closed-form CDF is known, Eq. 5 gives the exact solution for the extreme value distribution of the standard normal distribution. For a sample size p , the exact EVD of the standard normal distribution is:

$$F_{M_p}(z) = (F(z))^p \quad (15)$$

Cramer' Solution: The details of Cramer's asymptotic solution can be found in the literature [12]. The extreme value distribution of a standard normal distribution is the Gumbel distribution. The CDF of the Gumbel distribution is:

$$F(x) = \exp[-\exp[-a_p[x - u_p]]] \quad (16)$$

where, according to Cramer a_p , u_p , EVD mean, and EVD standard deviation for EVD of p samples at a time can be estimated as follows:

$$\begin{aligned} u_p &= \sqrt{2 \ln(p)} - \frac{\ln(\ln(p)) + \ln(4\pi)}{2\sqrt{2 \ln(p)}} \\ a_p &= \sqrt{2 \ln(p)} \\ \sigma &= \frac{\pi}{\sqrt{6}} \frac{1}{a_p} \\ \mu &= u_p + \frac{0.5772}{a_p} \end{aligned} \quad (17)$$

where, u_p and a_p are the constant from Eq. 16 and σ and μ indicate the standard deviation and mean of the EVD. These formulas were approximated after some assumptions and ignoring higher order terms in the derivation.

Results for Standard Normal Distribution: Figure 4 compares the results from the three approaches: the exact solution, MLE estimate, and Cramers solution. The MLE approach is in agreement with Cramer's approximation and the exact solution for the selected five cases. In fact, the MLE approach is closer to the exact solution as compared to Cramer's solution. The results are also shown in Table 5. The table shows that Cramer's solution and the likelihood approach are in agreement with each other.

3.2 Aircraft Example

A convergence study was performed to determine the number of data points needed to estimate the converged EVD parameters. Four parametric studies have been performed for parameters: number of samples, flight matrices, exceedance curves, and usage.

3.2.1 Convergence Study

It was essential to conduct a convergence study to determine how many flights (number of data points, N) were needed to obtain converged EVD parameters. The convergence study was conducted for EVD of maximum load per flight. This convergence study is performed for Instructional

usage (Single Engine Unpressurized Basic Instructional Usage). The exceedance curve was fixed at the mean.

In the convergence study, the number of flights were increased from 10 to 1,000,000. The input parameters for the load spectrum generation are given in Tables 2, 3, and 6. The results of the convergence study are shown in Table 7 and Figure 5. The fact that the distribution type changes as the number of data points were increased shows the importance of the convergence study. The optimum distribution was Weibull for 100 ad 1000 data points. The distribution was almost Gumbul (shape parameters close to zero) for 10,000 data points. The distribution was Frechet for data points 100,000 or more.

Figure 5 shows that the EVD plots are on top of each other when the number of data points are 10,000 or more. However, the results in Table 7 show that 100,000 flights were needed to estimate converged EVD parameters for a given load spectrum. The location and scale parameters converged in 10,000 data points but the shape parameter required a large number of data points. Based on this convergence study, 100,000 or more data points (N) were used to estimate EVD parameters in all the examples in this research.

3.2.2 Confidence Bounds and Model Checking:

The data points and the parameters computed using MLE approach are used to estimate variance-covariance matrix (V) of the distributions parameters $\theta = (\theta_1, \theta_2, \dots, \theta_d)$. The matrix can be used to approximate confidence bounds on the distribution parameters. These confidence bounds provide important information about the reliability of the predicted parameters. To estimate the variance-covariance matrix, the expected information matrix, also called the observed information matrix (O), was computed using following formula [14]:

$$O = \begin{bmatrix} -\frac{\partial^2}{\partial \theta_1^2} l(\theta) & \dots & \dots & -\frac{\partial^2}{\partial \theta_1 \theta_d} l(\theta) \\ \vdots & \ddots & -\frac{\partial^2}{\partial \theta_i \theta_j} l(\theta) & \vdots \\ \vdots & -\frac{\partial^2}{\partial \theta_j \theta_i} l(\theta) & \ddots & \vdots \\ -\frac{\partial^2}{\partial \theta_d \theta_1} l(\theta) & \dots & \dots & -\frac{\partial^2}{\partial \theta_d^2} l(\theta) \end{bmatrix}$$

This matrix is called the expected information matrix because it measures the expected curvature of the log-likelihood surface. This is also called the observed information matrix because the observed data is used to compute the matrix. The matrix O is a symmetric matrix and each component of the matrix is a double differentiation of the log-likelihood function with respect to the target distribution parameters. The inverse of the observed information matrix is the variance-covariance matrix. Denoting the terms if the inverse matrix with $\psi_{i,j}$, it follows that an approximate $(1 - \alpha)$ confidence interval for θ_i is:

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \sqrt{\psi_{i,i}} \quad (18)$$

where $z_{\frac{\alpha}{2}}$ represents the standard normal variate. For example, for 95% confidence bound $z_{\frac{\alpha}{2}} = \pm 1.96$. As shown in Table 7, the converged distribution parameters form are convergence study example are 11224, 1430, and 0.0189 for location, scale, and shape of the EVD distribution, respec-

tively. The variance-covariance matrix for the example is: $V = \begin{bmatrix} 26.1 & 7.72 & -0.0043 \\ 7.72 & 13.9 & -0.0024 \\ -0.0043 & -0.0024 & 5.64 \times 10^{-6} \end{bmatrix}$.

The 95% confidence upper bounds on the three parameters are 11234.00, 1437.31, and 0.0236. The 95% confidence lower bounds are 11213.99, 1422.69, and 0.0142. The confidence bounds show that the MLE estimated parameters are reliable.

Probability and quantile plots are graphical approaches to check the parameters estimated using MLE approach. In the probability plot, the probability of each observed data point is computed using two approaches: MLE estimated parameters and empirical function. In the quantile plots, the quantile of each observed data point is computed using the same two approaches. The probabilities or quantiles for the data points should be the same from the two approaches. Both plot should form a straight line if MLE results are accurate. The formula for the plots are given in the literature [14]. The probability and quantile plots for the convergence study are shown in Figure 6 and 7. The probability plot shows that MLE estimated parameters are accurate. The quantile plot shows that the MLE estimated parameters are accurate for majority of the data range but the accuracy reduces on the right tail of the distribution.

3.3 Parametric Investigation

3.3.1 EVD per p Samples/Flights

Typically, the EVD of the maximum load per flight is required in a PDTA analysis. For this typical case, the value of p (number of samples) is 1. However, the developed approach can be used to estimate the EVD of the maximum load per p flights. Here p is a positive integer number. The selected values of p for this study were 1, 2, 4, 5, and 10. At minimum, 100,000 samples were used to calculate the EVD parameters for a selected case. The results of this parametric study are shown in Figure 8 and Table 8.

The figure shows that the EVD location moved to the right side as the value of parameter p was increased. This was expected because higher load values are left in the data as compared to lower load values when number samples are increased. The table shows that location parameter was increasing with the sample numbers. The scale of the distribution did not change significantly. The shape parameter also changed with respect to p . In fact, EVD is Frechet if $p = 1$ and Weibull for all other values of p . A change in EVD can also be seen in Table 5 and Figure 4 for the standard normal distribution example. The reduction in scale was significant for the standard normal distribution example because the change in number of samples was multiple orders of magnitude.

3.3.2 Flight Matrices

To simplify this investigation, the flight length-velocity and flight length-weight were kept the same. The effects of the one matrix were assumed to be similar to the other matrix. The effects of the flight matrices were determined by making the matrices deterministic and random. In the deterministic case, the flight speed was fixed at 80% of the maximum speed of 148.50 knots and the flight weight was also fixed at the 80% of the maximum weight. In the random case, the matrices were the same and as shown in Table 9. These two cases were repeated for exceedance curve at

the mean and random exceedance curve. The remaining input parameters for this investigation were kept the same and are shown in Column 3 of Table 6.

The resulting EVDs of four cases are shown in Table 10 and Figure 9. The results indicate the the randomness in the matrices increases the standard deviation of the EVD. The highest standard deviation resulted for the case when both matrices and exceedance curves were random. The lowest standard deviation resulted for deterministic matrices and exceedance curves. The location of the EVD did not change for all the four cases.

3.3.3 Exceedance Curve

An exceedance curve in the load module has a lognormal distribution with 12% coefficient of variance. This randomness in the exceedance curve was incorporated into the load module. The usage selected for this investigation is Twin Engine Unpressurized General Usage. The effects of variations in exceedance curves were investigated in two parts: deterministic and random flight length-velocity and flight length-weight matrices. Four cases are selected in both parts of the study: an exceedance curve in the left tail of its distribution ($\mu - 1.96$), at the mean (μ), in the right tail ($\mu + 1.96$), and random exceedance curves. The area within the left and right points is 95% of the exceedance distribution. All other parameters were as shown in Column 2 of Table 6.

Table 11 and Figure 10 show the results of the first part of the investigation. Table 12 and Figure 11 show the results of the second part of the investigation. A summary plot of both investigations is shown in Figure 12. The first observation from the tables and figures is that the exceedance curve has a significant effect on the EVD parameters and distribution. An exceedance curve on the left tail of its distribution is less severe as compared to a curve on the right tail of the distribution. The standard deviation of the EVD increases as the exceedance curve moves from left to right. The EVD standard deviation of the random exceedance curve is higher as compared to deterministic exceedance curve at the mean.

3.3.4 Usage

Three type of usage were selected for this investigation: Single Engine Unpressurised Basic Instructional Usage (Instructional), Single Engine Unpressurised Executive Usage (Special), and Mixed Usage (50% Instructional + 50% Special). These usages were selected from the FAA report [20] in the literature. The other input parameters used to generate load spectrum for the three cases were kept the same.

Table 13 and Figure 13 show the results of this investigation. The Instructional usage has higher maximum load per flight as compare to the Special usage. The standard deviation of the Instructional usage EVD is higher as compared to the Special and Mixed usage EVDs. The EVDs converge to Frechet, Weibull, and Weibull for the instructional, special, and mixed usage, respectively.

It is known that severity of the flights (or maximum load per flight) is higher for higher number of flights as compared to lower number of flights. In the Mixed usage there are 500,000 Special and 500,000 Instructional flights. This means that the severity of these flights is going to be lower as compared to 1 Million flights of only one usage. This is a potential reason that the mixed usage

EVD looks closer to Special (less severe than Instructional) usage as compared to Instructional usage and converged to Weibull instead of Frechet distribution.

4 Summary Remarks and Conclusions

A probabilistic damage tolerance methodology was introduced and a case was made that the EVD is an important part of a PDTA approach. There is a need to develop a better understanding of the load spectrum parameters that affect the EVD parameters. An approach was developed that integrates generalized extreme value theory, maximum likelihood function, and particle swarm optimization. The developed approach was employed to estimate the converged EVD parameters for many cases. The effect of number of samples, flight length-velocity or flight length-weight matrix, exceedance curve, and usages was investigated in multiple parametric studies.

The developed approach accurately estimated the EVD parameters for all the cases considered. The results were verified for the problem where the exact solution was available and a good agreement was found between the estimated and the exact results. The results of the investigations provided valuable insight into the effect of load parameters in the EVD parameters.

Acknowledgment

This research is funded by the Federal Aviation Administration (FAA) under the grant number 09-G-016. The authors would like to thank Marvin Nuss, Michael Reyer, and Dr. Felix Abali of the FAA for their technical feedback in the research.

References

- [1] Berens, A., Hovey, P., and Skinn, D., "Risk Analysis for Aging Aircraft Fleets," Tech. Rep. WL-TR-91-3066, University of Dayton Research Institute, 1991.
- [2] Miedlar, P. C., Berens, A., Hovey, P., Boehnlein, T., and Loomis, J., "PRoF v3 PRobability Of Fracture," Tech. Rep. UDR-TR-2005-00240, University of Dayton Research Institute, December 2005 2005.
- [3] MIL-STD-1530, "Department of Defense Standard Practice - Aircraft Structural Integrity Program (ASIP)," Tech. Rep. MIL-STD-1530, US Air Force, November 2005.
- [4] Artley, M., "Probabilistic Damage Tolerance Method for Metallic Aerospace Structure," Tech. Rep. WRDC-TR-89-3093, Wright-Patterson Air Force Base, February 1989.
- [5] Berens, A., "Probability of Detection (POD) Analysis for the Advanced Retirement for cause (RFC)/Engine Structural Integrity Program NonDestructive Evaluation System Development," Report AFRL-ML-WP-TR-2001-4010, Wright-Patterson Air Force Base, 2001.
- [6] Shiao, M., Boyd, K., and Fawaz, S., "A Risk Assessment Methodology and Tool for Probabilistic Damage Tolerance-based Maintenance-based Maintenance Planning," *8th Joint NASA/FAA/DoD Conference on Aging Aircraft*, 31 January - 3 February 2005, pp. 1–18.

- [7] White, P., Molent, L., and Barter, S., “Interpreting Fatigue Test Results using a Probabilistic Fracture Approach,” *International Journal of Fatigue*, Vol. 27, 2005, pp. 752–767.
- [8] Tong, Y. C., Hou, J., and Antoniou, R. A., “Probabilistic Damage Tolerance Assessment: The Relative Merits of DARWIN, NERF AND PROF,” Tech. Rep. DSTO-TR-1789, Defense Science and Technology Organisation, Australia, 2005.
- [9] Liu, Y. and Mahadevan, S., “Stochastic fatigue damage modeling under variable amplitude loading,” *International Journal of Fatigue*, Vol. 29, October 2007, pp. 1149–1161.
- [10] Lin, K. Y., “Probabilistic Approach to Damage Tolerance Design of Aircraft Composite Structures,” *Journal of Aircraft*, Vol. 44, No. 4, 2007, pp. 1309–1321.
- [11] McFarland, J., Popelar, C., and Bichon, B., “Establishing Confidence Bounds on Reliability Assessments for Fatigue Crack Growth,” *Aircraft Airworthiness and Sustainment Conference*, San Diego, 2011, pp. 1–11.
- [12] Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, 1946.
- [13] Ang, A. H.-S. and Tang, W. H., *Probabilistic Concepts in Engineering Planning and Design, Volume II*, Birkhauser Verlag, first edition ed., 1984.
- [14] Coles, S., *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.
- [15] Yang, J. and Du, S., “An Exploratory Study into the Fatigue of Composites under Spectrum Loading,” *Journal of Composite materials*, Vol. 17, No. 6, 1983, pp. 511.
- [16] Smith, O. E. and Adelfang, S. I., “Extreme Value Statistical Analysis of Solar Activity for Space Station,” *33rd AIAA Aerospace Sciences Meeting and Exhibit, January 9-12, Reno, Nevada*, 1995-0056, AIAA, January 1995.
- [17] Bland, J. D., “Detail Fatigue Analysis Exceedance Spectra Determined from Random Multi-directional Force Components,” *AIAA Journal of Aircraft*, Vol. 34, No. 1, 1997, pp. 109–113.
- [18] Arrieta, A. J. and Striz, A. G., “Modeling of the Aircraft Fatigue Load Environment,” *38rd AIAA Aerospace Sciences Meeting and Exhibit, January 9-12, Reno, Nevada*, No. 2000-973 in AIAA, 10-13 January 2000.
- [19] AFS120-73-2, “Fatigue Evaluation of Wing and Associated Structure on small Airplanes,” FAA Report AFS120-73-2, Federal Aviation Administration, 1973.
- [20] AC23-13A, “Fatigue Fail-Safe, and Damage Tolerance Evaluation of Metallic Structure for Normal, Utility, Acrobatic, and Commuter Category Airplanes,” FAA Report AC23-13A, Federal Aviation Administration, 2005.
- [21] Coles, S. and Pericchi, L., “Anticipating Catastrophes through Extreme Value Modelling,” *Applied Statistics*, Vol. 52, No. 4, 2003, pp. 405–416.

- [22] Cober, S. G. and Isaac, G. A., “Estimating Maximum Aircraft Icing Environments Estimating Maximum Aircraft Icing Environment Using a Large Data Base of In-Situ Observations,” *44th AIAA Aerospace Sciences Meeting and Exhibit, January 9-12, Reno, Nevada*, No. 2006-266 in AIAA, January 2006.
- [23] Singh, G., Millwater, H., and Domyancic, L., “Extreme Value Modeling of Gust and Maneuver Loads for General Aviation,” *52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, No. 2011-2042 in AIAA, 4-7April, 2011, Denver, Colorado, 2011.
- [24] Reter, W. M., “Probability Basis of Safe-Life Evaluation in Small Airplanes,” *9th Joint FAA/DoD/NASA Conference on Aging Aircraft*, 6-9 March, Atlanta, GA, USA, 2006.
- [25] Ocampo, J., Millwater, H., Singh, G., Smith, H., Abali, F., Nuss, M., Reyer, M., and Shiao, M., “Development of a Probabilistic Linear Damage Methodology for Small Aircraft,” *AIAA Journal of Aircraft*, Vol. 48, No. 6, 2011, pp. 2090–2106.
- [26] Kennedy, J. and Eberhart, R. C., “Particle Swarm Optimization,” *Proceeding of IEEE conference on Neural Network, IV, Piscataway, NJ*, 1995, pp. 1942–1948.
- [27] Singh, G. and Grandhi, R. V., “Mixed-Variable Optimization Strategy Employing Multifidelity Simulation and Surrogate Models,” *AIAA Journal*, Vol. 48(1), 2010, pp. 215–223.

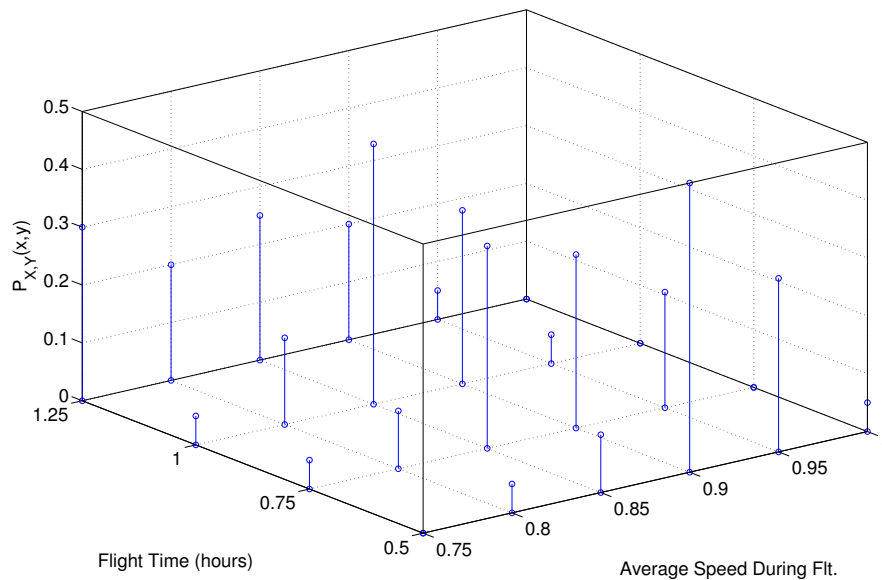


Figure 1: Flight length and velocity joint probability density function

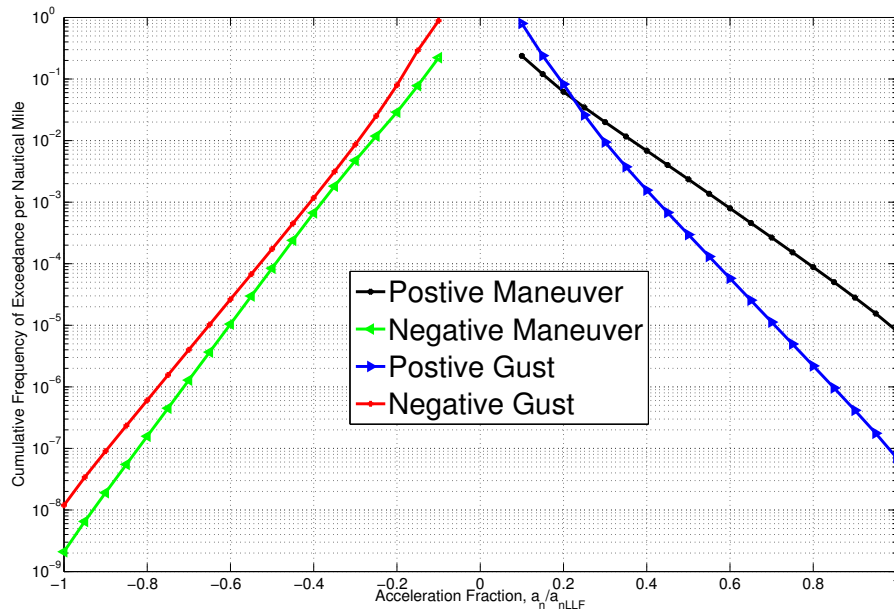


Figure 2: Gust and Maneuver spectra exceedance curves

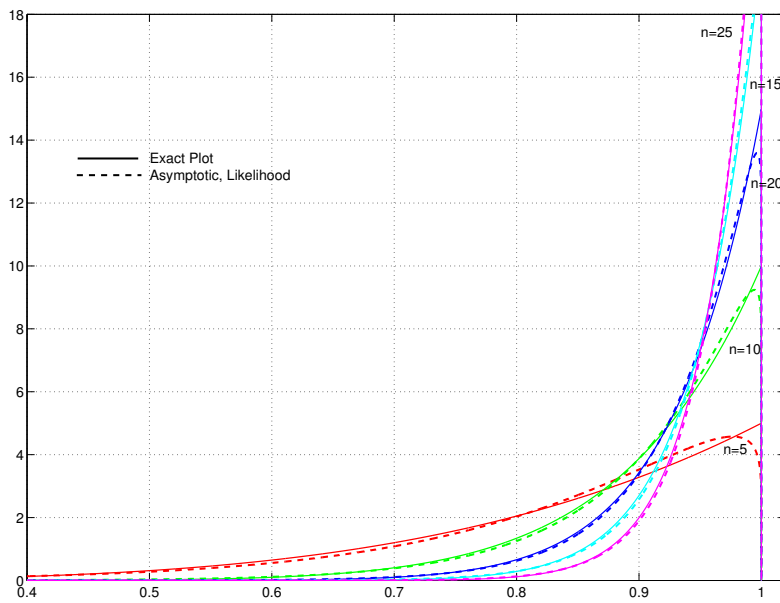


Figure 3: EVDs for uniform distribution comparison between exact and MLE approaches

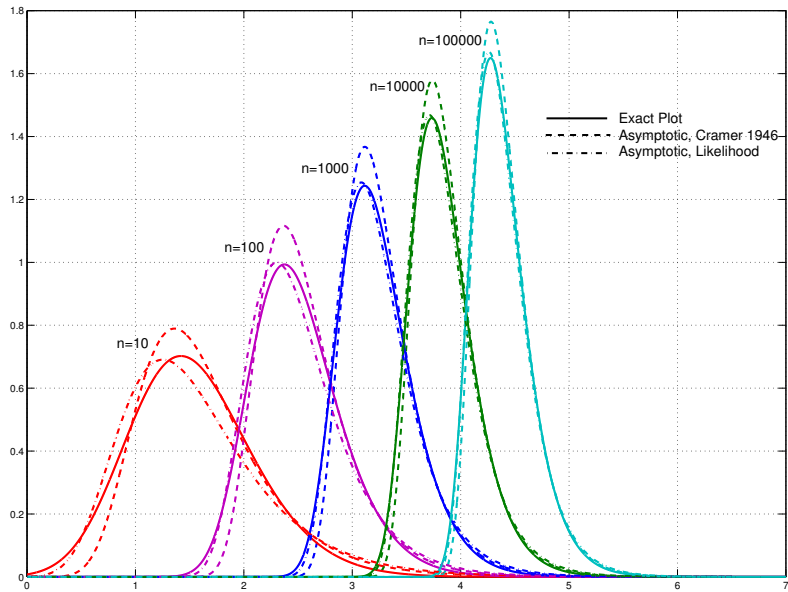


Figure 4: EVD for the normal distribution comparison between exact, asymptotic (Cramer, 1946), and asymptotic (maximum likelihood) approaches

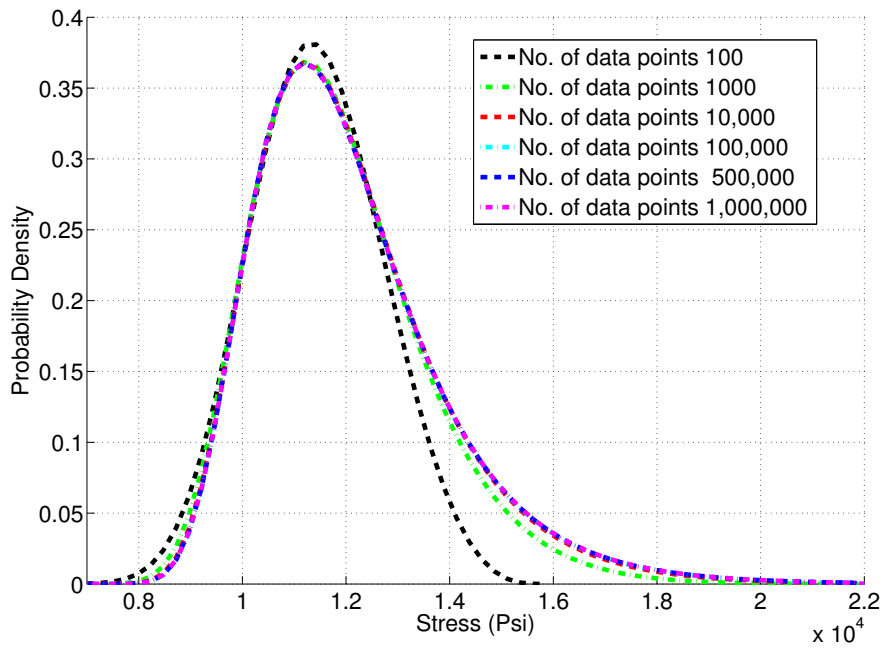


Figure 5: Convergence study to determine number of data points required for a converged EVD

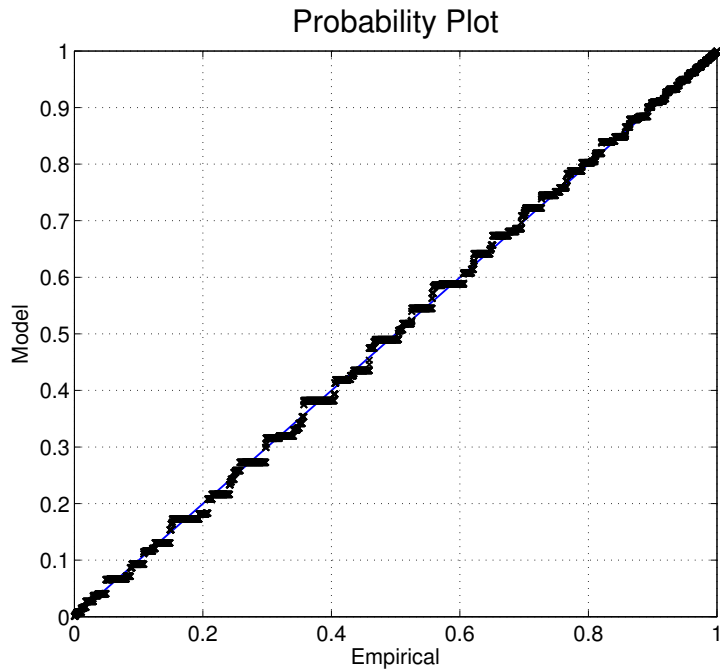


Figure 6: Probability plot for model checking

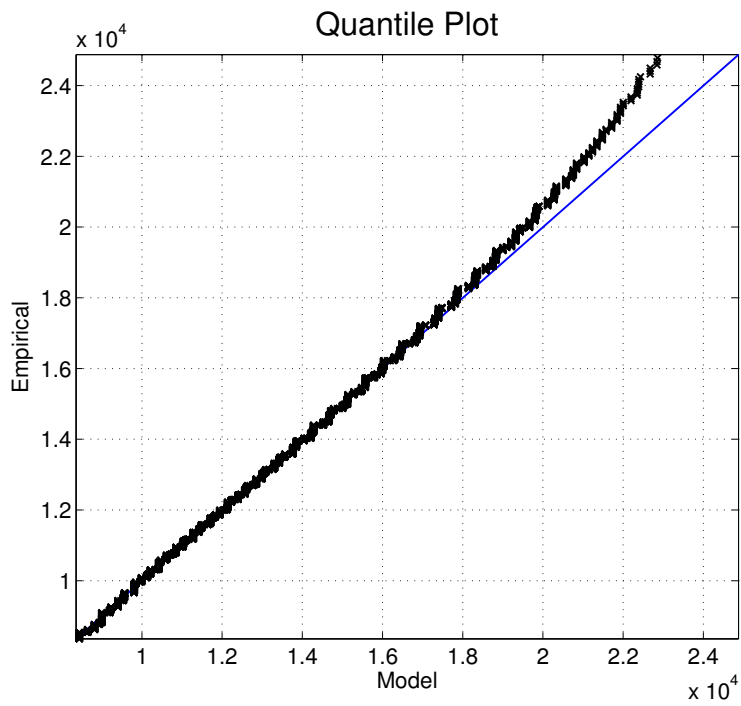


Figure 7: Quantile plot for model checking

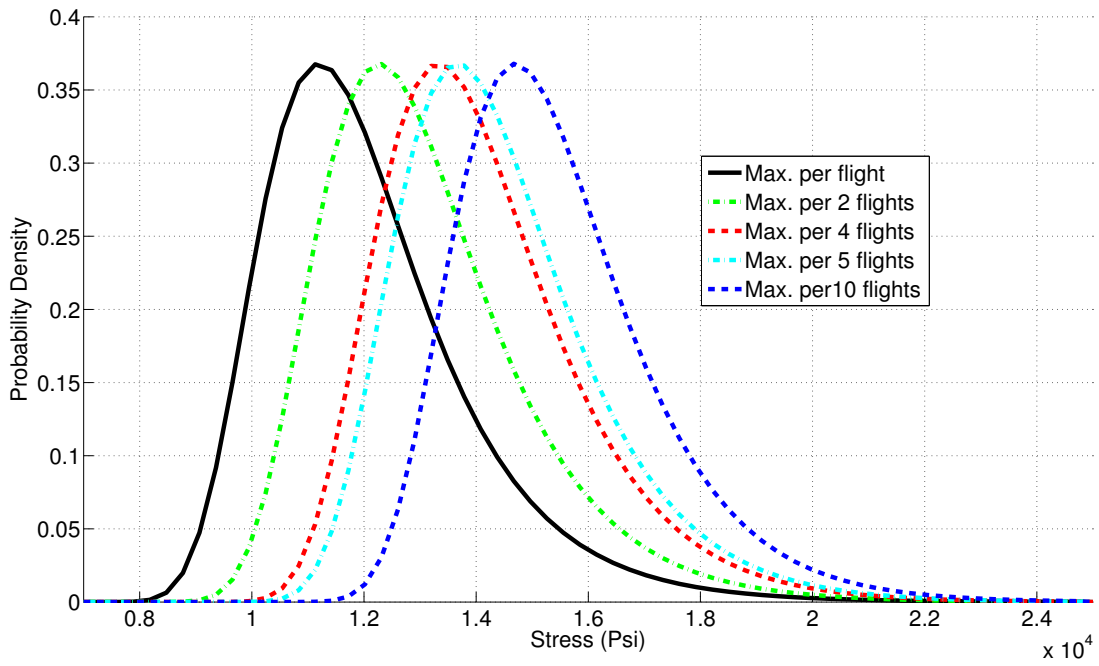


Figure 8: Effect of sample size (p) on an EVD

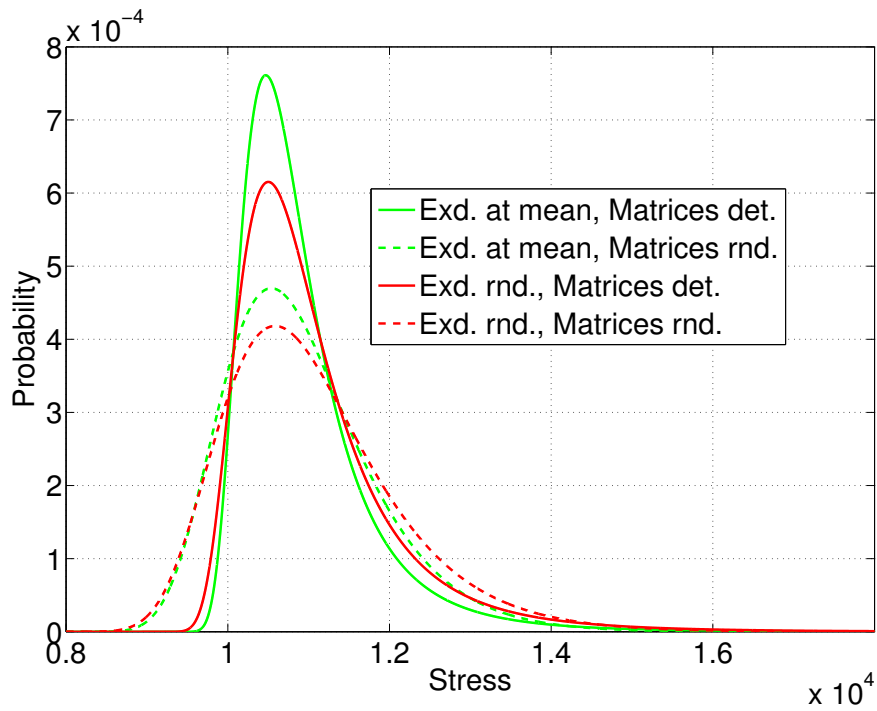


Figure 9: Effects of matrices with exceedance at mean and random

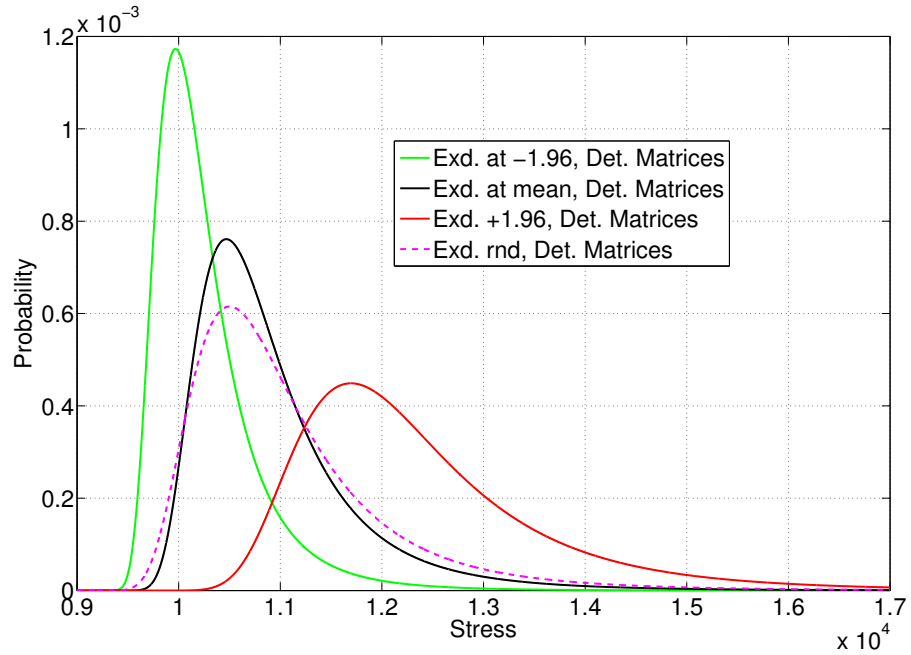


Figure 10: Effects of Exceedance curve with deterministic flight length-velocity and flight length-weight matrices

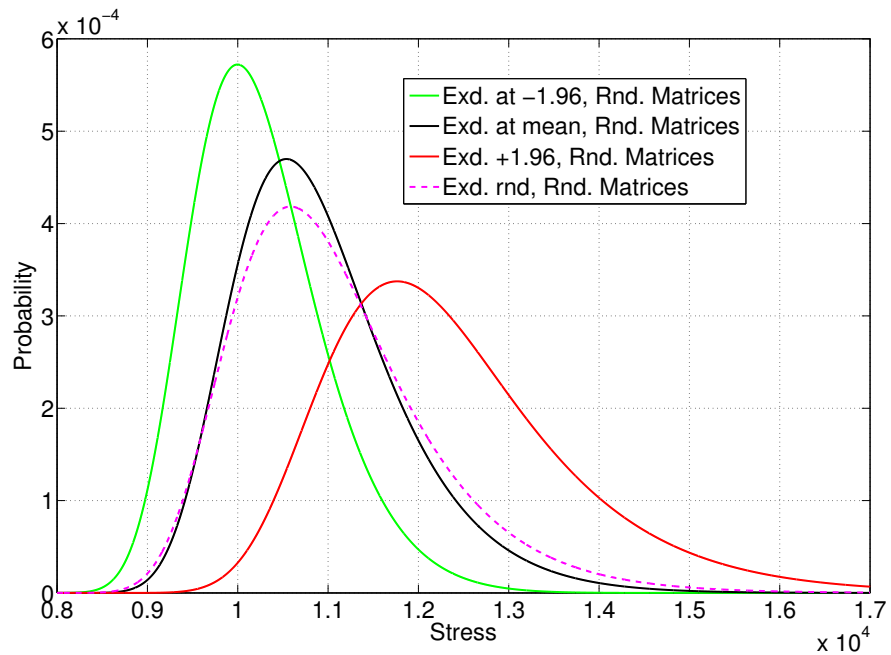


Figure 11: Effects of Exceedance curve with random flight length-velocity and flight length-weight matrices

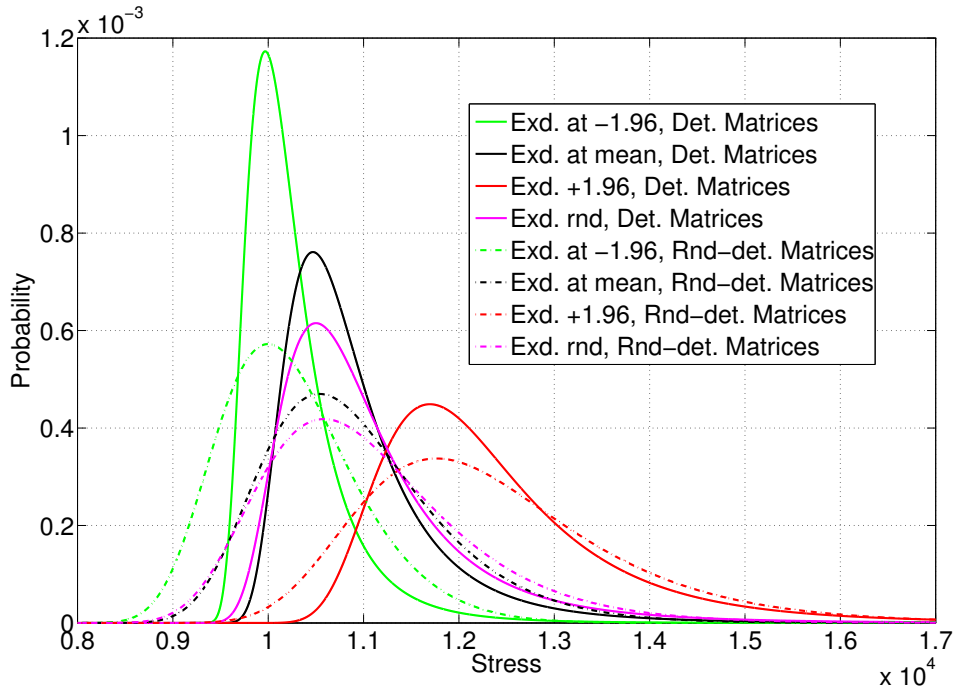


Figure 12: A summary of effects of Exceedance curves

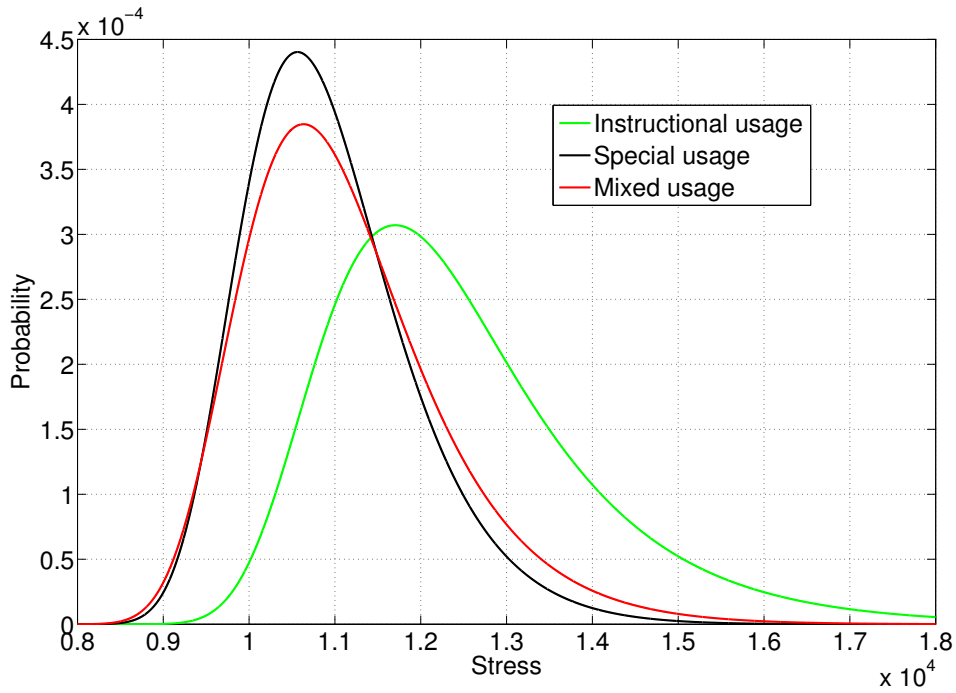


Figure 13: Effects of usage

Table 1: List of the load spectrum parameters

Variable	Description
Number of flights	Number of flight to be generated in the flight spectrums
Taxi Stage	If Taxi stage is generation in the flight spectrum
Usage	Airplane usage as per FAA Report AC23-13 A [20]
Exceedance curves	Usage exceedance curves
Maneuver Load Limit Factor	Maximum load limit factors for maneuver load
Gust Load Limit Factor	Maximum load limit factors for gust load
Maximum Ground Stress	Airplane ground stress in psi with negative sign
Maximum One g Stress	One g stress of an airplane in psi.
Maximum A/C Velocity	Average Speed During Flight, V_{NO} (Maximum aircraft safe cruise speed) or V_{MO} (Maximum operating limit speed). In nautical miles.
Flight Length-Velocity Matrix	Probabilistic flight length and airspeed data
Flight Length-Weight Matrix	Probabilistic flight length and weight data

Table 2: Flight length and airspeed data

		Average flight speed, % design velocity					
Flight time (Hours)	% of flights	1.000	0.950	0.900	0.850	0.800	0.750
0.500	0.450	0.050	0.300	0.500	0.100	0.050	0.000
0.750	0.400	0.000	0.200	0.300	0.350	0.100	0.050
1.000	0.100	0.000	0.050	0.300	0.450	0.150	0.050
1.250	0.050	0.000	0.050	0.200	0.250	0.200	0.300

Table 3: Flight length and weight data matrix

		Weight (one-g-stress and ground stress) percentage					
Flight time (Hours)	% of flights	1.000	0.950	0.900	0.850	0.800	0.750
0.500	0.450	0.050	0.300	0.500	0.100	0.050	0.000
0.750	0.400	0.000	0.200	0.300	0.350	0.100	0.050
1.000	0.100	0.000	0.050	0.300	0.450	0.150	0.050
1.250	0.050	0.000	0.050	0.200	0.250	0.200	0.300

Table 4: EVD parameters for uniform distribution using MLE approach

Number of samples (p)	Location parameter (μ)	Scale parameter (σ)	Shape parameter (ξ)
5	0.8268	0.1547	-0.8931
10	0.9071	0.0880	-0.9475
15	0.9364	0.0606	-0.9518
20	0.9524	0.0467	-0.9817
25	0.9619	0.0377	-0.9898

Table 5: Results for normal distribution comparison between Sampling, Cramer [12] and MLE solutions

p	From Samples		Cramer [1946]		MLE Approach	
	a_p	u_p	a_p	u_p	a_p	u_p
10	2.1794	1.2786	2.1459	1.3619	1.8776	1.2552
100	3.0022	2.3130	3.0349	2.3663	2.7173	2.3040
1000	3.6302	3.0891	3.7175	3.1165	3.4092	3.0841
10000	4.2009	3.7161	4.2919	3.7384	3.9914	3.7123
100000	4.7222	4.2637	4.7985	4.2802	4.5372	4.2611

Table 6: Load spectrum input parameter values

Description	Parameter value	
	Case I	Case II
Case #	100,000	100,000
Number of flights (data points)	100,000	100,000
Maneuver load limit factors	3.800, -1.520	3.800, -1.520
Gust load limit factors	3.155, -1.155	3.155, -1.155
Ground stress (psi)	-4520.0	-4100.0
One-g-stress (psi)	7410.0	6674.0
Aircraft velocity (knots)	165.0	148.50

Table 7: Convergence study for number of data points (N)

Number of data points (N)	Location (μ)	Scale (σ)	Shape (ξ)	EVD Distribution
100	10957	1257	-0.2610	Weibull
1000	11140	1395	-0.0645	Frechet
10,000	11216	1430	0.0005	Frechet
100,000	11224	1430	0.0189	Frechet
500,000	11222	1434	0.0178	Frechet
1,000,000	11224	1433	0.0179	Frechet

Table 8: Effects of number of samples on EVD parameters

Number of samples (p)	Location (μ)	Scale (σ)	Shape (ξ)	EVD Distribution
1	11224	1433	0.0179	Frechet
2	12263	1459	-0.0012	Weibull
4	13308	1457	-0.0126	Weibull
5	13642	1454	-0.0155	Weibull
10	14671	1429	-0.0191	Weibull

Table 9: Flight length and weight data matrix

		Weight and flight speed percentage										
Hours	Flt. %	1.00	0.98	0.96	0.94	0.92	0.90	0.88	0.86	0.84	0.82	0.80
0.50	0.02	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.55	0.04	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.60	0.06	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.65	0.08	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.70	0.10	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.75	0.12	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.80	0.16	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.85	0.12	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.90	0.10	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
0.95	0.08	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
1.00	0.06	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
1.05	0.04	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02
1.10	0.02	0.02	0.05	0.08	0.10	0.12	0.26	0.12	0.10	0.08	0.05	0.02

Table 10: Effects of exceedance curve with deterministic matrices

Matrices	Fixed	Random	Fixed	Random
Exceedance	Fixed	Fixed	Random	Random
# of data points	100,000	100,000	1000 × 10,000	1000 × 10,000
Distribution	Frechet	Weibull	Frechet	Weibull
Location	10551	10501	10600	10563
Shape	491	784	607	880
Scale	0.1830	-0.0467	0.1788	-0.0248
Mean	10943	10919	11080	11050
STD	863	949	1057	1093

Table 11: Effects of exceedance curve with deterministic matrices

Random #	-1.96	0.0	1.96	Random
# of data points	100,000	100,000	100,000	1000 × 10,000
Distribution	Frechet	Frechet	Frechet	Frechet
Location	10017	10551	11799	10600
Shape	317	491	827	607
Scale	0.1674	0.1830	0.1333	0.1788
Mean	10263	10942	12401	11080
STD	540	863	1311	1057

Table 12: Effects of exceedance curve with random matrices

Random #	-1.96	0.0	1.96	Random
# of data points	100,000	100,000	100,000	1000 × 10,000
Distribution	Weibull	Weibull	Weibull	Weibull
Location	9931	10501	11752	10563
Shape	646	784	1090	880
Scale	-0.0994	-0.0467	-0.0129	-0.0248
Mean	10246	12368	11752	11050
STD	740	1376	1090	1093

Table 13: Effects of usage

Usage	Instructional	Special	Mixed
Distribution	Frechet	Frechet	Frechet
# of data points	1,000,000	1,000,000	1,000,000
Location	11736	10507	10596
Shape	1198	837	957
Scale	0.0274	-0.0662	-0.0538
Mean	12462	10939	11113
STD	1596	993	1170