Analyzing fatal accidents in aviation using extreme value theory

Kumer Pial Das, Ph.D.* and Asim Kumer Dey†

Abstract: Even though air travel is considered a safe means of transportation, when aviation accidents do occur they often result in fatalities. Fortunately, the most extreme accidents occur rarely. However, 2014 was the deadliest year in the past decade causing 111 plane crashes, and among them worst four crashes cause 298, 239, 162, 116 deaths. In this study we want to assess the risk of the catastrophic aviation accident by studying historical aviation accidents. Applying a generalized Pareto model we predict the maximum fatalities from an aviation accident in future. The fitted model is compared with some of its competitive models. The uncertainty in the inferences are quantified using simulated aviation accident series, generated by bootstrap resampling and Monte Carlo simulation.

Keywords: Aviation accident, Extreme value model, Generalized Pareto distribution, Return level, Uncertainty, Bootstrap sampling , Monte Carlo simulation.

1 Introduction

Aviation accidents cause enormous loss of life and massive monetary costs worldwide. In 2002-2011 there were a total of 250 worldwide fatal accidents, which resulted in 7,148 fatalities (CAA, 2014). And among these 250 fatal accidents around 20% accidents each cause more than 50 fatal injuries. Although there is an overall decreasing trend in the number of fatal accidents (Tolan et al., 2015), there is no trend in the number of extreme fatal accidents which cause huge fatality. Quantification of the large accidents which have far reaching effect (fatality) would provide objective guidance in long-term planning and response for manufacturers, insurers and re-insurers.

Every year different organizations like Federal Aviation Administration (FAA), International Civil Aviation Organization (ICAO), Civil Aviation Author-

*Associate Professor, Department of Mathematics, Lamar University, Texas, USA. Email: kumer.das@lamar.edu

†Graduate Student, University of Texas at Dallas, Texas, USA. Email: adey@utdallas.edu

ity (CAA) of UK as well as different manufacturer corporations like Boeing, Airbus give valuable reports on aviation accident. These reports mainly focus on descriptive accident statistics like number as well as rate of worldwide fatal accident and fatalities by year, nature of flight, aircraft age and weight group (Boeing, 2013 and Airbus, 2014). Sometimes they make analysis on different causal factors for these accidents (CAA, 2014). Some research on aviation accident focus on behaviors that are associated with Loss-of-Control (LOC) events (Lasek et al., 2010). Some others calculate the occurrence probabilities of serious incidents using safety critical measures runway overrun, which indicate the actual landing distance (Wang et al., 2014). But there is no well known research which focuses on modeling the number of fatal injuries in individual aviation accident. Using different tail models (power law, stretched exponential and log-normal), one can assess the likelihood of a large event like serious aviation accident (Clauset and Woodard, 2013). The classical extreme value theory, for example, generalized Pareto distribution (GPD) can also be used to model extreme events like serious aviation accident (Reich and Porter, 2013).

In this paper we study aviation accidents from 1982 to 2014, their pattern within the period, the number of fatalities from them, etc. A prediction of possible number of fatal injuries for an extreme aviation accident in the future is made using peaks over threshold method (i.e., generalized Pareto distribution approach). We compare the the fitted generalized Pareto distribution (GPD) model with other long tail models to select better model for aviation accident using different measure of goodness of fit. We also quantify the uncertainty in the inferences using data generated by bootstrap resampling and Monte Carlo simulation. The overall aim of this paper is to measure the risk of aviation accident in terms of fatal injury. The challenge is in selecting reasonable tail model for this data as well as in measuring uncertainty in the inferences.

Rest of the paper is organized into five sections: the Section 2 introduces a brief description of the extreme value models and the generalized Pareto distribution (GPD); Section 3 proposes a GPD model for aviation fatal injury; Section 4 describes different uncertainty measurements in the inference of extreme fatalities through simulated accident series; Section 5 analyze goodness of fit of the fitted model and also made a comparison among different possible model in this situation; and finally, conclusions appear in Section 6.

2 Extreme Value Model and Generalized Pareto Distribution(GPD)

In many statistical applications, the interest is centered on estimating some population characteristics such as average or median of a process based on random samples taken from a population under study. However, in extreme value analysis, we are not interested in estimating the average but rather we want to quantify the behavior of the process at unusually large or small levels, such as we are interested in estimating the maximum or the minimum. Extreme value theory (EVT) deals with the extreme deviations from the median of probability distributions and seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than a certain large value. Usual bulk statistics tries to describe main part of distribution; may ignore outliers. But EVT tries to characterize the tail of the distribution; keeps only the extreme observations (Fig. 1).

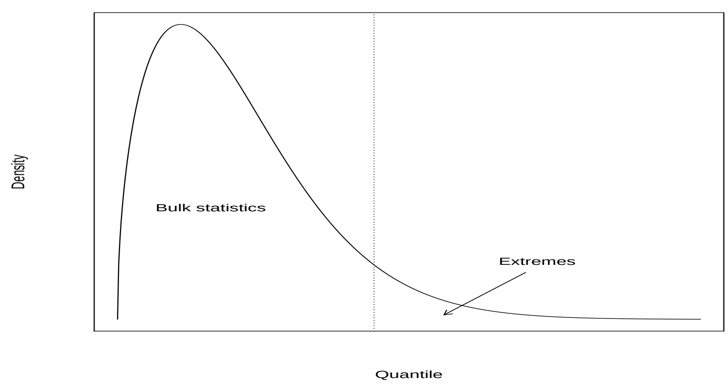

Figure 1: Extreme Value Model

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed (iid) random variables with common distribution function F. Extreme value analysis focuses on the statistical behavior of the maximum value observed, i.e.,

$$M_n = \max\{X_1, X_2, \cdots, X_n\}$$

In applications, the X_i usually represent values of a process measured on a regular time-scale at time i such as hourly measurements of sea level, or daily mean temperature so that M_n represents the maximum of the process over n time units of observation (Coles, 2001). If n is the number of observations in a year, then M_n corresponds to the annual maximum.

Using the fact that $X_1, X_2, \cdots, X_n$ are iid random variables

$$\begin{aligned} Pr(M_n \leq z) &= Pr\{X_1 \leq z, \cdots, X_n \leq z\} \\ &= Pr\{X_1 \leq z\} \times \cdots \times Pr\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned} \tag{2.1}$$

In practice, we might not know the distribution function F but according to the extremal types theorem (Fisher and Tippett, 1928), if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$Pr\{(M_n - b_n)/a_n \leq z\} \longrightarrow G(z) \quad \text{as} \quad n \longrightarrow \infty$$

with G being a non-degenerate distribution function, then G belongs to the following family of of models having a distribution function of the form :

$$G(z) \quad = \quad \exp\left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \tag{2.2}$$

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. This is the generalized extreme value (GEV) family of distributions. The model has three parameters: a location parameter, μ; a scale parameter, σ; and a shape parameter, ξ. The shape parameter ξ governs the tail behavior of the distribution. The subfamilies defined by $\xi \to 0$, $\xi > 0$ and $\xi < 0$ correspond, respectively, to the Gumbel, Frechet and Weibull families (Coles, 2001). This method is commonly known as block maxima (BM) approach.

However, BM is a wasteful method because maxima in some blocks (years) can be much below several high-order statistics in other blocks (years). Thus, in many practical applications an important part of the information such as large values other than the maxima occurring during the same year would be lost if we use only annual maximum value (Castillo et al., 2004). Consequently, the BM method was extended (Smith and Weissman, 1994; Smith, 1986; Arns et al., 2013) in order to include more than one independent observations from each block such as $r > 1$/year into the sample. However, if the inter-block variability is large such as if one year contains more extremes than another then incorporating more of the observed extreme data (r-largest observation per block) in the analysis can be wasteful (Coles, 2001).

In contrast to BM approach, it is more useful (efficient) to analyze the values of random variables that exceed a given threshold value if an entire time series of, say, hourly or daily observations is available. This method is commonly known as peaks over threshold (POT) Approach. Hence, a POT derived sample comprises not only one or a fixed number of events per year. It rather allows for a more rational selection of events fulfilling the criteria of being “extreme”.

In POT approach observations that exceed a given threshold μ are called exceedances over a threshold. Pickands (1975) demonstrates that for large enough μ the observations x, provided $x > \mu$, approximately follow a generalized Pareto distribution(GPD) with distribution function:

$$F(x) = 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} \tag{2.3}$$

for some $\sigma > 0$ and ξ, where σ and ξ are scale and shape parameters, respectively. If $\xi < 0$, the distribution of excesses has an upper bound of $-\sigma/\xi$; if $\xi > 0$ the distribution has no upper limit. And when $\xi \to 0$, GPD reduces to an exponential distribution with mean σ.

Different graphical approaches are used to select a threshold (μ) for a particular data set. Among them the mean residual life (MRL) plot is commonly used. If the tail data follow a GPD with a lower bound of μ, then the MRL plot should be approximately linear for values above μ. Therefore, the recommendation is to select the smallest μ which gives a linear MRL plot. Having determined a threshold, the parameters of the GPD can be estimate by the maximum likelihood method.

It can be derived from Eq. 2.3 (Coles, 2001) that for $x > \mu$,

$$\bar{F} = Pr\{X > x\} = \zeta_\mu \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \tag{2.4}$$

where $\zeta_\mu = Pr(X > \mu)$. Hence, the level x_m that is exceeded on average once every m observations is the solution of

$$\zeta_\mu \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m} \tag{2.5}$$

Rearranging,

$$x_m = \begin{cases} \mu + \frac{\sigma}{\xi}[(m\zeta_\mu)^\xi - 1] & \text{for } \xi \neq 0, \\ \mu + \sigma \log(m\zeta_\mu) & \text{for } \xi = 0, \end{cases} \tag{2.6}$$

provided m is sufficiently large to ensure that $x_m > \mu$. x_m is the $m-$observation return level. However, it is often more convenient to give return levels on an annual scale, so that the $N-$year return level is the level expected to be exceeded once every N years. If there are n_y observations per year, this corresponds to the $m-$observation return level, where $m = N \times n_y$. Hence, the $N-$year return level is defined by

$$x_N = \begin{cases} \mu + \frac{\sigma}{\xi}[(Nn_y\zeta_\mu)^\xi - 1] & \text{for } \xi \neq 0, \\ \mu + \sigma \log(Nn_y\zeta_\mu) & \text{for } \xi = 0, \end{cases} \tag{2.7}$$

Estimation of return levels requires the substitution of parameter values by their estimates. For μ and ξ this corresponds to substitution by the corresponding maximum likelihood estimates, but an estimate of ζ_μ, the probability of an individual observation exceeding the threshold μ, is also needed. This has a natural estimator of $\hat{\zeta}_\mu = k/n$, the sample proportion of points exceeding μ. Since the number of exceedances of μ follows the binomial $Bin(n, \zeta_\mu)$ distribution, $\hat{\zeta}_\mu$ is also the maximum likelihood estimate of ζ_μ (Coles, 2001).

3 Modeling Aviation Fatalities with Generalized Pareto distribution (GPD)

The objective of this study is to develop a predictive model which will give estimates of potential fatal injuries for an extreme aviation accident in the future. Data are obtained from the National Transportation Safety Board (NTSB). The NTSB aviation accident database contains information within the United States, its territories and possessions, and in international waters. The data consists of 15,187 aviation accidents from 1982 to 2014 which causes one or more fatal injuries. Figure 2 demonstrates time series plots of fatal injuries from aviation accident from from 1982 to 2014. The plot shows that 27 accident caused more than 150 fatalities and 5 of which exceeds 250 fatalities. The plot does not show any obvious trend in the data.

The histogram of fatal injuries (Fig. 3) shows that 15144 accidents, which is 99.72% of total accidents, are responsible for between 1 to 25 fatalities. There are 28 accidents with 26 to 50 fatal injuries, and after that frequencies gradually decrease with the number of fetal injuries. The histogram also indicates that the data follow a long tailed distribution. GPD is commonly used to model a long tailed distribution. The observations that exceed a given threshold, μ, in the tail of the distribution approximately follow GPD. However, a crucial step in this analysis is to select an appropriate threshold μ. If μ is too small, then the GPD will not fit the tail distribution and the estimates of the parameters (σ, ξ) may be biased. On the other hand, if μ is too large, the GPD may fit well, but fewer observations will be available to estimate the parameters and the estimates will suffer from increased variance.

We want to select the smallest μ in mean residual life (MRL) plot (Fig. 4) above which the MRL plot will be approximately linear. The MRL plot shows that the reasonable threshold should be around 25. Parameter stability plots (Fig. 5) have also been used to identify an appropriate threshold. In practice, the threshold μ should be chosen where the shape and modified

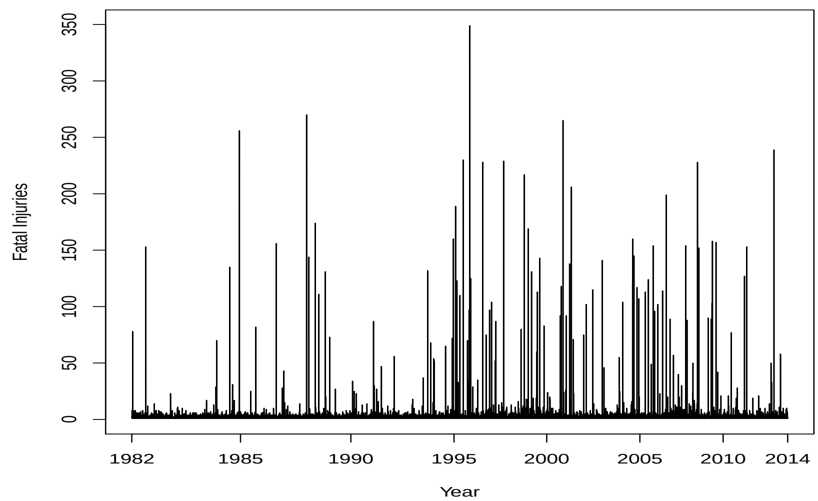

Figure 2: Numbers of fatal injuries from aviation accidents, 1982 - 2014.

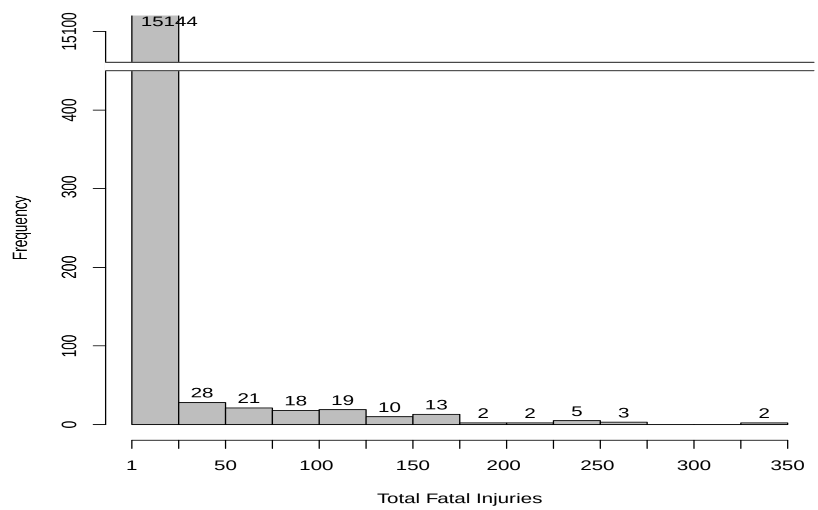

Figure 3: Histogram of fatal injuries from aviation accidents.

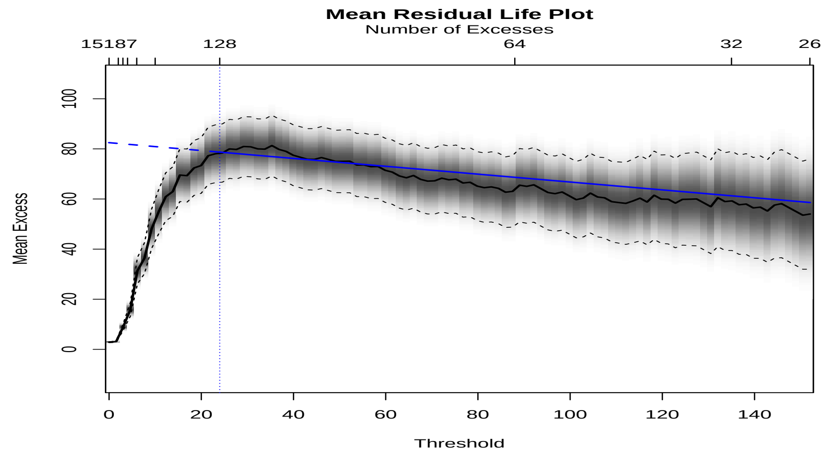

Figure 4: Mean residual life plot.

scale parameters remain constant in parameter stability plot after taking the sampling variability into account. Parameter stability plots also suggest to select threshold around 25.

However, instead of considering one threshold we fit the GPD model with different thresholds within the range of 15 to 50. The fitted models are summarized in Table 1, which shows, for different thresholds, the number of observations and proportion of the data beyond the threshold as well as the maximum likelihood estimates (MLE) of the GPD parameters (μ and ξ). However, Fig. 6 shows that all the QQ-plots for $\mu = 24$, 35 and 50 are approximately 45−degree line but the plot for $\mu = 15$ has a clear deviation from 45−degree line. Therefore Fig. 6 suggests that the models for $\mu = 24$, 35 and 50 appear to provide a better fit than $\mu = 15$.

Table 3.1: Table 1: Estimates from the GPD model for different threshold values μ.

Threshold (μ)	15	24	35	50
$\#intail$	162	127	107	95
$Pr(X > \mu)$	0.011	0.008	0.007	0.006
$\hat{\sigma}$	72.57	93.39	99.89	90.34
$\hat{\xi}$	-0.039	-0.186	-0.232	-0.198

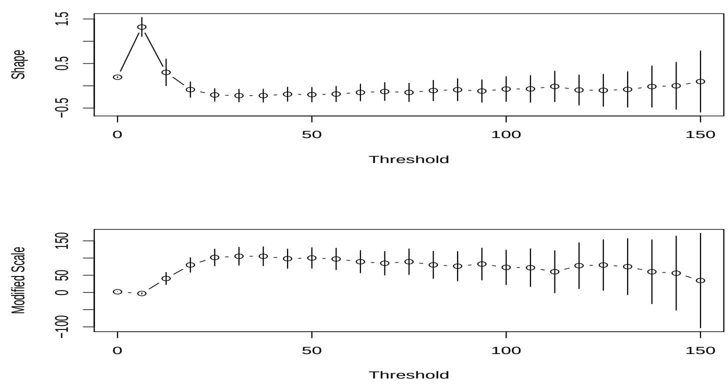

Figure 5: Parameter stability plots.

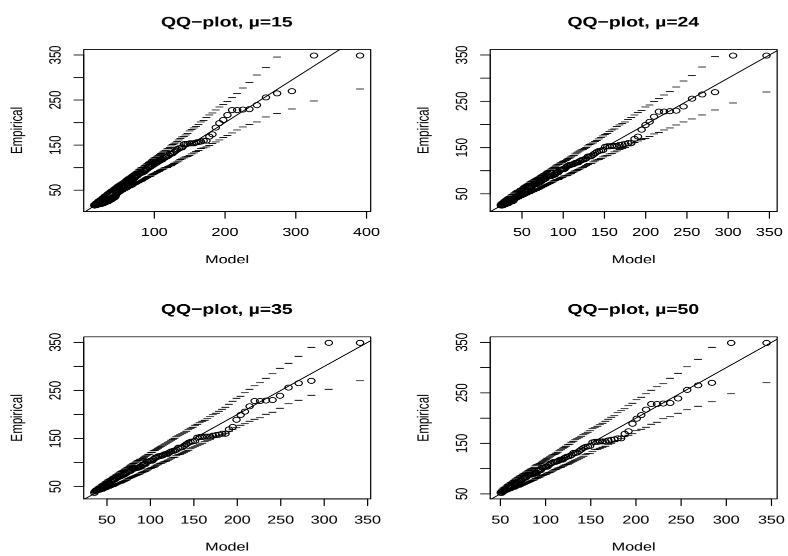

Figure 6: QQ-plots of the GPD models for different thresholds.

Now considering 24 as the threshold the extreme quantile (x_N) which is expected to exceed on average once every N y years can be obtained by Eq. 2.7. Table 3 summarizes the $N-$year return levels ($\hat{x}_N$) for different values of N. For example, 10 year return level is 273 fatal accident, which means that 273 or more fetal injuries is likely to occur once every 10 years.

Table 3.2: Table 2: Estimated fatal injuries expected to be exceeded once every m years.

m (Year)	Fatal injury
5	237
10	273
20	303
30	319
50	338

4 Uncertainty quantification

Estimating the probabilities of the extreme aviation accidents or estimating the extreme quantiles is very important to manufacturers, insurers and re-insurers. However, models do not provide a perfect representation of reality, and the inference what we draws from the models are not certain. Thus in many cases assessing uncertainties is very important, especially when we extrapolate the observed data.

Several methods have been proposed to quantify different sources of uncertainties in the extreme value statistics (Muller et al., 2015; Wehner, 2010). In this paper, we focus on uncertainties in model parameters and model misspecification due to threshold selection.

4.1 Uncertainty from estimate of the model parameters

A main source of uncertainty in quantifying extreme aviation accident statistics comes from uncertainty in the fit of the GPD parameters. This uncertainty is usually measured by standard error, coefficient of variation (CV) and confidence level of the parameter of interest. Coefficient of variation (CV), is defined as the ratio of the standard deviation to the mean, is a standardized measure of dispersion of a distribution. The estimated confidence level is claimed to include the true parameter value with a specified probability.

In the previous section we obtained the estimates of an $N-$year return level for different values of N. Now we want to calculate their uncertainties on the basis of their standard error, coefficient of variation (CV) and confidence interval. However, a standard 95 percent confidence interval for x_N depend on the asymptotic normality of $(\hat{x}_N)$. This asymptotic normality assumption is questionable for data with small observations (Carpenter and Bithell, 2000). Accordingly, we may want to construct a confidence interval that does not depend on this assumption. Bootstrapping provides a ready, reliable way to do this. Here confidence intervals for x_N are obtained by non-parametric bootstrap approach (Dey and Das, 2015).

Table 4.1: Table 3: Mean, standard Error, Coefficient of Variation and 95% Confidence Intervals for Fetal Injury from Bootstrap Resampling.

Year (m)	Mean Fatal injury	Standard error	Coefficient of variation	95% Lower Limit	95% Upper Limit
5	234.9109	15.77122	6.713702	202.3180	264.3762
10	268.9248	20.27671	7.539917	225.5305	304.3111
20	298.7887	25.73977	8.614707	242.7753	344.2439
30	314.5683	29.28162	9.308511	251.3096	367.3089
50	332.8647	34.01564	10.219058	259.4316	394.0819

In brief, the original data set consists of 127 observations above the threshold $\mu = 24$. A random sample of size 127 is selected with replacement from the original data set to obtain a bootstrap data set. From this bootstrap data set the maximum likelihood estimates $\hat{\sigma}$ and $\hat{\xi}$ of the parameters are calculated and using these estimates the return level X_N is determined for each return period N =5, 10, 20, 30, and 50.
This procedure is repeated 2,000 times. Therefore, for each return period (N) we have 2,000 values of return level. From this set of 2,000 values the mean return level, standard error, coefficient of variation, and 95% confidence intervals are calculated. The information is summarized in Table 5. Results in Table 5 shows that the standard error, the coefficients of variation and 95% confidence intervals for return level increase with the return period. This means uncertainty increases with the model extrapolation.

4.2 Uncertainty from model misspecification

The GPD approach can fail because the assumptions of strict independent and identically distributed random variables are violated, but even if they are fulfilled, the chosen threshold parameter μ may be inappropriately small, thereby leading to a poor extreme-value approximation.

Although in this study we have chosen 24 as the threshold, from section 3 we can see 35 and 50 are also good candidates. So the main potential source of misspecification here is threshold selection (Süveges and Davison, 2010). Here our aim is to assess the biases incurred by the estimated parameters and when a GPD distribution with threshold 24 is fitted to data from GPD distribution with threshold 15, 24, 35, and 50 (Dupuis and Tawn, 2001).

A suitable method to analyze such misspecification problems is through a simulation study. Data, generated from different generalized Pareto models (Table 1) using Monte Carlo simulation, are fitted GPD with threshold

$\mu = 24$. Standard bias and standard root mean square error (RMSE), are defined by

$$\text{Standard bias} = \frac{E(\hat{\theta}) - \theta}{\theta} \tag{4.1}$$

$$\text{Standard RMSE} = \frac{\sqrt{E((\hat{\theta} - \theta)^2)}}{\theta} \tag{4.2}$$

where, $\hat{\theta}$ is an estimate of θ (parameter or return level).

Table 4.2: Table 4: Standard Bias and Standard RMSE (Bracket) on parameters and return levels (10 & 20 years).

Model	σ	ξ	x_{10}	x_{20}
GPD$(15, 72.57, -0.039)$	0.017[0.122]	0.574[−2.297]	−0.023[0.103]	−0.023[0.123]
GPD$(24, 93.39, -0.186)$	0.024[0.123]	0.129[−0.471]	−0.018[0.076]	−0.018[0.086]
GPD$(35, 99.89, -0.232)$	0.235[0.272]	0.523[−0.648]	0.028[0.071]	0.009[0.073]
GPD$(50, 90.34, -0.198)$	0.556[0.578]	1.148[−1.238]	0.110[0.131]	0.069[0.104]

Bias and RMSE (Bracket) on parameters and return levels (10 & 20 years) are calculated from 2,000 replications for each case. For example, in the first case, a sample of size 162 is generated from GPD(15,72.57,-0.039). GPD with threshold 24 is fitted from this data. Using this estimated parameters we calculate return levels for 10 & 20 years. This procedure is repeated 2,000 times. Therefore for each parameter and return level we have 2,000 values. Then the standard bias and standard RMSE on each parameter and return level can be easily calculated from 2,000 repeated values. It is seen from Table 6 that bias and RMSE for parameters and return levels are negligible for all given cases. This implies the fitted GPD model with threshold $\mu = 24$ is robust for GPD model with threshold 15, 24, 35, and 50.

5 Goodness of Fit

A models is only an approximation of reality. So it is very important to determine whether our model is good enough to fit the aviation accident data. The most common graphical methods that are used for model validation are $Quantile - Quantile$ (Q-Q) plot and the $Probability - Probability$ (P-P) plot. Figure 7 shows that both the P-P plot and Q-Q plot are approximately $45-$degree line. This implies that the GPD model is a reasonable fit for modeling this data set.

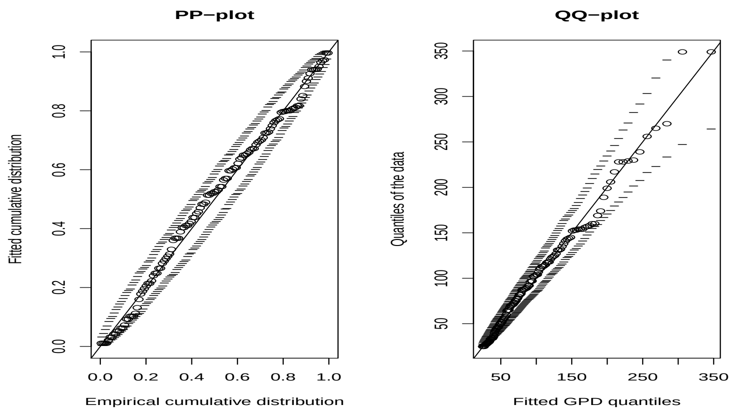

Figure 7: Diagnostic plots for the fitted GPD model.

One useful technique to check the goodness of fit of a model is to compare the model with its competitors. The two special cases of GPD is exponential distribution and Pareto distribution. So it is important to check whether or not the GPD model performs better than it's special cases. We can fit the Pareto model and exponential model from the data above threshold 24. For Pareto distribution the MLE of shape and scale parameters are 0.827 and 25, respectively. Again for exponential distribution the MLE of rate parameter is 0.00972. Figure 8 compares the data density with different model densities and it is seen from the figure that exponential and Pareto model understate the probabilities of fatal injuries in the near tail and overstate in the far tail. But GPD shows better fit in both near tail and far tail. Figure 9 shows that distribution function (cdf) of exponential and Pareto model overstates the empirical distribution function in near tail and understates in far tail. Again GPD gives better fit in both near tail and far tail.

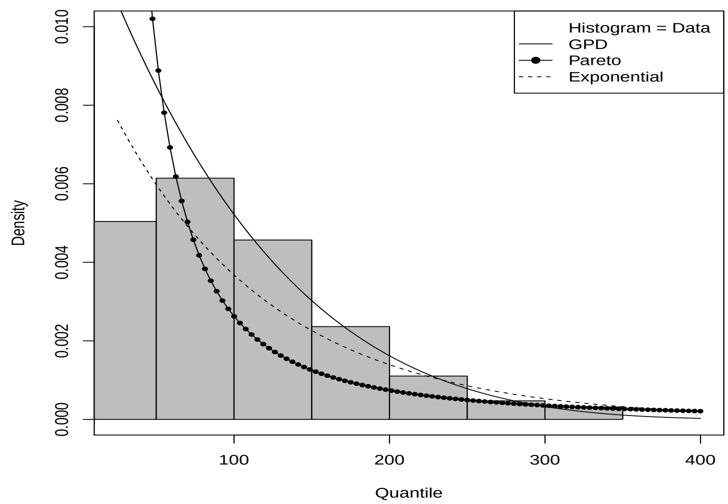

Figure 8: Models vs. data density plot.

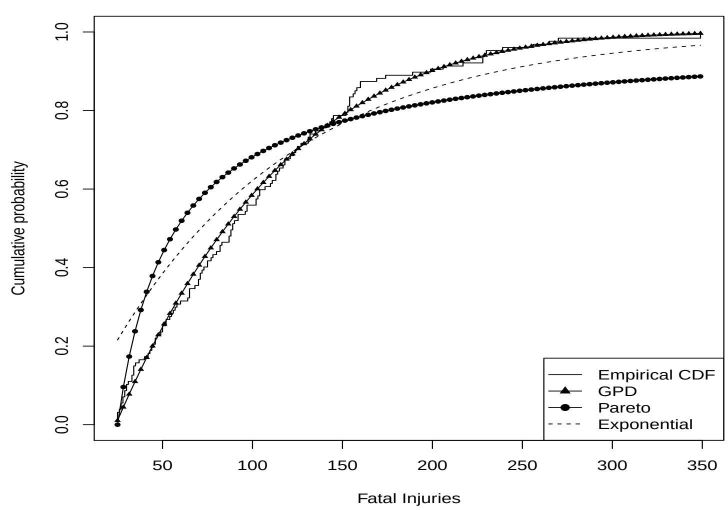

Figure 9: Models vs. data cdf plot.

Another approach to the model selection problem is to formulate the problem into a hypothesis testing framework. Table 7 summarize the results of two commonly used goodness of fit tests. The $p-$values of Kolmogorov-Smirnov (KS) test and Anderson-Darling test (Klugman et al., 2008) for both exponential and Pareto model are less than 0.001. This implies that exponential and Pareto model are not appropriate for this data. And because the $p-$values for GPD are greater than 0.05, at a 5% level of significance GPD is a plausible model for this data.

Table 5.1: Table 5: $p-$values of goodness of fit tests for different models.

Model	Kolmogorov-Smirnov (KS) Test	Anderson-Darling (AD) Test
$\mathrm{GPD}(24, -0.186, 93.39)$	0.1667	0.2174
$\mathrm{Pareto}(24, 0.827, 25)$	< 0.001	< 0.001
$\mathrm{Exponential}(0.00972)$	< 0.001	< 0.001

Though we have illustrated our methodology in detail only for the threshold 24, we have also applied the technique to data for thresholds μ =35, and 55. These analysis are summarized in Table 8 and Table 9. It is seen that for thresholds 35 and 50, the probabilities of exceeding different fatal injuries (Table 8) and return level estimates (Table 9) do not differ greatly. This implies the model gives similar results for any threshold value between 24 and 50.

Table 5.2: Table 6: Estimated fatal injuries expected to be exceeded once every m years based on GPD model for different thresholds (μ).The 95% bootstrap confidence intervals are given in the brackets.

m(Year)	$\mu = 35$	$\mu = 50$
5	$240[207, 267]$	$238[204, 265]$
10	$274[228, 304]$	$272[226, 304]$
20	$302[243, 337]$	$302[243, 342]$
30	$317[251, 356]$	$318[249, 363]$
50	$333[258, 376]$	$336[258, 389]$

The entire data analysis is done by statistical programming language R with commonly used R packages for extreme value analysis such as **evmix** (Scarrott, 2014) , **POT** (Ribatet, 2012), **extRemes** (Gilleland, 2011).

6 Conclusion

The assessment of risks of extreme aviation accident is highly interesting issue and challenging question. This study highlights the formulation of a model to predict the possible fatal injuries associated with a future aviation accident. We presented the GPD approach that is able to obtain the relevant figures based on NTSB aviation accident data. The fitted model gives probable fatal injuries from an aviation accident which is expected to be exceeded once in a certain period (years). Bootstrap resampling and Monte Carlo simulation are used generate data to quantify the uncertainties in the estimates. It is demonstrated from different measures of goodness of fit that GPD gives better approximation of the observed data that other models.

References

Airbus (2014). *Commercial Aviation Accidents, 1958 - 2013. A Statistical Analysis.* Retrieved from http://www.airbus.com/

Boeing (2013).*Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations—1959-2013.* Retrieved from http://www.boeing.com/

Civil Aviation Authority (CAA) (2014).*Global Fatal Accident Review, 2002 to 2011* Retrieved from http://publicapps.caa.co.uk/

Carpenter, J. and Bithell, J., (2000). "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians", *Statistics in Medicine*, **19**, 1141-1164.

Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M., (2004). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley-Interscience.

Clauset, A. and Woodard, R. (2013). " Estimating the Historical and Future Probabilities of Large Terrorist Events.", *Annals of Applied Statistics*, **7(4)**, 18381865.

Coles, S. and Simiu, E. (2003). "Estimating Uncertainty in the Extreme Value Analysis of Data Generated by a Hurricane Simulation Model.", *Journal of Engineering Mechanics*, **1288**, 0733-9399.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag.

Dey, A. K., and Das, K., (2016). "Modeling Extreme Hurricane Damage Using the Generalized Pareto Distribution", *American Journal of Mathematical and Management Sciences*, **35(1)**, 55-66.

Dupuis, D.J. and Tawn, J.A. (2001). “Effects of mis-specification in bivariate extreme value problems.” *Extremes*, **4**, 315330.

Fisher, R. and Tippet, L. (1928). “Limiting Forms of the Largest or Smallest Member of a sample.”, Proceedinds of the Cambridge Philosophical Society, **24**, 180-190.

Gilleland, E. and Katz, R. W. (2011). “New software to analyze how extremes change over time.” Eos, **92(2)**, 13–14.

Klugman, S. A., Panjer, H. H., Willmot. G. E. (2008). *Loss Models: From Data to Decisions*, Wiley.

Lasek, M., Sibilski, K., ?urek, J. (2010). “ Aviation Accident Aetiology from catastrophe Theory Point of View”, 27th International Congress of the Aeronautical Sciences.

Muller, A., Arnaud, P., Lang, M., Lavabre, J. (2015). “ Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized Pareto distribution / Incertitudes des quantiles extrmes de pluie estims par un modle stochastique d’averses et par une loi de Pareto gnralise”, *Hydrological Sciences Journal*, **54(3)**, 417.

Pickands, J. (1975). “Statistical inference using extreme order statistics.”, *Annals of Statistics*, **3**, 119-131.

Reich, B. J. and Porter, M. D. (2013). “Discussion of “Estimating the Historical and Future Probabilities of Large Terrorist Events” by Aaron Clauset and Rayan Woodard.” *Annals of Applied Statistics*, **7(4)**, 18711875.

Ribatet M. (2012). “POT: Generalized Pareto Distribution and Peaks Over Threshold.”, R package version 1.1-3. http://CRAN.R-project.org/package=POT/.

Scarrott C. J. and Hu Y. (2014). “evmix: Extreme Value Mixture Modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation.”, Available on CRAN. University of Canterbury, New Zealand. http://www.math.canterbury.ac.nz/ c.scarrott/evmix/.

Smith, R.L.,Weissman, I. (1994). “Estimating the extremal index.” *Journal of the Royal Statistical Society: Series B: Methodological*, **56**, 515528.

Smith, R.L. (1986). “Extreme value theory based on the r largest annual events.” *Journal of Hydrology*, **86**, 2743.

Suveges, M. and Davison, A.C., (2010). “Model misspecification in peaks over threshold analysis”. *Annals of Applied Statistics*, **4**,201-221.

Tolan, C., Patterson, T., Johnson, A. (2015, July 28). “Is 2014 the deadliest year for flights? Not even close.” *CNN*. Retrieved from http://edition.cnn.com/interactive/2014/07/travel/aviation-data/

Wang, W., Drees, L., Gissibl, N., Hhndorf, L., Sembiring, J., Holzapfel, F. (2014). "Quantification of Incident Probabilities Using Physical and Statistical Approaches", Sixth International Conference on Research in Air Transportation, Istanbul, Turkey.

Wehner, W. (2010). "Sources of uncertainty in the extreme value statistics of climate data." *Extremes*, **19**, 205217.