Statistical Modelling Immoderate Weather Event by Using R and SAS: A Case Study of Minneapolis/St Paul Region in Minnesota, USA

Mayooran Thevaraja and Deepak Sanjel

Department of Mathematics and Statistics, Minnesota State University, Mankato, USA

***Abstract*:-Climate projections suggest the frequency and intensity of some environmental extremes will be affected in the future due to a changing climate. Ecosystems and the various sectors of human activity are sensitive to extreme weather events, such as heavy rains and floods, droughts and high and low temperatures, especially when they occur over prolonged periods. In 1985 Wigley studied about extreme events dangerously affected human society which is included among others agriculture, water resources, energy demand and mortality. In this paper, extreme elevated temperature events for nearly 117 years from the Minneapolis/St Paul, Minnesota State, and area are analyzed from the major international airport [St. Paul] and popular city in Minnesota. The main aim of this study is to find the best fitting distribution to the extreme daily temperature measured over the Minneapolis region for the years 1900-2016 by using the maximum likelihood approach. The study also predicts the extreme temperature for return periods and their confidence bands. In this paper, extreme temperature events are defined by two different methods based on (1) the annual maximums of the daily temperature, (2) the daily temperature exceeds some specific threshold value and (3) Bayesian Model using Markov chain Monte Carlo (MCMC). The Generalized Extreme Value distribution and the Generalized Pareto distribution are fitted to data corresponding to the methods 1 and 2 to describe the extremes of temperature and to predict its future behavior. Finally, we find the evidence to suggest that the Frechet distribution provides the most appropriate model for the annual maximums of daily temperature after removing an outlier and the Generalized Pareto Distribution (GPD) gives the reasonable model for the daily temperature data over the threshold value of 96°F for the Minneapolis location. Further, we derive estimates of 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150 and 200 years return levels and its corresponding confidence intervals for extreme temperature.**

***Keywords:* Annual maximum, Threshold, Generalized Extreme Value distribution (GEVD), Generalized Pareto Distribution (GPD), Maximum likelihood estimation, Return period, Bayesian**

I. INTRODUCTION

Global climate change is generally considered a result of increasing atmospheric concentrations of greenhouse gases, mainly due to human activity. Climate change (i.e., global temperature increases) then in turn can modify the frequency (and intensity) of extreme weather and climate events (e.g., heat waves and sea-level rise). This research paper will take a case study approach by identifying observed extreme temperature events in the Minneapolis region in Minnesota. Information on these observed events will be discussed with future climate change projections on temperature. Approaches to analyse observed data with consideration of climate change and potential challenges for adoption into engineering practice will be discussed. The body of the paper will begin by covering background on temperature extremes, and then is a discussion of the data and methodologies used. Analysis using the methodologies is discussed subsequently. The final portion of the paper talks about conclusions for these paper and future steps to be taken. The statistical analysis of extreme value analysis has been done by the scholars in various locations all over the world. Hirose (1994) have found Weibull distribution is the best fit for the annual maximum of daily rainfall in Japan by considering the maximum likelihood parameter estimation method. Nadarajah and Choi (2007) have studied annual maxima of daily rainfall for the years 1961–2001 and modelled for five locations in South Korea. They found the Gumbel distribution provides the most reasonable model for four of the five locations considered using maximum likelihood estimation and they derived estimates of 10, 50, 100, 1000, 5000, 10,000, 50,000 and 100,000-year return levels for daily rainfall and described how they vary with the locations. Chu and Zhao (2008) have applied the Generalized Extreme value distribution for annual maxima of daily rainfall data for Hawaii Islands using L-moments method and derived estimates for return periods. Husna B. Hasan et al. (2012) has used 10years' daily temperature data in Penang Malaysia, and studied Modelling of Extreme Temperature Using Generalized Extreme Value (GEV) Distribution.Nadarajah and Withers (2001) and Nadarajah (2005) provided the application of extreme value distributions to rainfall data over sixteen locations spread throughout New Zealand and fourteen locations in West Central Florida, respectively. Varathan et al. (2010) has used 110 years' data in Colombo district, studied the annual maximums of rainfall by using the GEV distribution and found Gumbel is the best fitting distribution. Mayooran and Laheetharan(2014) have used 110 years' data in Colombo district, Srilanka and identified best fit probability distribution revealed that the probability distribution pattern for different data set are identified out of a very large number of commonly employed probability distribution models by using different goodness of fit tests.

In Minnesota, Sanjel and Wang (2014) have considered maximum gage height for 110 years recorded in Minnesota River at Mankato 1903 to 2013. They analysed of Minnesota River flood level data has been performed using traditional Block Maxima Model, relatively new Pick over Threshold (POT) model, and nonparametric Bayesian MCMC technique.

In this paper, the following objectives were considered, to find the best fitting distribution for annual maximums of daily temperature data by considering the common Generalized Extreme Value distribution and estimate the return-periods & their confidence bands. To find the best fitting distribution for daily temperature (peaks over a threshold) data by considering the common Generalized Pareto distribution and estimate the return-periods & their confidence bands. To estimate the return-periods & their confidence bands by using nonparametric Bayesian MCMC technique. Finally, in Section 5 contains some concluding remarks and future work. To facilitate the exposition, the R, SAS programming codes and figures of the section 5's results are relegated to the Appendix.

II. STUDY AREA

The data which consists of daily temperatures measured (in Fahrenheit) at the Minneapolis, Minnesota weather station, is obtained from the Minnesota, Department of Natural Resources webpage. We consider the years 1900 to 2016 December 6, Minneapolis–Saint Paul is a major metropolitan area built around the Mississippi, Minnesota and St. Croix rivers in east central Minnesota. The area is commonly known as the Twin Cities after its two largest cities, Minneapolis, the city with the largest population in the state, and Saint Paul, the state capital. It is an example of twin cities in the sense of geographical proximity. Minnesotans often refer to the two together (or the seven-county metro area collectively) as The Cities. There are several different definitions of the region. Many refer to the Twin Cities as the seven-county region which is governed under the Metropolitan Council regional governmental agency and planning organization. The United States Office of Management and Budget officially designates 16 counties as the Minneapolis-St. Paul–Bloomington MN-WI Metropolitan Statistical Area, the 16th largest in the United States. The entire region known as the Minneapolis-St. Paul MN-WI Combined Statistical Area, has a population of 3,866,768, the 14th largest, according to 2015 Census estimates.

Owing to its northerly latitude and inland location, the Twin Cities experience the coldest climate of any major metropolitan area in the United States. However, due to its southern location in the state and aided further by the urban heat island, the Twin Cities is one of the warmest locations in Minnesota. The average annual temperature at the Minneapolis–St. Paul International Airport is 45.4 °F; 3.5 °F colder than Winona, Minnesota, and 8.8 °F warmer than Roseau, Minnesota. Monthly average daily elevated temperatures range from 21.9 °F in January to 83.3 °F in July; the average daily minimum temperatures for the two months are 4.3 °F and 63.0 °F respectively. Minimum temperatures of 0 °F or lower are seen on an average of 29.7 days per year, and 76.2 days do not have a maximum temperature exceeding the freezing point. Temperatures above 90 °F occur an average of 15 times per year. Elevated temperatures above 100 °F have been common in recent years; the last occurring on July 6, 2012. The lowest temperature ever reported at the Minneapolis–St. Paul International Airport was −34 °F on January 22, 1936; the highest, 108 °F was reported on July 14 of the same year. Early settlement records at Fort Snelling show temperatures as low as −42 °F. Recent records include −40 °F at Vadnais Lake on February 2, 1996 (National Climatic Data Centre)

The Twin Cities area takes the brunt of many types of extreme weather, including high-speed straight-line winds, tornadoes, flash floods, drought, heat, bitter cold, and blizzards. Hail and Wind damage exceeded $950 million, much of it in the Twin Cities. Other memorable Twin Cities weather-related events include the tornado outbreak on May 6, 1965, the Armistice Day Blizzard on November 11, 1940, and the Halloween Blizzard of 1991. In 2014, Minnesota experienced temperatures below those in areas of Mars when a polar vortex dropped temperatures as low as −40 °F in Brimson and Babbitt with a wind-chill as low as −63 °F in Grand Marais. (Source: Wikipedia)

III. METHODOLOGY

3.1 Univariate Extreme Value Theory

Classical extreme value theory is used to develop stochastic models towards solve real life problems related to unusual events. Classical theoretical results are concerned with the stochastic behaviour of some maximum (minimum) of a sequence of random variables which are assumed to be independently and identically distributed.

There are three models that are commonly used for extreme value analysis. These are the Gumbel, Frechet, and Weibull distribution functions. The Gumbel is easier to work with since it requires only location and scale parameters, while the Weibull and Frechet require location, scale, and shape parameters. These three models may be unified in what is sometimes called the Unified Extreme Value model (Reiss and Thomas, 1997). The Generalized Extreme Value (GEV) distribution function is,

$$H(x) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (1)$$

where $\sigma > 0$ –scale, ξ- shape and μ-location parameter

According to the value of, H(x) can be divided into following three standard types of distributions: This concept stated without detailed mathematical proof by Fisher and Tippett (1928), and later rigorously derived by Gnedenko (1943).

1. If $\xi \to 0$(Gumbel Distribution)

$$H(x) = exp\left[-exp\left\{\left(\frac{x-\mu}{\sigma}\right)\right\}\right] \quad \text{for} -\infty < x < \infty (2)$$

2. If $\xi > 0$ (Frechet Distribution with $\alpha = {}^{1}/_{\xi}$)

   $H(x) = exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right]$ if $x > \mu$ and $H(x) = 0$ if $x \leq \mu$ (3)

3. If $\xi < 0$ (Weibull Distribution with $\alpha = {}^{-1}/_{\xi}$)

   $H(x) = exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{\xi}\right]$ if $x < \mu$ and $H(x) = 1$ if $x \geq \mu$ (4)

3.2 Sample Selection

In every research problems or experiments, the sample selection procedure contributes a key role of statistics. Sanjel and Wang (2014) have considered maximum gage height for 110 years recorded in Minnesota river at Mankato 1903 to 2013. Begueria and Vicente-Serrano (2006) were applied the threshold technique to model the extreme daily rainfall in Spain by considering 43 daily precipitation series from 1950 to 2000. In this study, two different methods are used to select the sample of extreme daily temperature values. The First method is, by considering the annual maximums of daily temperature and the second method is by considering the exceedance over some specific threshold. The data consists of daily temperatures for the years from 1900 to 2016 for the Minneapolis/St Paul, Minnesota location. The data were collected from the Minnesota, Department of Natural Resources webpage, which lists the daily Maximum and Minimum temperatures in Fahrenheit. The extreme values were selected from the tabulated daily data(from approximately 42700 data points).

3.2.1 Annual maxima

In this procedure, the annual maximums of daily temperature for 117 years are taken as a sample of extreme temperature and the modeling is done by under Univariate extreme value theory. Coles (2001), Sanjel and Wang (2014) theoretically explains this topic in their papers, Extremes of temperatures are best expressed in terms of statistical variation, rather than in Fahrenheit of temperatures. Evaluating temperature events as standard deviations above the mean provides a truer measure of maximum temperatures. When extreme temperatures are thus normalized, the highest values often are shown to have occurred at stations other than those that received the highest temperatures. If $X_1, X_2 \ldots . X_{365}$ are daily temperature values then our data selection point *(extreme point)* $=Max\ \{X_1, X_2\ \ldots . X_{365}\}$; where X_i is the daily temperature data in degrees Fahrenheit of any year. $i= 1, 2, 3 \ldots\ldots 365$

3.2.2 Peaks over threshold

The Peaks over threshold (POT) approach generates a subset of data points from a parent set by only considering those events (data peaks) above a defined threshold. By only considering peaks above a threshold the data is more than likely to be from the same distribution. This intrinsically assists with obtaining an identically distributed data set. In addition to this, provided the data peaks can be considered statistically independent, thus the *i.i.d.* condition is satisfied, the distribution of the peak events should have a Generalized Pareto distribution.

Generalized Pareto distribution

In general, we are interested not only in the maxima of observations, but also in the behavior of large observations that exceed a high threshold. Given a high threshold u, the distribution of excess values of x over threshold u is defined by

$$F_u(y) = P\{X - u \leq y | X > u\} = \frac{F(y+u)-F(u)}{1-F(u)} \quad (5)$$

Which represents the probability that the value of x exceeds the threshold u by at most an amount y given that x exceeds the threshold u. A theorem by Balkema and de Haan (1974) and Pickands (1975) shows that for sufficiently high threshold u, the distribution function of the excess may be approximated by the generalized Pareto distribution (GPD) such that, as the threshold gets large, the excess distribution $F_u(y)$ converges to the GPD, which is

$$G(x) = 1 - \left(1 + k * \frac{x}{\beta}\right)^{-1/k} \text{ if } \quad k \neq 0$$

$$\text{and} \quad 1 - e^{-x/\beta} \text{ if } \quad k = 0 \quad (6)$$

; Where k is the shape parameter. The GPD embeds several other distributions; When k >0, it takes the form of the ordinary Pareto distribution. This case is the most relevant for financial time series analysis, since it is a heavy-tailed one. For $k >0$, $E\ [X_r]$ is infinite for $r \geq 1/k$. For instance, the GPD has an infinite variance for k= 0.5 and, when k = 0.25, it has an infinite fourth moment. For security returns or high-frequency foreign exchange returns, the estimates of k are usually less than 0.5, implying that the returns have finite variance (Jansen and devries, 1991; Longin, 1996; Muller, Dacorogna, and Pictet, 1996; Dacorognaet al. 2001). When $k = 0$, the GPD corresponds to exponential distribution, and it is known as a Pareto II–type distribution for k <0. The importance of the Balkema and de Haan (1974) and Pickands (1975) results is that the distribution of excesses may be approximated by the GPD by choosing k and β and setting a high threshold u. The GPD can be estimated with various methods, such as the method of probability-weighted moments or the maximum-likelihood method.10 For $k >-0.5$, which corresponds to heavy tails, Hosking and Wallis (1987) presents evidence that maximum-likelihood regularity conditions are fulfilled and that the maximum-likelihood estimates are asymptotically normally distributed. Therefore, the approximate standard errors for the estimators of β and k can be obtained through maximum-likelihood estimation.

3.3 Parameter Estimation

There are several well-known methods which can be used to estimate distribution parameters based on available sample data. For every supported distribution one of the following parameter estimation methods:

- Method of moments (MOM);
- Maximum likelihood estimates (MLE);
- Least squares estimates (LSE);
- Method of L-moments.

Since the detailed description of these methods goes beyond the scope of this manual, we will just note that, where possible, we use the least computationally intensive methods. Thus, it employs the method of moments for those distributions whose moment estimates are available for all possible parameter values, and do not involve the use of iterative numerical methods.

For many distributions, we use the MLE method involving the maximization of the log-likelihood function. For some distributions, such as the 2-parameter Exponential and the 2-parameter Weibull, a closed form solution of this problem exists. For other distributions, we implement the numerical method for multi-dimensional function minimization. Given the initial parameter estimates vector, this method tries to improve it on each subsequent iteration. The algorithm terminates when the stopping criteria is satisfied (the specified accuracy of the estimation is reached, or the number of iterations reaches the specified maximum). The advanced continuous distributions are fitted using the MLE, the modified LSE, and the L-moments methods.

3.3.1 *Maximum Likelihood Estimation*

Maximum likelihood estimation begins with the mathematical expression known as a likelihood function of the sample data. Loosely speaking, the likelihood of a set of data is the probability of obtaining that set of data given the chosen probability model. This expression contains the unknown parameters. Those values of the parameter that maximize the sample likelihood are known as the maximum likelihood estimates.

The advantages of this method are:

- Maximum likelihood provides a consistent approach to parameter estimation problems. This means that maximum likelihood estimates can be developed for a large variety of estimation situations. For example, they can be applied in reliability analysis to censored data under various censoring models.
- Maximum likelihood methods have desirable mathematical and optimality properties. Specifically,

They become minimum variance unbiased estimators as the sample size increases. By unbiased, we mean that if we take (a very large number of) random samples with replacement from a population, the average value of the parameter estimates will be theoretically exactly equal to the population value. By minimum variance, we mean that the estimator has the smallest variance, and thus the narrowest confidence interval, of all estimators of that type.

They have approximate normal distributions and approximate sample variances that can be used to generate confidence bounds and hypothesis tests for the parameters.

Several popular statistical software packages provide excellent algorithms for maximum likelihood estimates for many of the commonly used distributions. This helps mitigate the computational complexity of maximum likelihood estimation. The advantage of the specific MLE procedures is that greater efficiency and better numerical stability can often be obtained by taking advantage of the properties of the specific estimation problem. The specific methods often return explicit confidence intervals. So, we used MLE method in this analysis.

Suppose we have observations $X_1, X_{2,} \dots X_N$ which are annual maximum temperature values for each of N year, for which the Generalized Extreme Value (GEV) distribution is appropriate.

The corresponding log likelihood is,

$$l(\mu,\sigma,\xi) = -Nlog(\sigma) - \left(\frac{1}{\xi}+1\right)\sum_i log\left(1+\xi\frac{X_i-\mu}{\sigma}\right) - \sum_i\left(1+\xi\frac{X_i-\mu}{\sigma}\right)^{\frac{-1}{\xi}} \qquad (7)$$

Where $1+\ \xi\left(\frac{X_i-\mu}{\sigma}\right) > 0$ for all i

3.4 Likelihood Ratio (LR) Test for the Gumbel Model

Under the Generalized Extreme Value (GEV) distribution, we should test the hypothesis Testing whether the shape parameter ξ = 0 or not. (ie: The data fits the Gumbel Distribution or not) with unknown location and scale parameters. Thus, the Gumbel distributions are tested versus other type of GEV distributions for a given vector $X = (X_1, X_{2,} \dots X_n)$ of data. The likelihood ratio (LR) test statistics is given by,

$$\chi^2 = 2\log\left(\frac{\prod_{i\le n} h(X_i:\hat{\xi},\hat{\mu},\hat{\sigma})}{\prod_{i\le n} h(X_i:0,\tilde{\mu},\tilde{\sigma})}\right) \qquad (8)$$

Where $(\hat{\xi},\hat{\mu},\hat{\sigma})$ and $(\tilde{\mu},\tilde{\sigma})$ are MLEs in the GEV distribution, because he parameter sets have dimensions 3 and 2 respectively. But theoretically, under the null hypothesis likelihood ratio (LR) test statistics is asymptotically distributed according to the chi square distribution with one degree of freedom. Therefore P-value is given by

$$P - \text{Value} = 1 - \chi^2(\text{test statistics value}) \qquad (9)$$

Moreover, suppose p value greater than our significance level fail to reject null hypothesis, his implies that there is enough evidence the data fits the Gumbel Distribution

3.5 Bayesian Method

Annual maximum and Peak over threshold methods are all assume limiting distributions. Since the amount of data available is low in extreme value analysis, often asymptotic limiting distribution may not be correct. Alternatively, Bayesian approach can be used. Bayesian methods are based on specifying a density function for the unknown parameters, (prior density), and then computing a posterior density for the parameters given the observed data (likelihoods). Using Bayesian inferences allow us to use additional prior information about the processes.

3.6 Return Period

Return period (T): Once the best probability model for the data has been determined, the interest is in deriving the return levels of temperature. The T year return level, say x_T, is the level exceeded on average only once in T years. For example, the 2-year return level is the median of the distribution of the annual maximum daily temperature.

Probability of occurrence (p) is expressed as the probability that an event of the specified magnitude will be equaled or exceeded during a one-year period. If n is the total number of values and m is the rank of a value in a list ordered descending magnitude ($x_1 > x_2 > x_3 \ldots > x_m$), the exceeding probability of the m^{th} largest value, x_m, is

$$P(X \geq x_m) = \frac{m}{n}. \tag{10}$$

(See Rao. A and Hamed. K, page 6-7). However, a relationship between the probability of occurrence of a level $x_T$and its return period T are expressed as follows. A given return level x_T with a return period T may be exceeded once in T years. Hence the probability of exceedance is

$$P(X \geq x_T) = \frac{1}{T} \tag{11}$$

If the probability model with CDF, F is assumed then on inverting

$$F(x_T) = P(X \leq x_T) = 1 - P(X \geq x_T) = 1 - \frac{1}{T} \tag{12}$$

3.7 Outliers

An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs. Outliers can occur by chance in any distribution, but they are often indicative either of measurement error or that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the distribution has high kurtosis and that one should be very cautious in using tool or intuitions that assume a normal distribution. In larger samplings of data, some data points will be further away from the sample mean than what is deemed reasonable. This can be due to incidental systematic error or flaws in the theory that generated an assumed family of probability distributions, or it may be that some observations are far from the center of the data. Outlier points can therefore indicate faulty data, erroneous procedures, or areas where a certain theory might not be valid. However, in large samples, a small number of outliers are to be expected. In this project, we will use box plot for identify (graphically) any outlier points.

3.8 SAS and R

The SAS system is a widely-used resource for statistical analysis and data mining. It is rare to find a job advert for a data mining practitioner that does not ask for SAS skills. The main positive points of SAS are its ability to handle large files transparently, the ease and comprehensive way that standard analyses can be done, the interactive way that analyses can be built alongside a systematic programming environment, and the data handling capabilities. Its main negative points are its graphical capabilities, and that adding your own extensions to the techniques using macros and the interactive matrix language are slightly more cumbersome than other languages (e.g. R) and then more modern language constructs.R is a computer language for statistical computing like the S language developed at Bell Laboratories. The R software was initially written by Ross Ihaka and Robert Gentleman in the mid-1990s. Since 1997, the R project has been organized by the R Development Core Team. R is open-source software and is part of the GNU project. R is being developed for the UNIX, Macintosh, and Windows families of operating systems.

In this research project, initially we used SAS for select extreme points (annual maximum and Peak over threshold) from 117 years' daily temperature data (approximately 42700 observations) and after analyzed the data used by R.

IV. RESULTS AND DISCUSSION

The data consists of daily temperatures for the years from 1900 to 2016 for the Minneapolis/St Paul, Minnesota location. The data were collected from the Minnesota, Department of Natural Resources webpage, which lists the daily Maximum and Minimum temperatures in Fahrenheit. The extreme values were selected from the tabulated daily data (from approximately 42700 data points).

Firstly, we have applied the Univariate Extreme Value Theory to fit the for the 117-years (1900-2016) annual maximums of daily temperatures in Minneapolis/St Paul by using the statistical software "R". When we observe the box plot (Figure 4.7) most of the points are lie within the IQR box, only three points fall outside. Among these three points, twopoints' falls far from the IQR box, so we can conclude these points as an outlier (88,106 and 108).After removing this outlier points, we should check normality assumption, According to QQ plot (Figure 4.8) and Shapiro.test value = 0.98692, and corresponding p-value = 0.3445> 0.05, so we can say data satisfy normality assumption.

The Table 4.1 gives the estimates of the parameters of the GEV distribution using maximum likelihood method after removing the outlier.

Table 4.1: Estimated parameters by MLE.

Parameter	Estimate	Standard Error
μ	95.8401	0.3183
σ	3.0650	0.2231
ξ	-0.2453	0.0603

The Figure 4.1 in appendix section shows, the fitted density seems a reasonable fit to the histogram of maximums of temperature. After fitting the GEV distribution, we check the whether the shape parameter (ξ) is zero or not. So, we consider the following statistical hypotheses,

Ho: The data fits the Gumbel distribution (ie: $\xi = 0$*)*

H_1*: Not Ho*

Under Ho,

Likelihood ratio test statistic value = 15.195

Chi-square critical value = 3.8415

Chi-square P-Value = 0.0004792 < 0.05

Reject the null hypothesis at 5% level of significance.

ie) The data do not fit the Gumbel distribution.

In R output indicated, alternative hypothesis: greater, so we can say there is enough evidence fail to reject $\xi > 0$ .Therefore, the data fits the Frechet distribution. Secondly, we identified the threshold value or cuts-off value for the daily temperature using the Mean Residual Life plot. The Mean Residual life plot (Figure 4.2) for the daily temperature from 1900 to 2016 shows approximate linearity above a threshold of 96°F. So, we select 96°F as the threshold value of the daily temperature. The threshold value of 96°F was found using the Mean Residual life plot. Initially 229 data points were collected using the threshold value of 96°F and after removing the outlier (Based on boxplot (Figure 4.10), identified 18outlier points 108,106,105,104,103) points were collected as extreme points, by using this collected data, first we fit the Generalized Pareto Distribution (GPD).

The cumulative distribution function of the GPD distribution is,

$$H(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{\frac{-1}{\xi}} \qquad \text{for all } \frac{1+\xi x}{\sigma} > 0 \tag{13}$$

Where $\sigma > 0$ is the scale parameter and $\xi -$ is a shape parameter

Table 4.2 Maximum Likelihood Parameter Estimation for GPD

	Estimate	Standard Error
σ - scale parameter	3.8489	0.3232
ξ - shape parameter	-0.6109	0.0637

The Table 4.2 gives the estimates of the parameters of the GPD distribution using maximum likelihood method. The figure 4.11shows (see the Appendix 1), the fitted density shows a precise fit to the observed data. After fitting the GPD distribution, we need to check the whether the shape parameter (ξ) is zero or not (i.e.: the data fits the Exponential distribution or not). So, we consider the following statistical hypotheses,

H_o*: The data fits the Exponential distribution (ie:* $\xi = 0$*)*

H_1*: Not Ho*

Under Ho,

Likelihood ratio test statistic value = 65.59

Chi-square critical value = 3.8415

Chi-square P-Value = 4.229e-16< 0.05

So, reject the null hypothesis at 5% level of significance. ie) The data fits the Generalized Pareto Distribution (GPD) distribution.

Finally, we will use nonparametric Bayesian MCMC technique, so first we will Estimate parameters.

Table 4.3 Quantiles of MCMC Sample from Posterior Distribution

	Estimate	Standard Error
μ -location	95.8876	0.1603
σ - scale parameter	3.1384	0.0693
ξ - shape parameter	-0.2347	0.0049

The Table 4.3 gives the estimates of the parameters of the nonparametric Bayesian MCMC technique method. The Table 4.6 and 4.7 gives the return values of the annual maximum temperature daily and their 95% confidence levels for the return periods 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150 and 200 years respectively.

The computed return levels for each data set are listed in Table 4.6. It has been predicted that the 2-year return period's return level is approximately 96.9144°F in GEVD method, which means temperature of 96.9144°F or more, should occur at that location on the average only once every two years. In other way round, the average 101 °F or more daily extreme temperature event occur for the period of every ten-year with the occurrence probability 0.1000. According to the table, the 50-year return period is 103.9498 °F in Bayesian method, which means every 50 year we can expect in average 103.9498°F or more daily extreme temperature with the probability 0.05.Among the various methods considered, the using nonparametric Bayesian MCMC techniqueappears to be associated with the highest return levels. As notice that, the GEVD technique's estimated return levels have approximately equal to Bayesian method's estimations.

V. CONCLUSION

In this study, we have performed a statistical modeling of extreme daily temperature over 117 years in Minneapolis/St Paul, Minnesota using extreme value distributions under three different approaches. However our original collected data of annual maximum daily temperature data fits the GEV distribution, the distribution converges to the Frechet distribution and the predicted values for different return periods and their confidence levels decrease following the removal of the single outlier identified using boxplot. Therefore, the outlier is more important in this analysis. The identified outlier is 105°F, 106°F and 108°F, which occurred on the 07/31/1988, 5/31/1934 and 07/14/1936 respectively. The return period of the outlier points105°F, 106°F and 108°F are nearly 250, 1200 and more than 50000 years respectively. So, we can't predict this return value using the identified results shown in Tables 4.6 and 4.7. Therefore, more sophisticated analysis is needed to establish its true return period. We have established the Frechet and Generalized Pareto distribution (GPD) are suitable models for extreme daily temperature by considering annual maximums of daily temperatures and daily temperatures greater than 96°F. The predicted return values and the confidence levels are very similar in sampling techniques, annual maxima and peaks over threshold. For example, the 50-year return value of extreme daily temperature using annual maxima is 103.536°F and using peaks over threshold is 103.4017°F and their corresponding confidence levels are (102.3288,104.7445) and (101.9243, 102.6443). Finally, we used, Bayesian MCMC technique for annual maximum daily temperature data and estimated parameters and return levels. For example, the 100-year return value of extreme daily temperature using annual maxima is 104.2918°F and using Bayesian MCMC technique is 104.7865°F and their corresponding confidence levels are (102.8251, 105.7585) and (103.3612, 107.4158).

This research project only provides an initial study of extreme daily maximum temperature in Minneapolis region. This study can be extended in several ways. We can consider annual maximums of the 2-day, 4-day, 7-day & 10-day maximum temperature by using the GEV distribution and similarly if we consider daily minimum temperature data, which is very useful for this region. The other is a more sophisticated analysis of the actual return period of the identified outlier to assess its relevance for design.

REFERENCES

[1]. Balkema, A. A. and L. de Haan (1974), Residual lifetime at great age, Annals of Probability, 2,792–804.

[2]. Begueria, S. and Vicente-Serrano, S.M. (2006). Mapping the Hazard of Extreme Rainfall by Peaks over Threshold Extreme Value Analysis and Spatial Regression Techniques. Journal of Applied Meteorology 45(1), 108-124.

[3]. Chu, P. S, Zhao. X, Ruan. Y and Grubbs. M, (2009). "Extreme Rainfall Events in the Hawaiian Islands" journal of applied meteorology and climatology volume 48, American Meteorological Society.

[4]. Coles, S. 2001. An Introduction to Statistical Modeling of Extreme Values. Springer.

[5]. Fisher, R. A. and Tippet, L.H.C. 1928. On the estimation of the frequency distributions of the largest or smallestmember of a sample. Proceeding of the Cambridge philosophical society, 24, 180-190.

[6]. Frechet. M, 1927. "Sur la loi de probabilité de l'écart maximum." Ann. Soc. Polon. Math. 6, 93

[7]. Gnedenko B.V., 1943. Sur la distribution limite du terme maximum d'uneseriealeatoire, Annals of Mathematics,44, 423453.

[8]. Gnedenko, B. "Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire." Annals of Mathematics, Second Series, 44, no. 3 (1943): 423-53.

[9]. Gumbel, E. (1958), Statistics of Extremes, Colombia University Press, New York.

[10]. Hirose, H. (1994). Parameter Estimation in the Extreme-Value distributions using the Continuation Method, Information Processing Society of Japan Vol 35, 9.

[11]. Husna. B, Noor. B. Ahmad. R, and Suraiya. B (2012) Proceedings of the World Congress on Engineering 2012 Vol I WCE 2012, July 4 - 6, 2012, London, U.K.

[12]. Mayooran. T and Laheetharan.A (2014), "The Statistical Distribution of Annual Maximum Rainfall in Colombo District"Sri Lankan Journal of Applied Statistics, Vol (14-2)

[13]. Muller, U. A, M. M. Dacorogna, O. V. Pictet (1996), Heavy tails in high frequency financialdata, Olsen & Associates Discussion Paper.

[14]. Nadarajah S., and Withers, C. S. (2001). Evidence of trend in return levels for daily rainfall in New Zealand, Journal of Hydrology New Zealand, Volume 39, p.155-166.

[15]. Nadarajah, S. (2005), Theexponentiated Gumbel distribution with climate application, Environmetrics, 17(1), 13-23.

[16]. Nadarajah. S and Choi. D, (2007). Maximum daily rainfall in South Korea J. Earth Syst. Sci. 116, No. 4, pp. 311–320.

[17]. Pickands, J. (1975), Statistical inference using extreme order statistics, Annals of Statistics, 3,119–131.

[18]. Ramachandra Rao. A andHamed. K. H, (2000). Flood Frequency Analysis, CRC Press, Boca Raton, Florida, USA.

[19]. Reiss, R.-D. and Thomas, M. 2007. Statistical Analysis of Extreme Values: with applications to insurance, finance, hydrology and other fields. Birkhauser, 530pp., 3rd edition

[20]. Sanjel, D., Wang, Y. G., (2014). Extreme Value Modeling of Minnesota River Flood, American Statistical Association,JSM 2014 proceeding. (Section for Statistical Programmers and Analysts) pp: 2594-2604.

[21]. Varathan, N, Perera, K and Nalin, (2010). Statistical modeling of daily extreme rainfall in Colombo, International Conference on Sustainable Built Environment (ICSBE-2010) Kandy, 13-14 December 2010 pp 144-151, Sri Lanka.

[22]. Wigley, T.M.L., 1985: Impact of extreme events. Nature, 316, 106-107.

Appendix I:

Table 4.4: *Annual Maximums of Daily Temperature from 1900-2016 in Minneapolis*					
Obs	**Date**	**Maximum**	**Obs**	**Date**	**Maximum**
1	July 30, 1900	95	35	May 31, 1934	106
2	July 20, 1901	102	36	July 27, 1935	98
3	July 29, 1902	88	37	July 14, 1936	108
4	July 7, 1903	92	38	July 10, 1937	100
5	July 16, 1904	92	39	July 12, 1938	95
6	August 10, 1905	95	40	September 14, 1939	98
7	August 16, 1906	93	41	July 22, 1940	103
8	August 31, 1907	94	42	July 24, 1941	104
9	July 10, 1908	94	43	July 16, 1942	96
10	August 2, 1909	93	44	June 26, 1943	96
11	June 20, 1910	96	45	June 25, 1944	96
12	July 1, 1911	99	46	July 23, 1945	96
13	September 5, 1912	95	47	August 16, 1946	95
14	August 15, 1913	100	48	August 4, 1947	102
15	July 26, 1914	96	49	July 6, 1948	101
16	July 12, 1915	88	50	July 3, 1949	100
17	July 28, 1916	97	51	August 16, 1950	96
18	July 28, 1917	99	52	July 15, 1951	91
19	July 20, 1918	94	53	July 19, 1952	93
20	July 26, 1919	96	54	June 18, 1953	98
21	June 13, 1920	94	55	July 19, 1954	95
22	June 30, 1921	99	56	July 26, 1955	100
23	June 23, 1922	99	57	June 13, 1956	100
24	July 9, 1923	97	58	July 11, 1957	97
25	August 26, 1924	92	59	June 29, 1958	95
26	May 22, 1925	99	60	July 29, 1959	96
27	July 16, 1926	102	61	July 21, 1960	95
28	June 28, 1927	96	62	June 28, 1961	98
29	July 7, 1928	94	63	June 28, 1962	95
30	July 26, 1929	97	64	June 30, 1963	99
31	July 10, 1930	98	65	August 1, 1964	98
32	July 27, 1931	104	66	July 23, 1965	95
33	July 20, 1932	101	67	July 10, 1966	99
34	June 19, 1933	100	68	July 21, 1967	91
69	June 4, 1968	96	96	July 13, 1995	101
70	August 29, 1969	96	97	June 28, 1996	96
71	June 29, 1970	97	98	June 23, 1997	94
72	August 22, 1971	97	99	July 13, 1998	94
73	August 16, 1972	97	100	July 25, 1999	99
74	June 10, 1973	98	101	June 8, 2000	94

75	July 8, 1974	101	102	August 6, 2001	99
76	July 29, 1975	98	103	June 30, 2002	97
77	July 13, 1976	100	104	August 24, 2003	97
78	July 19, 1977	100	105	June 7, 2004	95
79	May 26, 1978	96	106	July 16, 2005	97
80	August 6, 1979	96	107	July 31, 2006	101
81	July 11, 1980	100	108	July 7, 2007	98
82	July 8, 1981	91	109	July 29, 2008	94
83	July 5, 1982	100	110	May 19, 2009	97
84	August 7, 1983	97	111	August 8, 2010	96
85	July 22, 1984	94	112	June 7, 2011	103
86	June 8, 1985	102	113	July 6, 2012	102
87	June 19, 1986	93	114	May 14, 2013	98
88	June 13, 1987	99	115	July 21, 2014	92
89	July 31, 1988	105	116	August 14, 2015	94
90	July 5, 1989	97	117	July 22, 2016	97
91	July 3, 1990	100			
92	June 26, 1991	95			
93	June 12, 1992	92			
94	August 9, 1993	89			
95	June 14, 1994	95			

Table 4.5: ***Daily Temperature Over 96°F from 1900-2016 in Minneapolis/St Paul***

Obs	Date	Maximum	Obs	Date	Maximum
1	July 13, 1901	98	45	July 14, 1931	98
2	July 14, 1901	98	46	July 15, 1931	101
3	July 20, 1901	102	47	July 16, 1931	100
4	July 23, 1901	101	48	July 25, 1931	98
5	July 24, 1901	101	49	July 26, 1931	99
6	June 22, 1911	98	50	July 27, 1931	104
7	June 30, 1911	97	51	July 28, 1931	99
8	July 1, 1911	99	52	August 4, 1931	99
9	July 30, 1913	97	53	September 8, 1931	99
10	August 15, 1913	100	54	September 10, 1931	104
11	September 1, 1913	97	55	July 12, 1932	97
12	September 5, 1913	97	56	July 14, 1932	98
13	July 28, 1916	97	57	July 18, 1932	97
14	July 29, 1916	97	58	July 19, 1932	97
15	August 6, 1916	97	59	July 20, 1932	101
16	July 28, 1917	99	60	June 16, 1933	97
17	June 30, 1921	99	61	June 17, 1933	97
18	July 9, 1921	97	62	June 18, 1933	97
19	July 10, 1921	98	63	June 19, 1933	100

20	July 11, 1921	98	64	June 20, 1933	98
21	June 23, 1922	99	65	June 26, 1933	98
22	September 5, 1922	98	66	June 27, 1933	99
23	September 6, 1922	98	67	June 28, 1933	97
24	July 9, 1923	97	68	July 29, 1933	98
25	July 22, 1923	97	69	July 30, 1933	100
26	May 22, 1925	99	70	May 28, 1934	98
27	July 11, 1925	97	71	May 30, 1934	98
28	September 3, 1925	97	72	May 31, 1934	106
29	September 4, 1925	98	73	June 23, 1934	97
30	July 16, 1926	102	74	June 25, 1934	98
31	July 20, 1926	98	75	June 27, 1934	104
32	August 27, 1926	99	76	July 14, 1934	97
33	July 26, 1929	97	77	July 19, 1934	98
34	July 10, 1930	98	78	July 21, 1934	105
35	July 25, 1930	97	79	July 22, 1934	105
36	July 26, 1930	97	80	July 23, 1934	105
37	July 27, 1930	98	81	August 18, 1934	97
38	August 2, 1930	97	82	July 27, 1935	98
39	August 3, 1930	98	83	July 31, 1935	98
40	June 26, 1931	99	84	July 6, 1936	104
41	June 27, 1931	97	85	July 7, 1936	101
42	June 28, 1931	102	86	July 8, 1936	101
43	June 29, 1931	102	87	July 10, 1936	106
44	June 30, 1931	100	88	July 11, 1936	106
89	July 12, 1936	106	135	June 10, 1956	99
90	July 13, 1936	105	136	June 13, 1956	100
91	July 14, 1936	108	137	July 11, 1957	97
92	July 15, 1936	98	138	July 19, 1957	97
93	July 16, 1936	98	139	June 28, 1961	98
94	July 17, 1936	99	140	June 30, 1963	99
95	August 15, 1936	103	141	July 19, 1964	97
96	June 23, 1937	99	142	July 23, 1964	97
97	July 10, 1937	100	143	August 1, 1964	98
98	August 5, 1937	97	144	August 5, 1964	97
99	September 2, 1937	97	145	July 10, 1966	99
100	September 14, 1939	98	146	July 11, 1966	99
101	September 15, 1939	98	147	June 29, 1970	97
102	July 18, 1940	101	148	August 22, 1971	97
103	July 19, 1940	100	149	August 16, 1972	97
104	July 21, 1940	99	150	August 20, 1972	97
105	July 22, 1940	103	151	June 10, 1973	98
106	July 23, 1940	103	152	July 7, 1974	97
107	July 22, 1941	98	153	July 8, 1974	101

108	July 23, 1941	100	154	July 13, 1974	99
109	July 24, 1941	104	155	July 29, 1975	98
110	July 25, 1941	99	156	July 30, 1975	97
111	July 28, 1941	97	157	July 9, 1976	99
112	August 3, 1941	99	158	July 10, 1976	99
113	July 26, 1947	98	159	July 13, 1976	100
114	August 4, 1947	102	160	August 18, 1976	98
115	August 5, 1947	100	161	August 19, 1976	97
116	August 10, 1947	101	162	August 21, 1976	97
117	August 11, 1947	97	163	September 7, 1976	98
118	August 17, 1947	100	164	July 19, 1977	100
119	August 21, 1947	98	165	July 7, 1980	98
120	July 5, 1948	98	166	July 10, 1980	98
121	July 6, 1948	101	167	July 11, 1980	100
122	July 7, 1948	98	168	July 14, 1980	99
123	July 8, 1948	99	169	July 4, 1982	99
124	August 23, 1948	97	170	July 5, 1982	100
125	August 24, 1948	98	171	August 2, 1982	98
126	June 30, 1949	99	172	August 7, 1983	97
127	July 3, 1949	100	173	June 8, 1985	102
128	July 4, 1949	100	174	July 7, 1985	97
129	July 5, 1949	98	175	June 13, 1987	99
130	August 7, 1949	97	176	June 14, 1987	98
131	June 18, 1953	98	177	June 19, 1988	98
132	July 26, 1955	100	178	June 20, 1988	97
133	July 28, 1955	100	179	June 24, 1988	101
134	August 1, 1955	98	180	July 5, 1988	97
181	July 6, 1988	99	222	July 2, 2012	99
182	July 7, 1988	99	223	July 3, 2012	97
183	July 15, 1988	102	224	July 4, 2012	101
184	July 27, 1988	97	225	July 6, 2012	102
185	July 28, 1988	97	226	July 16, 2012	98
186	July 31, 1988	105	227	May 14, 2013	98
187	August 1, 1988	101	228	August 26, 2013	97
188	August 2, 1988	99	229	July 22, 2016	97
189	August 15, 1988	98			
190	August 16, 1988	99			
191	August 17, 1988	97			
192	July 5, 1989	97			
193	July 3, 1990	100			
194	July 12, 1995	97			
195	July 13, 1995	101			
196	July 15, 1999	97			
197	July 24, 1999	98			

198	July 25, 1999	99			
199	July 29, 1999	98			
200	June 25, 2001	97			
201	July 31, 2001	98			
202	August 5, 2001	98			
203	August 6, 2001	99			
204	August 7, 2001	98			
205	June 30, 2002	97			
206	August 24, 2003	97			
207	July 16, 2005	97			
208	July 17, 2005	97			
209	August 2, 2005	96			
210	May 28, 2006	97			
211	July 15, 2006	99			
212	July 28, 2006	98			
213	July 30, 2006	99			
214	July 31, 2006	101			
215	July 7, 2007	98			
216	May 19, 2009	97			
217	June 6, 2011	97			
218	June 7, 2011	103			
219	July 1, 2011	99			
220	July 18, 2011	98			
221	July 19, 2011	97			

In above tables, outlier points were indicated in red color

Table 4.6: Estimated return levels based on fixed return periods.

Probability of Occurrence	Return Period (*T* in years)	Estimated Return Level		
		GEVD	POT	Bayesian
0.5000	2	96.9144	102.1878	96.9893
0.2000	5	99.6864	102.2359	99.8541
0.1000	10	101.1404	102.2580	101.3778
0.0500	20	102.3048	102.2725	102.6153
0.0333	30	102.8872	102.2786	103.2423
0.0250	40	103.2636	102.2820	103.6511
0.0200	50	103.5367	102.2843	103.9498
0.0167	60	103.7484	102.2860	104.1827
0.0143	70	103.9199	102.2872	104.3722
0.0125	80	104.0631	102.2882	104.5311
0.0111	90	104.1854	102.2890	104.6675
0.0100	100	104.2918	102.2897	104.7865
0.0067	150	104.6760	102.2920	105.2197
0.0050	200	104.9259	102.2933	105.5048

Table 4.7: 95% Confidence bands of Estimated return levels based on fixed return periods.

Probability of Occurrence	Return Period (*T* in years)	95% Confidence interval of Return Level		
		GEVD	POT	Bayesian
0.5000	2	(96.2731, 97.5558)	(101.8822, 102.4934)	(96.2031, 97.7967)
0.2000	5	(98.9798, 100.3930)	(101.9071, 102.5647)	(98.9881, 100.7952)
0.1000	10	(100.3572, 101.9235)	(101.9164, 102.5997)	(100.4361, 102.4934)
0.0500	20	(101.3795, 103.2301)	(101.9213, 102.6237)	(101.5618, 104.0634)
0.0333	30	(101.8480, 103.9263)	(101.9230, 102.6341)	(102.1189, 104.9212)
0.0250	40	(102.1324, 104.3948)	(101.9238, 102.6402)	(102.4588, 105.5536)
0.0200	50	(102.3289, 104.7445)	(101.9243, 102.6443)	(102.7011, 106.0448)
0.0167	60	(102.4752, 105.0217)	(101.9246, 102.6473)	(102.8908, 106.4015)
0.0143	70	(102.5897, 105.2501)	(101.9249, 102.6496)	(103.0461, 106.7075)
0.0125	80	(102.6824, 105.4437)	(101.9250, 102.6514)	(103.1754, 106.9723)
0.0111	90	(102.7596, 105.6112)	(101.9251, 102.6529)	(103.2724, 107.2070)
0.0100	100	(102.8251, 105.7585)	(101.9252, 102.6542)	(103.3612, 107.4158)
0.0067	150	(103.0487, 106.3033)	(101.9255, 102.6585)	(103.6774, 108.1543)
0.0050	200	(103.1827, 106.6691)	(101.9256, 102.6610)	(103.8679, 108.6844)

Appendix II

SAS and R Codes:

********************Extreme points Selection - SAS*******************

```
PROC IMPORT OUT= WORK.MinnesotaMSPdatamax
DATAFILE= "C:\Users\vp0011hr\Desktop\MinnesotaMSPdatamax.xlsx"
DBMS=xlsx REPLACE;
SHEET="DataFull";
GET NAMES=YES;
RUN;

PROC PRINT DATA = WORK.MinnesotaMSPdatamax;
TITLE'Minnesota MSP Tempdata';

Data Extremetempthreshold;
      Set WORK.MinnesotaMSPdatamax;
      where Maximum >96 ;
run;
proc print data=Extremetempthreshold;
run;
```

******** Annual Maximums of daily temperature data Analysis – R *********

```
library(graphics)
library(extRemes)
library(evd)
library(POT)
library(PASWR)
library(evir)
MSPTempmax<-read.table("C:/Users/vnuu/Desktop/MSPannualmax.txt",header=T)
names(MSPTempmax)
```

```
plot(MSPTempmax$Year,MSPTempmax$Maximum, type ="p", pch=20,xlab = "Year",ylab =
"Maximum temperature(in °F)",col ="red",lwd=0.5,cex.lab = 1.0,main="Scatter plot for
Annual Maximum Temperature in MSP",col.main= "blue",font.main= 6,col.lab=
"darkblue",font.lab= 6)
shapiro.test(MSPTempmax$Maximum)
qqnorm(MSPTempmax$Maximum,col="blue")
qqline(MSPTempmax$Maximum,col="red")
boxplot(MSPTempmax$Maximum,id.n=Inf,col = "lightpink",main="Box plot for Annual
maximum temperature in MSP region",col.main= "blue",ylab = "Maximum temperature(in
°F)",font.main = 6)
fit1 <- fevd(Maximum, MSPTempmax, units = "deg F")
fit1
distill(fit1)
summary(MSPTempmax$Maximum)
hist(MSPTempmax$Maximum,prob=T, main="Histogrm of Annual Maximum temperature data with
dencity", col=gray(0.8), xlab = "Maximum temperature(in °F)", col.main=
"blue",font.main= 6,col.lab= "darkblue",font.lab= 6)
lines(density(MSPTempmax$Maximum),col="red", lty=2)
curve(dgev(x,95.6563637,3.4752222,-0.2139041),col="blue", lwd=2,add=T)
leglebels<- c("Est pdf","Actual pdf" )
legend ("topright", legend=leglebels, lty=c(2,1), col=c("red","blue"),lwd=2 )
fit2 <-fevd(Maximum, MSPTempmax,type = "Gumbel",units = "deg F")
fit2
lr.test(fit1,fit2)
plot(fit1)
plot(fit1, "trace")
return.level(fit1)
return.level(fit1, do.ci = TRUE)
ci(fit1,return.period = c(2,5,10,20,30,40,50,60,70,80,90,100,150,200))
ci(fit1, type = "parameter")
```

***** Annual Maximums of daily temperature data Analysis (after removing outlier points) – R *****

```
MSPTempmaxrout<-read.table("C:/Users/vnuu/Desktop/MSPannualmaxrout.txt",header=T)
names(MSPTempmaxrout)
boxplot(MSPTempmaxrout$Maximum,id.n=Inf)
shapiro.test(MSPTempmaxrout$Maximum)
EDA(MSPTempmaxrout$Maximum)
qqnorm(MSPTempmaxrout$Maximum,col="blue")
qqline(MSPTempmaxrout$Maximum,col="red")
fitremout<- fevd(Maximum, MSPTempmaxrout, units = "deg F")
fitremout
hist(MSPTempmaxrout$Maximum,prob=T, main="Histogrm of Annual Maximum temperature data
with dencity", col=gray(0.8), xlab = "Maximum temperature(in °F)", col.main=
"blue",font.main= 6,col.lab= "darkblue",font.lab= 6)
lines(density(MSPTempmaxrout$Maximum),col="red", lty=2)
curve(dgev(x,95.8400740,3.0650074,-0.2453332),col="blue", lwd=2,add=T)
leglebels<- c("Est pdf","Actual pdf" )
legend ("topright", legend=leglebels, lty=c(2,1), col=c("red","blue"),lwd=2 )
fitremout2 <-fevd(Maximum, MSPTempmaxrout,type = "Gumbel",units = "deg F")
fitremout2
lr.test(fitremout,fitremout2)
plot(fitremout)
plot(fitremout, "trace")
return.level(fitremout)
```

```
return.level(fitremout, do.ci = TRUE)
ci(fitremout,return.period = c(2,5,10,20,30,40,50,60,70,80,90,100,150,200))
ci(fitremout, type = "parameter")
ci(fitremout,return.period = c(250,1100,1000000))
```

***** Daily extreme temperature data Analysis (by using Peaks Over a Threshold) – R *****

```
MSPTemp<-read.table("C:/Users/vnuu/Desktop/MSPmaximum.txt",header=T)
names(MSPTemp)
mrlplot(MSPTemp$Maximum,xlim=c(-20,120),col=c("blue","red","blue"))
MSPTempthreshold<-read.table("C:/Users/vnuu/Desktop/MSPthreshold.txt",header=T)
names(MSPTempthreshold)
plot(MSPTempthreshold$Year,MSPTempthreshold$Maximum, type = "p",pch=20, xlab =
"Year",ylab = "Maximum temperature(in °F)",col = "darkgreen", lwd = 0.5, cex.lab =
1.0,main="Scatter plot for Temperature in MSP by using threshold value 95
°F",col.main= "blue",font.main= 6,col.lab= "darkblue",font.lab= 6)
boxplot(MSPTempthreshold$Maximum,id.n=Inf,col = "lightpink",main="Box plot for Annual
maximum temperature in MSP region",col.main= "blue",ylab = "Maximum temperature(in
°F)",font.main = 6)
EDA(MSPTempthreshold$Maximum)

MSPTempthreshold<- read.table("C:/Users/vnuu/Desktop/MSPthreshold.txt",header=T)
names(MSPTempthreshold)
boxplot(MSPTempthreshold$Maximum,id.n=Inf,col= "lightblue",main="Box plot for
temperature in MSP region by using threshold value 96 °F",col.main= "blue",ylab =
"Maximum temperature(in °F)",font.main = 6)
plot(MSPTempthreshold$Year,MSPTempthreshold$Maximum, type = "p",pch=20, xlab =
"Year",ylab = "Maximum temperature(in °F)",col = "darkgreen", lwd = 0.5, cex.lab =
1.0,main="Scatter plot for Temperature in MSP by using threshold value 96
°F",col.main= "blue",font.main= 6,col.lab= "darkblue",font.lab= 6)

MSPTempthresholdremoveout<-
read.table("C:/Users/vnuu/Desktop/MSPthresholdremoveout.txt",header=T)
names(MSPTempthresholdremoveout)
boxplot(MSPTempthresholdremoveout$Maximum,id.n=Inf,col= "lightblue",main="Box plot for
temperature in MSP region by using threshold value 96 °F",col.main= "blue",ylab =
"Maximum temperature(in °F)",font.main = 6)
fitD1 <- fevd(Maximum,MSPTempthresholdremoveout, threshold = 96, type = "GP", units =
"deg F")
fitD1
hist(MSPTempthresholdremoveout$Maximum,prob=T, main="Histogrm of Maximum temperature
data with dencity by using threshold value 96 °F", col=gray(0.8), xlab= "Maximum
temperature(in°F)",col.main= "blue",font.main= 6,col.lab= "darkblue",font.lab= 6)
lines(density(MSPTempthresholdremoveout$Maximum),col="red", lty=1)
fitD2<-fevd(Maximum,MSPTempthresholdremoveout,threshold=96,type="Exponential", units =
"deg F")
fitD2
lr.test(fitD1,fitD2)
plot(fitD1)
plot(fitD1, "trace")
return.level(fitD1)
return.level(fitD1, do.ci = TRUE)
ci(fitD1,return.period = c(2,5,10,20,30,40,50,60,70,80,90,100,150,200))
ci(fitD1, type = "parameter")
```

```
fitB <- fevd(Maximum,MSPTempmax, method="Bayesian",verbose=TRUE)
fitB
plot(fitB)
plot(fitB, "trace")
return.level(fitB)
return.level(fitB, do.ci = TRUE)
ci(fitB,return.period = c(2,5,10,20,30,40,50,60,70,80,90,100,150,200))
ci(fitB, type = "parameter")
```

Appendix III

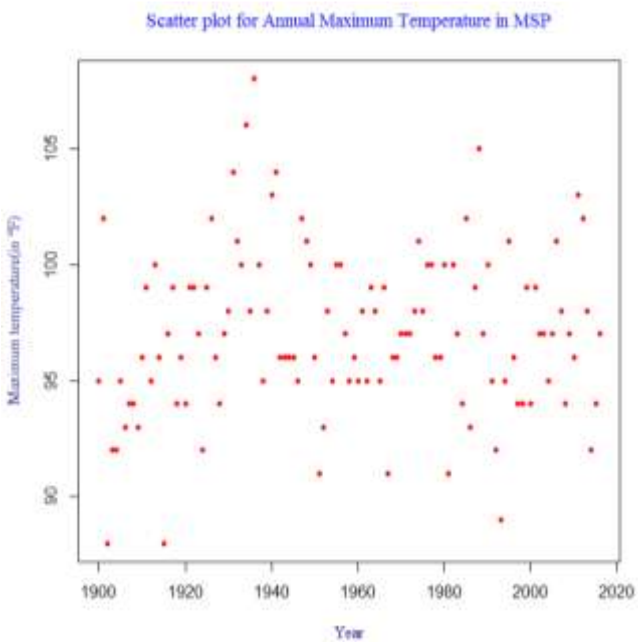

Figure 4.1

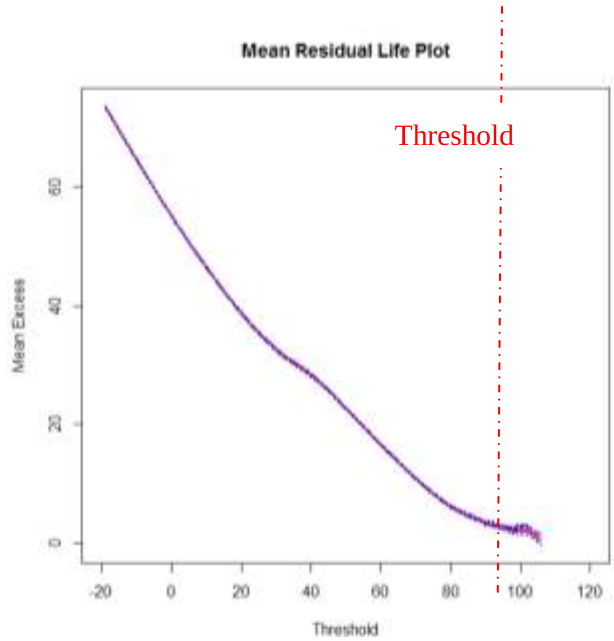

Figure 4.2

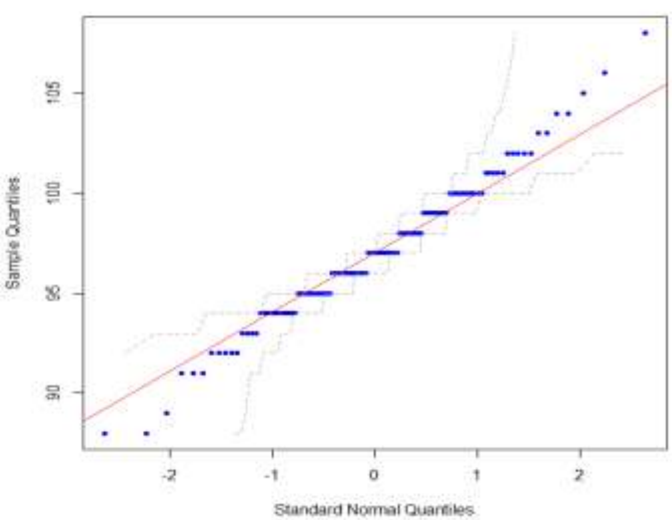

Figure 4.3

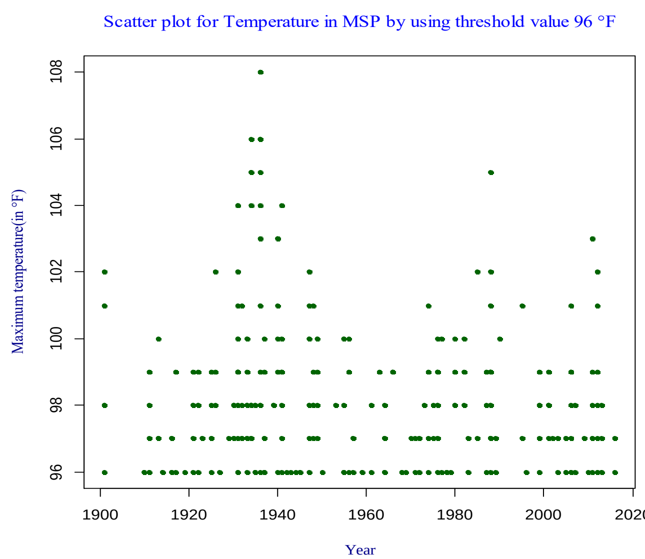

Figure 4.4

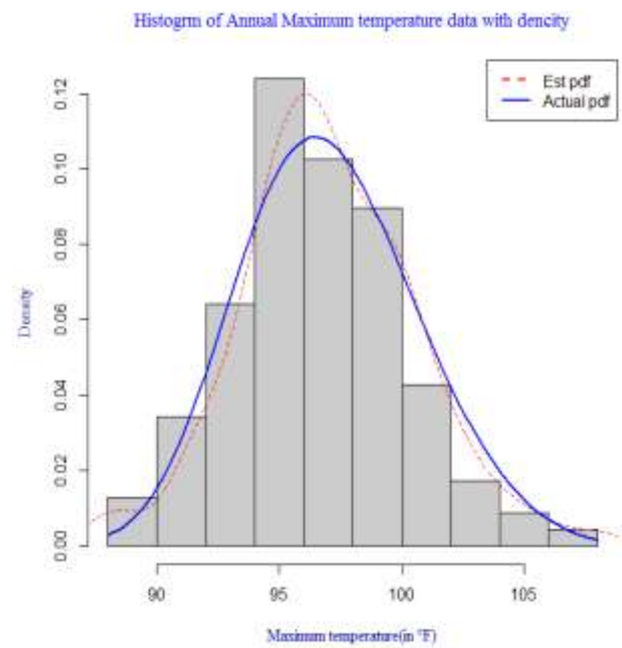

Figure 4.5

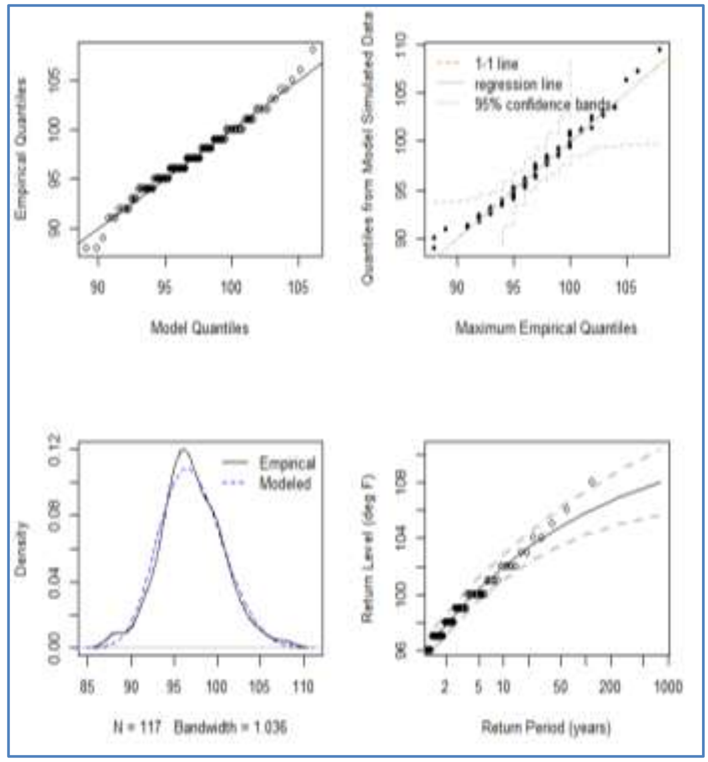

Figure 4.6

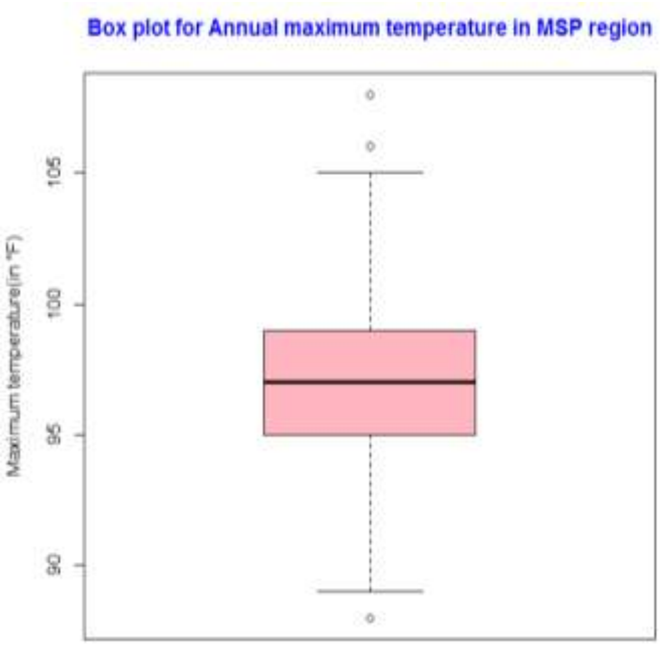

Figure 4.7

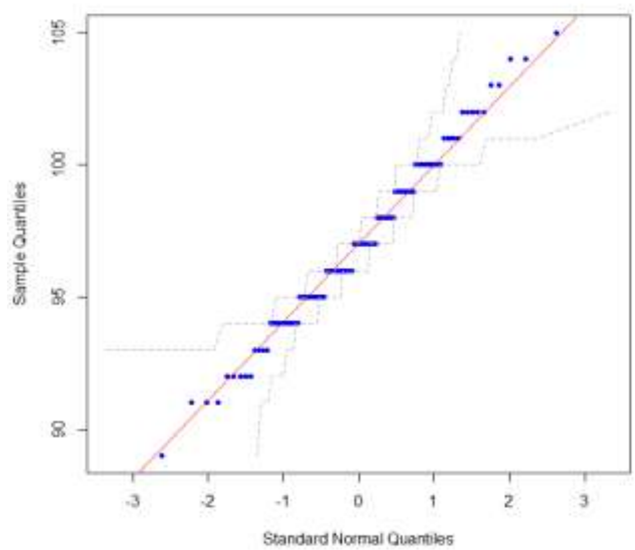

Figure 4.8

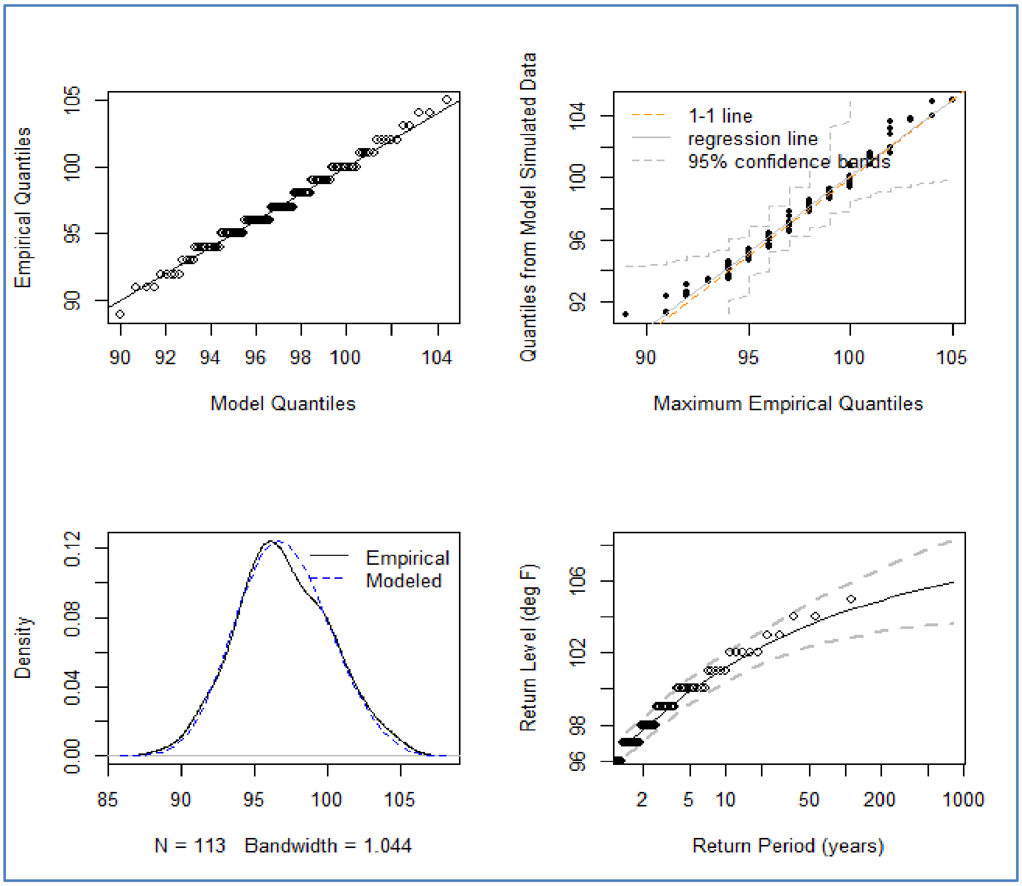

Figure 4.9

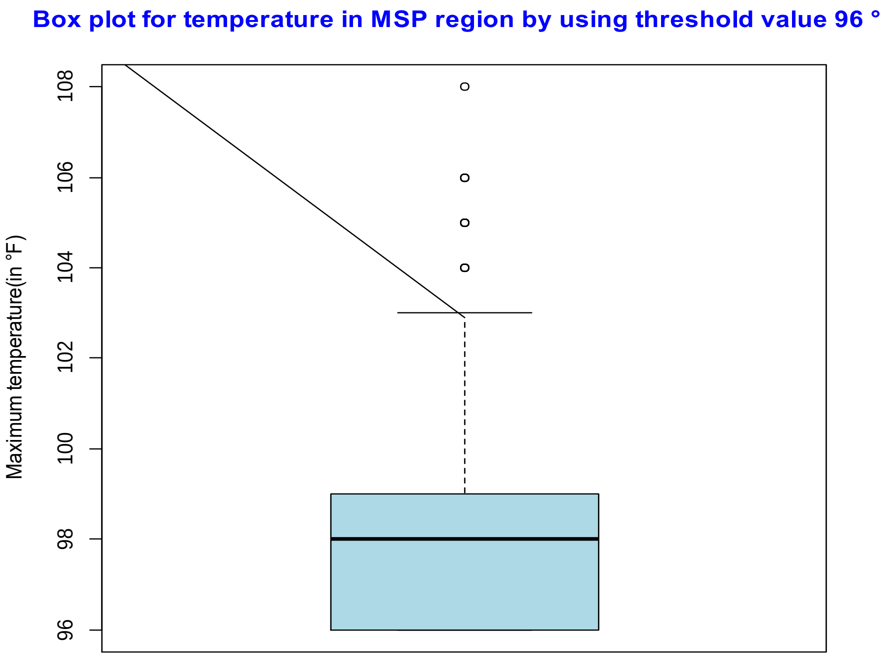

Figure 4.10

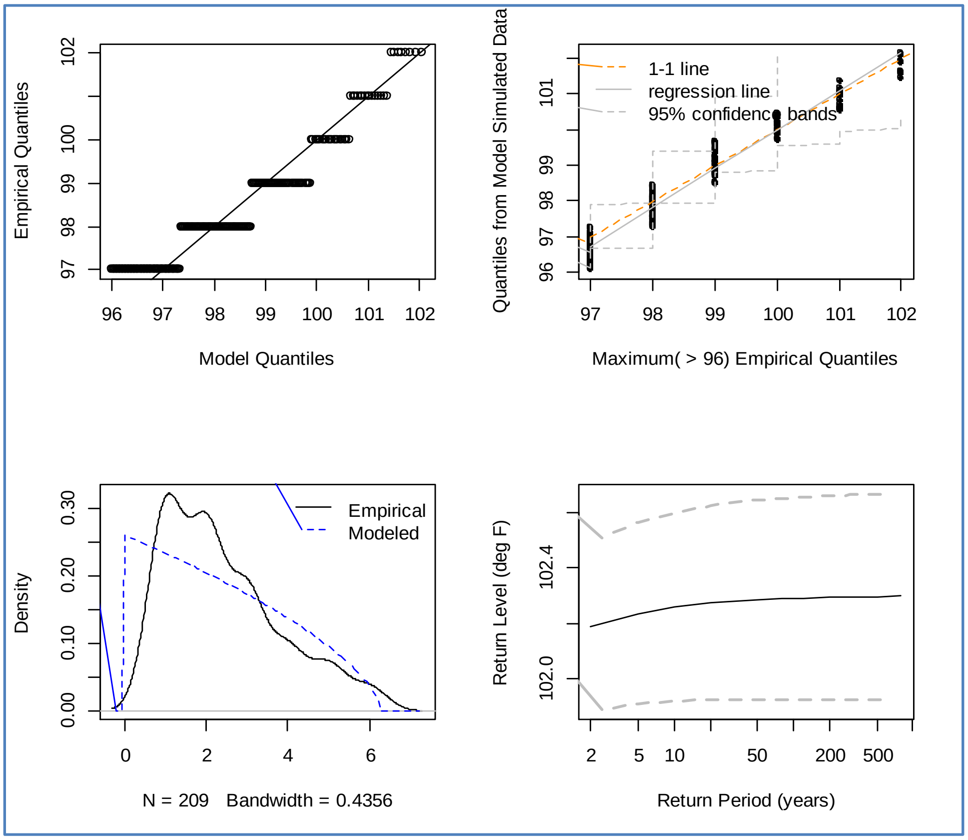

Figure 4.11

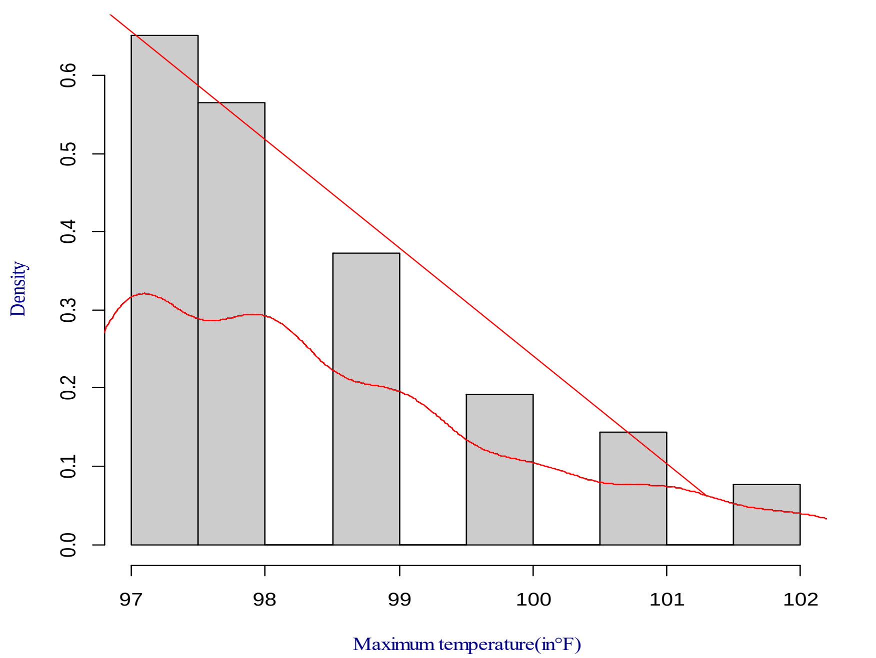

Figure 4.12

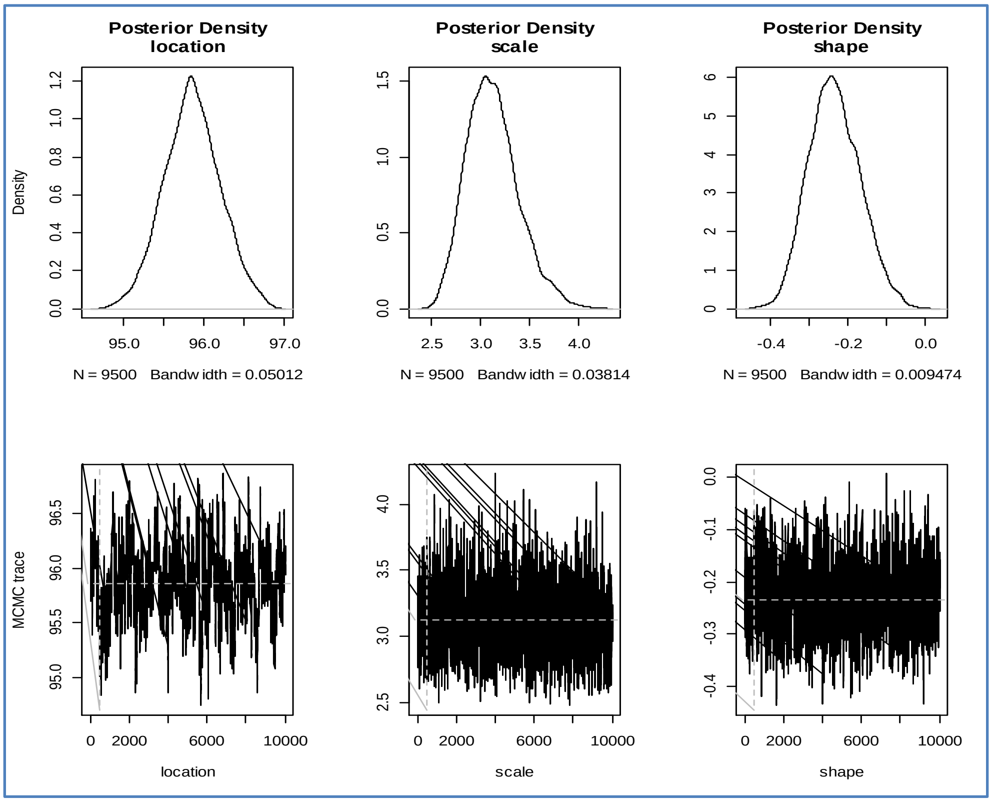

Figure 4.13

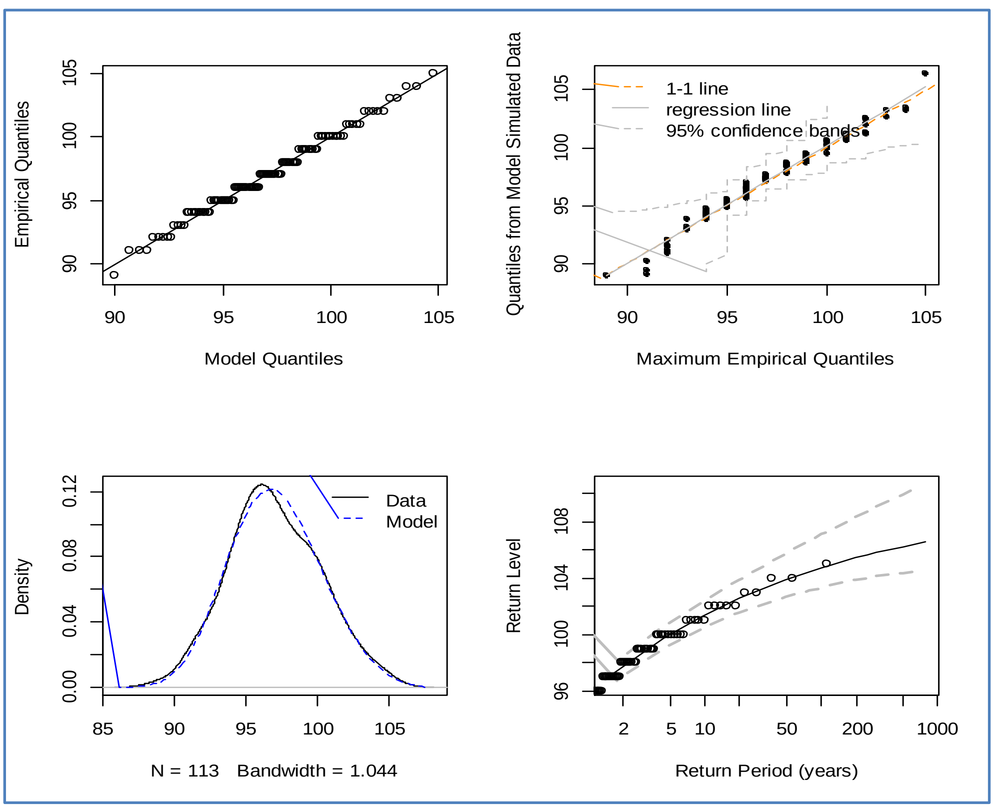

Figure 4.14