COMPUTATION OF PROBABLE MAXIMUM PRECIPITATION (PMP) AND ITS UNCERTAINTY

A Thesis

by

ABHISHEK SINGH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of the Committee, Vijay Singh
Committee Members, Ralph Wurbs
Srinivasula Ale
Head of the Department, Steve Searcy

December, 2016

Major Subject: Biological and Agricultural Engineering

ABSTRACT

Probable Maximum Precipitation (PMP) is used for estimating Probable Maximum Flood (PMF) which, in turn, is used for design of major hydraulic structures, such as dams and spillways, flood protection works, and nuclear power plants. One of the commonly used methods for estimating PMP is the statistical method that entails computation of frequency factor, adjustment of the frequency factor, construction of an enveloping curve of the frequency factor, estimation of PMP, choosing a probability distribution, and determination of the return period. This study determined the PMP values for different durations using data from the Brazos River basin, Texas. There are, however, uncertainties associated with the PMP values estimated using the statistical method. It was found that significant uncertainty in the PMP estimates can occur from the use of enveloping curve of the frequency factor, and uncertainty in sample mean and sample standard deviation. Hershfield's curve yielded higher PMP estimates, therefore, a basin specific-enveloping curve is suggested. The return period of a PMP value was obtained from frequency analysis. From 24 commonly used statistical distributions, 5 goodness of fit tests and the use of hazard rate, the Burr XII distribution was found to be the best frequency distribution. It was observed that the return period was not significantly higher than that obtained from the hydrometeorological reports (HMRs) of National Weather Service and other studies. For quantifying uncertainty, design risk estimates along with probability bounds on the PMP values were determined. The relative contribution of each random variable to the total uncertainty was also determined. Then, risk analysis of extreme precipitation was also done to assess the damage a PMP event can cause. The damage due

to a single PMP event of 12-hour duration can be as high as 2 billion U.S. dollars in Harris County, Texas.

NOMENCLATURE

List of Symbols

P - Probable maximum precipitation

X_m - Highest annual maximum precipitation values of a given duration

$\overline{X}$ - Mean of *n* annual maximum precipitation

n - Record length

σ_n - Standard deviation for a series of *n* annual maximum precipitation

$\overline{X}_{n-1}$ - Mean excluding the highest value from the series

S_{n-1} - Standard deviation excluding the highest value from the series

k_m - Frequency factor

α - Significant level

S_n - Sample standard deviation

μ - Population mean

σ - Population standard deviation

$E(\overline{X_n})$ - Expected value of the sample mean

$E(S_n)$ - Expected value of the sample standard deviation

$\Psi(a)$ - Incomplete gamma function

$E(P)$ - Expected value of the PMP estimate

$Var(P)$ - The variance of the PMP estimator *P*

$Var(S_n)$ - Variance of standard deviation

$Cov(\overline{X_n}, S_n)$ - Covariance of mean and standard deviation

α - First shape parameter of Burr distribution

κ - Second shape parameter of Burr distribution

β - Scale parameter of Burr distribution

P_d - Design risk PMP value

k_{mH} - Hershfield frequency factor value

k_{mB} - Basin specific frequency factor value

PMP_{HMR} - PMP values from the HMR documents

$f(t)$ - Probability density function

List of Abbreviations

PMP - Probable Maximum Precipitation

PMF - Probable Maximum Flood

MPP - Maximum Possible Precipitation

HMR - Hydrometeorological Reports

WMO – World Meteorological Organization

FAFP - Free-Atmospheric Forced Precipitation

FMP - Fractal Maximum Precipitation

NRC - National Research Council

GEV – Generalized Extreme Value

AMS - Annual Maximum Series

CV – Coefficient of Variation

FEMA – Federal Emergency Management Agency

HEC-FDA - Hydrological Engineering Center's Flood Damage Analysis

NWS - National Weather Service

NOAA - National Oceanic and Atmospheric Administration

USBR - United States Bureau of Reclamation

GOF - Goodness of Fit

K-S - Kolmogorov-Smirnov

A-D - Anderson-Darling

C-S - Chi-square

RMSE - Root Mean Square Error

PDF – Probability Density Function

CDF - Cumulative Distribution Function

NFIP - National Flood Insurance Program

GDP – Gross Domestic Product

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION AND BACKGROUND

1.1 Introduction

Probable Maximum Precipitation (PMP) is termed as "theoretically the greatest depth of precipitation for a given duration that is physically possible over a given size storm area at a particular geographic location at a given time of the year" (U.S National Weather Service, 1988). PMP is used for the calculation of Probable Maximum Flood (PMF) which is then used for design of hydraulic structures, such as large dams and spillways, flood control works, levees, and nuclear power plants. PMF is used to size the hydraulic structures such that the risk of their failure is minimized (Hershfield, 1965). There are uncertainties involved in PMP estimation regardless of the method used to calculate it. An upper bound with zero risk is not realistic, as there have been instances where storms in USA have exceeded the PMP estimates (Dooge, 1986) and the recorded floods have exceeded the estimated PMFs (Dawdy and Lettenmaier, 1987). Riedel and Schreiner (1978) concluded that the PMP estimates were too high east of 105th meridian where 18 storms out of 75 exceeded 70% of the PMPs.

Texas has a history of major floods that have caused huge losses of property and life. For example, in 1921 a tropical storm, that formed in the Bay of Campeche, caused 36.7 inches of precipitation within 36 hours, drenching San Antonio, causing the Thall flood in which 51 people were killed. Recently, on November 1st 2015 areas close to and near Houston received precipitation close to 12 inches causing damages to buildings,

property, and life. Brazos River basin also has a history of major floods due to extreme precipitation, like the Flood of 1899 causing damage to property over $9 million and killing 284 people. Under the specter of climate change, such catastrophic precipitation events are expected to increase and occur more frequently.

1.2 Background

PMP has been used to predict volume, timing, and peak flow of extreme flood events all around the world. Originally PMP was defined as Maximum Possible Precipitation (MPP), the value of precipitation that could not be exceeded. However, MPP values have been exceeded (Benson, 1973) and because of the complex atmospheric interactions contributing to extreme precipitation its name was changed to PMP. Since the 1940s, the National Weather Service has published a series of Hydrometeorological Reports (HMRs) that describe procedures for deriving the PMP values for the majority of United States. The generalized PMP studies currently used in the conterminous United States include HMR 49 (1977) for the Colorado River basin and Great Basin drainage; HMRs 51 (1978), 52 (1982) and 53 (1980) for the U.S. east of the 105th meridian; HMR 55A (1988) for the area between the Continental Divide and the 103rd meridian; HMR 57 (1994) for the Columbia River drainage basin; and HMRs 58 (1998) and 59 (1999) for California. Current HMRs and their application region are shown in Figure 1 (National Oceanic and Atmospheric Administration, 2012). Figure 2 shows an example of an HMR 51 PMP map and its coverage, as it covers all of the U.S east of the Front Range of the Rocky Mountains (National Weather Services, 1978). The main assumption in these procedures for PMP calculation is that there is the optimum combination of available moisture in the

atmosphere and efficiency of the causative mechanism in the storm that will cause maximum precipitation.

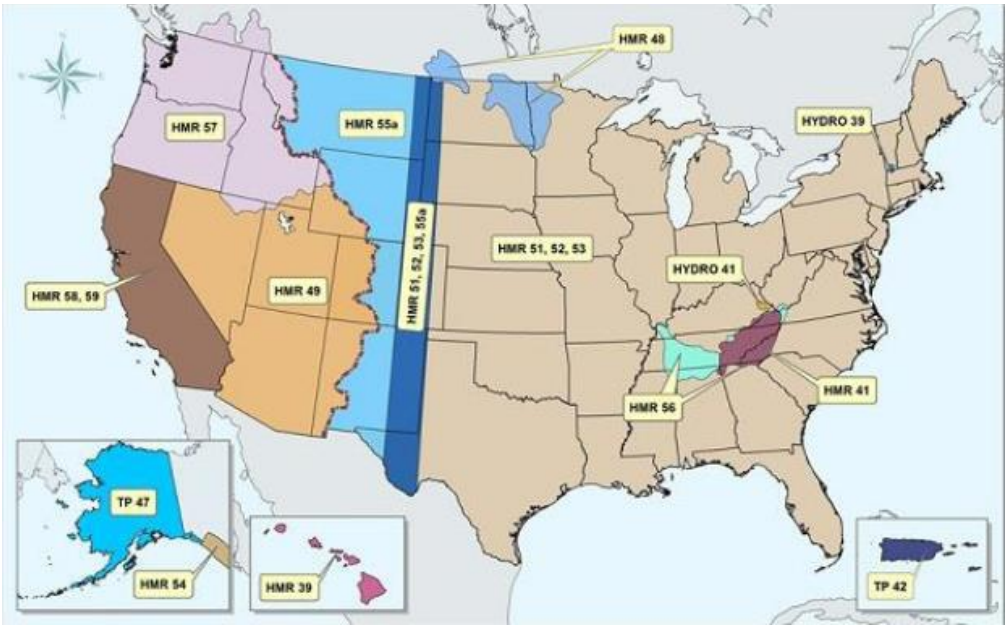


Figure 1. Regions for application for HMR PMP Reports. (http://www.nws.noaa.gov)

.

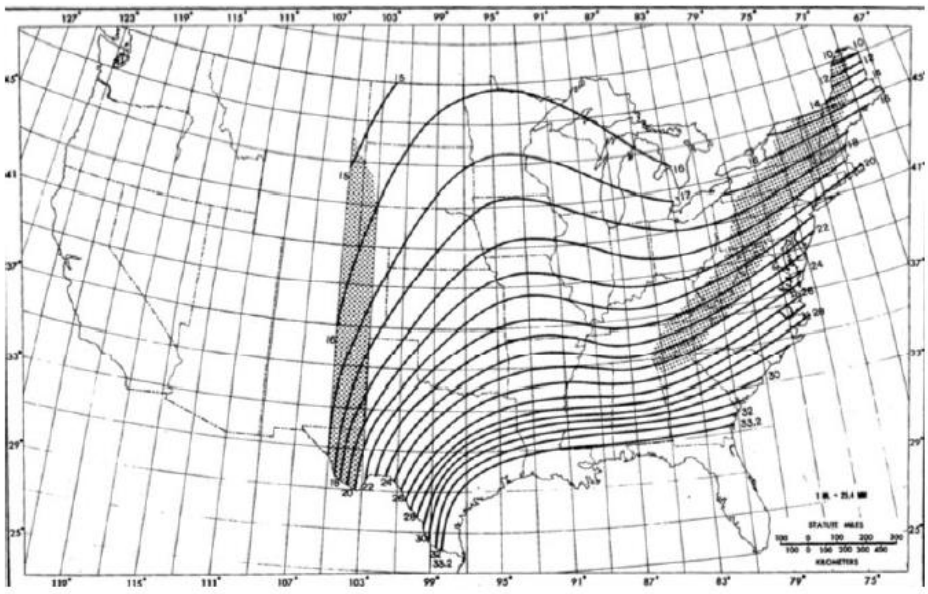


Figure 2. Example of a HMR 51 PMP map for 24-hour precipitation over 1000 square miles (Schreiner and Riedel, 1978)

2. LITRATURE REVIEW

2.1 Methods of Estimating Probable Maximum Precipitation (PMP)

There are different methods for PMP estimation which can be categorized as hydrometeorological and statistical. Common hydrometeorological methods include moisture maximization method, storm transposition method, and generalized method, and storm separation method. Statistical methods include Hershfield's method and multifractal method. Depending upon the watershed topography and data availability, some methods provide better PMP values in certain regions and other methods in other regions. In moisture maximization the storm precipitation is increased to such a value that is consistent with the maximum moisture in the atmosphere for the storm location and month of occurrence (Schreiner and Riedel, 1978). The basic assumptions in this method are that precipitation is linearly related to perceptible water. As the moisture available to the storm increases the precipitation efficiency of the storm does not change and the record of extreme storms is sufficiently large to represent the most efficient storm mechanisms but not the optimum available moisture that would accompany a PMP event (Tomlinson and Kappel, 2009). The efficiency with which storm converts moisture into precipitation and the amount of moisture content are considered important atmospheric conditions in most PMP studies and the moisture maximization procedure is used to approximate the highest moisture potential in the storm.

Storm transposition is associated with the relocation of storm precipitation within a region that is homogeneous relative to terrain and meteorological features important to the particular storm rainfall. The basic assumption behind the idea is that a

meteorologically homogeneous region exists such that a major storm occurring somewhere in the region could occur anywhere else in the region. The storms transported to a location could occur under similar meteorological conditions as the original location. It involves meteorological analysis of the storm to be transported, the determination of transposition limits, and the application of the appropriate adjustments for the change in storm location. The maximum observed storm precipitation data that is adjusted for moisture maximization is plotted on a map and is analyzed. The analysis allows the largest moisture maximized precipitation amounts to control the isolines within meteorologically homogeneous regions. This procedure implies transposition of the precipitation value and of the storm itself (U.S. Department of Commerce, 1960).

In the generalized method, maximum recorded rainfall depths of rainstorms over a large area and adjustment source are made in applying the maximum recorded rain depths to a particular catchment (Kulkarni, 2010). The generalized method has an advantage of using the maximum recorded rain depths for all combinations of area and duration and allowing for almost free transposition in space (Koutsoyiannis and Papalexiou, 2006). Rakhecha and Kennedy (1985) used a generalized method to estimate the PMP values for catchments of four large dam basins in India. It was assumed that the PMP values would result from the optimum combination of the available moisture in the atmosphere and the storm mechanism efficiency which was indirectly measured by observed precipitation.

The storm separation method is used particularly in orographic regions where storm transposition methods are inappropriate. It assumes that orographic and

convergence rainfall amounts can be explicitly determined. The convergence rainfall is referred to as the free-atmospheric forced precipitation (FAFP) (HMR, 57). HMR 36 is one of the earliest reports which discusses the development of PMP in terms of orographic and convergence components. Convergence precipitation is the product of atmospheric mechanisms acting independently from the terrain effect, and orographic precipitation is the precipitation that results from the terrain effect. It is recognized that atmosphere is not totally free from the terrain feedback, but cases can be found where the terrain feedback is either too small or insufficiently varied to explain the storm precipitation patterns, and in these cases precipitation is classified as pure convergence or non-orographic precipitation (U.S. National Weather Services, 1994).

Recently, multifractal analysis has been used for PMP estimation. Multifractal, also known as multiscaling, is widely used to describe the scaling behavior of precipitation and streamflow. Douglas and Barros (2003) used this technique to calculate the physically meaningful estimates of maximum precipitation from observations in the eastern United States. The multifractal approach has an advantage in that it provides a formal framework to infer the magnitude of extreme events, called the fractal maximum precipitation (FMP), independently of empirical adjustments, a site specific application of FMP in orographic regions. The method is constrained by the length of record, the spatial resolution of raingauge network, and the lack of uncertainty estimates.

Of all the methods, the statistical method, often called Hershfield method (1961), is more commonly used and can be applied, if long term precipitation data is available. The statistical method due to Hershfield (1961) has been used all over the world for

estimating the PMP values and for comparing with other methods and results obtained have been quite satisfactory (Casas et al., 2008; Tessier et al., 1992; Rakhecha et al., 1992; Srinivas and Chavan, 2015).

Since the Hersfield method is based on average precipitation and standard deviation of precipitation, it is similar to the Chow (1951) frequency factor method expressed as:

$$P = \overline{X} + k_m S_n \quad (1)$$

where, *n* is the number of annual maximum precipitation values corresponding to a given duration, $\overline{X}$ is the sample mean, S_n is the sample standard deviation, and k_m is the frequency factor. Hershfield (1961) used 15 as the maximum value of k_m for computing PMP. Later Hersfield (1965) found that an upper envelope of k_m had a tendency to decrease with the increasing precipitation amount. In other words, the frequency factor decreases with increasing mean annual maximum precipitation. The value of k_m varies from 5 to 20, depending upon the precipitation duration and average precipitation (Casas et al, 2011). This method was also used in this study. Hershfield (1965) analyzed over 95,000 station-years of annual maxima belonging to 2645 stations, about 90% data was from the United States and 10% from other parts of the world which included some of the heaviest precipitation regions. He then produced an empirical nomograph ranging from 5 minutes to 24 hours that have been standardized by WMO (1986) as a basis for estimating PMP (Koutsoyiannis, 1999). Figure 3 shows k_m as a function of mean annual maximum precipitation and duration.

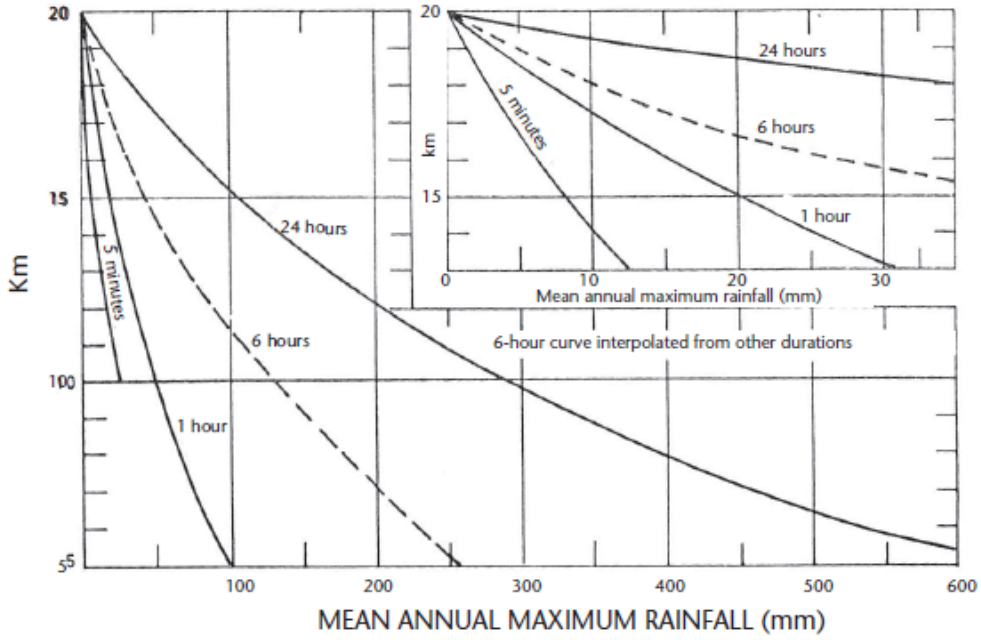


Figure 3. k_m as a function of mean annual maximum precipitation and of duration (WMO, 2009).

Using this method, enveloping curves were derived for particular areas and durations and these have been used to calculate the PMP values (Casas et al, 2008; Tessier et al, 1992). The enveloping frequency factor serves the purpose of transposition. Casas et al. (2008) used Hershfield's method to estimate the PMP values for one-day duration and their return periods, and spatial resolution over the Catalonia region. The Gumbel distribution with parameters estimated by the L-moments method was used to determine the return periods of calculated PMP values. They showed that 90% of the PMP values had return periods of 10^4 to 10^8 years.

2.2 Uncertainty in PMP Computations

Uncertainty with different methods of estimating PMP has been investigated by researchers who have been mainly concerned with maximizing and transposing actual storms using in-place moisture maximization (Micovic et al, 2014; Koutsoyiannis and Papalexiou, 2006). Micovic et al. (2014) identified 5 main sources of uncertainty, including horizontal transposition factor, factor for storm efficiency, factor of in-place moisture maximization, factor for centering the storm within the basin, and 24-hour precipitation for the controlling storm at the location. They found the operational PMP estimates to be lower than the theoretical upper limit by some variable deriving the estimates. There can occur uncertainty in the PMP estimate due to the way we define storm center at the location of storm occurrence. From analysis of atmospheric moisture, dewpoint temperature and maximized precipitation, Koutsoyiannis and Papalexiou (2006) concluded that no upper bound of PMP estimates was evident, and suggested finding the design values of maximum precipitation by using frequency analysis of observed data based on the GEV distribution. Uncertainty can also be categorized as natural uncertainty which represents the intrinsic variability of the physical system and the knowledge uncertainty which is due to insufficient data and lack of understanding of the system (NRC, 2000).

Studies focusing on uncertainties in the PMP estimates using the statistical method or Hershfield method have been limited (Salas et al, 2014; Koutsoyiannis, 1999). There can be two ways to quantify uncertainty in the PMP estimates using statistical method. First, uncertainty can be determined due to uncertainties in frequency factor,

and mean and standard deviation of extreme precipitation values. Second, frequency analysis of PMP can be used to quantify uncertainty. There exist uncertainties in the frequency factor (k_m) which is accounted for by using an enveloping function of the highest frequency factor values. Koutsoyiannis (1999) pondered whether the extreme precipitation data used in the Hershfield method suggested a deterministic upper limit of precipitation. He suggested unifying all classes of record length and adding the number of occurrences of all classes after ignoring the effect of record length on k_m. Considering it as a random variable, the probability of its non-exceedance can be estimated using the Weibull formula, assuming all records of standardized annual maximum precipitation k_m represented practically the same population.

There are also uncertainties in the sample mean and sample standard deviation which can affect the PMP estimation. Assuming that the extreme precipitation series followed the Gumbel distribution, Salas et al. (2014) considered the uncertainty of PMP estimates arising from the uncertainty of sample mean and sample standard deviation. They calculated the expected value and standard deviation of the PMP values obtained from the Hershfield method and then estimated the design risk values of PMP using Chebyshev's inequality.

2.3 Frequency Analysis of Extreme Precipitation

On the other hand, the uncertainty of PMP values can be quantified by frequency analysis of the annual maximum precipitation series. The first step is to determine the best fit probability distribution for the extreme precipitation series and return periods of PMP values. The exceedance probability of PMP values can be used to analyze risk. Although

the definition of PMP assumes an upper bound of precipitation, there are however no assigned probability levels and return periods to 'probable' events which might exceed the upper limits (Kites, 1988). There is the unknown risk of occurrence of such extreme events. However, by selecting an appropriate distribution for extreme precipitation values and ignoring the concept of upper limit, the return period can be calculated for the estimated PMP value.

Various probability distributions can be used to calculate the return periods of maximum precipitation of different durations or calculate the return period for extreme precipitation. The Gumbel distribution has been commonly used for extreme frequency analysis, because maximum annual precipitation series are relatively short, especially in developing countries, and outliers are observed. The traditional fitting method with the conventional moments, such as mean and standard deviation, can result in return periods shorter than the ones corresponding to a longer sample containing a large number of years (Casas et al, 2008). Adjustment of CV of the annual maximum precipitation series can be done to compensate for the effect of outliers (Rakhecha et al., 1992).

There is a considerable amount of uncertainty associated with finding the best-fit distribution for doing frequency analysis. Stations having limited quantity of data for frequency analysis introduce sampling uncertainty, in particular due to the presence of outliers, which make the estimates of higher order moments (like skewness) become unstable (Rahman and Mamoon, 2014). For daily time series, Koutsoyiannis (2004) found that the Generalized Extreme Value (GEV) type II (EV2) better described hydrological extremes than did the Gumbel distribution. Assuming the shape parameter of the EV2

distribution as constant (= 0.15) across Europe and North America, the distribution fitting was simplified. More recently Papalexiou and Koutsoyiannis (2012) used a three-parameter Generalized Gamma (GG) distribution and a four-parameter Generalized Beta distribution of the second order (GB2) to 11519 daily precipitation records across the globe. Results showed that these distributions described almost all empirical records satisfactorily.

Determining the best fit probability distribution is important to quantify the uncertainty in the PMP estimates. Asquith (1998) analyzed frequencies of annual maximum precipitation for durations of 15, 30, and 60 minutes; 1, 2, 3, 6, 12, and 24 hours; and 1, 2, 3, 5, and 7 days using L-moments like mean, L-scale, L-coefficient of variation, L-skew, and L-kurtosis. He found that the generalized logistic distribution, using L-moment ratio diagrams, was an appropriate probability distribution for modeling the frequency of annual maxima for durations of 15 minutes to 24 hours; whereas the generalized extreme-value distribution was appropriate for durations of 1 to 7 days (Asquith, 1998). However, the results were only based on the L-moments ratio and included only a few distributions like Generalized Logistic distribution and Generalized Extreme value (GEV) distribution, Generalized Pareto distribution, and Pearson Type III distribution. To our knowledge, the best-fit probability distributions for different durations like 2, 3, 6, 12, 24 hours have not been determined for the Brazos River basin.

2.4 Risk Analysis of Extreme Precipitation

Flooding from extreme precipitation can vary from upwelling groundwater levels which occur frequently to very large inundations (Koks et al, 2012). Large damage due to

extreme precipitation occurs in case of large inundations when there is more precipitation than the water system in a specific area can handle. For decision making the expected damage along with the probability of exceedance of extreme precipitation is also vital to quantify the amount of risk associated with PMP. Risk can be defined as the expected losses due to a damaging event. It is a combination of the amount of damage caused for a particular hazard and the probability associated with this particular hazard (Villalta et al, 2014). Assuming the probability of exceedance of a precipitation event E_i is p_i and the associated loss is L_i. The number of precipitation events per year is not limited to one and numerous events can occur in a given year. The expected loss for a given event, E_i in a given year is simply $E(L) = p_i \times L_i$ (Kunreuther et al, 2004). Spekkers et al. (2012) correlated peak precipitation intensity and precipitation volume with total damage per 1000 insurance policies for private property owners that were in the vicinity of raingauges on pluvial flooding in the Netherlands. They estimated the total damage within an assumed radius of the rain gauging location. Koks et al. (2012) compared the flood risk in terms of annual expected damage (AED) of inundation due to extreme precipitation and large floods from the sea or river. They formed an integrated model to compare different types of flood risks and calculated the damage due to extreme precipitation for different return periods and different land uses in the area. In the United States the depth-damage curves developed by the U.S. Army Corps of Engineers have been utilized to determine the impact of floods in monetary terms. The curves are a relationship between the depth of water above or below the first floor of the building and the amount of damage that can be attributed to that water (David and

Skaggs, 1992). Different models like Hydrological Engineering Center's Flood Damage Analysis (HEC-FDA), HAZUS, etc. have been used to study the risk based analysis methods for flood damage reduction studies. However, we are only concerned with the flooding due to extreme precipitation or pluvial floods. Therefore, the question arises: "What are the uncertainties associated with the PMP estimated from the statistical method, and what in monetary terms damage a single PMP event can cause?"

3. SIGNIFICANCE, STUDY AREA AND APPROACH

3.1 Study Area

The study area for the study is the Brazos River basin which is the second largest river basin by area within Texas. Brazos River is the 11th largest river in the United States. It is 2,060 km long with its headwater source at the head of Blackwater Draw, Curry County, New Mexico to its mouth at the Gulf of Mexico (Wiki). It has a drainage basin area of about 116,000 km^2. Within the basin there are different types of climate and precipitation producing mechanisms. The climates are subtropical humid close to the Gulf of Mexico, continental steppe close to New Mexico or on the western side of the basin, and subtropical sub-humid in the middle part of the Basin (Hao and Singh, 2011). Figure 4 shows the location of Brazos River basin.

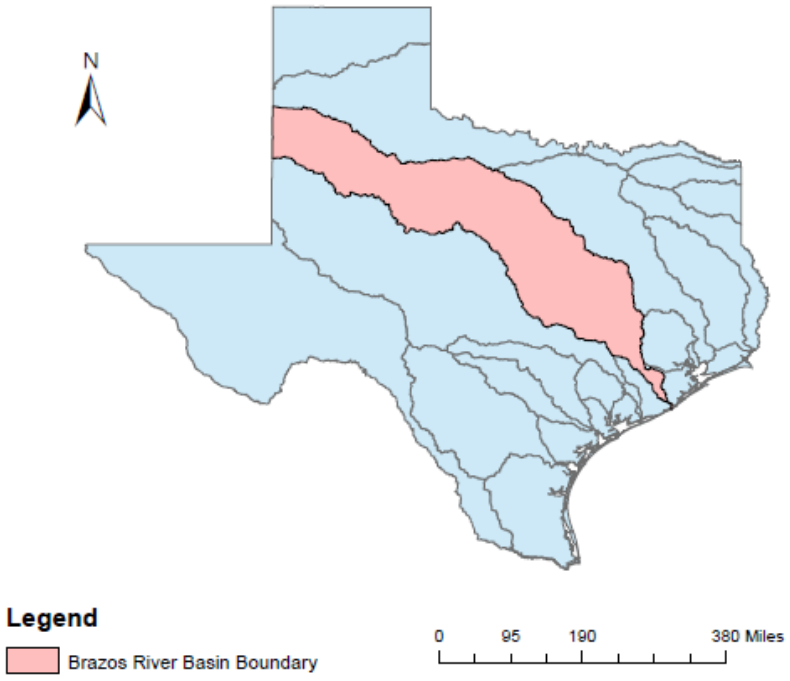


Figure 4 Brazos River Basin within Texas

3.2 Significance of the Work

There is a need to determine the site-specific PMP for the study area which can then be used for the calculation of PMF. It is because the site-specific PMP calculation can incorporate basin characteristics that are specific to the topography and local climate. Therefore, we ask the question: "What are the PMP estimates for Brazos River basin and what are the uncertainty and risk associated with those values." Our study calculated PMP for 1, 2, 3, 6, 12, and 24 hour durations and focused on uncertainties due to the use of frequency factor, enveloping curve, return period of PMP values, uncertainty in the

selection of best fit probability distribution, and uncertainties due to sample mean and sample standard deviation of annual maximum precipitation. We calculated the design risk estimate of the PMP values and probability bounds on the design PMP values. Such a design risk estimate gave a more conservative estimate of the PMP. Risk analysis of extreme precipitation was also performed and the total loss that can be expected from PMP events was calculated in Harris County.

It is also important to see how PMP values vary with the given duration and if there is any relation between the PMP values and the mean of extreme values, PMP values and the highest observed precipitation, or the mean and the standard deviation for different stations and durations. If there is any correlation between these statistics, then one statistic can be substituted for the other.

3.3 Objectives

The objective of this study therefore was to estimate PMP values for different durations and locations in the Brazos River basin using the statistical method and determine the associated uncertainty along with risk analysis of extreme precipitation. To achieve this objective, specific objectives were to:

(1) construct a basin-specific enveloping curve of frequency factor for the Brazos River basin and calculate the PMP values by using it and construct the Isohyetal maps of PMP values;

(2) determine the best-fit probability distribution for extreme precipitation and probability of exceedance and return period of PMP values for the Brazos River basin;

(3) compute the uncertainties associated which the statistical estimates of PMP values arising from the uncertainties from the choice of probability distribution, number of stations, and frequency factor, uncertainties in sample mean, sample standard deviation.; and

(4) compute design risk PMP values and probability bounds on the PMP estimates, estimate risk and assess the damage.

4. APPROACH AND HYPOTHESES

4.1 Approach

To fulfill our objectives following approaches were undertaken.

4.1.1 Estimation of PMP Values

The nomograph of varying frequency factor with sample mean was constructed and PMP values for different durations were estimated using statistical approach for the Brazos River basin. Assessment of uncertainty in the PMP estimate due to was performed along with comparing the results with the use of Hershfield's curve of frequency factor.

4.1.2 Frequency Analysis of Extreme Precipitation

Frequency analysis of extreme precipitation was performed for different durations. The best fit probability distribution was determined for the study area. The exceedance probabilities and return periods of PMP values were estimated. The return periods of PMP values were compared with the published HMR documents return periods, and the uncertainty introduced due to choice of probability distribution was determined.

4.1.3 Uncertainty Analysis of PMP Values

With the use of parameter values of best fit probability distribution, the uncertainty introduced in the PMP estimates due to sample mean and sample standard deviation was accounted for. Design risk estimates of PMP values were determined along with probability bounds on the PMP values. Using Taylor series expansion the relative contribution of each random variable to the total uncertainty was also determined.

4.1.4 Risk Analysis

Risk analysis of PMP estimates in Harris County was performed. Assessment of damage was done along with estimation of risk. The amount of damage that can be expected in a PMP event at the location was determined.

4.2 Hypothesis

HMR documents provide generalized precipitation values that are not basin specific. Hence, they tend to represent the largest PMP values across broad regions. Many site-specific studies in the past have produced different PMP values compared to HMR published values (Tomlinson and Kappel, 2009). The hypothesis is therefore as follows: By incorporating uncertainty in the PMP estimates the PMP values will differ widely from using Hershfield's method and published HMR documents. For a small number of stations in Brazos River basin as compared to 2645 stations used by Hershfield, the frequency factor will depend heavily on particular stations. Hence, we will get different PMP values when using Hershfield's original enveloping curve. Within Brazos River basin there exist different climate producing mechanisms for different areas. For example, in the eastern part of Texas or near the Gulf of Mexico there is fairly uniform seasonal precipitation, with slight maxima occurring in the summer season because the influence of the Gulf of Mexico is dominant (National Fibers Information Centre, 1987). In northwestern part of Texas the precipitation amount increases steadily through spring and reaches maximum in May and June, while thunderstorms are also high during spring (National Fibers Information Centre, 1987). Hence, for finding the best fit probability distribution it may not be possible to describe the whole data by a single best distribution. The best

distribution may vary depending upon the time duration, climatic zone, and distance from the gulf.

5. ESTIMATING PMP VALUES USING STATISTICAL APPROACH

5.1 Data Collection

Precipitation data for 1-hour duration were taken from the NCDC NOAA website (https://www.ncdc.noaa.gov/cdo-web/). Shapefiles of raingauging stations to be imported into GIS were prepared using the latitude and longitude of stations. Using the locations of stations and boundary of Brazos River basin it was found that the basin had more than 90 stations. The stations were selected based on the criteria of having at least 30 years of record length and having at least 9-month observations for each year (Singh and Hao, 2013). 39 stations were selected having an average record length of 50 years. Figure 5 shows the locations of 1-hour duration rain gauges. The recording time varied from 1940 to 2013. 17 stations had record lengths of more than 60 years. From the data of 1-hour duration the data for other durations 2, 3, 6, 12, and 24 hours were generated. Time series of stations with different durations were plotted to see if there was any trend in the precipitation records as a function of time. No time series plot showed any significant non-stationarity. Then, annual maximum precipitation series based on different durations were compiled for each station.

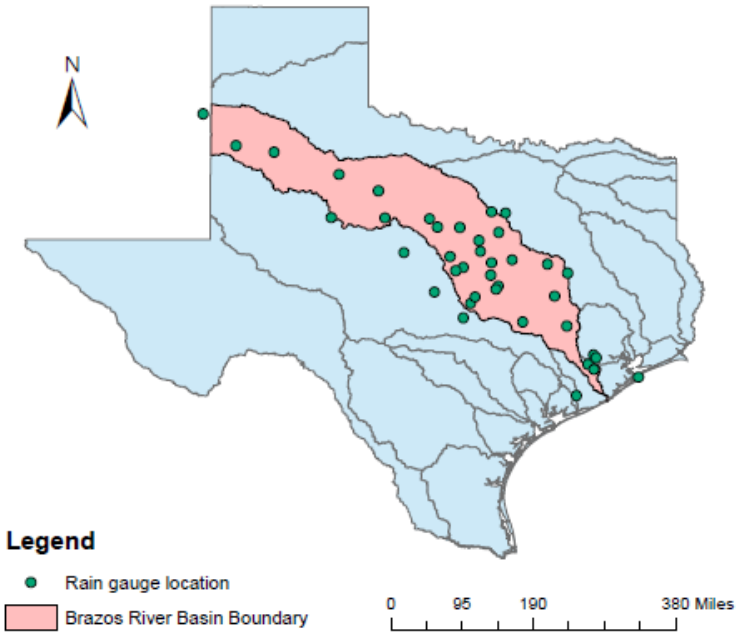


Figure 5. 1-hour durations rain gauge station locations

5.2 Estimation of Frequency Factor

The values of mean, standard deviation, and highest observed precipitation were calculated for annual maximum series of each station corresponding to each duration. Mean and standard deviation were adjusted for sample size and maximum observed event. Adjustments were made based on Figure A.1 and A.2 (Appendix) (WMO, 2009). The mean and standard deviation of the annual maximum series tend to increase with the length of record, because the frequency distribution of precipitation extremes is skewed to the right so that there is a greater chance of getting a larger than a small extreme as the length

of record increases. Hence, for smaller series of extreme precipitation *n*, adjustments were made to the mean and standard deviation for the length of record based on Figure A.3 (Appendix) (WMO, 2009). It may be pointed out that Hershfield (1965) derived his enveloping curve based on 204 stations (for 1-, 2-hour durations) and 2,700 stations (for 24-hour duration), whereas we used only 39 stations. The coefficient of variation (CV), the ratio of standard deviation, and mean of the annual maximum series was calculated for each station. Sometimes the inclusion of an outlier or an extraordinary extreme precipitation event, with a recurrence period much longer than the series, could cause an anomalous effect in the calculated mean and standard deviation values (Hershfield, 1961). The CV for each station was calculated and checked weather it did not differ too much from that of the neighboring stations. For stations whose CV value found to be too much different from the neighboring stations, it was adjusted to the nearest value as compared to the neighboring stations (Rakhecha et al., 1992).

The frequency factor k_m was calculated as:

$$k_m = \frac{X_m - \overline{X}_{n-1}}{S_{n-1}} \tag{2}$$

The highest value of frequency factor for 1 hour duration was found to be 10.1 at Santa Anna, Texas. This value of 10.1 is the single highest value but it cannot be used for the whole basin as value of frequency factor will vary depending upon the location of stations corresponding to climatic regions and different geographic locations. A similar procedure was applied for other durations of maximum precipitation series. Table 1 shows the maximum observed along with the station name for different durations. Histograms of

were also plotted for different durations (Appendix B, Figure B.1 to B.6) Since each station had its own frequency factor value, depending upon the magnitude of the mean, the values of 39 stations were plotted against the adjusted mean $\overline{X}$ in order to consider an appropriate enveloping curve that would give reliable estimates of 1-hour PMP rather than using the observed highest value.

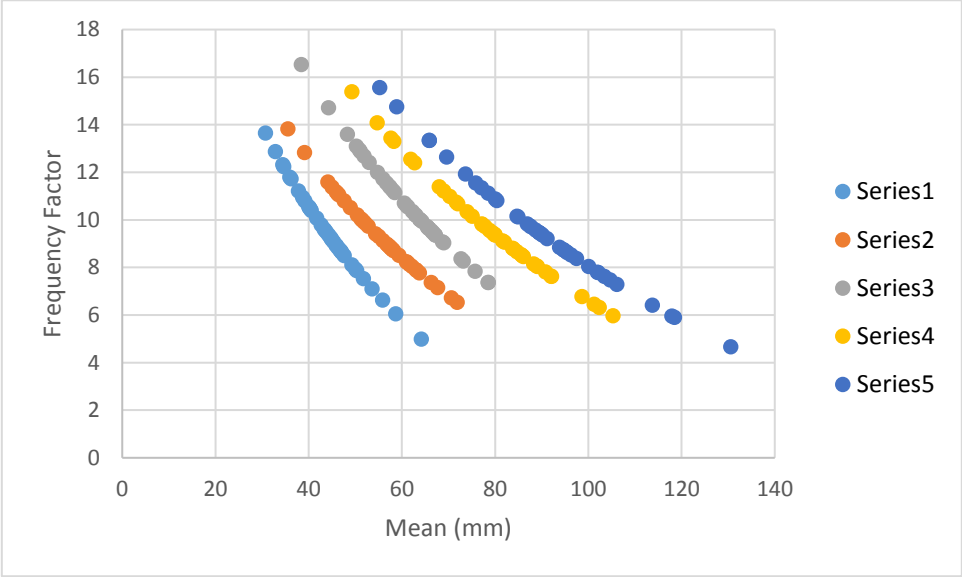
Table 1 Maximum observed frequency factor for different stations and durations

Stations	Duration					
	1-hour	2-hour	3-hour	6-hour	12-hour	24-hour
Albine	2.3	3.3	3.37	3.31	3.36	3.32
Bay City	9.1	5.4	4.68	3.58	4.06	4.15
Belton	3.7	2.8	3.51	4.91	4.1	3.9
Bertnam	4.8	3.9	3.92	3.28	3.54	3.09
Briggs	8.3	9.9	10.2	4.94	4.34	3.52
Burleson	3.8	3.33	3.42	3.24	2.7	2.22
Clovis	2.9	3.45	3.46	2.92	3.82	3.29
Coryell	6.3	6.18	5.63	5.32	4.06	5.87
Cranfills	4.08	3.65	3.27	2.69	3.76	4.17
Cherroke	3.47	3.28	2.68	6.06	5.43	5.43
Cresson	4.51	4.24	3.68	3.61	5.83	4.46
Eastland	4.36	3.14	2.89	4.14	3.59	3.79

Table 1 Continued

Stations	Duration					
	1-hour	2-hour	3-hour	6-hour	12-hour	24-hour
Evant	6.55	6.8	7.6	7.1	5.89	5.52
Santa Anna	10.1	10.9	11.4	11.4	10.5	4.9
Flat	2.9	3.65	4.07	3.5	2.97	2.71
Galveston	4.24	2.5	3.1	3.3	3.47	5.12
Gorman	6.53	4.55	3.62	3.52	3.69	3.4
Groesbeck	3.84	3.02	2.82	2.81	3.28	2.42
Houston Addicts	4.06	3.16	4	3.96	3.45	3.41
Houston Alife	4.5	3.52	3.09	3.97	6.28	5.25
Indian gap	2.8	3.31	3.06	2.46	3.7	5.97
Iredell	3.7	2.52	2.66	2.95	3.1	2.53
Jayton	7.9	7.56	7.48	8.9	5.23	4.52
Jewett	4.8	4.79	4.62	4.59	3.98	5.11
Kopperl	4.4	3.78	3.68	3.6	3.77	4.01
Lexington	9.01	6.01	4.91	3.64	4.61	6.02
Loraine	2.8	3.03	5.4	5.02	4.61	3.52
Lubbock	5.5	4.44	3.69	3.15	2.53	2.66
Moline	5.05	4.09	4.09	3.54	3.4	4.18
Pep	6	4.56	4.11	3.75	3.47	2.74
Richmond	2.7	2.84	2.33	4.94	3.39	3.09
Spicewood	3.4	2.94	2.88	4.24	5.09	4.81

An enveloping curve was drawn with the help of upper points for different durations. Figure 6 shows the enveloping curves for different durations. The curve seemed to be more sensitive for lower durations of precipitation, meaning changing the mean changed the value of the corresponding frequency factor by a considerable amount. However, all of the curves followed the same trend.



Series 1- 1-hour, Series 2- 3-hour, Series 3- 6-hour, Series 4- 12-hour, Series 5- 24-hour

Figure 6. Enveloping curve of k_m for all durations

5.3 Uncertainty in Frequency Factor

Figure 7 to 11 shows the enveloping curve of frequency factor based on 39 stations in the Brazos River basin and the enveloping curve provided by Hershfield for computing PMP based on different durations.

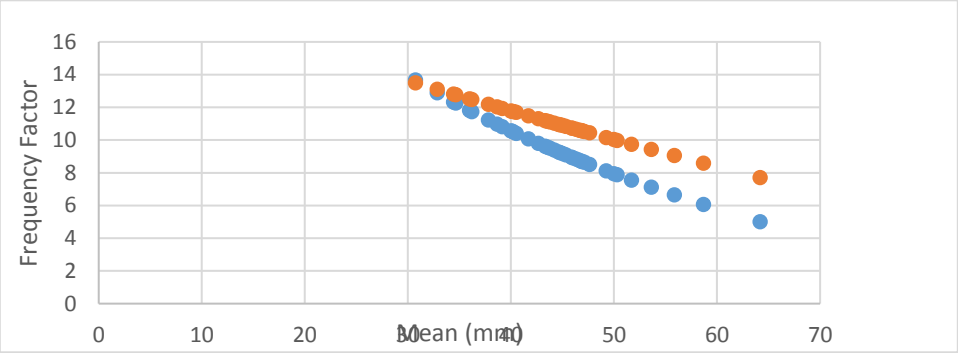
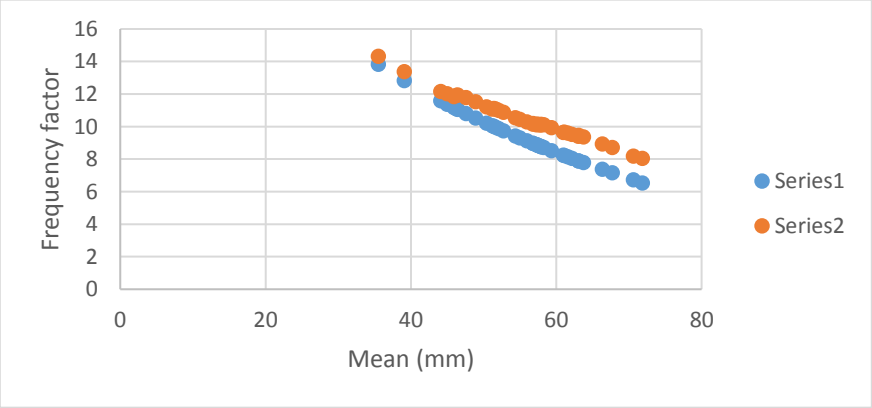
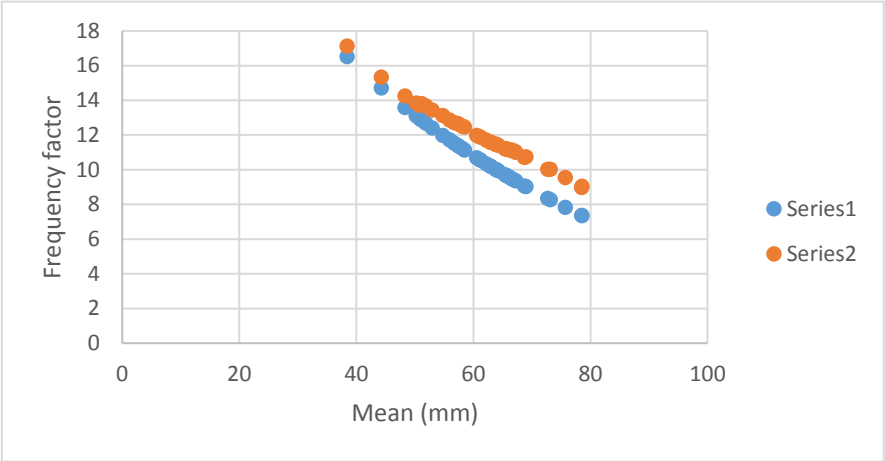


Figure 7 Comparison of Hershfield's enveloping curve of frequency factor with Brazos River basin enveloping curve based on 1 hour duration



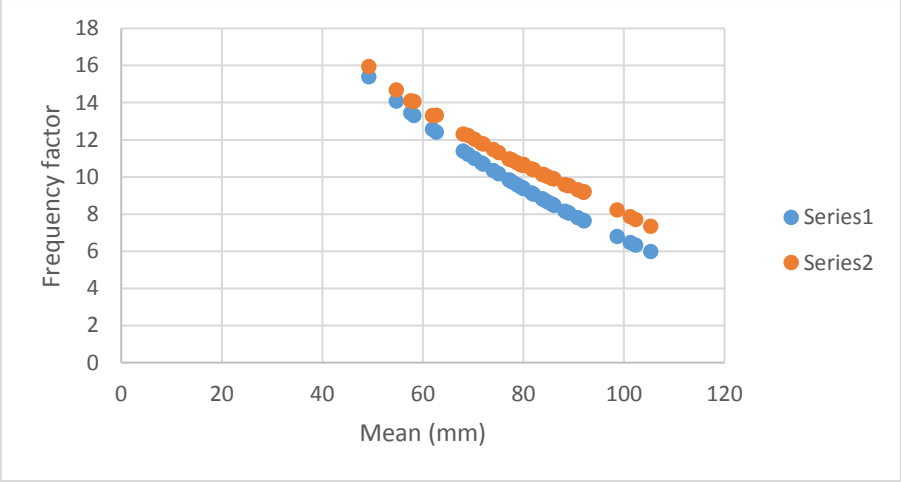
Series 1 – Basin-specific curve, Series 2- Hershfield curve

Figure 8 Comparison of Hershfield's enveloping curve of frequency factorwith Brazos River basin enveloping curve based on 2 hour duration.



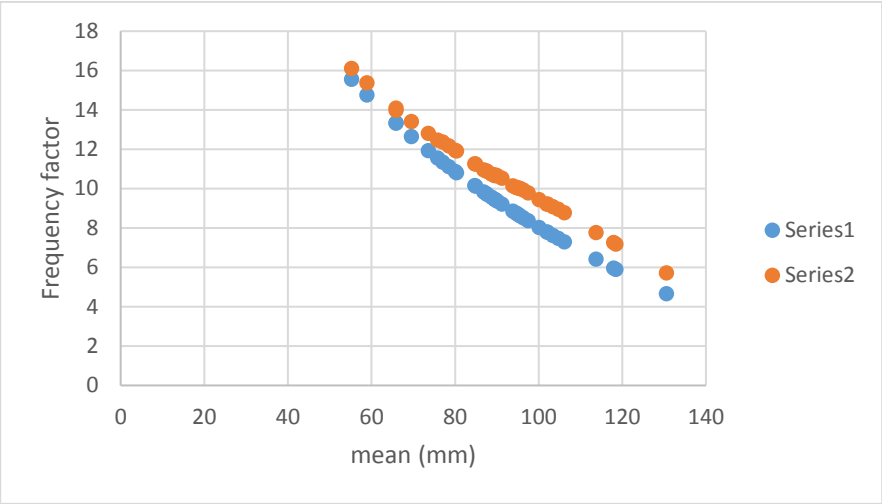
Series 1 – Basin-specific curve, Series 2- Hershfield curve

Figure 9 Comparison of Hershfield's enveloping curve of frequency factor with Brazos River basin enveloping curve based on 3 hour duration.



Series 1 – Basin-specific curve, Series 2- Hershfield curve

Figure 10 Comparison of Hershfield's enveloping curve of frequency factor with Brazos River basin enveloping curve based on 12 hour duration.



Series 1 – Basin-specific curve, Series 2- Hershfield curve

Figure 11 Comparison of Hershfield's enveloping curve of frequency factor with Brazos River basin enveloping curve based on 24 hour duration.

From the Figure, it is seen that both curves generally followed the same trends but did not match. Brazos River basin has a smaller number of stations than the 2645 stations that Hershfield (1965) used, hence the frequency factor markedly depends on the number of stations. The enveloping curve specific for the Brazos River basin is lower than the Hershfield curve, which was constructed using some of the highest precipitation producing regions with long term records. The Hershfield enveloping curve seems to give higher values of frequency factor as the mean increases. Figure B.7 to B.11 shows the algebraic relations between the difference between the basin-specific and Hershfield's curve in dimensionless terms. The relationship for different durations was based on the equations:

$$y = 0.0025x - 0.106 \quad (3)$$

$$y = 0.0031x - 0.1289 \quad (4)$$

$$y = 0.0036x - 0.1435 \quad (5)$$

$$y = 0.0045x - 0.159 \quad (6)$$

$$y = 0.0056x - 0.1328 \quad (7)$$

where, y is the Difference between the frequency factor from both the methods and x is the mean of annual maximum series. Hence, it is more conservative than the basic-specific one for different durations.

Figure 12 shows the difference between the frequency factor values for each station for 1-hour duration in dimensionless terms based on the formula:

$$\frac{k_{mH} - k_{mB}}{k_{mH}} \quad (8)$$

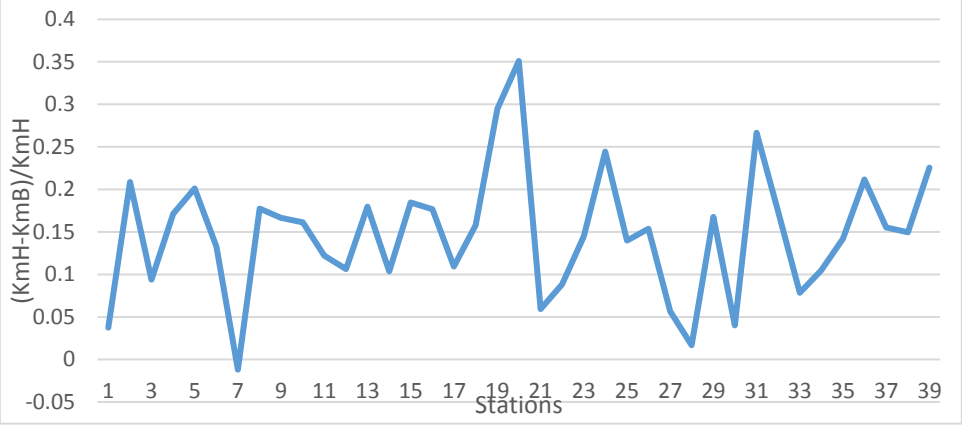


Figure 12 Uncertainty in the values of frequency factor in dimensionless terms

where k_{mH} is the Hershfield frequency factor value and k_{mB} is the basin specific frequency factor value. This difference is an indication of uncertainty that can be introduced when using the Hershfield curve rather than the basin-specific curve. The same procedure was applied for other durations and the same trend was observed. Using Hershfield's curve rather than basin specific can increase frequency factor 16% for 1 hour duration, 16.4% for 2 hour duration, 17.3% for 3 hour duration, 18.9% for 6 hour duration and 22.1% for 24 hour duration. k_m was also calculated by using the PMP values published in HMR documents (HMR, 51). The range of PMP values varied from 863.6 mm (station at Pep) to 1198.8 mm (station at Houston Alife) for 24-hour duration. The value of k_m was calculated as:

$$k_m = \left(\frac{PMP_{HMR} - \overline{X}}{S_n}\right) \tag{9}$$

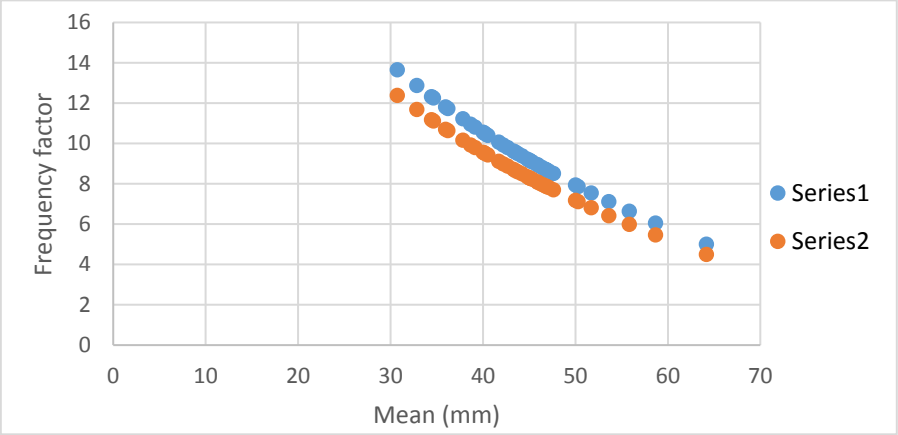
where PMP_{HMR} is the PMP values from the HMR documents. Using the PMP values and the mean and standard deviation of stations, the range of frequency factor was from 22.2 to 26.6. It was too high with a narrow difference between the highest and lowest values. It was because the PMP values published in HMR are too high as compared to the average precipitation amount and the PMP estimated using basic-specific enveloping curve. This shows the significance of constructing the basin-specific enveloping curve and then calculating PMP.

In order to quantify the uncertainty due to the number of stations, the enveloping curve was constructed by removing the top two stations (Lexington and Briggs) on an hourly basis. The curve changed, giving lower values of frequency factor that gave lower

PMP values (Figure 13). The frequency factor, on an average basis, decreased by 8.1%. Curve was also fitted by nondimensionalizing and calculating the difference between both curves. Figure 14 shows the curve and it was observed that the difference was more for stations having less mean. The difference can be calculated by the equation:

$$y = -0.0055x + 0.2088 \quad (10)$$

where, x is the mean (mm) for stations and y is the difference in dimensionless terms. Uncertainty can be introduced in the curve. Also, the inclusion of any outlier can increase the value of frequency factor which can change the shape of the curve.



Series 1- Original enveloping curve, Series 2- Curve made upon removing top two stations

Figure 13 Comparison of the original enveloping curve and the curve made upon removing top two stations

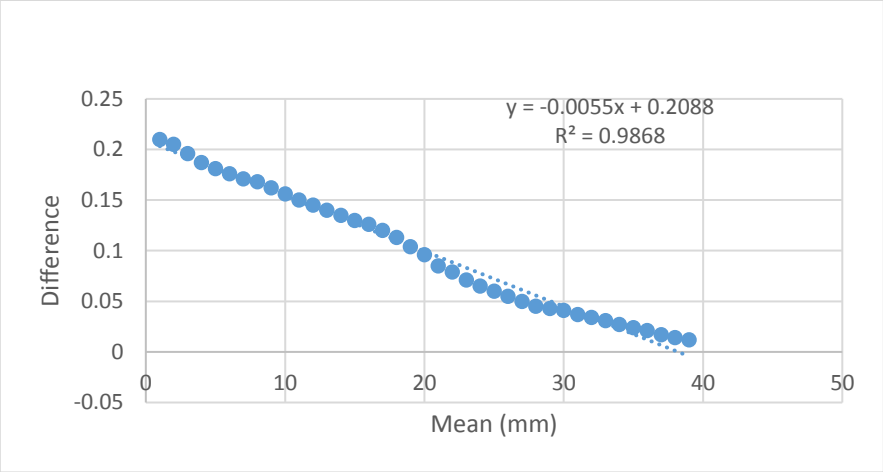


Figure 14 Difference between the original enveloping curve and the curve made upon removing top two stations in dimensionless terms

5.4 Computation of PMP

Then the PMP values for each station and duration was calculated based on equation (1):

$$P = \overline{X} + k_m S_n \qquad (1)$$

The calculated PMP values were adjusted for the fixed observational time interval. As precipitation data are usually given for fixed time intervals, for example 3 AM to 4 AM (hourly data), 6 AM to noon (6-hourly), or 8 AM to 8 PM (daily). The adjustment will yield values closely approximating those to be obtained from an analysis based on true maxima (Hershfield, 1961*a*, WMO, 2009). However, less adjustment is required when maximum observed amounts for various durations are determined from two or more fixed time intervals (Weiss, 1968; Miller, 1964). Recent studies indicate little higher values for the correction factor (Casas et al, 2008), but we used Figure A.4 (Appendix), as it is mostly used and has been generalized by WMO (2009). Table 2 shows the PMP values

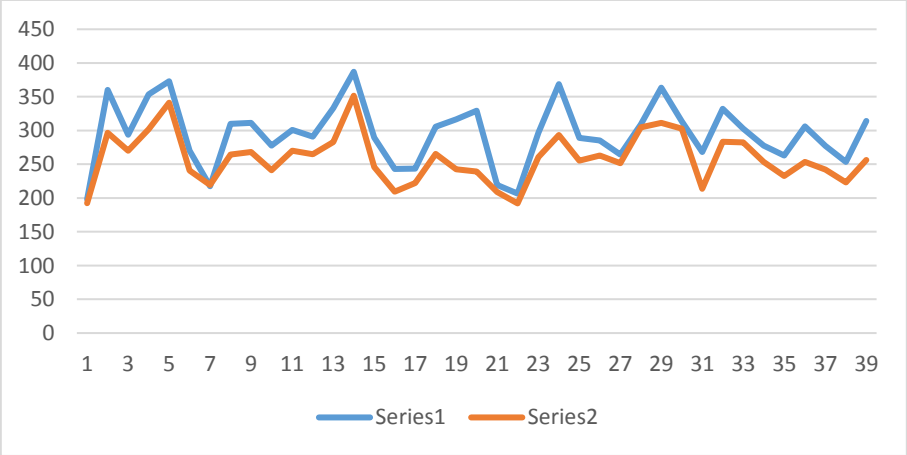
for the study area using basin-specific enveloping curve. Histograms of PMP values were also plotted for different durations (Appendix B, Figure B.12 to B17).

Table 2 Adjusted PMP values for different Stations and Durations (mm)

Station	Duration					
	1-hour PMP	2-hour PMP	3-hour PMP	6-hour PMP	12-hour PMP	24-hour PMP
Albine	192.2	242.9	271.1	314.1	364.06	421.8
Bay City	296.8	310.2	336.5	354.8	372.7	414.9
Belton	270.1	276.8	326.7	356.3	394.9	443.7
Bertnam	301.9	324.5	344.7	380.6	381.2	428.9
Briggs	341.02	344.6	355.6	385.4	390.5	447.3
Burleson	240.8	251.9	287.8	405.7	410.8	468.3
Clovis	219.6	240.1	265.6	284.09	323.6	374.1
Coryell	264.5	286.9	321.8	316.05	350.1	420.5
Cranfills	268.2	288.5	320.9	342.02	358.9	431.3
Cherroke	241.1	264.1	300.8	311.1	351.3	401.3
Cresson	270.1	282.1	307.2	339.8	365.3	408.2
Eastland	264.8	299.1	339.2	378.2	392.07	439.8
Evant	282.9	317.6	350.5	358.6	401.4	450.06
Santa Anna	351.6	358.4	374.8	369.8	371.9	419.5
Flat	246.03	325.5	350.8	393.2	392.9	425.8
Galveston	209.4	266.8	315.09	402.7	411.02	384.09

5.5 Uncertainty in PMP Values Due to Frequency Factor

To quantify the uncertainty that can be introduced in the PMP estimates by using Hershfield's curve, basin-specific PMP values were also calculated using Hershfield's enveloping curve. Figure 15 compares the PMP values based on both methods and shows that PMP from the Hershfield enveloping curve was higher than the basin-specific curve. For 1-hour duration the PMP values were 16.8% higher using Hershfield's curve rather than basin-specific, 17.2% for 2 hour duration, 17.7% for 3 hour duration, 18.5% for 6-hour duration, 20.9% for 12 hour duration and 23.4% for 24-hour duration.



Series 1 – Hershfield's PMP, Series 2- Own PMP

Figure 15 Comparison of Hershfield's PMP estimates against PMP estimates for Brazos River basin based on 1-hour duration (mm)

Plots between PMP values and mean of extreme values, PMP values and highest observed precipitation, and mean and standard deviation were also made. There was an

increasing correlation between mean and standard deviation, highest observed precipitation and PMP but not that significant. However, there was no significant correlation between the mean and PMP for different stations. It may be because the frequency factor comes in the multiplication with standard deviation which has a more effect on the values of PMP. Plots were also made for different durations for Eastland station (Figure 16 to 18), showing increasing correlation between PMP values and the mean of extreme values, PMP values and the highest observed precipitation, and the mean and standard deviation. The increasing correlation was based on the regression equations:

$$y = 1.6077x + 98.593 \quad (11)$$

where, y is the PMP values and x is highest observed precipitation.

$$y = 1.3647x + 3.4566 \quad (12)$$

where, y is the standard deviation and x is mean of annual maximum series.

$$y = 3.8075x + 114.86 \quad (13)$$

where, y is the standard deviation and x is mean of annual maximum series.

However, it may be noted that a highest observed precipitation for one duration can be the same for another duration.

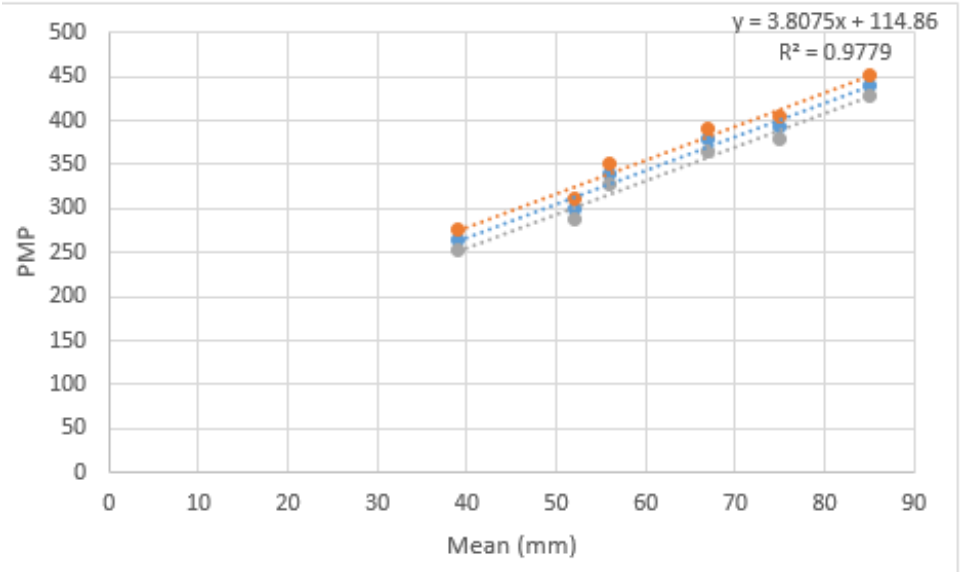


Figure 16 Plots between PMP values and mean of extreme precipitation

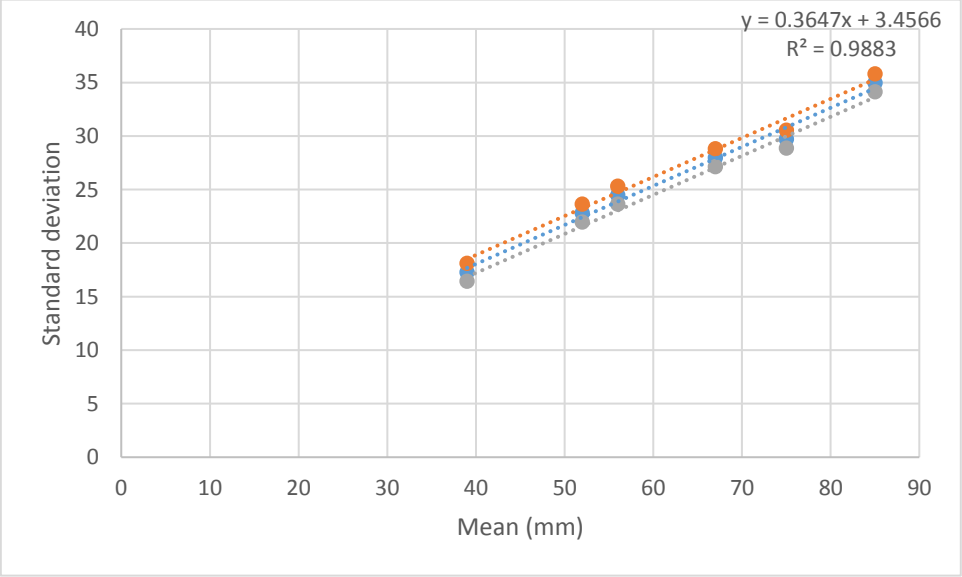


Figure 17 Plot between mean and the standard deviation of extreme precipitation

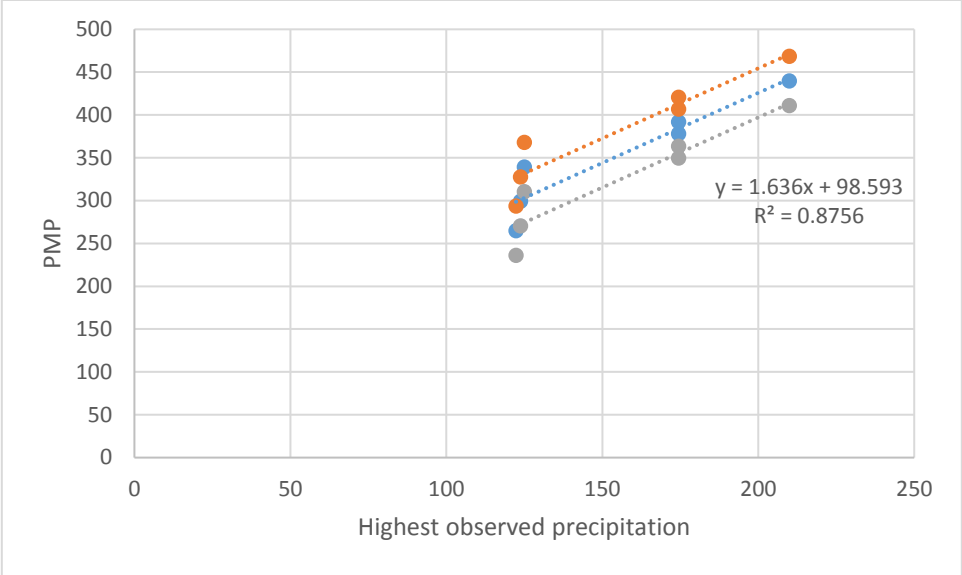


Figure 18 Plot between PMP values and the highest observed precipitation

6. FREQUENCY ANALYSIS OF EXTREME PRECIPITATION

6.1 Frequency Analysis

6.1.1 Probability Distributions

For frequency analysis of extreme precipitation for all durations and stations the same 39 stations were used as for calculating the PMP values along with 24 probability distributions that were: Generalized Extreme Value distribution, Burr XII distribution, Dagum, Log-logistics (3 parameter), Pearson 5 (3 parameter), Generalized Gamma (4 parameter), Pearson 6 (4 parameter), Log-normal (3 parameter), Generalized Gamma (3 parameter), Burr (4 parameter), Fretchet (3 parameter), Pearson 6 (2 parameter), Generalized Beat of the second order (4 parameter), Gumbel max, Log-Pearson 3, Log Gamma, Johnson SB, Inverse Gaussian (3 parameter), Dagum (4 parameter), Inverse Gaussian (2 parameter), Log-logistics (2 parameter), Frechet (2 parameter), Pearson 5 (2 parameter), and Log-normal (2 parameter). Table 3 shows probability distributions with Probability density function and Cumulative density function of distributions.

6.1.2 Goodness of Fit (GOF) Tests

Three goodness of fit tests (GOF), including Kolmogorov-Smirnov test, Anderson-Darling test, Chi-square test, were employed to check whether the hypothesized distribution function fitted the sample data (Chakravarti *et al.,* 1967*)*. The hypothesis of the GOF tests was:

H_0 = the precipitation data followed the specific distribution

H_1 = the precipitation data did not follow the specific distribution.

Kolmogorov-Smirnov (K-S) Test

The K-S test calculates the maximum difference D between the hypothesized distribution function and the empirical distribution. Let

$Z_{(i)} = F(x_i, \theta)$ (where x_i represents the order data and θ is the parameter sets) and $F_n(x_i)$ =empirical cumulative distribution function. Then,

$$D^{\pm} = \max_i \left(\frac{i}{n - Z_i}\right) \tag{14}$$

$$D^{\mp} = \max_i \left(\frac{-(i-1)}{n}\right) \tag{15}$$

where $D = \max(D^{\pm}, D^{\mp})$

For small samples, the K-S test is preferable to the chi-square test. This test is used to decide if a sample comes from a hypothesized continuous distribution.

Anderson-Darling (A-D) Test

The A-D test calculates the weighted square difference between the hypothesized distribution $Z_{(i)} = F(x_i, \theta)$ and empirical distribution $F_n(x_i)$. The weight function is described by $\{F(x_i, \theta)[1\text{-}F(x_i, \theta)]\}^{-1}$. The test static ($A$) can be defined as:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\{\ln F(x_i) + \ln[1 - F(x_{n-i+1})] \tag{16}$$

This test gives more weight to the tails than the K-S or chi-square test (Stephens, 1977).

Chi-Square (C-S) Test

The C-S test is based on the assumption that the number of observations is large enough so that the C-S distribution gives a good approximation to the distribution of the test static. It is defined as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{17}$$

where O_i is the observed frequency, and E_i is the expected frequency (calculated by $F(x_2)$ - $F(x_1)$). F is the CDF of the probability distribution that is being used.

The observed number of observations (k) in interval "I" is computed from the equation:

$$k = 1 + \log_2 n \tag{18}$$

where n is the sample size.

This test is used for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution (Sharma and Singh, 2010).

These tests were performed at the significant level ($\alpha = 0.05$) for choosing the best fit probability distribution (Sharma and Singh, 2010). Q-Q plot and Root Mean Square Error (RMSE) were also used to find the best fit probability distribution.

Q-Q Plot

It is a quantile-quantile plot of the input (observed) data values and theoretical (fitted) distribution quantiles against each other. It is based on the estimates of the quantiles. The pattern of points in the plot is used to compare the two distributions. In this the observed data x_i are ranked in ascending order ($x_{1:n}$ to $x_{n:m}$). A plotting position of the non-

exceedance probability ($G(x_{i:n})$) is computed for each $x_{i:n}$. There are many plotting positions in the literature but Cunnane (1978) plotting position formula yields approximately unbiased quantiles for a wide range of distributions:

$$G_{i:n} = \frac{(i-0.4)}{(n+0.2)} \tag{19}$$

The set of points from both the observed values and fitted quantiles is plotted on a normal graph with a 1:1 straight line extending from the origin. Theoretically, all points should fall on the 1:1 line if the assumed CDF is the true distribution (Tao et al., 2002).

Root Mean Square Error (RMSE)

It represents the sample standard deviation between the observed data and estimated quantiles. The RMSE serves to aggregate the magnitudes of the errors in the predictions for various times into a single measure of predictive power. It is given by

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{20}$$

where x_i is the observed data, y_i is the estimated quantile based on empirically derived CDF.

All the distributions were fitted to extreme precipitation data and parameters of the distributions were estimated by the maximum likelihood estimation. The probability density function (PDFs) were determined and plotted. Matlab and R-statistics were employed for fitting the probability distributions.

6.2 Frequency Analysis of Extreme Precipitation

To find the best fit probability distribution for each station and different durations, a three step process similar to Olofintoye et al. (2009) was used. It may be noted that our focus is on the right tail of the distribution where extreme precipitation occurs.

6.2.1 Step 1: Initial Processing

All 24 common statistical distributions were used in this step. For each station and duration the test statistic values of Kolmogorov Smirnov, Anderson Darling, and chi-square were calculated for every distribution. For each of the three tests the distributions were ranked according to the lowest test statistic value. The distribution having the 1st rank was assigned a score of 24, 2nd rank distribution a score of 23 and so on. The total scores from the three tests of each distribution were added to see which distribution had the highest score, the second highest, and so on. Then, the distributions having the 1st rank by each goodness of fit test were further analyzed. With that 2 or 3 distributions that had the highest total scores were also further analyzed. If the 1st rank distribution from a particular test also had the highest total score, then other distributions which had lower scores than the highest score distribution were considered. In any case, at least 5 to 6 distributions were considered for further analysis. The probability density function graph for the selected distribution was also compared with other remaining distributions to see that our results were legitimate. In most of the cases 3 parameter distributions were in the top distributions for different stations and durations.

6.2.2 Step 2: Using RMSE and Q-Q Plots

After selecting 5 to 6 best distributions for each station and duration, the Q-Q plots were considered and the Root Mean Square Error (RMSE) values for the selected distributions were determined. The stations were ranked according to the least RMSE value and best Q-Q plot (Here best means Q-Q plot will be linear or specified theoretical distribution is the correct model.). The PDFs of the selected 5 or 6 distributions were compared to see if our results were consistent with the PDF graph or not. We won't expect much difference between the selected distributions in fitting the data or visualizing the PDF graphs. We also tested the legitimacy of the two tests. If a particular distribution was clearly fitting the extreme precipitation data the best as seen from the PDF, then quantiles of the selected best distribution were compared with other distributions. If still the selected best fit distribution performed well, then it was selected as the best fit distribution. If not, then further analysis was done in which an average values of all test static were used to see which is the overall best distribution.

6.2.3 Step 3: Finding Best-fit Probability Distribution

In the last step, for those stations and durations for which there was not too much difference in the PDF graphs of selected distributions or there were contradicting results by observing the quantiles of the distributions with the observed values against the MSE and Q-Q plot results, the ranking system was used again. The top 5 or 6 distributions from step 1 were taken. The distributions were ranked according to the test statistic value from K-S, A-D, C-S, RMSE tests and visually seeing Q-Q plots. A score of 5 or 6 was assigned to the best distribution for a particular test and so on. The distribution having the highest

combined score from the 5 tests was regarded as the best distribution. In some cases more than 5 or 6 distributions were also taken depending upon the closeness of the test static values. After the best distribution was selected, it was analyzed which distribution fitted most of the stations and different durations overall. Other important things were also analyzed like is there a particular trend in the best fit distributions? What is the meaning of best distribution based on different climatic characteristics in the upstream, midstream and, downstream parts of the Brazos River basin?

Table 3 shows the overall best distributions for each station and 1-, 6-, and 24-hour durations, based on different GOF tests. For other durations Table C.2 (Appendix) shows the overall best fit distributions for other durations. Table C.3 to C.9 (Appendix) shows percentage coverage for different durations. The Anderson-darling GOF test performed better than did the other tests. It is because it focusses more on the tail of the distribution than the K-S test. The K-S test is distribution free in a sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. The log-logistic (3 parameter) distribution performed good in the right tail for higher quantiles for 1-, 2-, 3-, and 6-hour durations. But overall it did not perform as well as Burr XII or GEV for 2-, 3-, and 6 hour durations. For 12 and 24-hour durations of extreme precipitation, the generalized gamma (4 parameter) and Johnson SB performed better in the right tail.

Table 3 Overall best-fit distribution for different stations and durations

Station	Duration		
	1-hour	6-hour	24-hour
Albine	Log-logistics 3	GEV	Burr
Bay City	Log-logistics 3	Pearson 6 4p	Burr
Belton	Johnson SB	Log-Pearson 3	Johnson SB
Bertnam	GEV	Inverse-Gaussian 3	Inverse-Gaussian 3
Briggs	Log-logistics 3	Log-logistics 3	Log-logistics 3
Burleson	Burr	GEV	Log-Pearson 3
Clovis	GEV	GEV	Log-Pearson 3
Coryell	GEV	Burr	Burr
Cranfills	GEV	Burr	Burr
Cherroke	Burr	Log-logistics 3	Burr
Cresson	Burr	Log-logistics 3	Burr
Eastland	Log-logistics 3	Johnson SB	Johnson SB
Evant	Burr	Burr	Burr
Santa Anna	Log-logistics 3	Burr	Burr
Flat	Inverse-Gaussian 3	Burr	Inverse-Gaussian
Galveston	Burr 4p	Johnson SB	Inverse-Gaussian
Gorman	Burr	Johnson SB	Johnson SB
Groesbeck	Log-logistics 3	Burr	Burr
Houston Addict	Log-logistics 3	GEV	GEV
Indian Gap	Johnson SB	GEV	Log-logistics 3
Iredell	Burr	GEV	GEV
Jayton	Log-logistics 3	GEV	Burr 3

Table 3 Continued

Station	Duration		
	1-hour	6-hour	24-hour
Houston alife	GEV	Log-Pearson 3	Burr
Jewett	GEV	Dagum 4	Burr
Kopperl	Beta	GEV	Burr
Lexington	Burr	Gen Gamma 4p	Pearson 5 3p
Loraine	Log-logistics 3	GEV	Gumbel Max
Lubbock	Log-Pearson 3	Johnson SB	Log-Pearson 3
Moline	Burr	GEV	GEV
Pep 3	Dagum	Burr	Inverse-Gaussian
Richmond	Log-logistics 3	Log-Pearson 3	Johnson SB
Spicewood	Inverse-Gaussian 3	Burr	Burr
Stamphord	Log-logistics 3	Burr	GEV
Stephenville	Burr	Log-logistics 3	Log-logistics 3
Still house	Log-logistics 3	Burr	Burr
Thompson	Burr	GEV	Johnson SB
Waco	Log-logistics 3	Log-Pearson 3	Burr
Washington	GEV	GEV	GEV
Wheelock	Log-logistics 3	Johnson SB	Burr

*GEV = Generalized Extreme Value

6.3 Factors Affecting Frequency Distributions

Next, the effect of duration and distance from the Gulf on the histogram and best fit distribution was analyzed. It was observed that there was a general tendency for higher skewness for shorter durations of precipitation than for longer durations, as shown in Figure 19 to 23 for station at Evant TX for different durations. It is because for short durations such as 1-hour, a large amount of precipitation may occur within a short time in certain cases exhibiting large skewness, while for long durations, such as 24-hour, precipitation is averaged and thus exhibits less skewness.

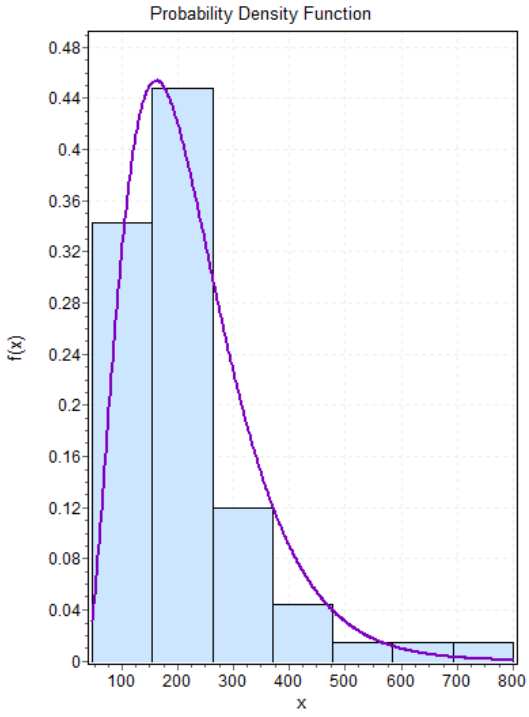


Figure 19 Histogram at Evant, TX for 2-hour duration

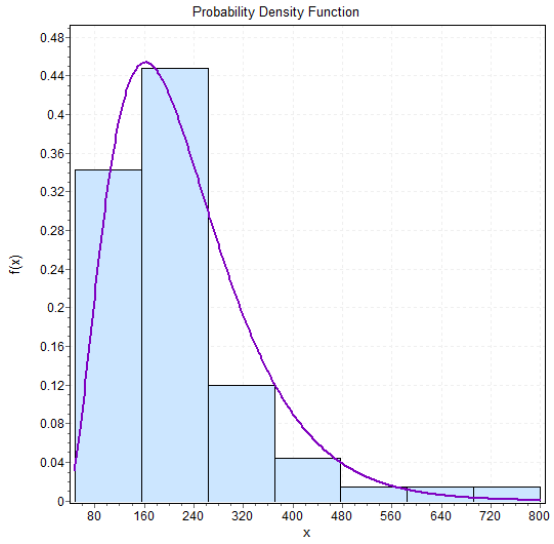


Figure 20 Histogram at Evant, TX for 3-hour duration

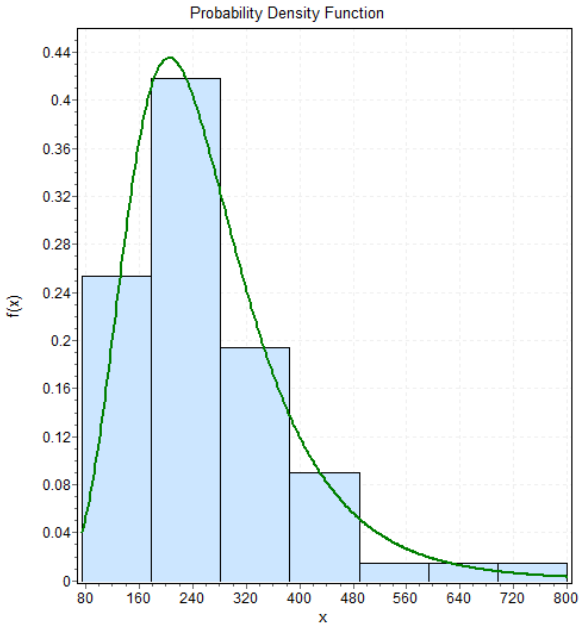


Figure 21 Histogram at Evant, TX for 6-hour duration

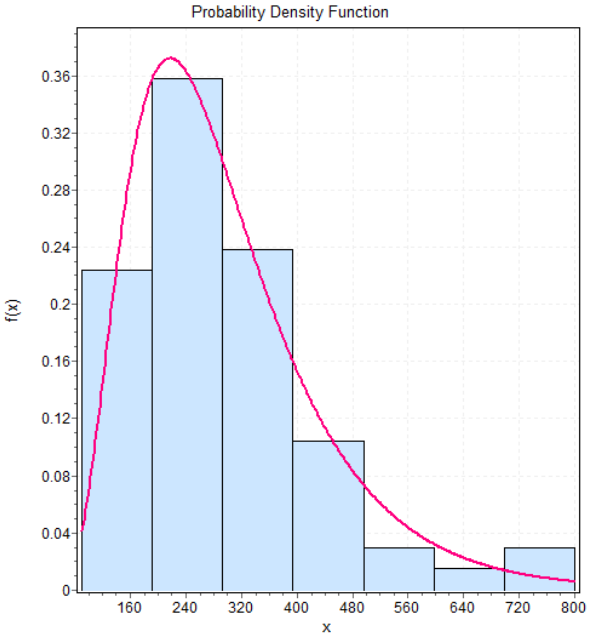


Figure 22 Histogram at Evant, TX for 12-hour duration

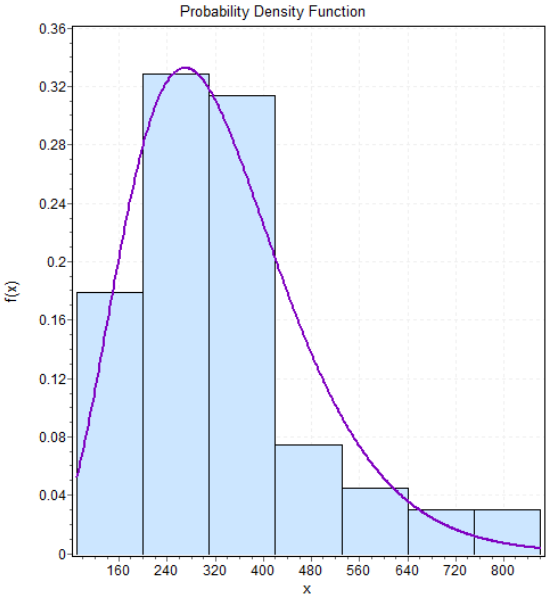


Figure 23 Histogram at Evant, TX for 24-hour duration

The Burr XII distribution performed better for less skewed distributions and log-logistic (3 parameter) performed better for more skewed distributions. Within Brazos River basin there exist different climate producing mechanisms for different areas. For example, in the eastern part of Texas or near the Gulf of Mexico there is fairly uniform seasonal precipitation, with slight maxima occurring in the summer season, because the influence of the Gulf of Mexico is dominant (National Fibers Information Centre, 1987). Hence, the effect of the distance from Gulf was analyzed. There was no systematic pattern but still it was observed that for stations close to the Gulf of Mexico, the histogram was smooth but had more variation. As the distance from the gulf increased the histogram began to become sharp with less variation. Figure 24 to 29 shows histograms for stations at Thompson and Lubbock for different durations. Thompson lies close to the gulf, whereas Lubbock lies in the north-western part of Texas. The reason for this pattern may be due to the moderating influence of the Gulf of Mexico. As we go farther from the gulf, in the northwest direction we come close to regions of High Plain division in which maximum precipitation comes from thunderstorms during the summer season. However, there was no preferable distribution which performed best near the Gulf or far away from it. However, Burr XII and GEV performed better for smooth histograms. Overall the Burr XII covered 30 to 40% of the stations for different durations. For other stations also, it was in most of the cases one of the top three best distributions.

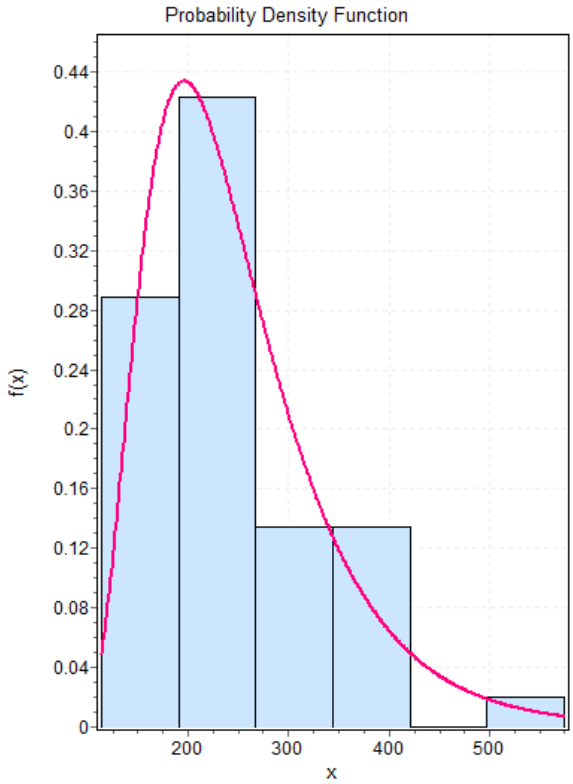


Figure 24 Histogram at Thompson, TX, for 2 hour duration

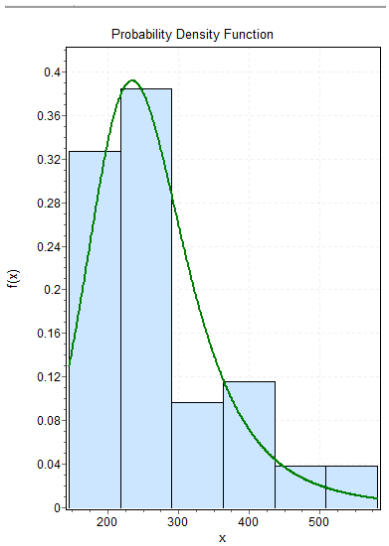


Figure 25 Histogram at Thompson, TX, for 3 hour duration

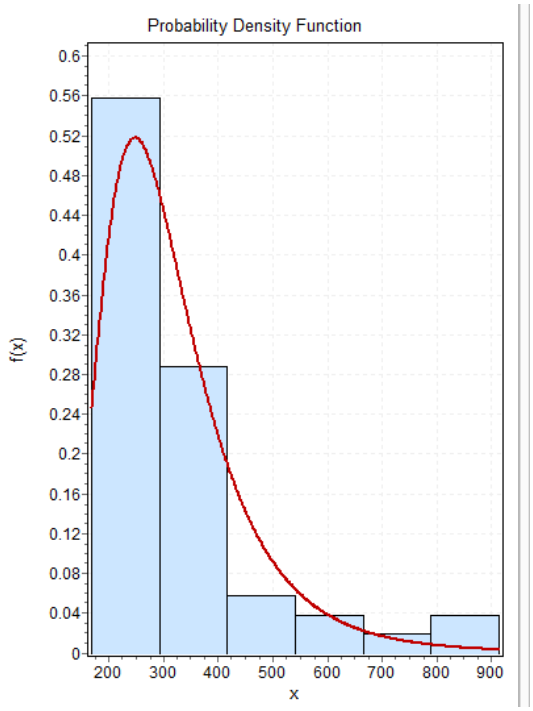


Figure 26 Histogram at Thompson, TX, for 6 hour duration

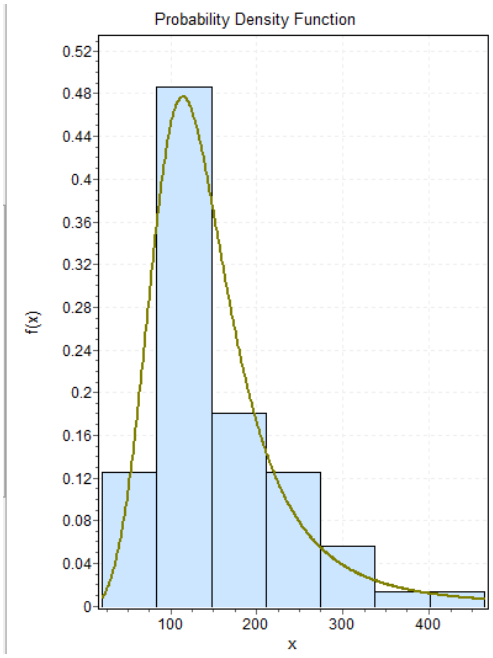


Figure 27 Histogram at Lubbock, TX, for 2 hour duration

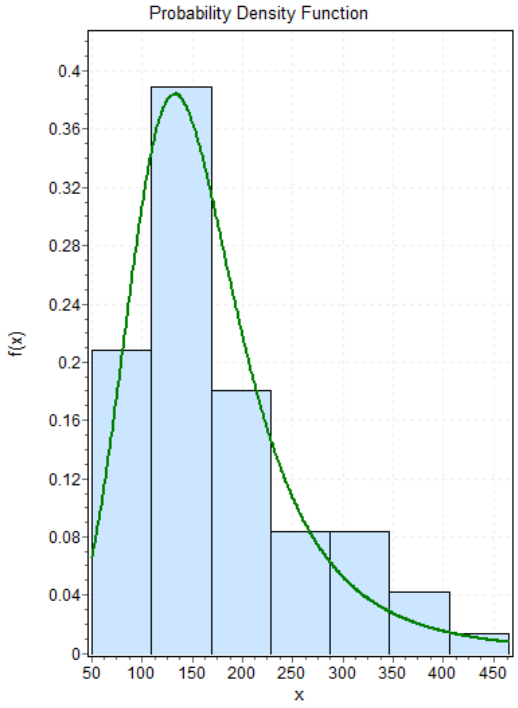


Figure 28 Histogram at Lubbock, TX, for 3 hour duration

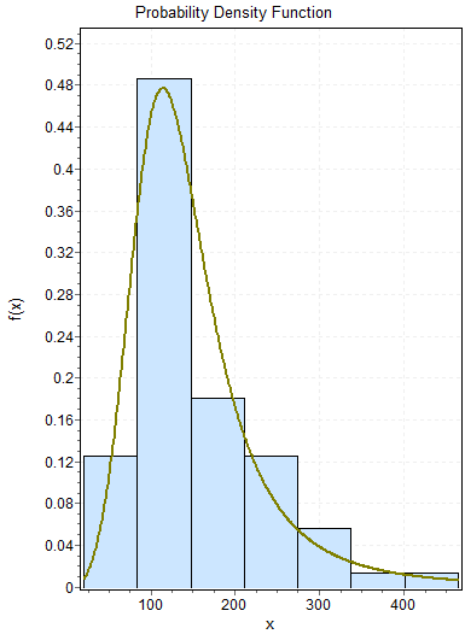


Figure 29 Histogram at Lubbock, TX, for 6 hour duration

6.4 Hazard Rate

Hazard rate is defined as the instantaneous rate of failure for the survivors to time *t* during the next instant of time. In flood frequency analysis it can be defined as the probability of extreme flooding in an infinitesimally small time period between *t* and $t+dt$ given that no flooding has occurred till time *t* or the rate of an event occurrence per unit of time. In choosing the best fit probability distribution it has more physical significance rather than just determining the best fit distribution based on different GOF tests as it is not a density or probability but a measure of risk. It was therefore decided to calculate the hazard rate for different distributions and durations to determine which distributions performed better. The hazard rate (*h(t)*) was calculated as:

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{21}$$

where $f(t)$ is the probability density function or the probability that the value will fall in a specific interval and $F(t)$ is the cumulative distribution function. $1 - F(t)$ can be defined as a survival function or the probability that something will survive up to a certain time *t*.

Based on equation (20), the hazard rate was calculated for different distributions and durations. For distributions with 2 parameters there was a general trend of increasing hazard rate. Figure 30 shows hazard rate for 5 common distributions for 2 hour duration at station Flat.

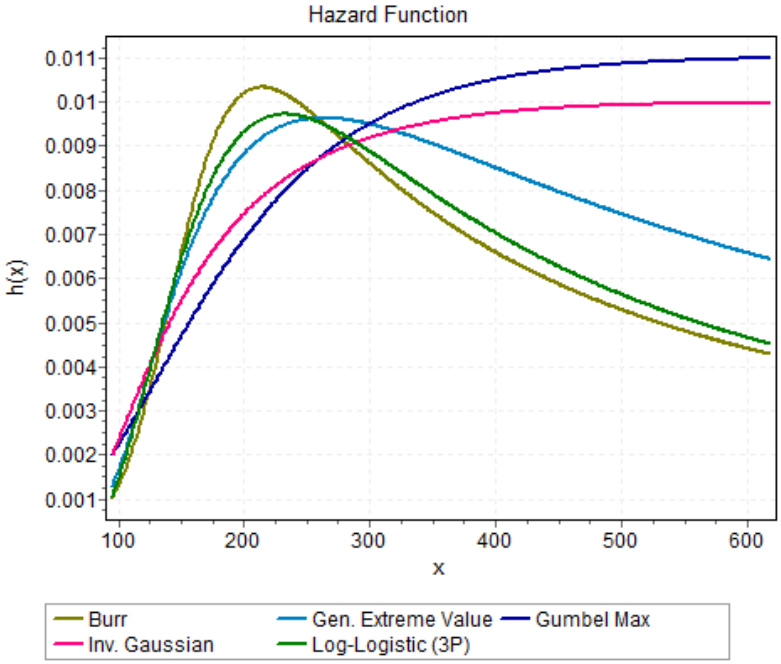
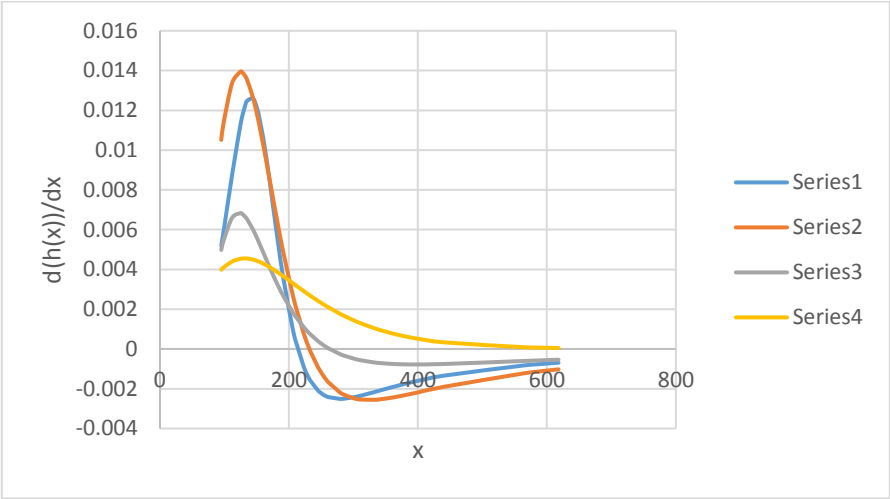


Figure 30 Hazard rate for different distributions for 2-hour duration at station Flat

It was observed that Gumbel Max and Inverse Gaussian (2 parameter) had an increasing hazard rate which then tended to become constant. While the Burr XII, log-logistics (3 parameter), and GEV had an increasing hazard rate which then became constant and started decreasing. It makes sense physically, as station at Flat lies in Post Oak Belt in Texas where Prairie grasslands are scattered throughout the area, hence, the damage rate due to flooding will increase to a certain extent and then will become constant and start decreasing. To know how the hazard rate was changing it was differentiated with respect to extreme precipitation *x*. Using the probability density function (PDF) and cumulative distribution function (CDF) of the distributions $d(h(x))/dx$ was calculated.

Figure 31 shows the rate of change of hazard rate for the same 2 hour duration at station Flat.



Series 1 – Burr, Series 2- Log logistics (3P), Series 3 – GEV, Series 4 – Gumbel max.

Figure 31 Rate of change of hazard rate for different distributions for 2-hour duration at station Flat

The rate of change of hazard rate generally followed the same trend for all four distributions, except that the rate of change was non-negative for Gumbel distribution. It shows how quickly hazard rate is increasing or decreasing with respect to extreme precipitation. To further study these characteristics it was also determined how hazard rate and rate of change of hazard rate varied according to different climatic regions in Brazos River basin. There was not any significant pattern but it was observed that as we moved near to the Gulf of Mexico there was an increasing hazard rate for most of the distributions. For stations away from the gulf, the hazard rate followed the trend of increasing first,

becoming constant and then decreasing. The increasing hazard rate is due to the fact that near the gulf or Costal Plains region damage is mostly urban but as we move away from the gulf close to High Plains which is a major farming area damage is mainly agricultural which saturates after a certain point of time. In Coastal Plains Regions where two thirds of the population lives and urban development is more the hazard rate was more as compared to Great Plains where urban development is less. Figure 32 to 34 shows the hazard rate for different distributions at three stations Lubbock (Upstream), Coryell (Middle) and Houston Addicts (Downstream) for 24-hour duration of precipitation. Figure C.1 to C.12 (Appendix) shows the hazard rate for other durations. For higher durations of precipitation the GEV distribution also began to reach a constant hazard rate. The rate of change of hazard rate was also calculated for the same distributions and durations. The log-logistic distribution consistently gave higher rate of change and Gumbel distribution gave non-negative rate of change. Figure 35 to 37 shows the rate of change of hazard rate for the same distributions and stations as above. Figure C.13 to C.22 shows the rate of change of hazard rate for other durations for the same stations.

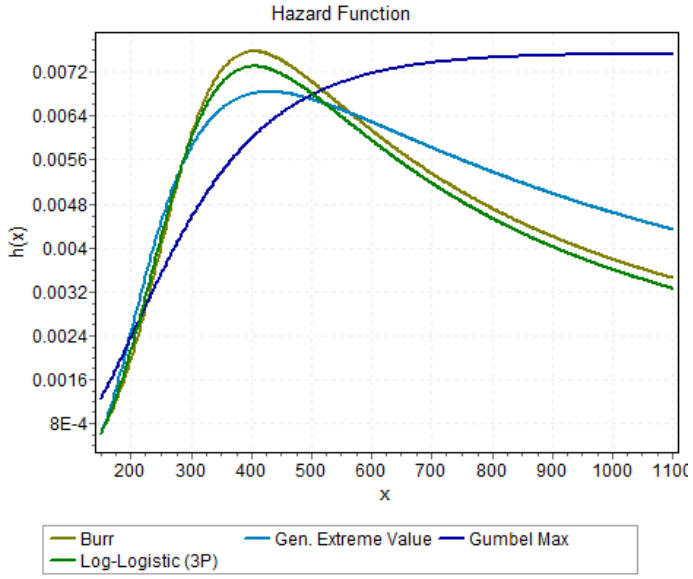


Figure 32 Hazard rate for different distributions for 24 hour duration at station Coryell

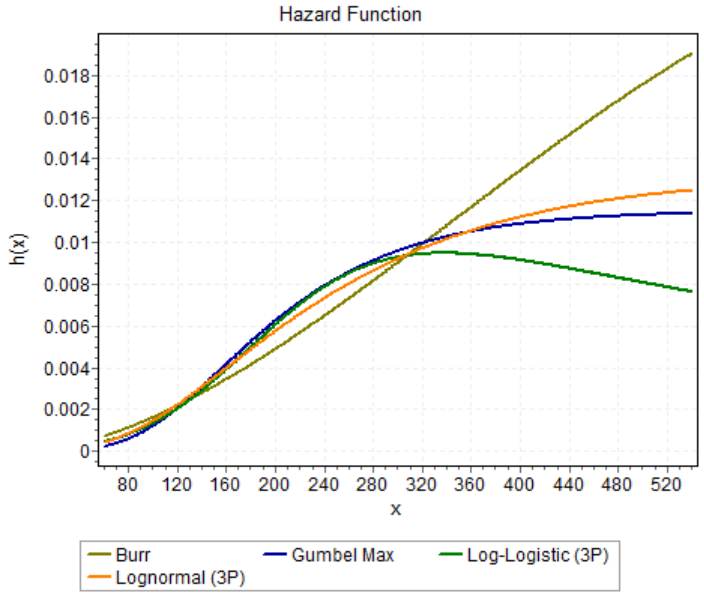


Figure 33 Hazard rate for different distributions for 24-hour duration at station Houston Alife

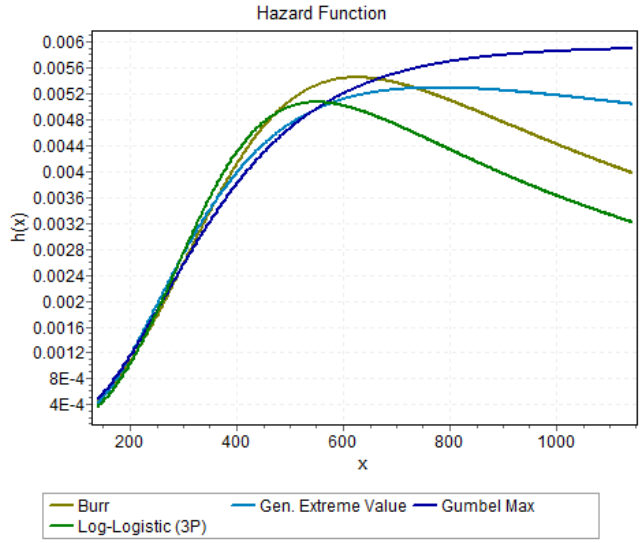
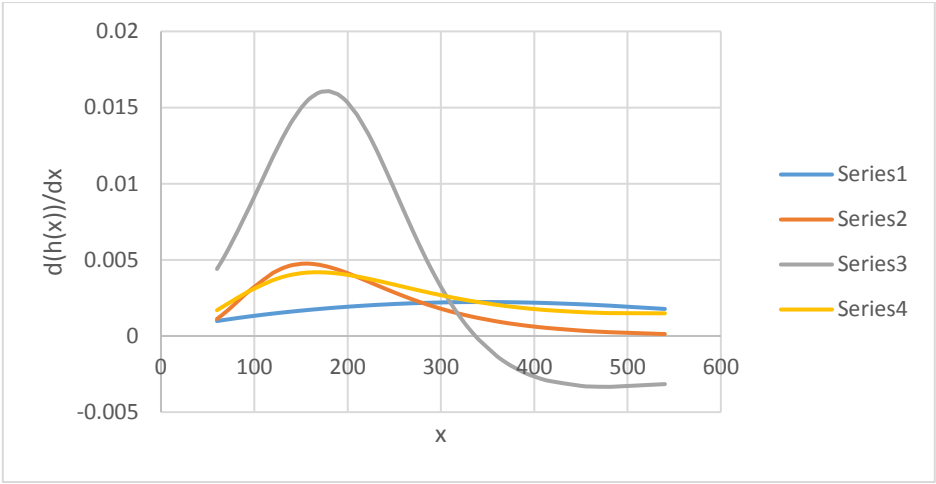
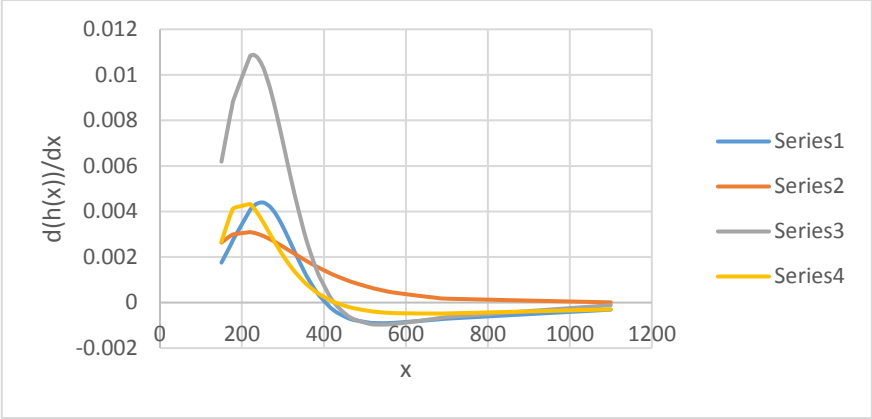


Figure 34 Hazard rate for different distributions for 24-hour duration at station Lubbock



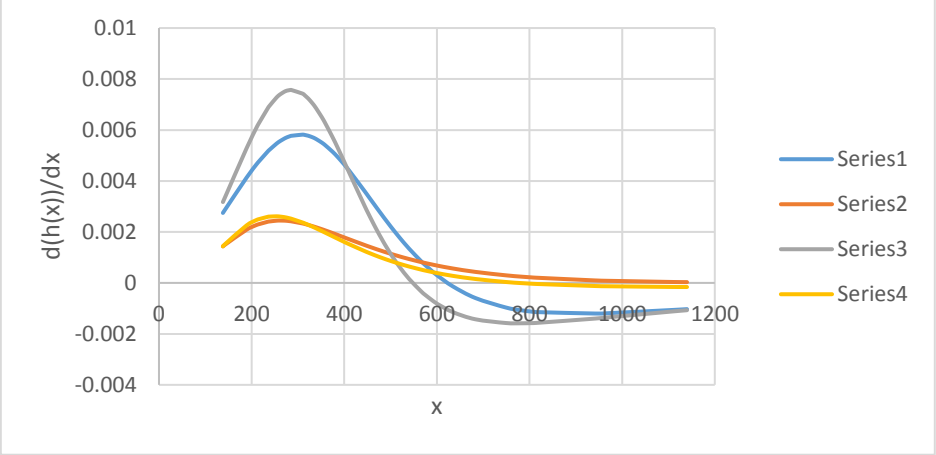
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – - Log logistics (3P), Series 4 – GEV

Figure 35 Rate of change of hazard rate for different distributions for 24-hour duration at station Houston Alife



Series 1 – Burr, Series 2 - Gumbel max, Series 3 – - Log logistics (3P), Series 4 – GEV.

Figure 36 Rate of change of hazard rate for different distributions for 24-hour duration at station Coryell



Series 1 – Burr, Series 2 - Gumbel max, Series 3 – - Log logistics (3P), Series 4 – GEV.

Figure 37 Rate of change of hazard rate for different distributions for 24 hour duration at station Lubbock

The pattern varied across the study area but Burr XII and GEV distribution performed better except in regions close to High Plain division in which the maximum precipitation comes from thunderstorms during the summer season.

6.5 Return Period of Estimated PMP Values

For quantifying uncertainty, return periods of the PMP values were determined for each duration (Table C.1, Appendix). For our study we used the PMP values derived from basin-specific enveloping curve of k_m as it was made only by using the data for the Brazos River basin and is more accurate. The return period was less than expected. For most of the Brazos River basin the return period of the PMP values was in the range of 1000 to 3000 years which was less than the range of 10^3 to 10^6 years reported in HMR 51. It shows the amount of risk associated with the PMP values. The difference between the two sets of values points to the uncertainty associated with the PMP values. To evaluate the uncertainty in the return period due to the choice of distribution, return periods for stations and durations were also calculated from the 4th best distribution. Table 4 shows the return period from the best and the 4th best distribution for 24 hour duration. Table C.10 and C.11 shows the return period from the best and the 4th best distribution for other durations. On an average basis the return period from the 4th best distribution was 55.1% lower than from the best distribution. Figure 38 shows the difference between the return periods of the 24-hour PMP values for selected stations from the best and the 4th best distribution in dimensionless terms as:

$$\frac{T_{best} - T_{4thbest}}{T_{best}} \qquad (22)$$

where T_{best} is the return period from the best distribution, and $T_{4thbest}$ is the return period from the 4th best distribution. As can be seen from the figure return periods were different, showing the importance of accurately determining the best-fit probability distribution. Figures 39 and 40 show the spatial distribution of the 1- hour PMP values and return period for those values calculated, based on the best fit probability distribution. The GIS spatial interpolation tool was employed for performing it. The spatial interpolation was done on the basis of inverse distance weighted interpolation.

Table 4 Return periods of PMP values from the best and the 4th best distribution for 24-hour duration

Return Period (years) Best Distribution	Return Period (years) from 4th Best Distribution
1111.1	7142.8
6579.8	3950.2
4347.8	1333.3
16666.6	3703.7
2222.2	2500
16666.6	25000
232552.7	20000
6136.4	6840.9
1886.7	16666.6
50796.6	1265.8

Table 4 Continued

Return Period (years) Best Distribution	Return Period (years) from 4th Best Distribution
33333.3	9090.9
12500	12500
1870.4	1149.4
1282	970.8

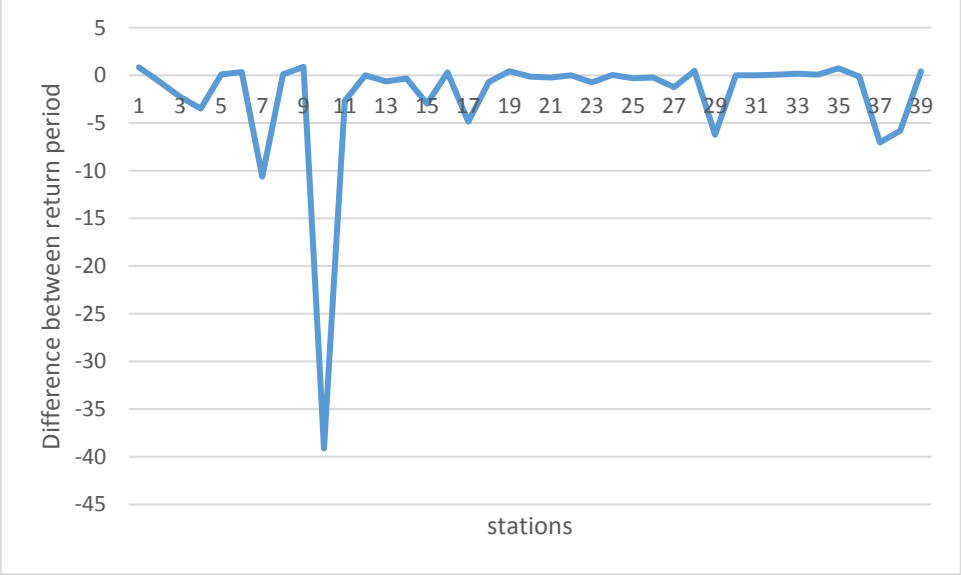


Figure 38 Difference between the return period of the 24-hour PMP values for selected stations from the best and the 4th best distributions in dimensionless terms

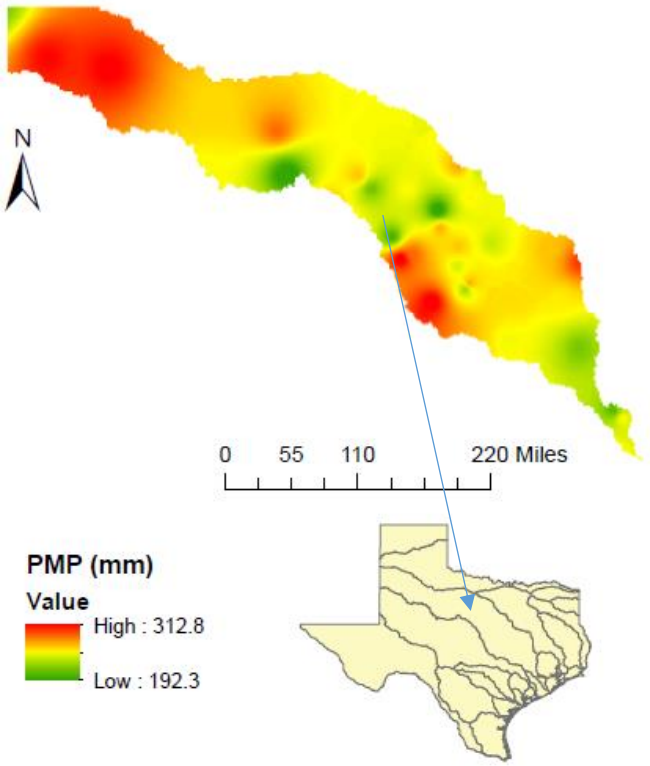


Figure 39 Spatial distribution of the PMP values in Brazos River Basin for 1-hour duration

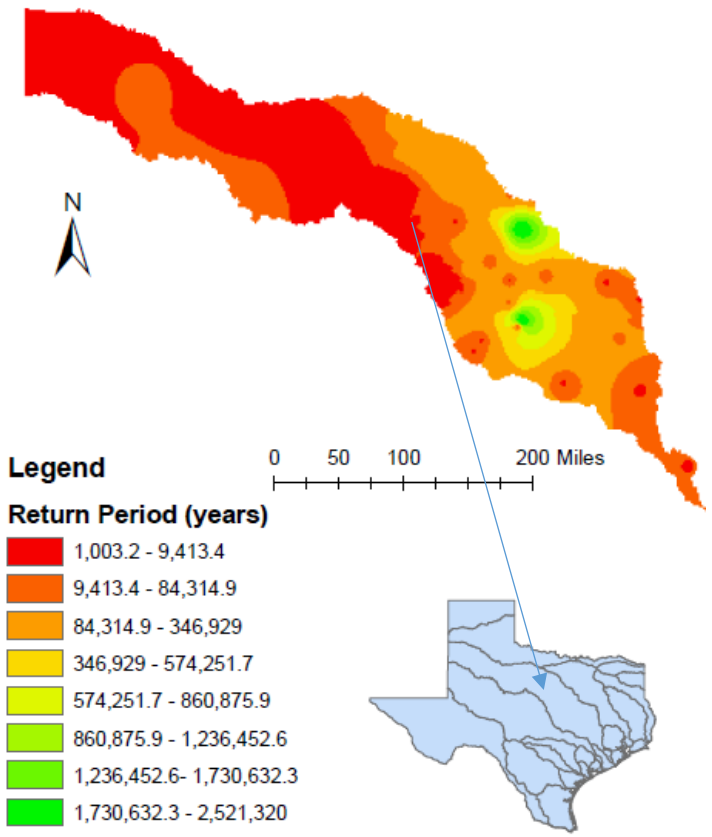


Figure 40 Spatial distribution of the Return period of 1-hour PMP values in Brazos River basin

The depth-duration-frequency curve was also constructed for PMP values. Log of 1, 2, 3, 6, 12, and 24 hour of precipitation and log of PMP values of different return period was taken. Figure 41 below shows the relation between PMP values and duration on log-log paper. It was observed that there was an increasing correlation between log of PMP values and log of duration for different return periods. The chosen return was the return period of different duration PMP values and for the same return period the depth of rainfall was calculated for different durations. The equation fitted to the regression line were:

$y = 0.0976x + 2.34$ (1 hour duration) (23)

$$y = 0.0909x + 2.3874 \quad \text{(2 hour duration)} \tag{24}$$

$$y = 0.0654x + 2.4942 \quad \text{(3 hour duration)} \tag{25}$$

$$y = 0.0513x + 2.5205 \quad \text{(6 hour duration)} \tag{26}$$

$$y = 0.0635x + 2.501 \quad \text{(12 hour duration)} \tag{27}$$

$$y = 0.0589x + 2.5385 \quad \text{(24 hour duration)} \tag{28}$$

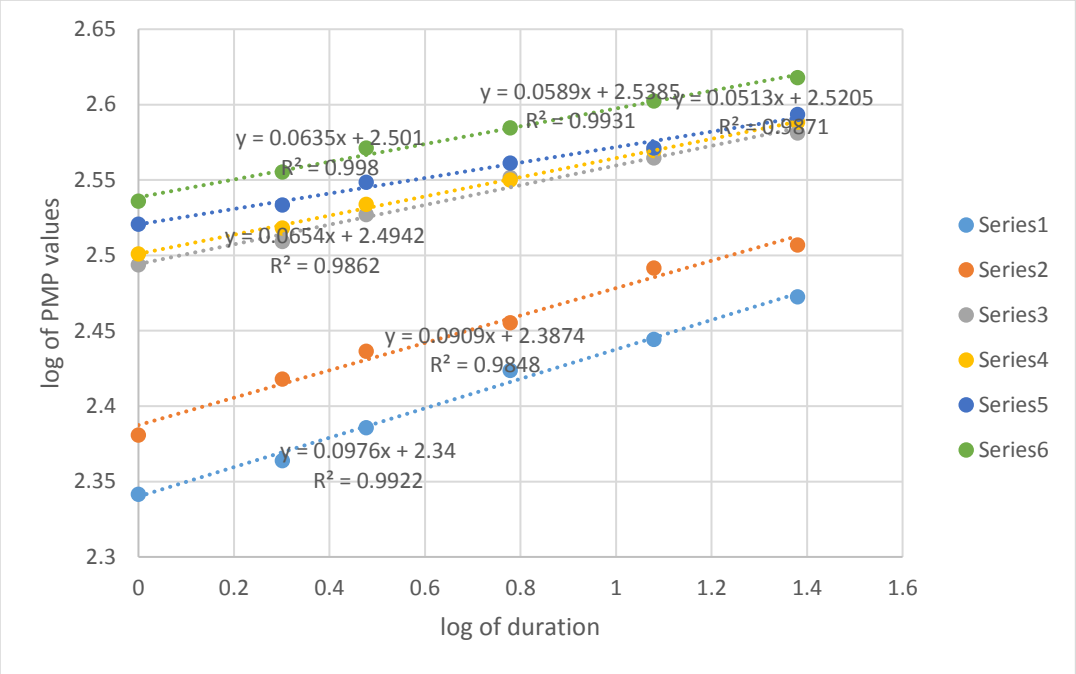


Figure 41 Depth-Duration-Frequency curve of PMP values

7. UNCERTAINTY ANALYSIS OF PMP ESTIMATES

7.1 Uncertainty Due to Mean and Standard Deviation

To quantify uncertainty due to sample mean and sample standard deviation the procedure similar to Salas et al. (2014) was followed. To quantify uncertainty one of the requirements was the calculation of Variance ($Var(P)$) of PMP estimate which can be calculated by (Mood et al, 1974):

$$Var(P) = Var(\overline{X_n}) + k_m{}^2 Var(S_n) + 2KCov(\overline{X_n}, S_n) \quad (29)$$

where $Var(\overline{X_n})$ is the variance of the mean, $Var(S_n)$ is the variance of standard deviation and $Cov(\overline{X_n}, S_n)$ is the covariance of mean and standard deviation. The requirement was to calculate $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$. Salas et al. (2014) used Gumbel distribution as the underlying distribution by assuming the location parameter value to 1 and calculating a correction factor for Gumbel to normal approximation for the calculation of $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$ using Monte Carlo analysis which is more generalized and can be used anywhere. However, for Brazos River basin using precipitation data the best fit distribution was Burr XII, hence, a different methodology was developed which would only be suitable for the study area.

The values of Burr XII distribution first shape parameter α, second shape parameter κ and scale parameter β were used along with Monte Carlo analysis. Matlab was used to simulate the results. The value of the first shape parameter α was in the range

of 3.5 to 6.5 and the value of the second shape parameter κ was in the range of 0.5 to 1.5 with a few exceptions like for 24-hour duration extreme precipitation at Groesbeck the value was 3.16, and for 1-hour of extreme precipitation at Jeweet the value was 3.33. Such values also shows that because of the unpexted high or low rainfall the value of the parameters can vary. But the value of scale parameter β varied readily from 150 to 400. It was around 500 in a few cases. The value of β tended to increase for higher durations of extreme precipitation. Since the values of shape parameters had narrower ranges, an average values of $\alpha = 4.8$ and $\kappa = 1$ were used by taking the average of all the values. 1000 samples of different sizes of n = 30, 50, 70, 100 and different scale parameter values of 150, 200, 300 and 400 were simulated and then the values of $\overline{X}_{(i)}$ and $S_n(i)$, i = 1,...,1000, were computed. $Var(S_n)$ and $Cov(\overline{X}_n, S_n)$ were estimated based on the pair of 1000 values , as shown in Table 5. The procedure was similar to Salas et al.(2014) method but we used different lengths of records and also simulated the values based on different values of scale parameter. It was observed that with increasing scale parameter $Var(S_n)$ and $Cov(\overline{X}_n, S_n)$ increased, but decreased with increasing record length (n). It can be expected because by increasing the scale parameter the spread increases, hence that leads to a large variance. Which leads to larger standard deviation in the results. The results are just an estimate but gives a good idea about the parameters of Burr distribution.

Table 5 Values of parameters for different record lengths and scale parameters

Record length (n)	β	$Var(S_n)$	$Cov(\overline{X_n}, \sigma_n)$
30	150	250	110
50	150	140	45
70	150	70	28
100	150	45	22
30	200	400	178
50	200	290	97
70	200	150	74
100	200	108	56
30	300	510	269
50	300	373	147
70	300	243	120
100	300	170	96
30	400	580	327
50	400	440	219
70	400	325	157
100	400	218	122

To check whether the values of α = 4.8 and κ = 1 would give reliable results, simulations were also run for shape parameter values α = 5.3 and κ = 0.8. These values were the average values for 24 hour duration extreme precipitation. The scale parameter value of 150 for the first simulation and 300 for the second simulation were used. It was because 150 was the least scale parameter value and 300 was the average value for 24-hour duration extreme precipitation series. The results for different record lengths of 30, 50, 70, and 100 were not very different from the one using the values of α = 4.8 and κ = 1, as shown in Table 6. Hence, it was decided to use the values of α = 4.8 and κ = 1 as the average values for calculating $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$. Then, a nomograph of varying

scale parameter β and varying record length was constructed, as shown in Figures 42 & 43. The equation fitted to different curves of different scale parameters for the estimation of variance were:

$y = 0.052x^2 - 9.6983x + 494.29$ for $\beta = 400$ (30)

$y = 0.052x^2 - 9.6983x + 494.29$ for $\beta = 300$ (31)

$y = 0.0533x^2 - 11.868x + 822.09$ for $\beta = 200$ (32)

$y = 0.0389x^2 - 10.239x + 852.86$ for $\beta = 150$ (33)

and for the estimation of covariance the equation were:

$y = 0.0312x^2 - 5.2541x + 236.98$ for $\beta = 400$ (34)

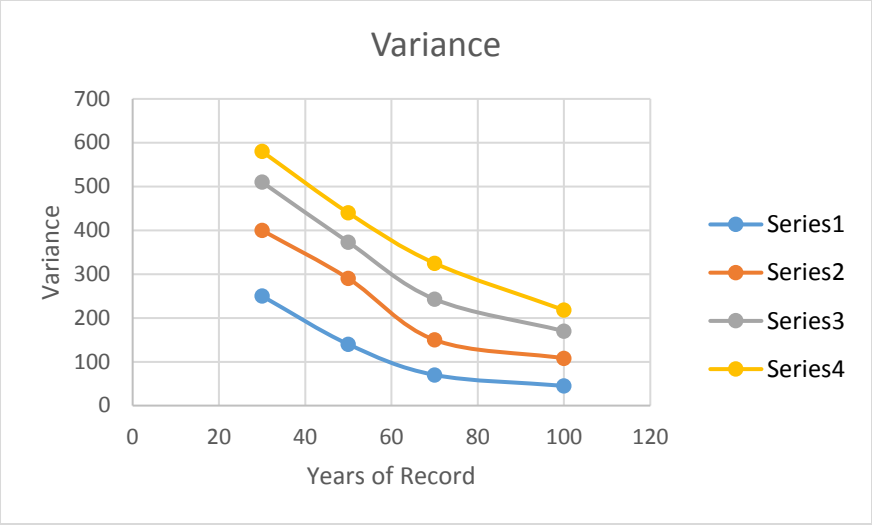
$y = 0.0348x^2 - 6.1915x + 329.06$ for $\beta = 300$ (35)

$y = 0.0527x^2 - 9.1902x + 491.38$ for $\beta = 200$ (36)

$y = 0.046x^2 - 8.8797x + 551$ for $\beta = 150$ (37)

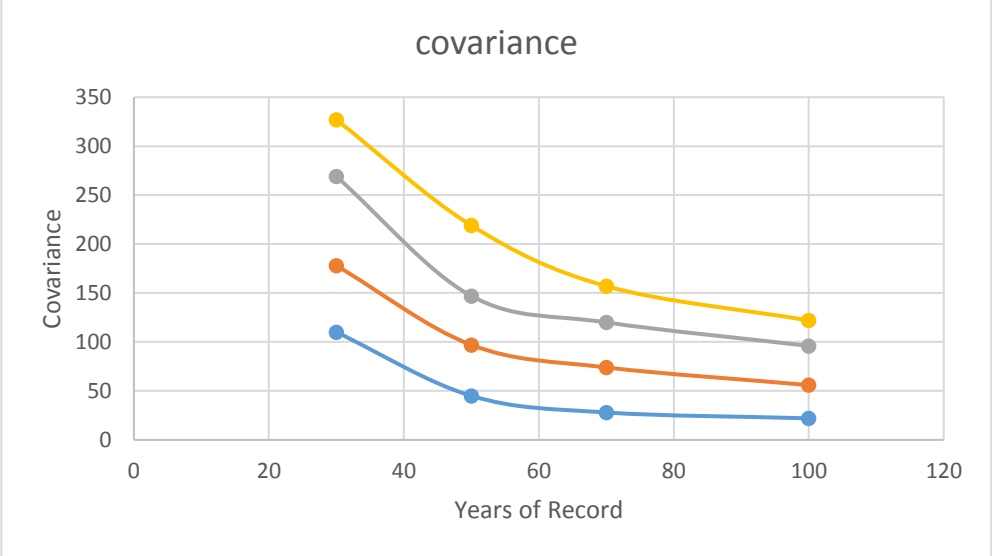
Table 6 Simulation results by changing record length and scale parameter

Record length	β	$Var(S_n)$	$Cov(\overline{X_n}, \sigma_n)$
30	150	263	121
50	150	168	51
70	150	87	32
90	150	56	25



Series 1: β = 150, Series 2: β = 200, Series 3: β = 300, Series 4: β = 400

Figure 42 Variance of standard deviation as a function of record length



Series 1: β = 150, Series 2: β = 200, Series 3: β = 300, Series 4: β = 400

Figure 43 Covariance of mean and standard deviation as a function of record length

From these Figures the values of $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$ were determined for all the stations with different durations having the Burr XII distribution as the best distribution. Then, $Var(P)$ was determined based on equation (28). The expected values of PMP estimates $E(P)$ were calculated by using the relationships shown below:

$$E(P) = E(\overline{X_n}) + k_m E(S_n) \tag{38}$$

where, $E(\overline{X_n})$ is the expected value of the sample mean and $E(S_n)$ is the expected value of the sample standard deviation. It can also be shown that:

$$E(\overline{X_n}) = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \tag{39}$$

and,

$$E(S_n) = \frac{\Psi(n/2)}{\Psi[(n-1)/2]\sqrt{(n-1)/2}}\sigma \quad \text{(Kendall and Stuart, 1963)} \tag{40}$$

where μ and σ are the population mean and standard deviation, $\Psi(a)$ is the incomplete gamma function with argument *a*, and *n* is the record length. Using equation (39) and (40) the expected value of PMP value were calculated by:

$$E(P) = \mu + \frac{\Psi(n/2)}{\Psi(n-1)/2\sqrt{(n-1)/2}}\sigma \tag{41}$$

where μ and σ are the population mean and standard deviation which were are replaced by the corresponding sample estimates after appropriate adjustments for outliers as needed. $\Psi(a)$ is the incomplete gamma function given as:

$$\int_0^\infty t^{a-1}e^{-t}dt \tag{42}$$

where *a* is the argument and *n* is the record

Following Salas et al. (2014) the design risk PMP values P_d were calculated by:

$$P_d = E(P) \pm c\sigma(P) \tag{43}$$

where $\sigma(P)$ is the standard deviation of PMP estimates and *c*>1 is the parameter. P_d is the value of uncertain PMP value *P* whose distribution is not known but only the estimates of its mean $E(P)$ and standard deviation $\sigma(P)$ are known. Table 7 compares 1-hour PMP values at Thompson, Texas, obtained using the statistical method and the method considering the Burr XII distribution for calculating $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$.

Table 7 Comparison of Hershfield PMP and Design Risk PMP values at Thompson, TX, for 1-hour duration (mm)

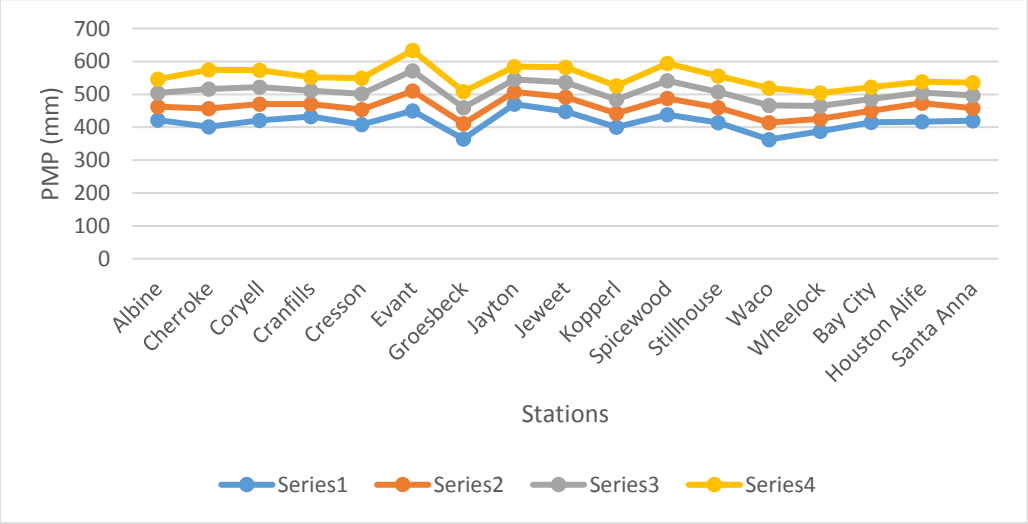
Hershfield PMP	Design Risk PMP using $P_d = E(P) + c\sigma(P)$		
	c= 1	*c*= 2	*c* = 3
253.45	251.88	280.4	309.06

The best-fit distribution at the station was the Burr XII distribution with scale parameter β = 165.43 and record length (*n*) = 52. From figures 23 & 24, the values of $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$ were obtained as 190 and 50. Using equations (40) and (28) $E(P)$

and $Var(P)$ were 223.29 mm and 3218.1 mm, respectively. $\sigma(P)$ was calculated by taking the square root of $Var(P)$. For different values of c the design risk PMP P_d were calculated. With c = 1 and +sign $P_d = E(P) + c\sigma(P)$ gave P_d = 251.88 mm. The value of c = 3 which was a more conservative estimate than using c = 1 gave P_d 309.06 mm. Using Hershfield's original method along with basin-specific enveloping curve the PMP estimate P was 253.45 mm after multiplying the PMP value with 1.13 for fixed observational time interval. The PMP values were also calculated using the Salas et al. (2014) method. It gave a PMP value of 241.2 mm and using normal distribution and it gave PMP value of 231.65 mm. The lower value of PMP using the Salas et al. method may be because of the use of Gumbel distribution which was found inappropriate for modelling the annual maximum precipitation series (Koutsoyiannis, 2004). Results showed major differences between PMP values using all the three methods in terms of value. It can be concluded that a bigger value of PMP must be selected, considering the associated exceedance probabilities (risk) of PMP. The analysis was performed for other durations and stations having the Burr XII distribution as the best distribution. Design risk values were obtained for different durations. Table 8 shows the 24 hour design risk values for different values of *c*. Figure 44 shows the comparison of original PMP values using Hershfield's method and method with takes into account the uncertainty due to mean and standard deviation with different values of *c*.

Table 8 24-hour design risk PMP values (mm) values up to two decimal places

Station	Original PMP	Design risk PMP		
		c = 1	*c* = 2	*c = 3*
Albine	421.81	462.52	504.54	546.57
Cherroke	401.33	457.07	515.71	574.34
Coryell	420.5	470.26	522.05	573.85
Cranfills	432.35	470.78	511.43	552.08
Cresson	408.29	454.25	501.61	548.98
Evant	450.06	510.11	571.94	633.76
Groesbeck	364.26	410.80	459.53	508.25
Jayton	470.2	507.18	545.68	584.19
Jeweet	447.98	491.33	536.74	582.15



Series 1 – Original PMP values, Series 2 – design risk PMP values using *c* = 1, Series 3 – design risk PMP values using *c* = 2, Series 4 – design risk PMP values using *c* = 4.

Figure 44 Design risk PMP values for different values of c for 24 hour duration

Chebyshev's inequality (Mood et al., 1974) was used to have a bound of the probability on the PMP estimates using equation (18):

$$P\{(E(P) - c\sigma(P) < PMP < E(P) + c\sigma(P)\} \geq 1 - \frac{1}{c^2} \quad (44)$$

Using the same data for 1-hour PMP values at Thompson, Texas, it can be shown that:

For $c = 1$ $\quad P\{194.70 < P < 251.88\} \geq 0.0$

For $c = 2$ $\quad P\{166.11 < P < 280.47\} \geq 0.75$

For $c = 3$ $\quad P\{137.52 < P < 309.06\} \geq 0.89$

It shows that there is a less than 11% probability that the PMP estimate P was bigger than 309.06 mm and smaller than 137.52 for $c = 3$. The probability bound suggested that the value of 251.88 mm had a higher chance to be exceeded because of the uncertainty associated with the estimates of $\overline{X}$, S_n and record length of 52. The value of 309.06 mm corresponded to a more conservative estimate that was less likely to be exceeded because of the uncertainty associated with $\overline{X}$ and S_n. It took into account the effect of uncertainty and the associated probability of exceedance.

7.2 Uncertainty in PMP Estimates Using Taylor Series Expansion

The Taylor series expansion was used to propagate the uncertainty introduced in the PMP estimates due to various parameters. Considering PMP as a function of mean, standard deviation and frequency it can be shown that

$$PMP = g(x), \qquad x : (x_1, x_2, x_3)$$

In terms of PMP formula, $g(x)$ can be shown as:

$$g(x) = x_1 + x_2 \times x_3 \quad (45)$$

where x_1 = the mean of PMP estimates, x_2 = the standard deviation of PMP estimates, and x_3 = the frequency factor of PMP estimates.

Taking the first-order approximation, the expected value of PMP estimates were determined based on (Ang and Tang, 1975):

$$E(PMP) = g(\overline{x})$$

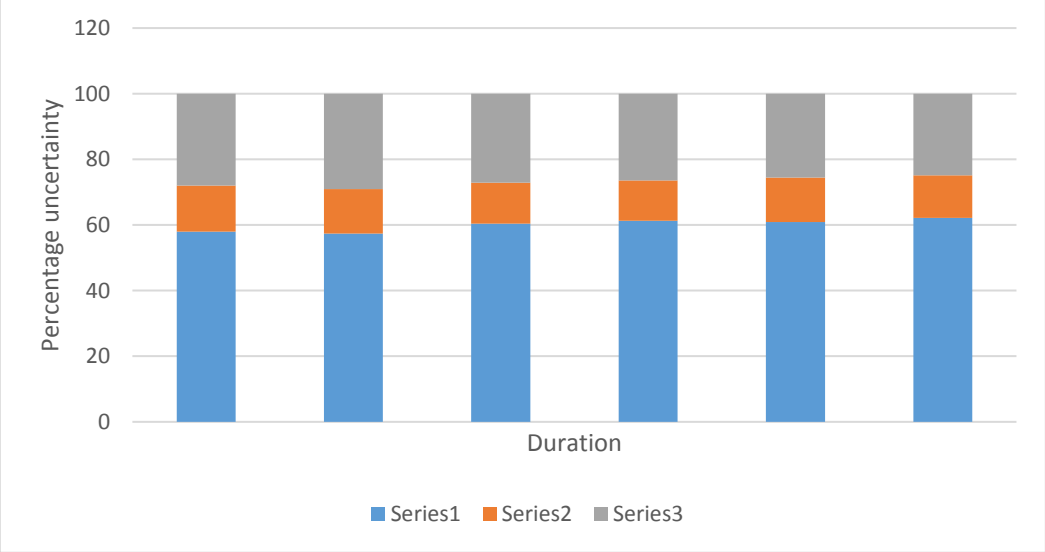
where $\bar{x}$ = $(\overline{x_1}, \overline{x_2}, \overline{x_3})$ and $\overline{x_1}$ is the mean value of mean of PMP estimates, $\overline{x_2}$ is the mean value of standard deviation, and $\overline{x_3}$ is the mean value of frequency factor. The variance of PMP due to each random component around the mean was calculated based on:

$$Var(PMP) = \sum_{i=1}^{i=k} \left(\frac{dg(x)}{dx_i}\right)^2 \sigma_i^2 \quad (46)$$

where $i = 1 \ldots k$ are the random variables, and σ_i is the standard deviation of each component. The expected value of PMP estimates was 276 mm for 1-hour duration with the variance of 32 mm due to mean, 22 mm due to frequency factor and 16 mm due to standard deviation. For the 6-hour duration the expected value was 346 mm with the variance of 39 mm due to mean 26.2 mm due to frequency factor and 17.9 mm due to standard deviation. For 24-hour duration the expected value was 423 mm with the variance of 45.3 due to mean 29.4 mm due to frequency factor and 21.4 mm due to standard deviation. To determine the relative contribution of each random component to the total uncertainty of PMP coefficient of variation was calculated based on:

$$Cov = \frac{\sigma_i^{\,2}}{\mu_i^{\,2}} = \sum_{i=1}^{i=k} \left(\frac{dg(x)}{dx_i}\right)^2 \frac{\sigma_i^{\,2}}{\mu_i^{\,2}} \tag{47}$$

It was observed that for 1-hour duration 58% uncertainty was introduced due to mean, 27% due to frequency factor and 14% due to standard deviation. For 6 hour duration it was 60.4% due to mean, 27.2% due to frequency factor and 12.5% due to standard deviation. For 24-hour it was 60.95 due to mean, 25.5% due to frequency factor and 13.5% due to standard deviation (Figure 45).



Series 1- Mean, Series 2 – Standard deviation, Series 3- Frequency factor

Figure 45 Relative uncertainty of each component to the total uncertainty in PMP estimates

8. RISK ANALYSIS OF EXTREME PRECIPITATION

8.1 Data

In the United States National Flood Insurance Program (NFIP) was established in 1968, in response to a lack of private sector availability of flood coverage (Kousky and Kerjan, 2015). Flood data for the Harris County was requested from FEMA which maintains the NFIP. The data comprised the flood event, amount of damage in dollars for a particular flood event in a particular area which was characterized by 5 digit zip code, the date of the event and the amount paid by NFIP. The flood events were from the period of year 1978 to 2002. Table 9 shows the flood events and the start date of the events.

Table 9 Flood events in Harris County with the start date

Event name	Start date
TX FLOOD APRIL 1979	4/18/1979
TX FLOOD SEPTEMBER 1979	9/19/1979
TX FLOOD MAY 1981	5/3/1981
TX FLOOD EVENT JUNE 1981	6/5/1981
TX FLOOD AUGUST 1981	8/31/1981
TX FLOOD MAY 1983	5/20/1983
TX FLOOD SEPTEMBER 1983	9/6/1983
TX FLOOD EVENT OCTOBER 1984	10/19/1984

In the Harris County there were two rain gauges Houston Alife and Houston Addicts. For gathering precipitation events at the two stations associated with flood events hourly precipitation data from NCDC NOAA was used. To compensate for the effect of wet catchment before a storm, the antecedent precipitation index (API) was used. Wet catchment and heavy rain can lead to flooding. API was calculated as (Ball et al, 2016):

$$API_n = P_n + kP_{n-1} + k^2 P_{n-2} + \ldots \quad (48)$$

where API_n is the Antecedent Precipitation Index for day *n*, *k* is the empirical decay factor less than one, and P_n is the precipitation for day *n*. The catchment wetness declines each day by the factor *k*. API increases again due any rain. A value of 0.90 was used for *k* (U.S. Department of Commerce, 1977).

Precipitation events associated with the flood date were selected. It may be noted that two precipitation events were considered independent when there was a certain dry period in between without or with minor precipitation. Urban drainage systems have a lagged response to precipitation and need a certain time to restore the equilibrium state. The flood damage can be completely related to a particular precipitation event if systems are in the equilibrium state before a precipitation event. A typical time for sewer systems to restore equilibrium state is between 10 and 20 hours (Spekkers et al, 2012). Hence, the precipitation events were selected that had at least a 10- to 12-hour difference between them. Also, the precipitation amount less than 1 mm/hour was treated as no precipitation. That prevented unrealistic long precipitation events to exist because of very small precipitation volumes between precipitation events. The damage amounts in dollars for

the flood events were arranged and adjusted for Consumer Price Index Inflation (Smith and Matthews, 2015). The damage data was given for zip areas in Harris County, hence for calculating the total damage due to a precipitation event the damages in all the zip areas up to which the event extended needed to be added. A weighted average function was developed, depending upon the amount of precipitation in both the rain gauges, the duration of precipitation and the closeness of zip area to the rain gauge. The total damage values were distributed among the zip areas for precipitation events occurring at both the stations and were added to have the total damage due to particular precipitation event.

8.2 Risk Analysis and Assessment of Damage

The exceedance probability was calculated for precipitation events associated with flood events from frequency analysis. The expected loss or risk associated with the precipitation events was calculated based on the formula:

$$E(L) = p_i \times L_i \tag{49}$$

where E_i is the expected loss for a given event, p_i is the probability of exceedance of a precipitation event, and L_i is the associated loss. Table 10 shows the exceedance probability of precipitation events, total damage and the expected loss or the associated risk in a given year for Houston Alife station. As can be seen, the expected loss due to precipitation events was very high. Figure 46 shows the correlation between the total damage and exceedance probability along with confidence bounds. The total damage was related to the exceedance probability by the power equation:

$$y = 491351x^{-0.817} \tag{50}$$

where, y is the total damage in U.S dollars and x is the exceedance probability.

Table 10 Expected loss due to precipitation events

Total Damage (Dollars)	Exceedance Probability	Expected Loss (Dollars)
5273021	0.047	247832
2786696	0.063	175561.8
74343336	0.0026	193292.7
16499400	0.02395	395160.6
7264661	0.02798	203265.2
27063964	0.00795	215158.5
1194753	0.1598	190921.5
1691062	0.07746	130989.6
31907133	0.009	287164.2
268499.5	0.0705	18929.2
9214725	0.00808	74454.9
5260444	0.0231	121516.3
40911349	0.0029	118642.9
45063692	0.00226	101843.9
6356198	0.01782	113267.5

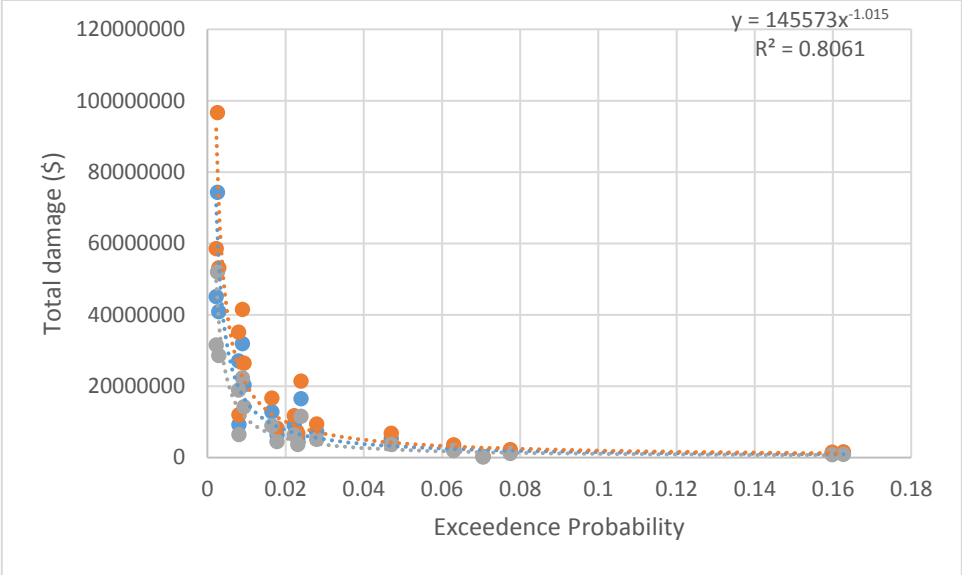


Figure 46 Plot between total damage and exceedance probability along with probability bound

Due to the limitation of data there were not many precipitation events of different durations so that the probability of exceedance can be correlated with damage amount. The exceedance probability was plotted against the total damage for 12-hour flood events at Houston Addicts station and 6-hour flood events at Houston Alife station. Figures 47 and 48 show the correlation curves for both stations along with the fitted regression equation and confidence bounds. The power equation fitted to both the curves were:

$$y = 491351x^{-0.817} \tag{51}$$

for 12-hour flood events at Houston Addicts station

$$y = 50234x^{-1.129} \tag{52}$$

for 6-hour flood events at Houston Alife station

where, *y* is the total damage in U.S dollars and *x* is the exceedance probability. The exceedance probability of PMP was fitted to the equations for both stations. Table 11 shows the exceedance probability of PMP values and total damage. As can be seen, the total damage amount is very high. The damage due to a single PMP event of 12-hour duration can be as high as 2 billion. It shows how much damage a single PMP event can cause. It may be noted that more events would have given better results.

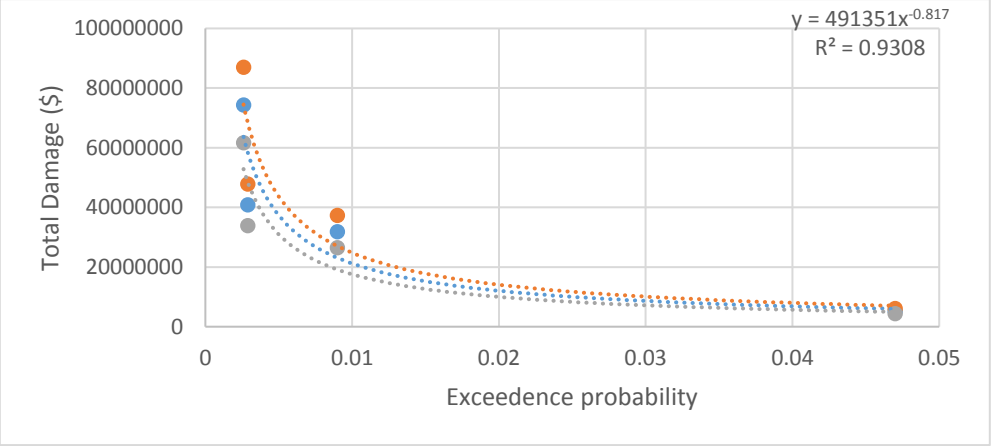


Figure 47 Correlation curve between exceedance probability of PMP and total damage for 12 hour duration precipitation at Houston Alife

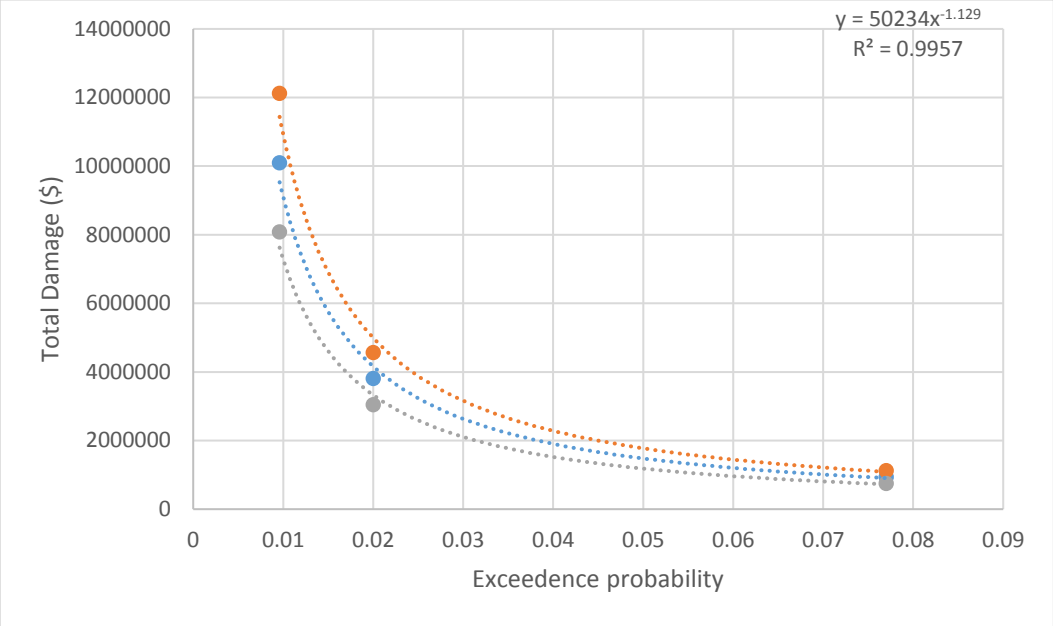


Figure 48 Correlation curve between exceedance probability of PMP and total damage for 6 hour duration precipitation at Houston Addicts

.

Table 11 Expected damage due to PMP events in both stations

Station	Duration (hour) and exceedance probability of PMP event		Expected Damage (Dollars)
	Duration	Probability	
Houston Alife	12	0.0000300	2425829620
Houston Addicts	6	0.00064	202684515

8.3 Limitations

The damage amount from PMP is just an expected value which we got by correlating a few precipitation events and amount of damage. More events would have given better results. In quantifying risk vulnerability is also an important term which involves damage cost functions of GDP (assets), population density, etc. Normalizing flood loss should also include GDP deflator, population, wealth per capita, etc. However, our focus was more on quantifying the uncertainty in the PMP estimates. These things were omitted in performing risk analysis. The amount of damage due to PMP is an approximate amount to show the hazardous nature of flooding.

9. CONCLUSION

9.1 Conclusion

It can be seen from our study that the PMP values are highly uncertain. The key conclusions from the study were:

(1) The PMP estimates derived from the statistical method depend largely on the frequency factor. Removing or adding any one station can change the shape of the curve which can result in highly uncertain PMP values.

(2) Hershfield enveloping curve yielded higher values of k_m compared to basin-specific curve which resulted in higher PMP estimates. Hershfield's statistical method can approximate the PMP values generally but for a specific area priority should be given to use the specific precipitation data for the area and derive the enveloping curve for the specific area.

(3) Frequency analysis was done using 24 commonly used probability distributions, 5 GOF tests, hazard rate and also using the PDFs of the distributions. The Burr XII performed better for less skewed distributions and the log-logistic distribution (3 parameter) performed better for more skewed distributions. On an average basis Burr type XII came out to be the best distribution for the study area.

(4) The return period of the PMP values was in the range of 1000 to 3000 years which was less than the range of 10^3 to 10^6 years reported in HMR 51. Considerable uncertainty can be introduced in return period of the PMP values because of the choice of probability distribution

(5) Design risk estimates were also calculated using Burr XII distribution parameters which gave a more conservative estimate of PMP by incorporating uncertainty. Probability bounds on PMP estimates suggested that PMP values had higher chance of being exceeded considering the uncertainty due to sample mean and standard deviation. It was also found that mean of extreme precipitation series has more relative effect on the combined uncertainty of PMP estimates.

(6) The expected damage due to a single PMP event of 12-hour duration can be as high as 2 billion in Harris County, Texas

9.2 Recommendation for Future

As can be seen from the study the PMP values derived from the basin-specific enveloping curve should be given higher priority than Hershfield's method or the published HMR documents PMP's. Quantifying uncertainty due to mean and standard deviation was possible because shape parameters of Burr XII distribution had narrower ranges hence construction of nomograph was possible by taking the average of both shape parameters and varying the record length and scale parameter. It may not be possible for other distributions. It would be interesting to see how the $Var(S_n)$ and $Cov(\overline{X_n}, S_n)$ will vary for other distributions.

REFERENCES

Ang, A.H.S., and Tang, W.H. (1975). *"Probability Concepts in Engineering Planning and Design*". Vol. 1, Basic Principles, John Wiley, New York.

Asquith, W.H. (1998). "Depth-Duration Frequency of Precipitation for Texas." *Water-Resources Investigations Report*. 98–4044.

Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E., Retallick, M., and Testoni, I. (2016). "Australian Precipitation and Runoff: A Guide to Flood Estimation". *Commonwealth of Australia* (Geoscience Australia), 2016.

Benson, M. A. (1973). "Thoughts on the Design of Design Floods." *Floods and Droughts, Processing 2nd International Symposium in Hydrology*, pp. 27-33, Water Resources Publications, Fort Collins, Colorado, 1973.

Brebbia, C.A. (2013). "*River Basin Management VII.*" Ashurust Lodge, Ashrust, Southampton, SO40 7AA, UK. WIT press.

Casas, M.C., Rodriguez, R., Redano, A., Prohom, M., and Gazquez, A. (2011). "Estimation of the Probable Maximum Precipitation in Barcelona (Spain)". *International Journal of Climatology,* 10.1002.

Casas, M.C., Rodriguez, R., Nieto, R., and Redano, A. (2008). "The Estimation of Probable Maximum Precipitation: The Case of Catolonia." *Annals of the New York Academy of Sciences*, 1146: 291–302 (2008).

Chakravarti, Laha., and L, Roy. (1967). *"Handbook of Methods of Applied Statistics."* Volume I, John Wiley and Sons, pp. 392-394.

Chow, V.T. (1951). "A General Formula for Hydrologic Frequency Analysis." *Transactions- American Geophysical Union,* 32, 231-237.

David, S.A., and Skaggs, L.L. (1992). "Catalog of Residential Depth-Damage Functions used by the Army Corps of Engineers in Flood Damage Estimates". *U.S Army Corps of Engineers.* IWR Report 92-R-3.

Dawdy, D.R., and Lettenmaier, D.P. (1987). "Initiative for Risk-based Flood Design". *ASCE Journal of Hydraulic Engineering* 113(8), 1041‑1051.

Douglas, E.M., and Barros, A.P. (2003). "Probable Maximum Precipitation Estimation Using Multifractals: Application in the Eastern United States." *Journal of Hydrometeorology*, 4, 1012-1024.

Dooge, J. C. I. (1986). "Looking for Hydrologic Laws." *Water Resources Research*, 22(9) pp. 46S-58S.

Hao, Z., and Singh, V.P. (2013). "Entropy-based Method for Extreme Precipitation Analysis in Texas." *Journal of Geophysical research: Atmospheres*, Vol. 10, 1-11.

Hershfield, D.M. (1965). "Method for Estimating Probable Maximum Precipitation." *American Water Works Association*, 57(8), 965-972.

Hershfield, D.M. (1961). "Estimating the Probable Maximum Precipitation." *Journal of Hydraulics Division*, 87(HY5), 99-106.

Kendall, M., and Stuart, A. (1963). "*The Advanced Theory of Statistics."* Vol. 1, Distribution Theory, 2nd edition., Hafner Publishing Company, Inc., New York.

Kite, G.W. (1998). "*Frequency and Risk analysis in Hydrology"*. Water resources publications, Littleton, CO, USA.

Koks, E.E., Moel, H., and Koomen, E. (2012). "Comparing Extreme Precipitation and Large-Scale Flooding Induced Inundation Risk– Evidence from a Dutch Case-Stud". *VU University Amsterdam*, the Nederland's.

Koutsoyiannis, D. (1999). "A Probabilistic view Of Hershfield's method for Estimating Probable Maximum Precipitation." *Water Resources Research*, 35(4), 1313-1322.

Koutsoyiannis, D. (2004). "Statistics of Extremes and Estimation of Extreme Precipitation II: Empirical Investigation of Long Precipitation Records." *Hydrological Sciences Journal*, 49(4): 591–610.

Koutsoyiannis, D., and Papalexiou, S.S. (2006) "A probabilistic approach to the concept of Probable Maximum Precipitation." *Advances in Geosciences*, 7, 51-54.

Koutsoyiannis, D., and Papalexiou, S.S. (2012). "Battle of Extreme Value distributions: A Global Survey on Extreme Daily Precipitation." *Water resources research*, vol. 49, 187–201.

Kousky, C., and Kerjan, E.M. (2015). "Examining Flood Insurance Claims in the United States: Six Key Findings". *The Journal of Risk and Insurance*, doi:10.1111/jori.12106.

Kulkarni, B.D. (2010). "Generalized Physical Approach of Estimating Areal Probable Maximum Precipitation for Plain Region of the Godavari River Basin (India)." *Journal of Spatial Hydrology*, 2(2).

Kunreuther, H., Meyer, R., and Bulte, C.V.D. (2004). "Risk Analysis for Extreme Events: Economic Incentives for Reducing Future Losses". *National Institute of Standards and Technology, Gaithersburg*, MD 20899-8603.

Mamoon, A.A., and Rahman, A. (2014). "Uncertainty in Design Precipitation Estimation: A Review." *Journal of Hydrology and Environment Research*, Vol 2, No 1.

Micovic, Z., Schaefer, M.G., and Taylor, G.H. (2014). "Uncertainty Analysis for Probable Maximum Precipitation Estimates." *Journal of Hydrology*, 521(2015), 360-373.

Miller, J.F. (1964). "Two- to Ten-Day Precipitation for Return Periods of 2 to 100 Years in the Contiguous United States." *Technical Paper No. 49*, Weather Bureau, United States Department of Commerce, Washington, DC.

Mood, A., Graybill, F., and Boes, D.C. (1974). *"Introduction to the Theory of Statistics*." 3rd edition, McGraw Hill, New York.

National Fibers Information Center. (1987). *"The Climate of Texas Counties."* University of Texas, Austin and Texas A&M University, College Station, Texas.

NRC (National Research Council). (2000). "Risk Analysis and Uncertainty in Flood Damage Reduction Studies*." National Academy Press,* Washington, DC, 202 pp.

Olofintoye, O., Sule, B. F., and Salami, A. W. (2009). "Best-fit Probability distribution Model for Peak Daily Precipitation of Selected Cities in Nigeria", *New York Science. Journal*, 2(3).

Rakhecha, P.R., Deshpande, N.R., and Soman, M.K. (1992). "Probable Maximum Precipitation for a 2-day duration over the Indian Peninsula." *Theory of Applied Climatology,* 45: 277–283.

Rakhecha, P.R., and Kennedy, M.R. (1985). "A Generalized Technique for the Estimation of Probable Maximum Precipitation in India". *Journal of Hydrology*, 78 (1985), 345—359.

Salas, J.D., Gavilan, G., Salas, F.R., Julien, P., and Abdullah, J. (2014). "Uncertainty of the PMP and PMF." *Handbook of engineering hydrology*, Ch 57, Vol 2.

Schreiner, L. C., and Riedel J.T. (1978). "Probable Maximum Precipitation Estimates, United States East of the 105th Meridian *HMR No. 51." National Weather Service, National Oceanic and Atmospheric Administration*, United States Department of Commerce, Washington, DC.

Sharma, M.A., and Singh, J.B. (2010). "Use of Probability Distribution in Precipitation Analysis", *New York Science. Journal*, 3(9), pp. 40-49.

Smith, A.B., and Matthews J.L. (2015). "Quantifying Uncertainty and variable sensitivity within the US billion-dollar Weather and Climate Disaster Cost Estimates". *Natural Hazards* 77: 1829. doi:10.1007/s11069-015-1678-x.

Spekkers, M.H., Kok, M., Clemens, F.H.L.R., ten Veldhuis, A.E. (2012). "A statistical Analysis of Insurance Damage Claims Related to Precipitation Extremes". *Hydrology and. Earth System Science*, 17, 913–922.

Srinivas, V.V., and Chavan, S.R. (2015). "Probable Maximum Precipitation Estimation for Catchments in Mahanadi River Basin". *Aquatic Procedia* 4 892 – 899.

Tessier, Y., Lovejoy, S., and Schertzer, D. (1993). "Universal multifractals: Theory and observations for rain and clouds". *Journal of Applied Meteorology*., 32, 223–250.

Tomlinson, E.M., and Kappel, W.D. (2009). "Dam Safety: Revisiting PMPs."

U.S. Department of Commerce. (1960). "Generalized Estimates of Probable Maximum Precipitation for the United States West of the 105th Meridain." *Technical Paper number 38*.

U.S. National Weather Service. (1988). "Probable Maximum Precipitation – United States between the Continental Divide and the 103rd Meridian." *Hydrometeorological Report No. 55A (HMR-55A)*, Silver Spring, MD, USA.

U.S Department of Commerce, National Oceanic and Atmospheric Administration. (1994). "Probable Maximum Precipitation- Pacific Northwest States". *Hydrometeorogical Report number 57*, October 1994, Silver Spring.

Villalta, D.E., Guenni, B., Sajo-Castelli, A.M., and Campos, J.S. (2014). 'Risk analysis to extreme precipitation: A retrospective approach'. Presentation.

Wiesner, C. (1970). "*Hydrometeorology, 232".* Chapman and Hall. Londres.

Weiss, L.L. (1968). "Ratio of True to Fixed-Interval Maximum Precipitation." *Journal of Hydraulics Division*, 90: 77–82.

Willeke, G.E. (1980). "Myths and uses of Hydrometeorology in Forecasting", *in Proceedings of March 1979 Engineering Foundation Conference on Improved Hydrological Forecasting- Why and How*, pp. 117-124, American Society of Civil Engineers, New York

World Meteorological Organization. (2009). "Manuel on Estimation of Probable Maximum Precipitation." *WMO-No 1045, Geneva, Switzerland*.

World Meteorological Organization. (1986). "Manual for Estimation of Probable Maximum Precipitation" *Operational Hydrology Report* 1, 2nd edition, Publication 332, World Meteorological Organization, Geneva.

APPENDIX A

INFORMATION FOR ESTIMATING PMP VALUES

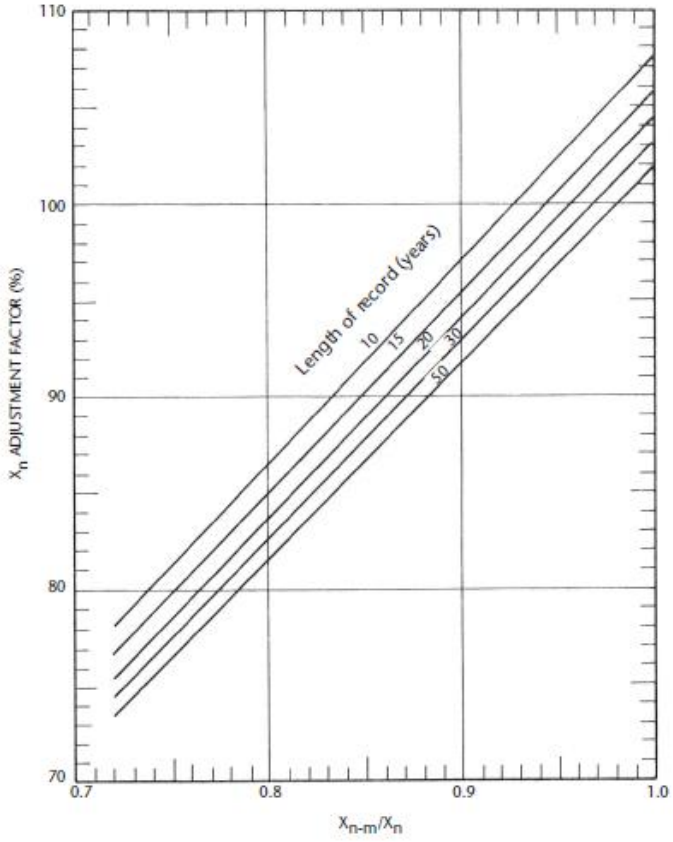


Figure A-1 Adjustment of mean of annual series for maximum observed precipitation (Hershfield, 1961b)

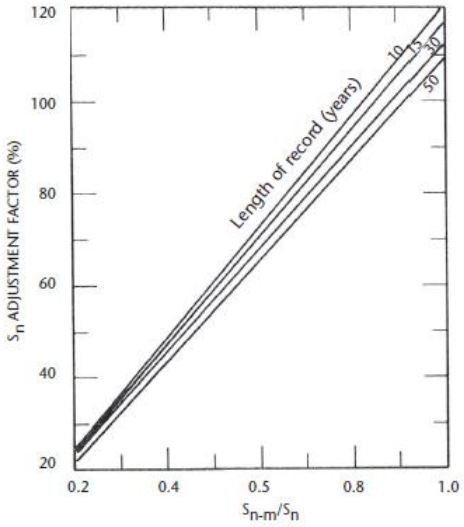


Figure A-2 Adjustment of standard deviation of annual series for maximum observed precipitation (Hershfield, 1961b)

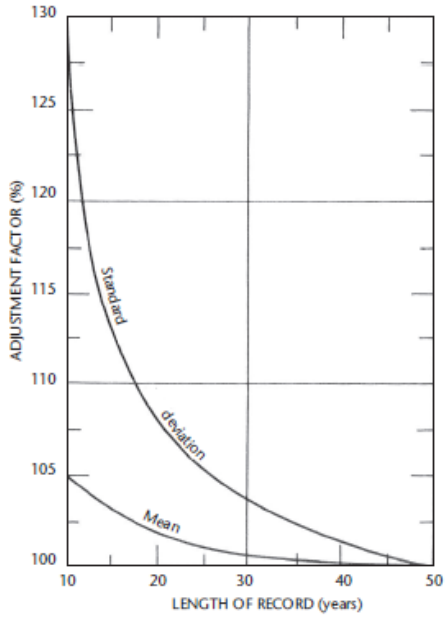


Figure A.3 Adjustment of mean and standard deviation of annual maximum series for length of record (Hershfield, 1961b) where X_{n-m} corresponds to the mean of the annual maximum series excluding the highest observation X_n is the mean of the annual maximum series, S_{n-m} is the standard deviation of the annual maximum series excluding the highest observation, and S_n is the standard deviation of the annual maximum series.

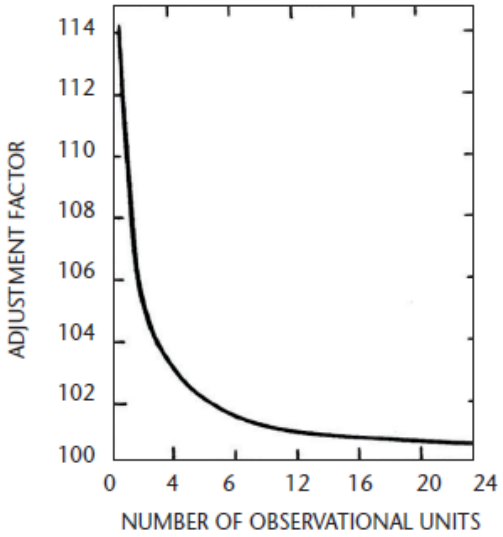


Figure A.4 Adjustment of fixed-interval precipitation amounts for a number of observational units within the interval (Weiss, 1968)

APPENDIX B

PMP ESTIMATION

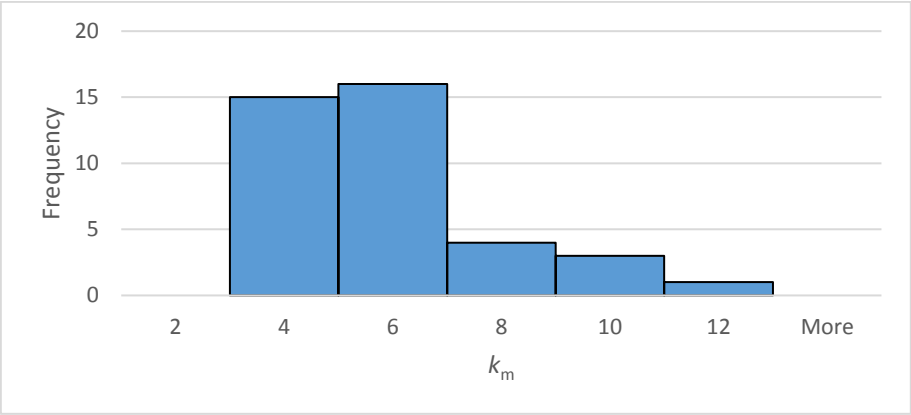


Figure B.1 Histogram of observed k_m for 1 hour duration of precipitation

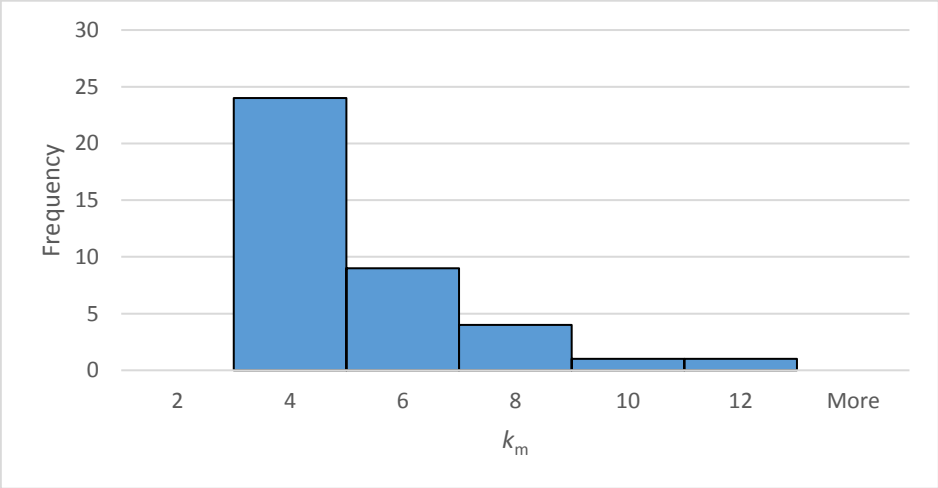


Figure B.2 Histogram of observed k_m for 2 hour duration of precipitation

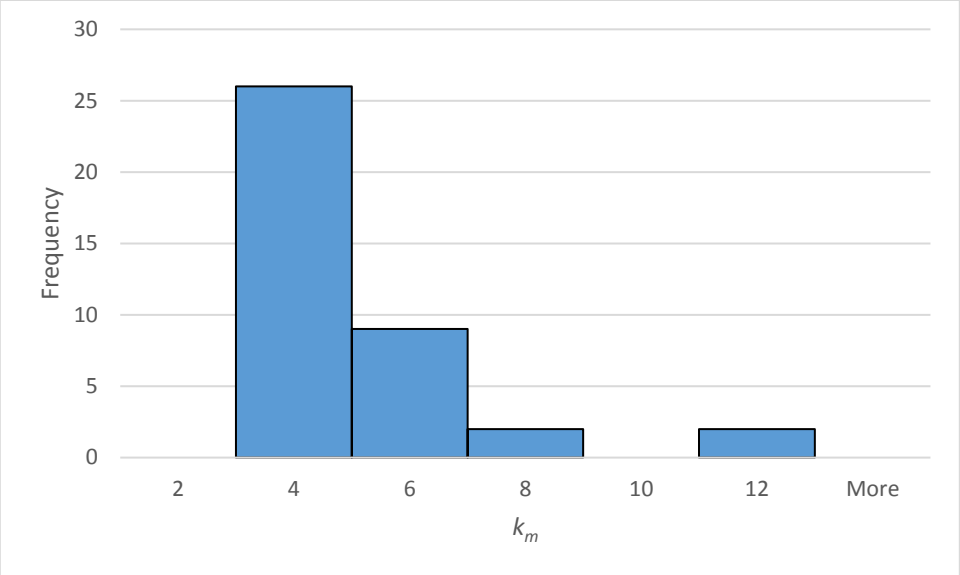


Figure B.3 Histogram of observed k_m for 3 hour duration of precipitation

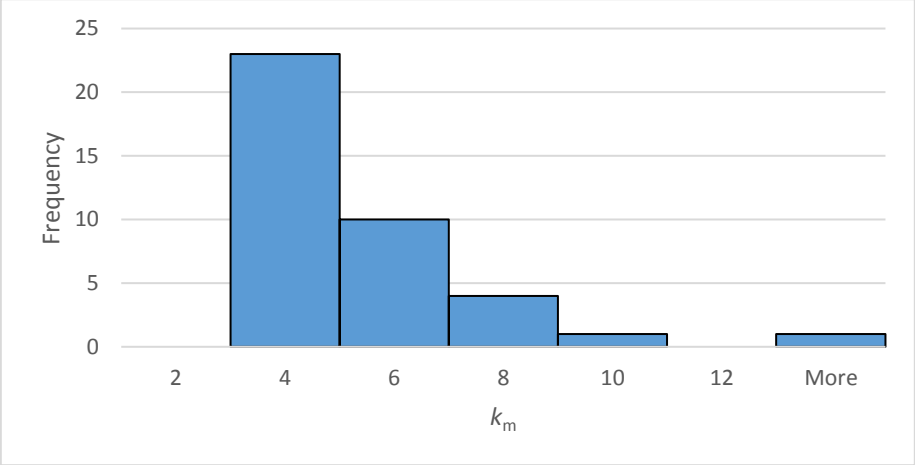


Figure B.4 Histogram of observed k_m for 6 hour duration of precipitation

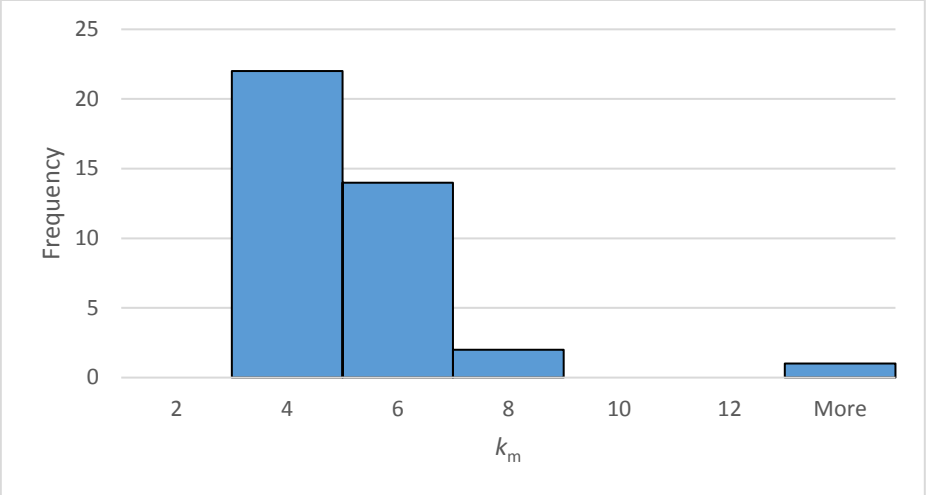


Figure B.5 Histogram of observed k_m for 12 hour duration of precipitation

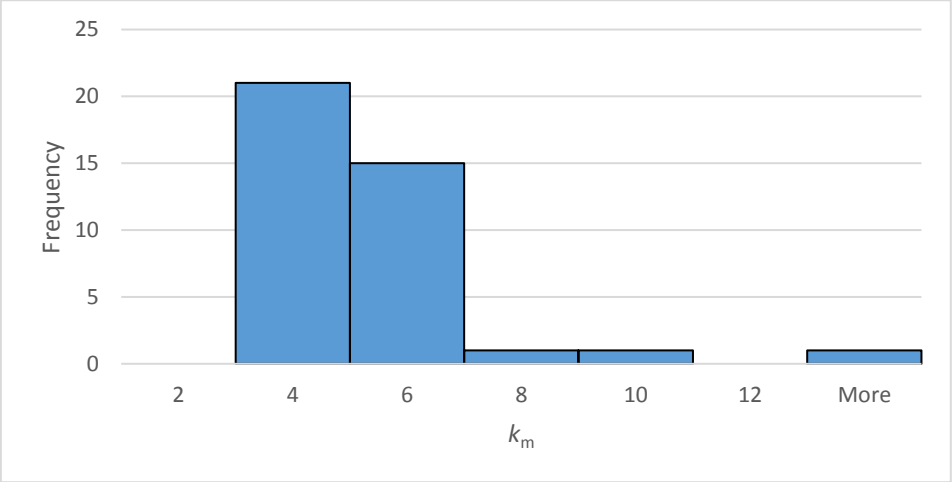


Figure B.6 Histogram of observed k_m for 24 hour duration of precipitation

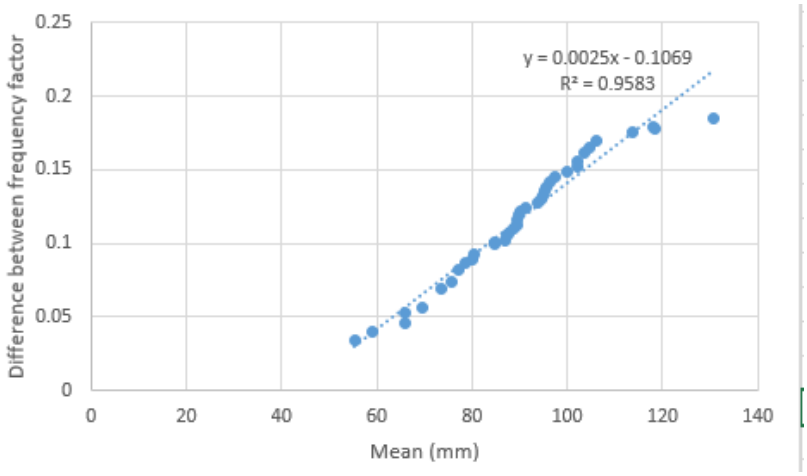


Figure B.7 Difference between the Hershfield enveloping curve and Basin-specific curve for 24 hour duration in dimensionless terms.

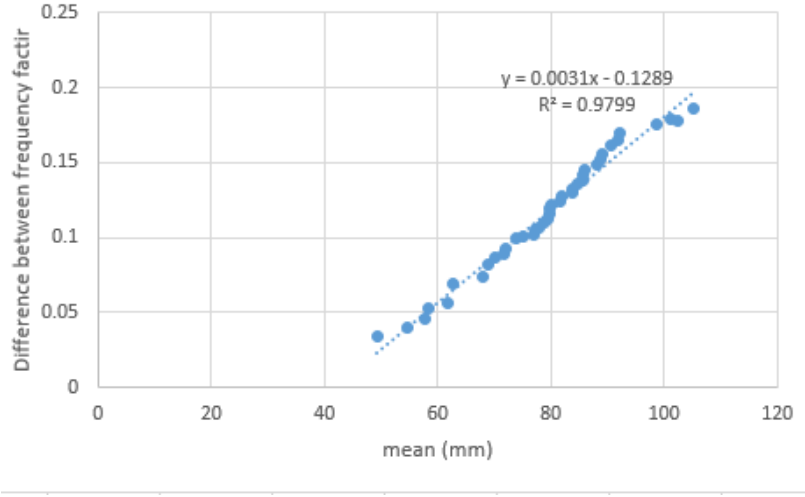


Figure B.8 Difference between the Hershfield enveloping curve and Basin-specific curve for 12 hour duration in dimensionless terms.

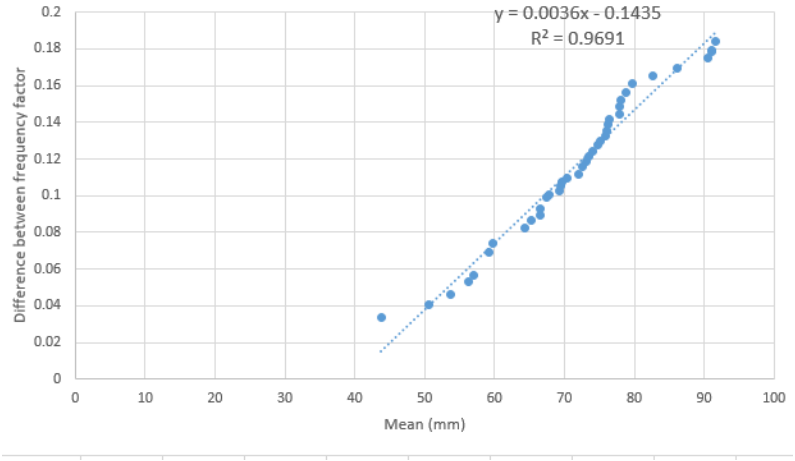


Figure B.9 Difference between the Hershfield enveloping curve and Basin-specific curve for 6 hour duration in dimensionless terms.

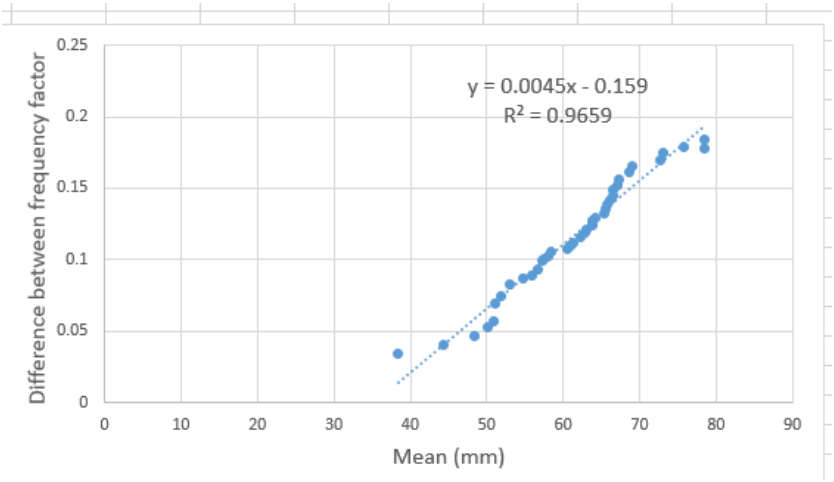


Figure B.10 Difference between the Hershfield enveloping curve and Basin-specific curve for 3 hour duration in dimensionless terms.

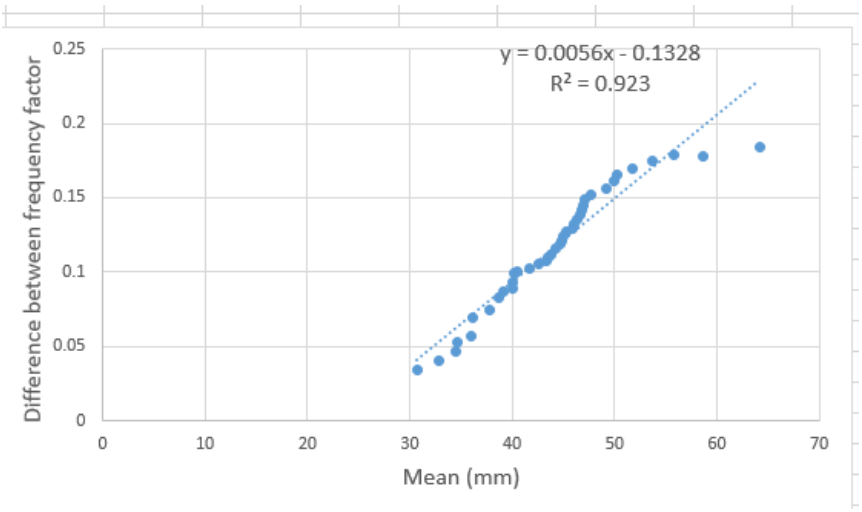


Figure B.11 Difference between the Hershfield enveloping curve and Basin-specific curve for 1 hour duration in dimensionless terms.

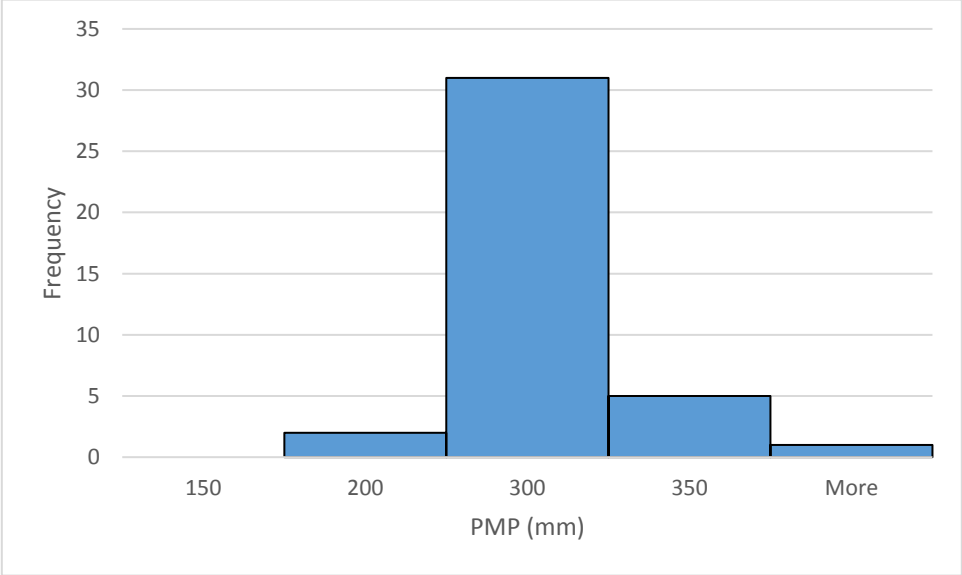


Figure B.12 Histogram of adjusted PMP values for 1 hour duration

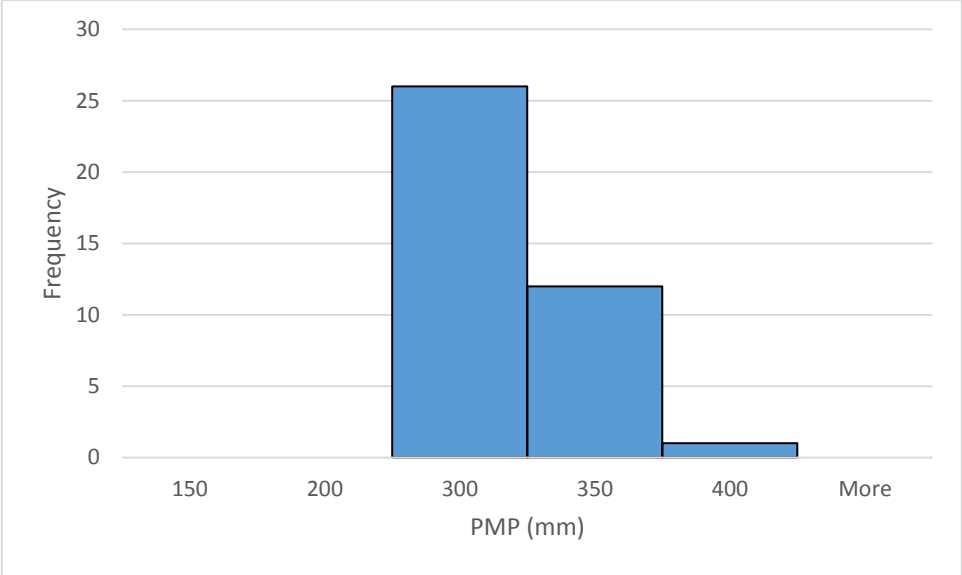


Figure B.13 Histogram of adjusted PMP values for 2 hour duration

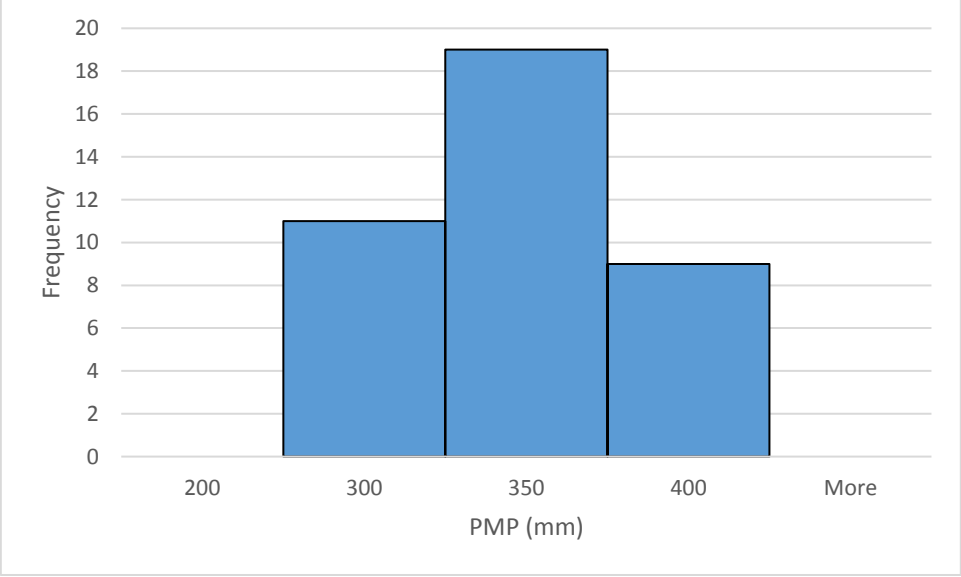


Figure B.14 Histogram of adjusted PMP values for 3 hour duration

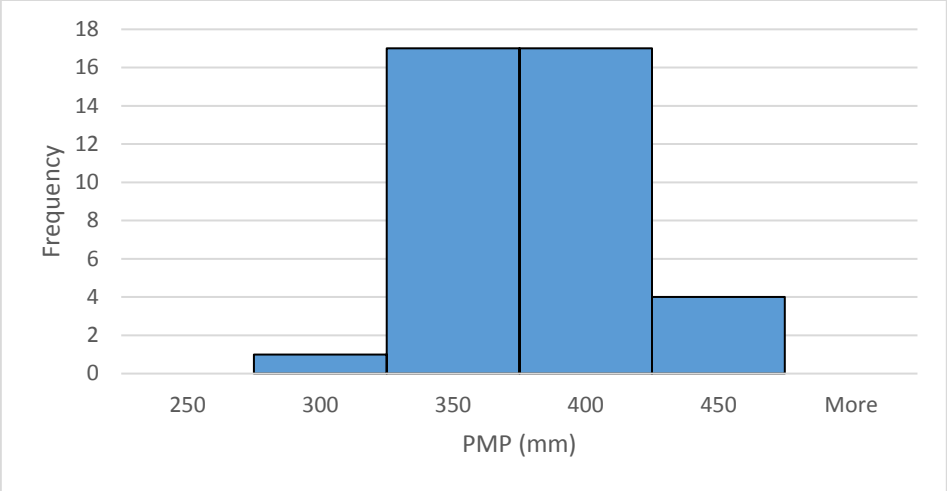


Figure B.15 Histogram of adjusted PMP values for 6 hour duration

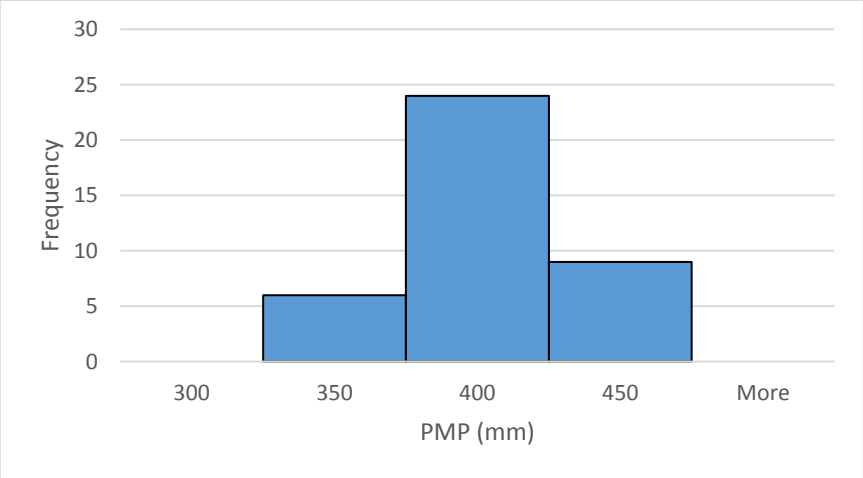


Figure B.16 Histogram of adjusted PMP values for 12 hour duration

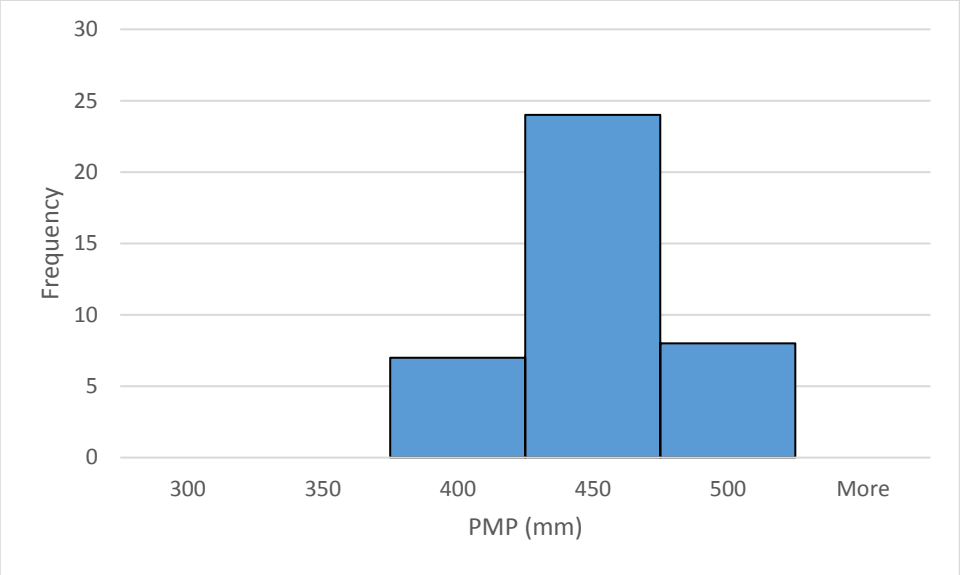


Figure B.17 Histogram of adjusted PMP values for 12 hour duration

APPENDIX C

FREQUENCY ANALYSIS OF EXTREME PRECIPITATION

RESULTS

Table C.1 Return Periods for all stations and durations in Brazos River basin

Station Name	1-hour	2-hour	3-hour	6-hour	12-hour	24-hour
Abilene	6666.667	1265.823	76923.08	5921364	33333.33	1111.111
Bay city	14285.71	2083.333	1538.462	14285.71	20000	6579.895
Belton	2557414	100000	100000	7142.857	2040.816	4347.826
Bertnam	6320.911	4505.45	8190.672	14285.71	33333.33	16666.67
Briggs	3125	6666.667	6250	1818.182	2941.176	2222.222
Burleson	7891.25	1306.986	3557.872	2127.66	3225.806	16666.67
Clovis	2564.103	1333.333	16666.67	12500	33333.33	232552.7
Coryell	568.1818	3448.276	2380.952	2631.579	1075.269	6136.497
Cranfills	2325.581	16666.67	125000	50000	5263.158	1886.792
Cherokee	24631.55	25000	353319.4	64551.16	70989.22	50796.61
Cressson 3 NW	66296.25	58842.35	25000	10000	50000	33333.33
Eastland	1265.823	5626.316	2941.176	9090.909	50000	12500
Evant 1 SSW	3500.877	3757.94	9710.874	1886.792	1470.588	1870.49
Santa anna	4761.905	2222.222	5000	2857.143	3333.333	1282.051

Table C.2. Overall best-fit distribution for different stations and durations

Station	Duration		
	2-hour	3-hour	12-hour
Albine	Burr	GEV	GEV
Bay City	Log-logistics 3	Burr	Gen Gamma (4p)
Belton	Log-Pearson 3	Inverse-Gaussian 3	GEV
Bertnam	GEV	GEV	Gen Gamma (4p)
Briggs	Dagum	Burr	Log-logistics 3
Burleson	Burr	GEV	Johnson SB
Clovis	GEV	Pearson 5 (3p)	GEV
Coryell	Gen Gamma (4p)	GEV	Burr
Cranfills	Inverse Gussain (3p)	GEV	Log-logistics 3
Cherroke	Inverse Gussain (3p)	Inverse-Gaussian 3	Log-logistics 3
Cresson	Burr	Log-logistics 3	Burr
Eastland	Burr	GEV	Johnson SB
Evant	Burr	Dagum	Burr
Santa Anna	Gen Gamma (4p)	Burr	Burr
Flat	Inverse-Gaussian 3	Burr	GEV
Galveston	Burr 4p	Johnson SB	Inverse-Gaussian 3
Gorman	Burr	Burr	Johnson SB
Groesbeck	GEV	GEV	Burr
Houston Addicts	GEV	Log-pearson3	Frechet (3p)
Houston alife	Burr	GEV	Gen Gamma (4p)
Indian Gap	GEV	GEV	Log-logistics 3

Table C.2 Continued

Station	Duration		
	2-hour	3-hour	12-hour
Iredell	Johnson SB	Johnson SB	Inverse Gussain (3p)
Jayton	Burr	Burr	GEV
Jewett	Burr	Burr	Burr
Kopperl	GEV	Burr	GEV
Lexington	Gen Gamma (4p)	GEV	Log-pearson3
Loraine	Gen Gamma (4p)	Gen Gamma (4p)	GEV
Lubbock	GEV	Frechet (3p)	Gen Gamma (4p)
Moline	Burr	Burr	GEV
Pep	GEV	Burr	Burr
Richmond	Johnson SB	Johnson SB	Frechet (3p)
Spicewood	GEV	Burr	Burr
Stamphord	Log-logistics 3	Log-logistics 3	GEV
Stephenville	Log-logistics 3	Log-logistics 3	Log-logistics 3
Still house	Burr	Burr	Burr
Thompson	Pearson 5 (3p)	Inverse Gussain (3p)	GEV
Waco	GEV	GEV	Log-logistics 3
Washington	Log-logistics 3	Burr	Burr
Wheelock	Dagum (4p)	GEV	Gen Gamma

*GEV = Generalized Extreme Value

Table C.3 Best probability distribution coverage for 1-hour duration in Brazos River basin

Distribution	% Coverage
GEV	17.95
Burr	23.08
log-logistic(3p)	38.46
Burr (4p)	2.56
Beta	2.56
log pearson 3	2.56
Johnson Sb	5.13
Inv Gussian (3p)	5.13
Dagum	2.56
Frechet (3p)	2.56

Table C.4 Comparison of test static values for different GOF tests at Iredell, TX for 3-hour duration

Distribution	Goodness of fit test			
	K-S	A-D	C-S	RMSE
Burr	0.0705	0.32	1.527	8.25
GEV	0.0691	0.24	0.631	8.21
Log-Pearson 3	0.0758	0.27	1.318	9.14
Johnson SB	0.0644	0.19	0.871	7.39

Table C.5 Test Static values for different GOF tests for 3-hour duration at Thompson, TX

Distribution	Goodness of fit test			
	K-S	A-D	C-S	RMSE
GEV	0.053	0.178	0.261	19.17
Log-logistics	0.091	0.597	1.41	33

Table C.6 Best probability distribution coverage for 2-hour duration in Brazos River basin

Distribution	% Coverage
Gev	28.21
Burr	30.77
Log-logistics 3	7.69
Dagum	2.56
Pearson 5 (3p)	2.56
Gen Gamma (4p)	10.26
Log-pearson3	2.56
Johnson SB	5.13
Inverse Gussain (3p)	7.69
Dagum (4p)	2.56

Table C.7 Best probability distribution coverage for 3-hour duration in Brazos River basin

Distribution	% Coverage
Gev	33.33
Burr	33.33
Log-logistics 3	5.13
Dagum	2.56
Pearson 5 (3p)	2.56
Gen Gamma (4p)	2.56
Log-pearson3	2.56
Johnson SB	7.69
Inverse Gussain (3p)	7.69
Frechet (3p)	2.56

Table C.8 Best probability distribution coverage for 12-hour duration in Brazos River basin

Distribution	% Coverage
Gev	25.64
Burr	23.08
Log-logistics 3	17.95
Gen Gamma (4p)	12.82
Log-pearson3	2.56
Johnson SB	7.69
Inverse Gussain (3p)	5.13
Frechet (3p)	5.13
Pearson 6 (4p)	2.56
Frechet (3p)	2.56

Table C.9 Best probability distribution coverage for 24-hour duration in Brazos River basin

Distribution	% Coverage of stations
GEV	12.8
Burr	43.6
Log-Logistics 3	7.69
Log-Pearson 3	7.69
Johnson SB	12.8
Inverse Gaussian (3p)	10.3
Pearson 5 (3p)	2.56
Gumble Max	2.56

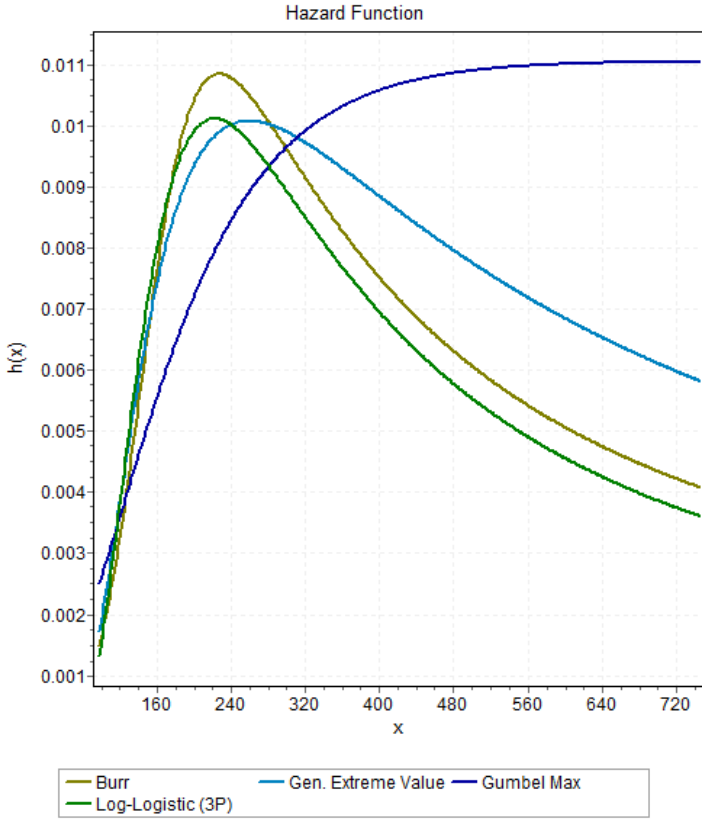


Figure C.1 Hazard rate for different distributions for 2 hour duration at station Coryell

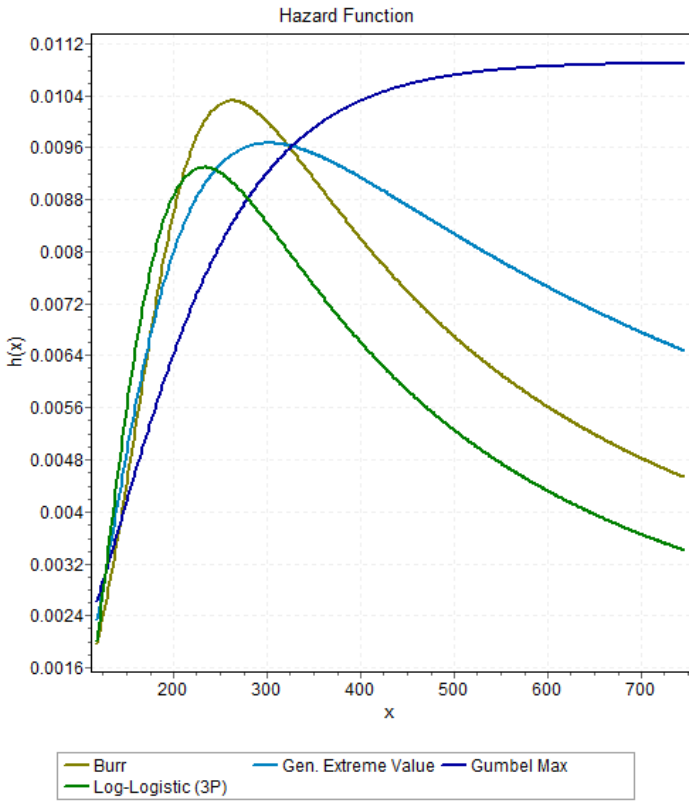


Figure C.2 Hazard rate for different distributions for 3 hour duration at station Coryell

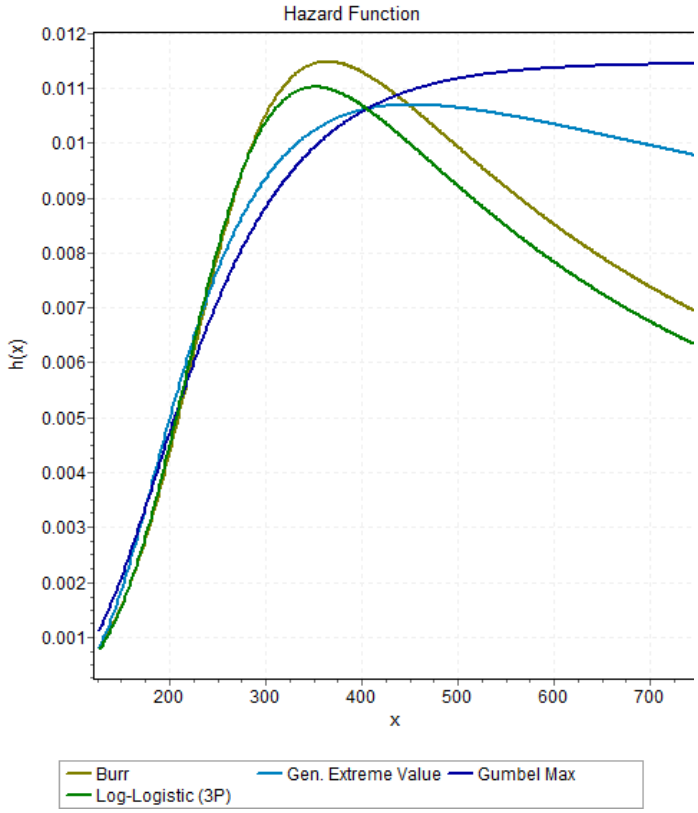


Figure C.3 Hazard rate for different distributions for 6 hour duration at station Coryell

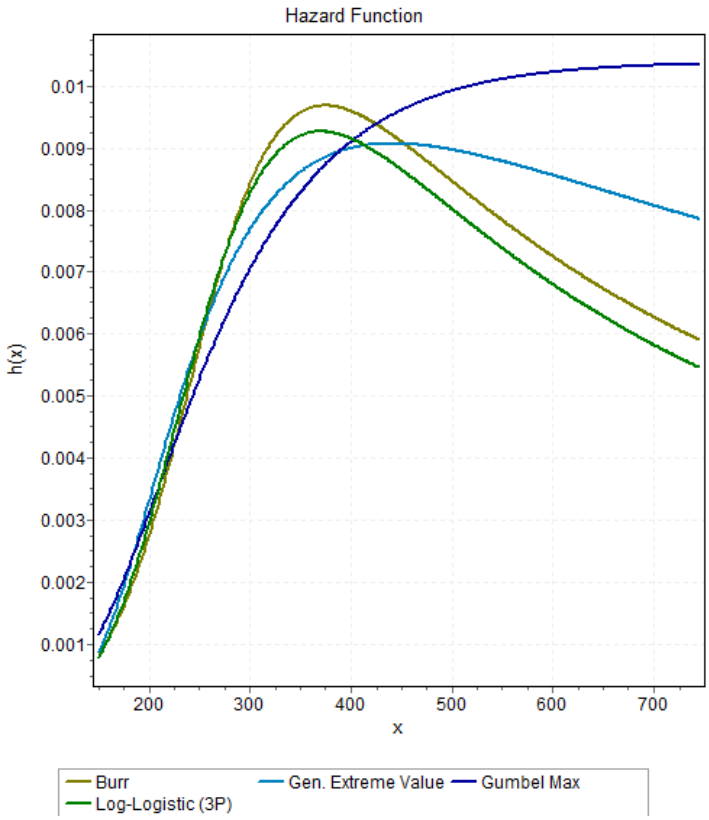


Figure C.4 Hazard rate for different distributions for 12 hour duration at station Coryell

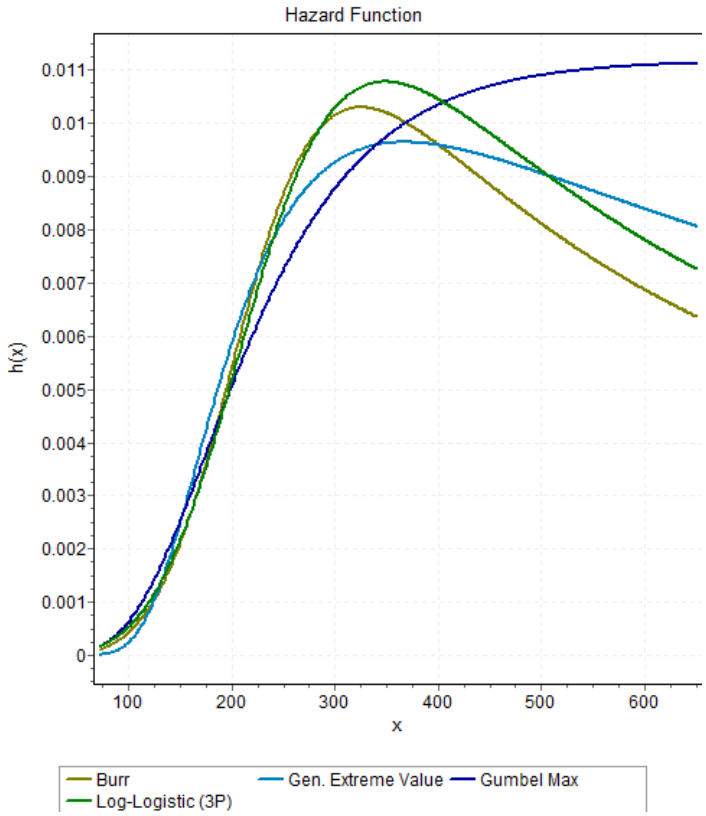


Figure C.5 Hazard rate for different distributions for 2 hour duration at station Houston Alife

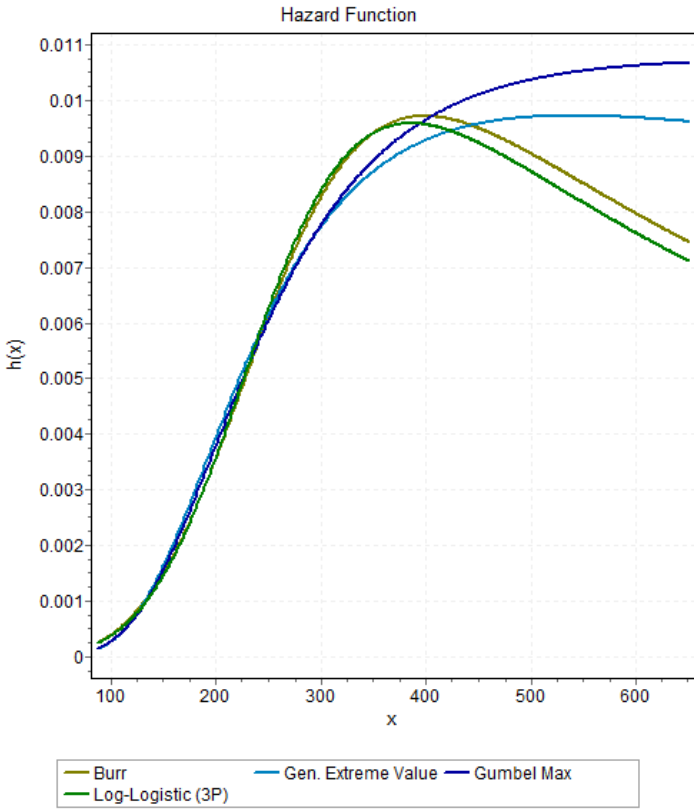


Figure C.6 Hazard rate for different distributions for 3 hour duration at station Houston Alife

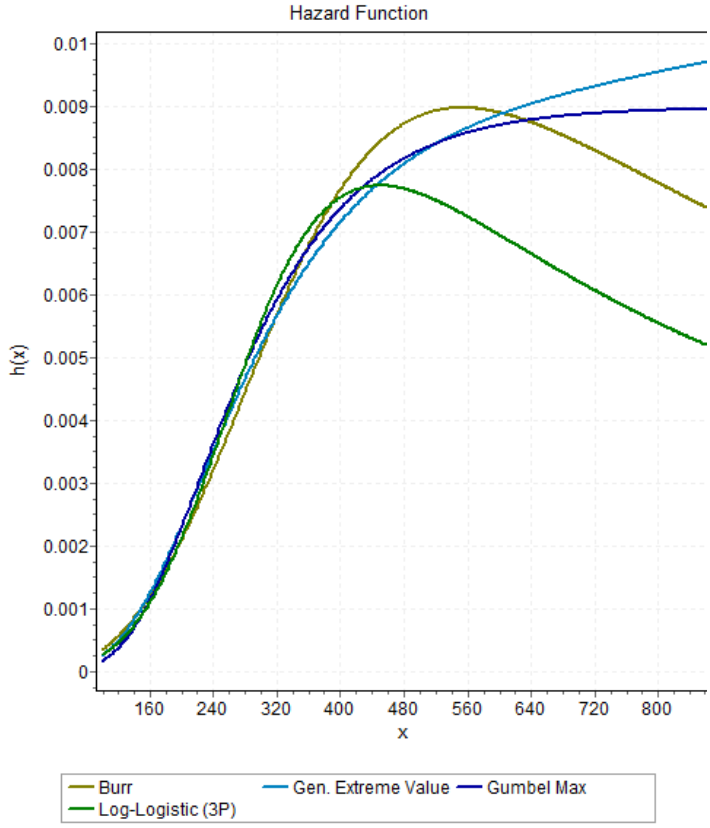


Figure C.7 Hazard rate for different distributions for 6 hour duration at station Houston Alife

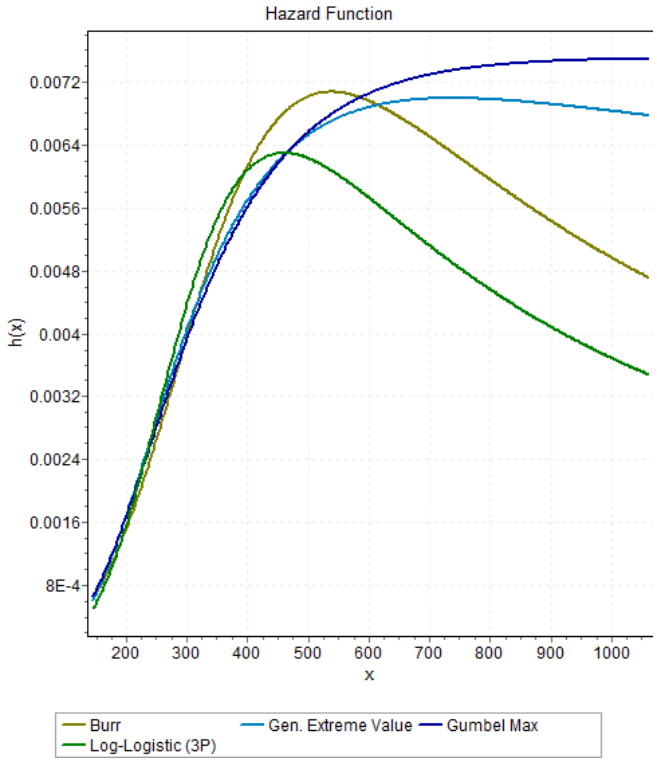


Figure C.8 Hazard rate for different distributions for 12 hour duration at station Houston Alife

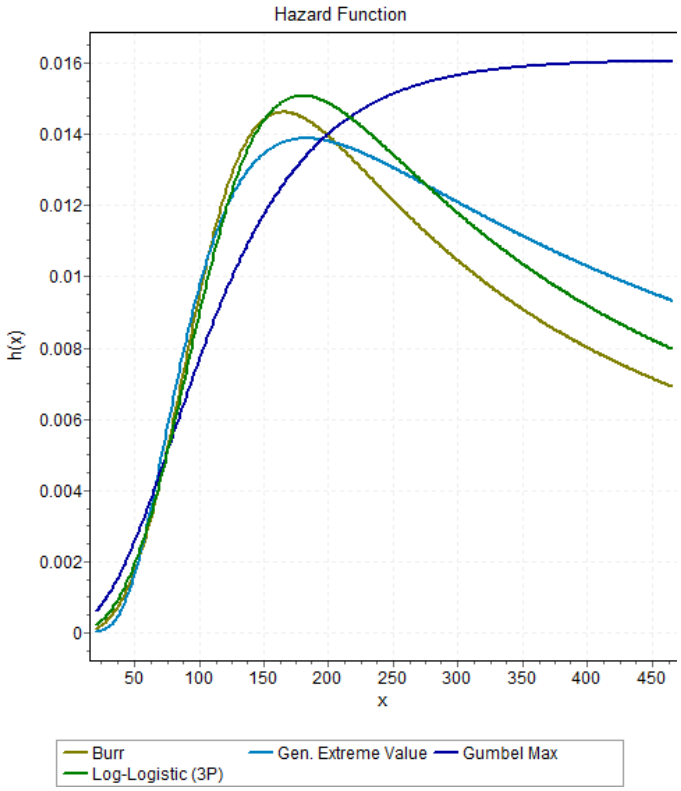


Figure C.9 Hazard rate for different distributions for 2 hour duration at station Lubbock

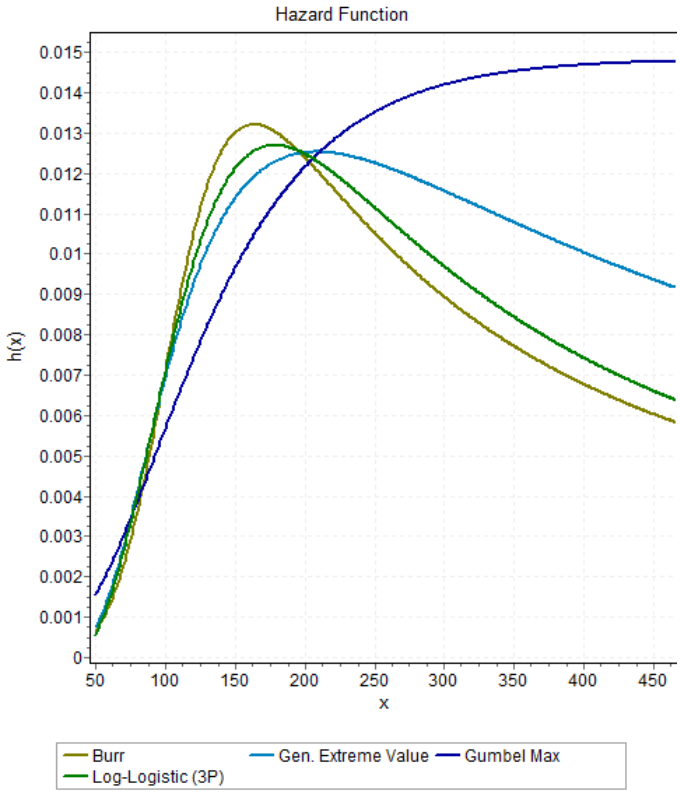


Figure C.10 Hazard rate for different distributions for 3 hour duration at station Houston Alife

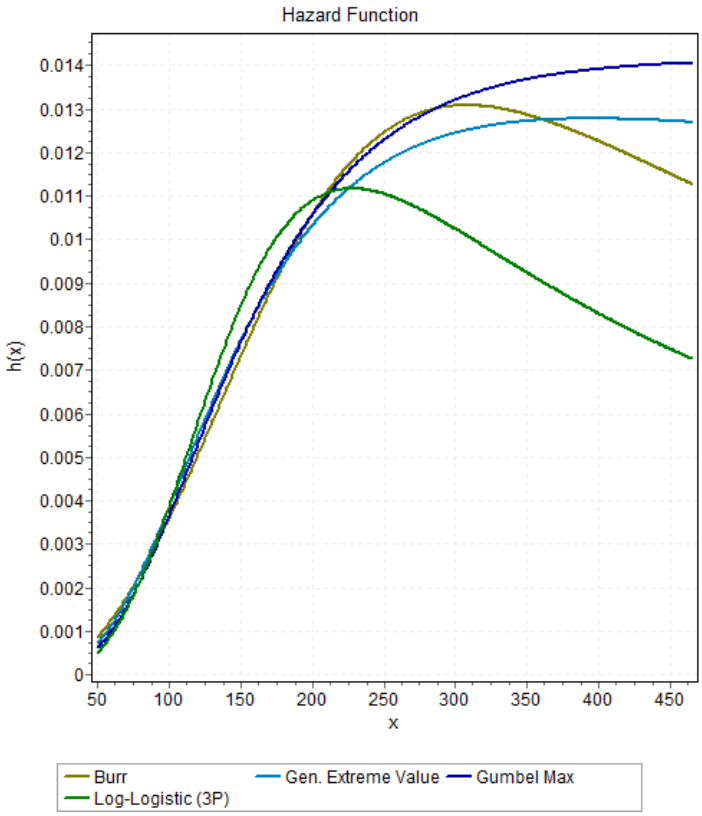


Figure C.11 Hazard rate for different distributions for 6 hour duration at station Houston Alife

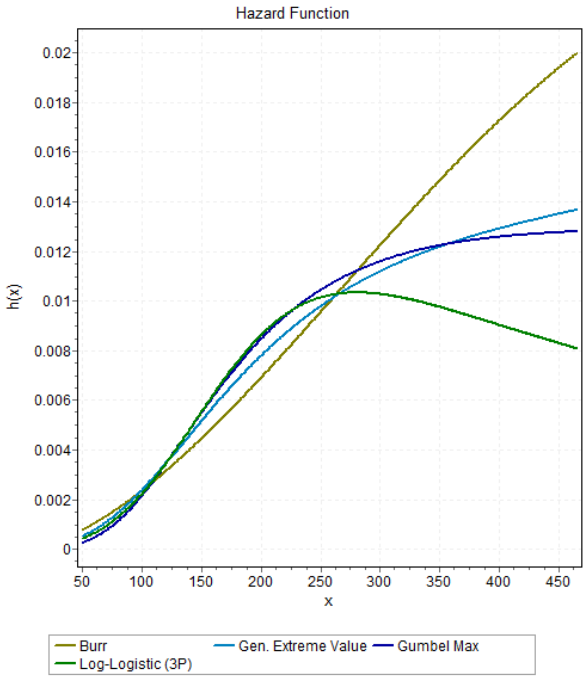
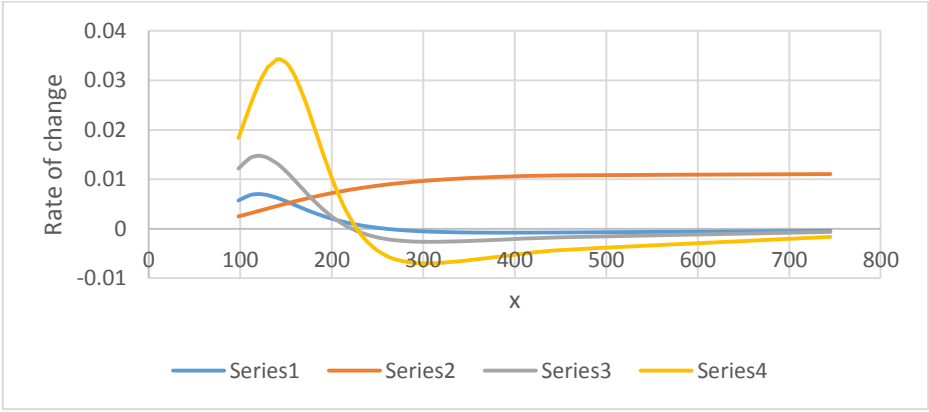
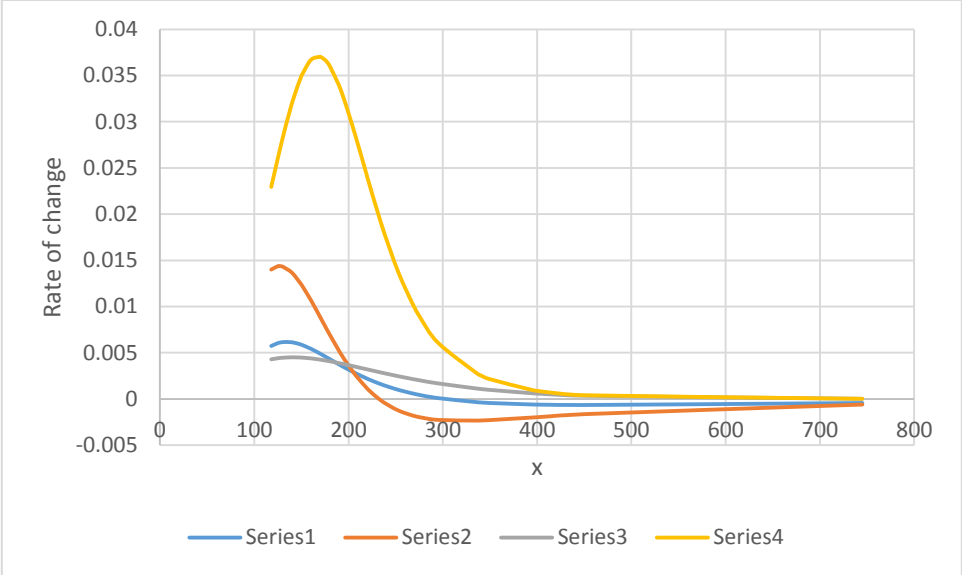


Figure C.12 Hazard rate for different distributions for 24 hour duration at station Houston Alife



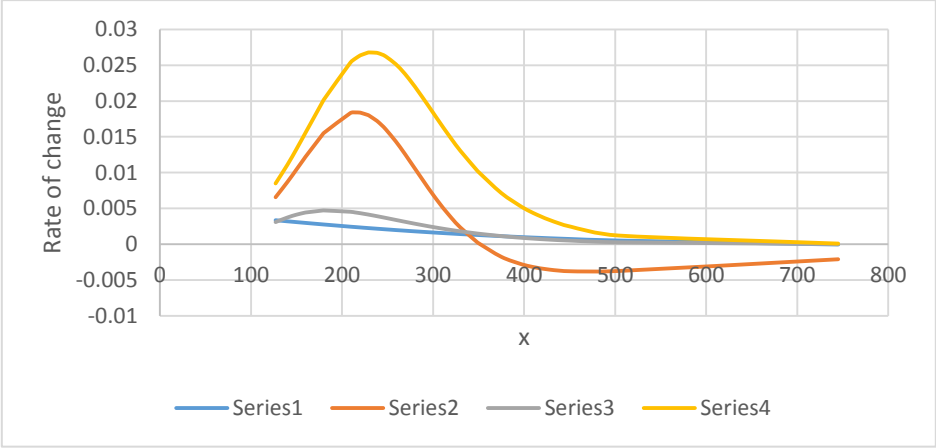
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV

Figure C.13. Rate of change of hazard rate for different distributions for 2 hour duration at station Coryell



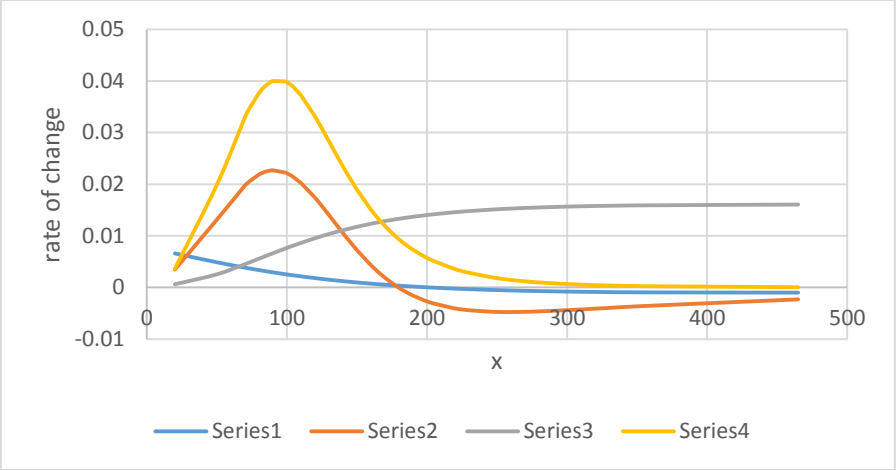
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 GEV.

Figure C.14. Rate of change of hazard rate for different distributions for 3 hour duration at station Coryell



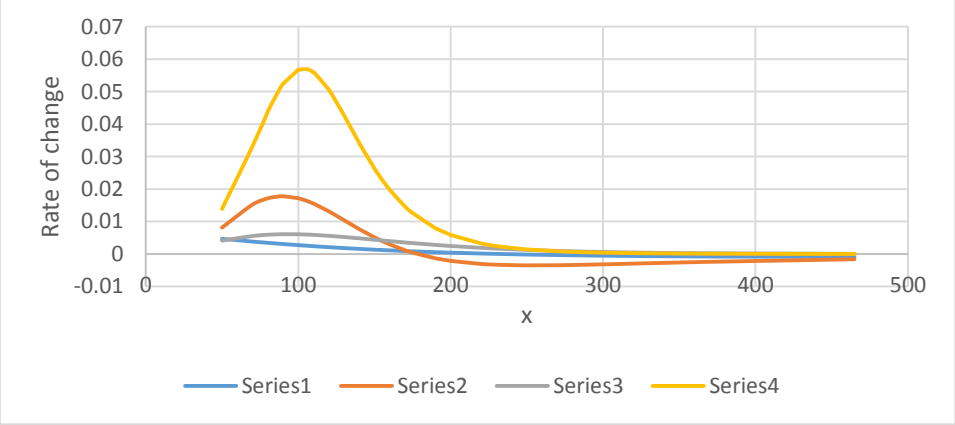
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.15. Rate of change of hazard rate for different distributions for 6 hour duration at station Coryell



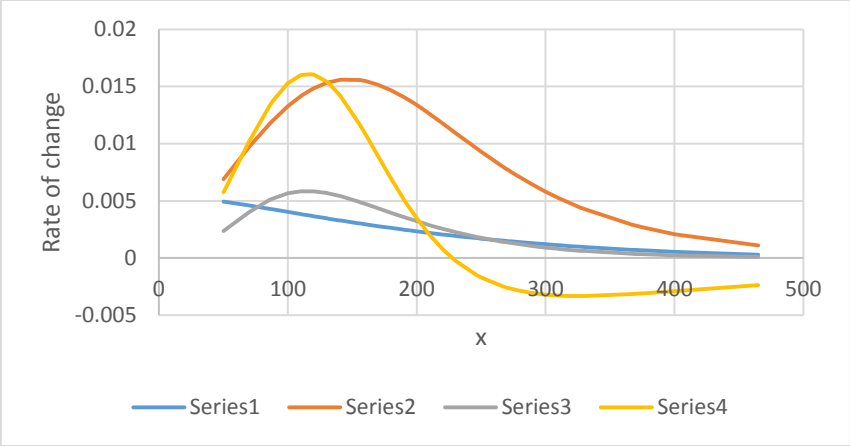
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.16. Rate of change of hazard rate for different distributions for 2 hour duration at station Lubbock



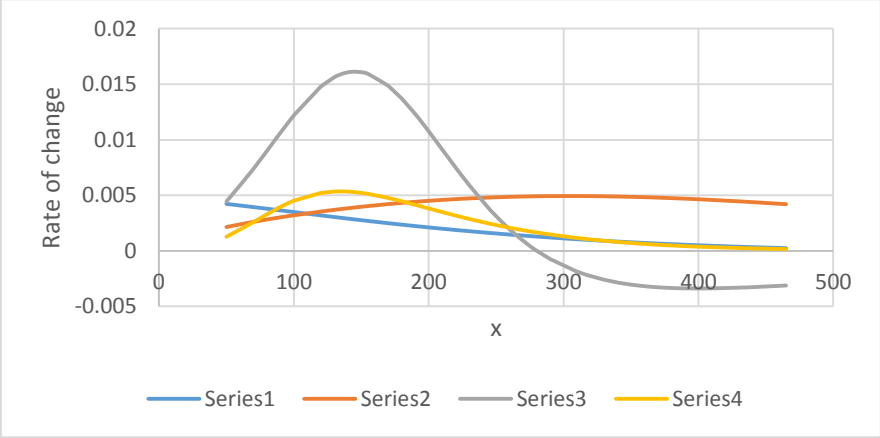
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.17. Rate of change of hazard rate for different distributions for 3 hour duration at station Lubbock



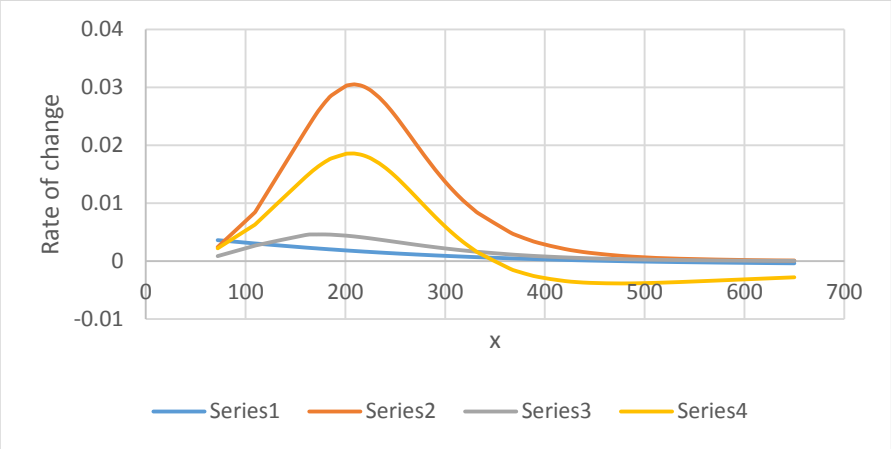
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.18. Rate of change of hazard rate for different distributions for 6 hour duration at station Lubbock



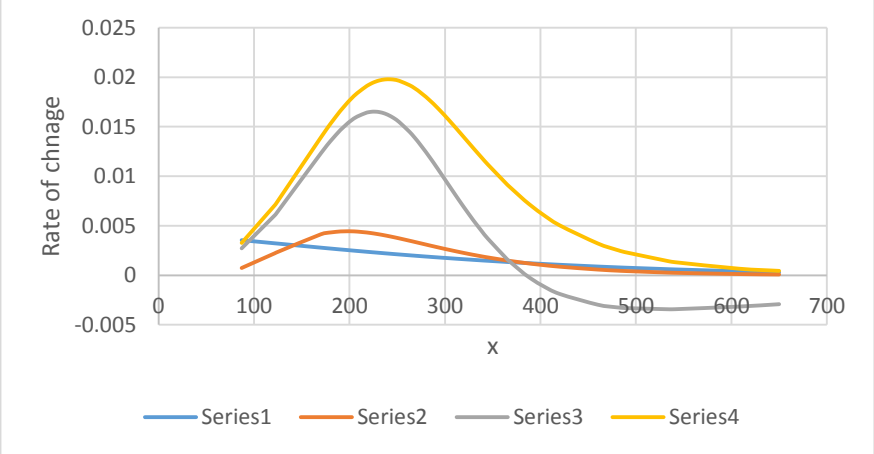
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.19. Rate of change of hazard rate for different distributions for 12 hour duration at station Lubbock



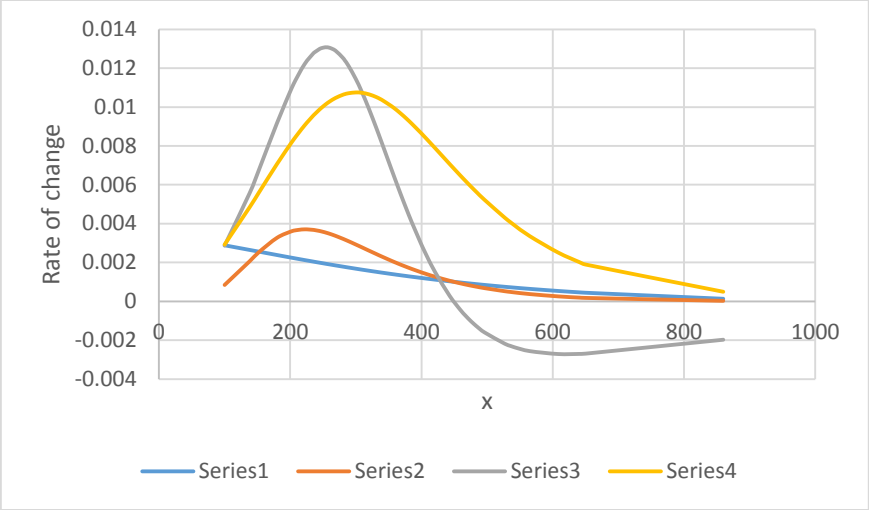
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.20. Rate of change of hazard rate for different distributions for 2 hour duration at station Houston Alife



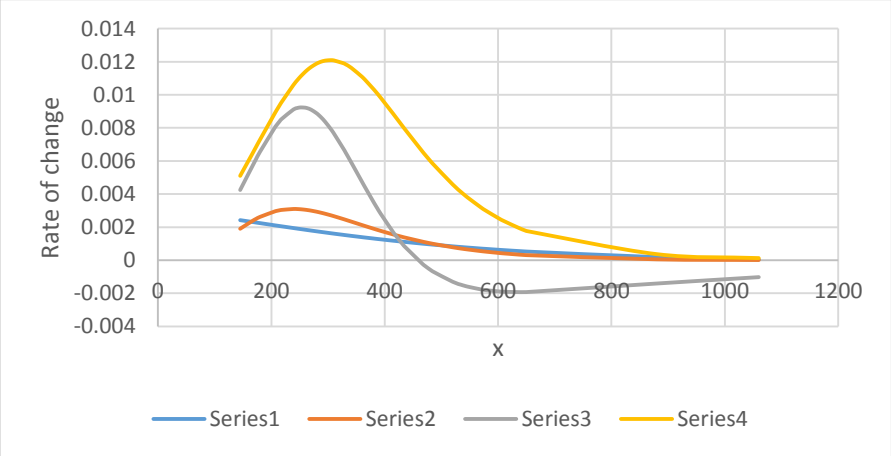
Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.20. Rate of change of hazard rate for different distributions for 3 hour duration at station Houston Alife



Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.21. Rate of change of hazard rate for different distributions for 6 hour duration at station Houston Alife



Series 1 – Burr, Series 2 - Gumbel max, Series 3 – Log logistics (3P), Series 4 – GEV.

Figure C.22. Rate of change of hazard rate for different distributions for 12 hour duration at station Houston Alife.

Table C.10 Return periods of PMP values from the best and the 4th best distribution for 1-hour duration

Return Period (years) Best Distribution	Return Period (years) from 4th Best Distribution
16666.6	15247.2
58842.3	45896.2
18586.7	17895
100000	965874
50000	50000
4398.5	3692.3
7693.2	4136.1
100000	845693.5
12500	11457.4
14285.7	1000
26588.1	19751.9
20000	18524
166666.7	25000

Table C.11 Return periods of PMP values from the best and the 4th best distribution for 2-hour duration

Return Period (years) Best Distribution	Return Period (years) from 4th Best Distribution
76923.8	85426.3
100000	853241.2
8190.6	7412.6
16666.6	13325.6
125000	136548.9
353319.4	201355
9710.8	9882.2
76929.3	76125.3
10519450	96547812
3448.2	6547.2
2068851	1698752
55157198	49875421
14285.7	13698
392834.7	336874
250000	225487.2