MODELLING PRECIPITATION VOLUMES USING A WEIBULL MIXTURE AND THE GAMMA GENERALIZED LINEAR MODEL

by

Vineet A Vora

(Under the Direction of Lynne Seymour)

Abstract

A novel approach is used to model the distribution of precipitation volumes using Weibull mixture model and gamma model as a function of Convective Available Potential Energy (CAPE). Seasonal Weibull mixture model is fit to precipitation volumes to determine the distributions of convective and stratiform precipitation for Lakewood, Fort Collins and Boulder, Colorado. This was achieved by implementing Nelder-Mead Algorithm to minimize the negative log-likelihood. We find that season is a significant factor in determining the mixture distribution. In addition, seasonal gamma regressions with log link were estimated to model the precipitation volumes as a function of CAPE and location. The models accurately predict rainfall/snowfall events with low or medium amount of precipitation in general. The Fall model also predicts events with high precipitation.

Key Words: Precipitation, Convective rainfall, Stratified rainfall, Weibull Mixture, Gamma regression, CAPE, Colorado, Nelder Mead algorithm, Kolmogorov Smirnov Test

MODELLING PRECIPITATION VOLUMES USING A WEIBULL MIXTURE AND THE GAMMA GENERALIZED LINEAR MODEL

by

Vineet A Vora

B.B.A., Temple University, 2009

M.S., Temple University, 2010

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Master of Science

Athens, Georgia

2018

MODELLING PRECIPITATION VOLUMES USING A WEIBULL MIXTURE
AND THE GAMMA GENERALIZED LINEAR MODEL

by

Vineet A Vora

Major Professor:	Lynne Seymour
Committee:	Cheolwoo Park Jaxk Reeves

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2018

Acknowledgments

Firstly, I want to express my deep appreciation to my advisor, Dr. Lynne Seymour for her overwhelming generosity of time and advice on developing this thesis. Her guidance has been an essential part of my professional development.

I would also like to thank my committee members Dr. Cheolwoo Park and Dr. Jaxk Reeves for their support and insight.

Finally, I would like to give my gratitude to my family and Haiyan Liu. Without their invaluable support, it would have been a difficult journey.

Table of Contents

List of Figures

List of Tables

Introduction

Convection is the movement caused within gases or fluids due to the transfer of heat. The transfer of heat makes the hotter and less dense material rise and the colder denser material fall. This process affects the movement in the atmosphere, oceans and even land. In the atmosphere, hot air rises on convection currents and leads to the formation of winds and clouds. In the ocean, the convection process keeps the water temperature stable throughout the ocean. On land, the convection process is the driving force for the movement of tectonic plates.

Understanding the convection process of the Earth's atmosphere is of particular importance to researchers in the field of meteorology and climatology. This thesis first attempts to identify the proportion of convective rainfall from precipitation volumes for three cities of Colorado, and then attempts to model the intensity of rainfall volume using the Convective Available Potential Energy (CAPE) measurements.

Precipitation is any form of water particles that falls from the atmosphere as rain, snow, sleet and hail. Occasionally, higher levels of precipitation are associated with the formation of extreme storms. These storms associated with heavy rainfall often lead to severe flooding and have an adverse affect on the infrastructure, landscape and ecology. Thus numerous studies have been focused on predicting the severity and occurrences of these extreme events. Precipitation can been classified into two basic types based on different cloud dynamics, convective and stratiform (Gaal et al., 2014). Convective precipitation is marked by convective cloud formations, that is vertical cloud development (cumulus congestus and cumulonimbus). Convective precipitation is mostly in the form of large intensity rainfall accompanied by thunderstorm. Sometimes, the precipitation in clouds condense very fast leading to the formation of hailstorm (Houze Jr, 2014).

Stratiform precipitation has extensive horizontal development and is from nimbostratus clouds. Nimbostratus clouds are produced by widespread (large area) vertical air movement carrying precipitation elements. These vertical air draft are relatively small and allows the precipitation elements normally to grow to the size of raindrops and snowflakes and are relatively uniform in intensity(i.e steady rain) (Qi et al., 2013).

Colorado is a state with a diverse landscape where most of it is made up of mountains, foothills and desert land. Due to the combination of mountains and the surrounding valleys, extreme weather changes including thunderstorms are quite common. Some of these extreme storms bring about catastrophic damage to infrastructure and society as seem with the Colorado flooding in September 2013. These storms are a result of convective rainfall and thus, a lot of research is done to predict the occurrence of convective rainfall, in an attempt to identify future severe storms. Rulfová and Kyselỳ (2014) examines trends in characteristics of convective and stratiform precipitation in the Czech Republic over 1982-2010. They found that in Spring and Summer, the observed increase in total precipitation are mainly due to increases in convective precipitation.

A collection of daily precipitation measurements from the Community Collaborative Rain, Hail and Snow (CoCoRaHS) network for Boulder, Fort Collins, and Lakewood, Colorado, from January 1,2005 through December 31,2014 were used by Kriebel (2016) to estimate precipitation volumes for these cities. The precipitation volumes were computed only for the days when all weather stations in a city recorded rainfall. Thereby we have a total of 601, 416 and 629 observations of precipitation volumes from Boulder, Fort Collins, and Lakewood respectively.

Convective Available Potential Energy (CAPE) is the maximum buoyancy of an undiluted air parcel (a block of air), related to the potential updraft strength of thunderstorms.That is, it measures the amount of precipitation present in a parcel of air. The higher the CAPE (precipitation present in the air parcel) the higher the chances of a thunderstorm, thereby leading to convective rainfall. A weather balloon is launched at KDNR facility in Denver

Colorado, which records the CAPE measurements. CAPE values are measured in Joules per kilogram (J/kg). To put it in perspective, when the CAPE value is less than 1000 J/kg, it is “weakly convective”, for CAPE values from 1000 to 2500 J/kg it is “moderately convective” and with more than 3000 J/kg, it is “highly convective” which leads to tornado and damaging thunderstorm. The weather balloon measure three different types of CAPE:

1. Surface Based CAPE (SBCAPE) uses the surface air and dewpoint temperatures to determine the convection in the parcel of air.

2. Mixed Layer (MLCAPE) is calculated by averaging temperature and moisture variables in the lowest 100 mb (millibars) of atmospheric pressure.

3. Most Unstable CAPE (MUCAPE) is the parcel trajectory that produces the largest value of CAPE. To measure this value, a parcel is lifted from the large number of pressure surfaces and the trajectory that produces the maximum amount of J/kg is the MUCAPE.

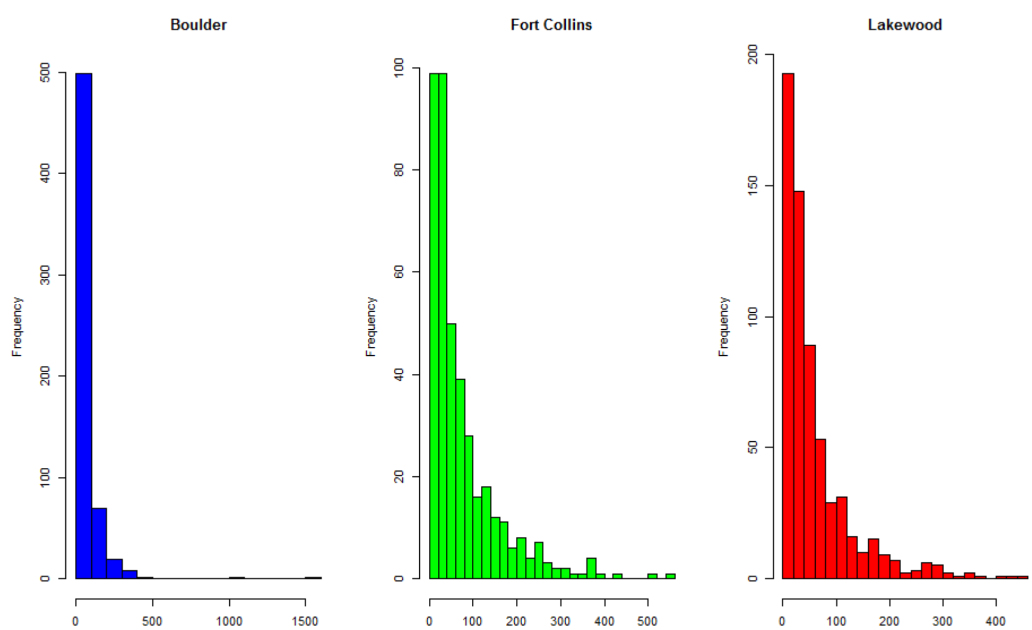

Figure 1.1: Histogram of Precipitation volumes by Season

The weather balloon released from the official weather station in Denver (KDNR) records the measurements of all three CAPE types twice a day at 00 AM (midnight) and 12 PM(afternoon) Greenwich Mean Time (GMT). This means that 00 AM and 12 PM weather balloon launch on 5/15/2017 was actually launched around 6∼7 PM (based on daylight saving) on 5/14/2017 and 6∼7 AM on 5/15/2017 respectively. Thus we have a total of 6 readings per day SBCAPE0, MLCAPE0 and MUCAPE0 measurements taken at 6∼7 PM and SBCAPE12, MLCAPE12 and MUCAPE12 measurements taken at 6∼7 AM. It is justifiable to use one set of CAPE measurements for all three locations, as Lakewood, Fort Collins and Boulder are within 60 mile radius of KDNR.

Figure 1.1 provides the histogram for the precipitation volumes from the three cities. All three histograms are skewed right with high density of values being close to 0. Furthermore, there are large outliers to the right for each histogram. These huge precipitation volumes are from September, 2013 when Colorado experienced historical rainfall with widespread flooding.

Figure 1.2 displays the seasonal precipitation volume for each city. The two highest observed volumes with more than 10 billion liters for Boulder Fall season are from Sept 2013 floods. Figure 1.3 displays the seasonal precipitation volumes for each city without the Sept 2013 flood volumes. Winter volumes are the lowest with the least amount of variability. Spring and Fall volumes have the highest variability of precipitation volumes. In terms of cities' precipitation volumes, Fort Collins is generally lower followed by Boulder and Lakewood with the highest variability. Thus it can be seen that seasonal and location effects are significant for precipitation volumes.

Similar to precipitation volume, all the six CAPE measurements are right skewed with most of the values less than 100 which is seen from Figure 5.39 and 5.40. We explore further by looking at the boxplots of each CAPE by season in Figures 1.4,1.5 and 1.6. The CAPE values differ a lot between seasons. Lowest amount of convection is present in CAPE during Winter season, followed by Fall, Spring and lastly Summer. Summer is highly volatile

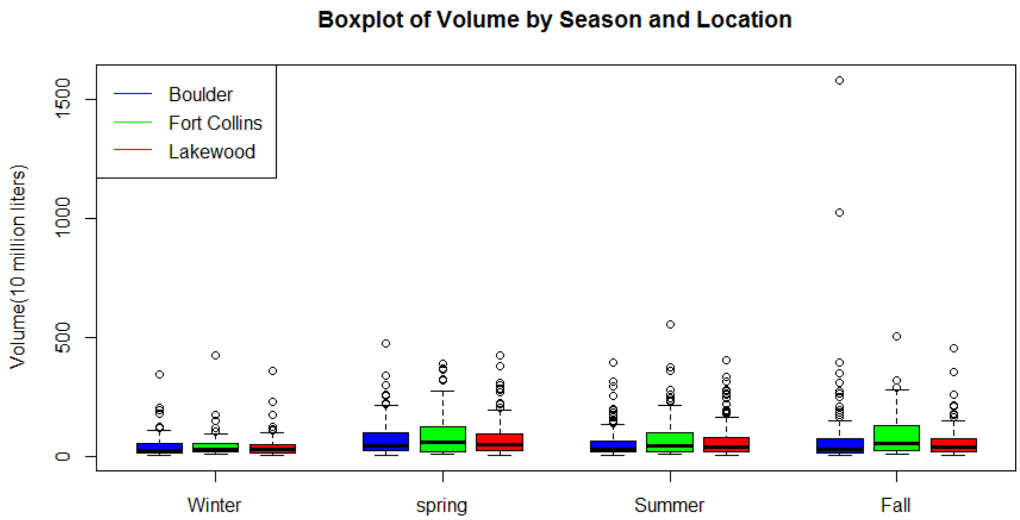

Figure 1.2: Boxplot of Precipitation volume by Season and Location

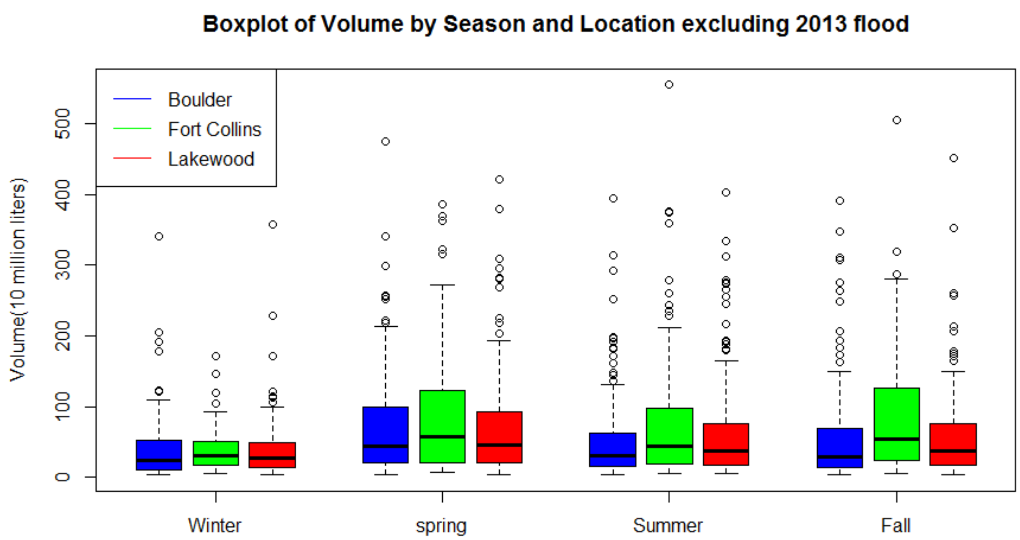

Figure 1.3: Boxplot of Volume by Season and Location excluding Colorado Flood,2013

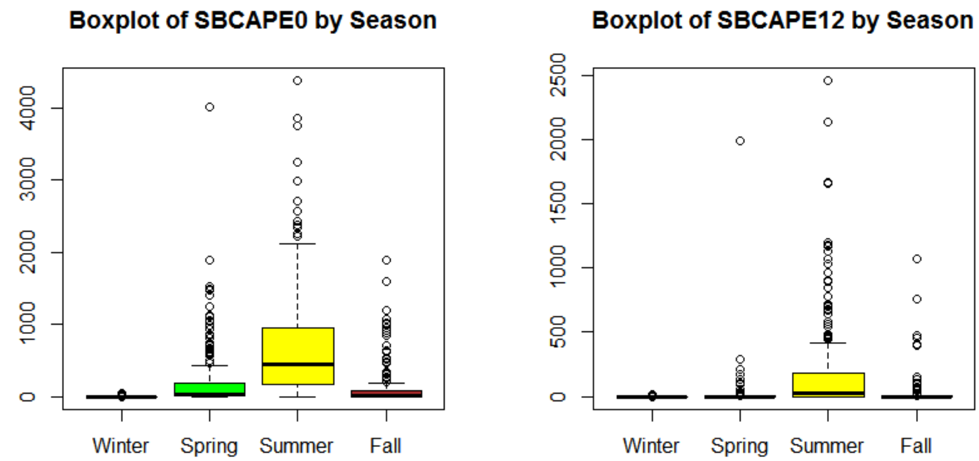

Figure 1.4: Boxplot of SBCAPE by Season

and has the highest amount of convection present in the parcel of air of all seasons. Another thing of importance is that CAPE0 measurements are higher than their counterpart CAPE12 measurements for all three seasons. The rationale for this difference is that, CAPE0 measurements are recorded at evening time after the atmosphere has been heated all day whereas CAPE12 measurements are recorded in morning, when the atmosphere has been cooling all night. This indicates that there is a higher chance of thunderstorms and convective rainfall to occur at night times compared to day time.

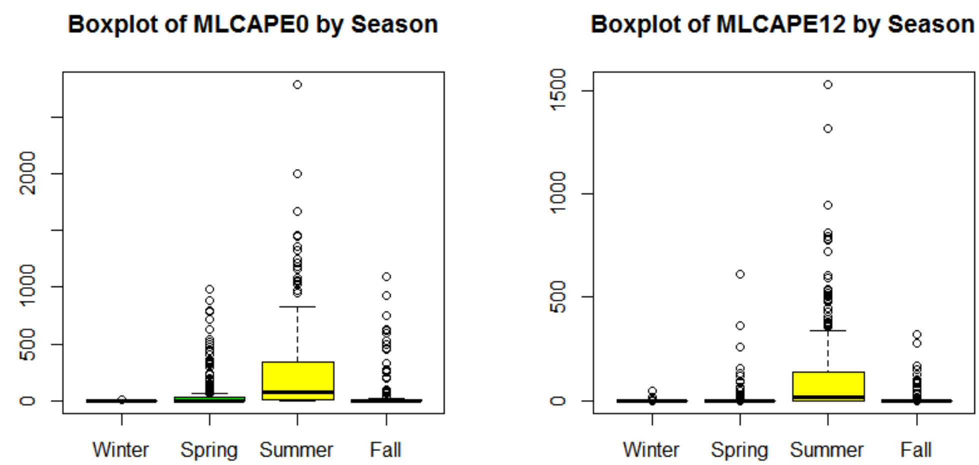

Figure 1.5: Boxplot of MLCAPE by Season

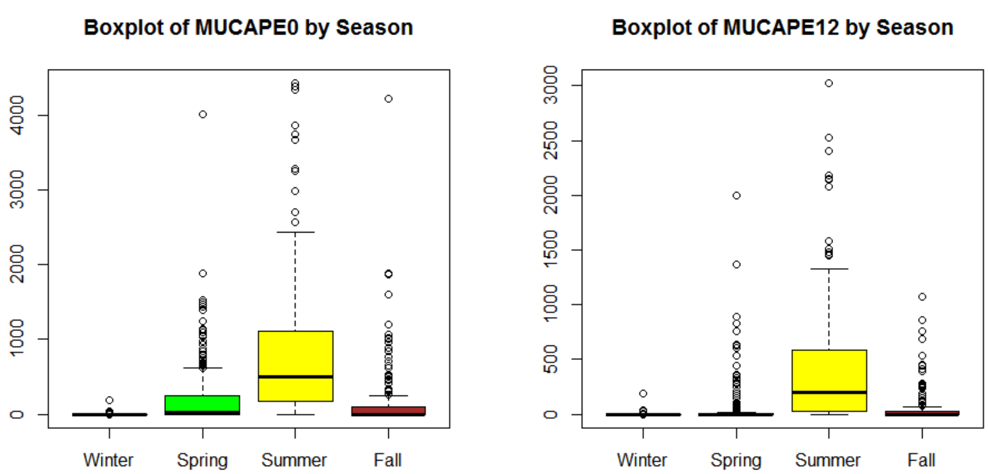

Figure 1.6: Boxplot of MUCAPE by Season

2.1 Method

The goal of this research is to model the distribution of precipitation volume that results from convection. We are particularly interested in the likelihood that a large volume of precipitation will be released during a storm which can lead to catastrophic damage to the infrastructure. We know that precipitation volumes are classified into two namely convective rainfall and stratiform rainfall. Thus we decided to fit a mixture distribution for precipitation volume comprised of two distributions. The precipitation volumes are positive and right skewed. For the distribution of the components we have decided to use the flexible and versatile Weibull distribution which can take on characteristics and shape of other distributions based on the value of its shape parameter.

$$f_j(x|\theta_j) = \frac{\alpha_j}{\beta_j}\left(\frac{x}{\beta_j}\right)^{\alpha_j - 1} e^{(x/\beta_j)^{\alpha_j}} \qquad \alpha_j, \beta_j > 0 \tag{2.1}$$

where α_j is the shape parameter of the Weibull distribution. When the α is 1, it becomes an exponential distribution and when α is 3, it becomes a normal distribution. This can be seen in Figure 2.7. β_j is the scale parameter which determines the variability in the distribution.

A mixture distribution is defined as a weighted sum of component distributions:

$$f(x|\theta) = \sum_{j=1}^{m} p_j f_j(x|\theta_j) \tag{2.2}$$

where the parameter $\theta = (p_1,...,p_j,\ \theta_1,...,\theta_j)$ are such that $p_j > 0$ for $j = 1,...,m$ and $\sum_{j=1}^{m} p_j = 1$. Constant p_j is called a weight and f_j is the component density function. In one set of our analyses, in which we categorize precipitation into "convective" and "stratiform",

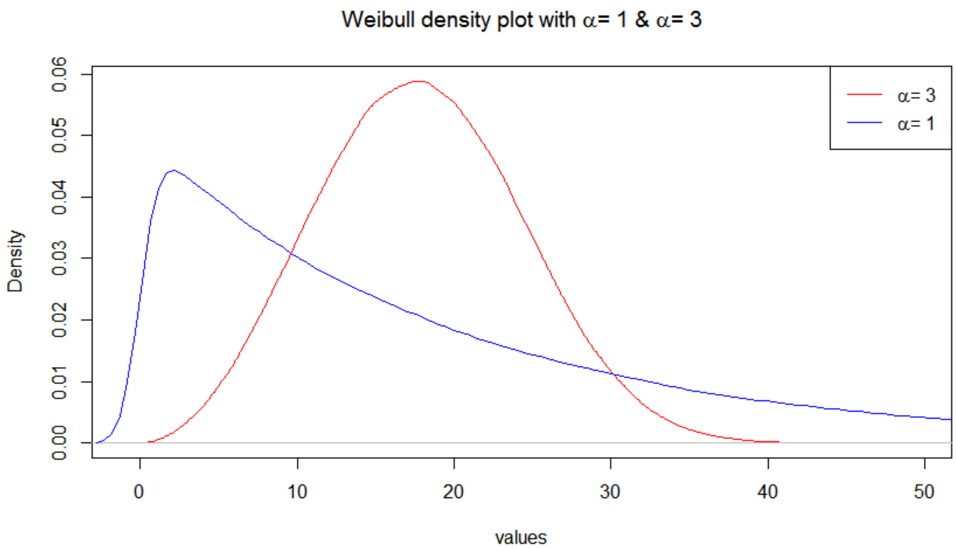

Figure 2.7: Weibull Density Plot with $\alpha = 1$ & $\alpha = 3$

$j = 1, 2$. Thus the distribution used to model precipitation volume is:

$$f(x|\theta) = p_1 * f_1(x|\theta_1) + (1 - p_1) * f_2(x|\theta_2) \tag{2.3}$$

which is

$$f(x|\theta) = p_1 \frac{\alpha_1}{\beta_1} \left(\frac{x}{\beta_1}\right)^{\alpha_1 - 1} e^{(x/\beta_1)^{\alpha_1}} + (1 - p_1) \frac{\alpha_2}{\beta_2} \left(\frac{x}{\beta_2}\right)^{\alpha_2 - 1} e^{(x/\beta_2)^{\alpha_2}} \tag{2.4}$$

where $\theta = (p_1, \alpha_1, \beta_1, \alpha_2, \beta_2)$

In order to estimate the parameters of the convective rainfall we try to maximize the likelihood of the precipitation volumes. This is achieved by finding the parameter values θ using an optimization technique, to maximize the probability or likelihood of getting the data we observed. The likelihood of the mixture model is:

$$L(\theta/x) = \prod_{i=1}^{n} \left[p_1 \frac{\alpha_1}{\beta_1} \left(\frac{x}{\beta_1}\right)^{\alpha_1 - 1} e^{(x/\beta_1)^{\alpha_1}} + (1 - p_1) \frac{\alpha_2}{\beta_2} \left(\frac{x}{\beta_2}\right)^{\alpha_2 - 1} e^{(x/\beta_2)^{\alpha_2}} \right] \tag{2.5}$$

The log likelihood is:

$$l(\theta/x) = \sum_{i=1}^{n} log[(p_1)\frac{\alpha_1}{\beta_1}\left(\frac{x}{\beta_1}\right)^{\alpha_1 - 1} e^{(x/\beta_1)^{\alpha_1}} + (1-p_1)\frac{\alpha_2}{\beta_2}\left(\frac{x}{\beta_2}\right)^{\alpha_2 - 1} e^{(x/\beta_2)^{\alpha_2}}] \tag{2.6}$$

To find the parameters θ we use the FMINSEARCH function in MatLab which uses the Nelder Mead optimization algorithm to minimize the log likelihood.

2.1.1 Nelder Mead Algorithm

The Nelder-Mead algorithm provides a means of minimizing an objective function of dimension n. The algorithm iterates on a simplex, a geometric figure in n dimensions that is the convex hull of a $n + 1$ vertices. In a 2 dimensional plane, the simplex consists of 3 points forming a triangle, and in 3 dimensions the simplexs consist of 4 points forming a tetrahedron. It specifies a sequence of steps for iteratively updating the worst point in the simplex in order to converge to the smallest value of the object function (Gavin, 2013). Wright (2012) shows the different ways a new point on a simplex is generated on $\mathcal{R}^2$.

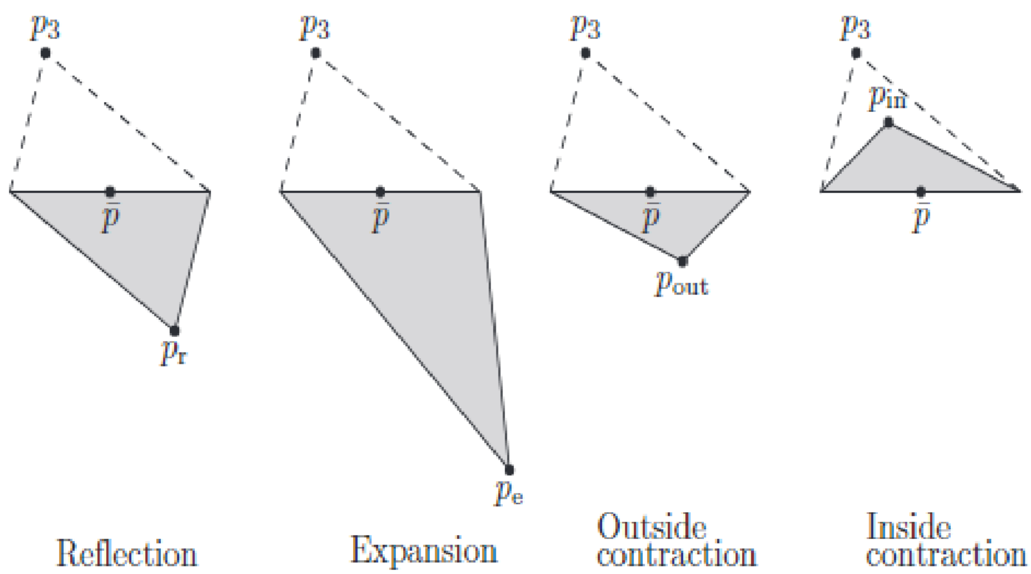

Figure 2.8: Nelder Mead Algorithm different ways to generate new point

The steps for a single iteration of Nelder Mead Algorithm (Byatt, 2000) are as follows.

1. Create an initial simplex in $\mathcal{R}^n$. The initial estimate is used as one of the vertices of the simplex. The n remaining vertices of the simplex are found by perturbing each of the coordinates of the initial guess. The perturbation used is:

   (a) If the coordination is non-zero then set the perturbated coordinate to 105% of its current value; else

   (b) If the coordinate is zero then set the perturbed coordinate to 0.00025.

2. Evaluate the objective function at each of the $n + 1$ vertices of the simplex.

3. Order the vertices $v_0, v_1, , v_n$ of the current simplex so that
   $f(v_0) \leq f(v_1) \leq ... \leq f(v_n)$

4. Calculate the reflect point, x_r of the worst performing vertices v_n.

5. Accept the reflect point if $f(v_0) \leq f(x_r) \leq f(v_{(n-1)})$.

6. If $f(x_r) < f(v_0)$, then calculate the expansion point.

7. If $f(x_r) \geq f(v_{(n-1)})$ then perform a contraction:

   (a) If $f(v_{(n-1)}) \leq f(x_r) \leq f(v_n)$ then contract outside; else

   (b) If $f(v_n) \leq f(x_r)$ then contract inside.

By iterating this process several times, the Nelder-Mead algorithm finds the parameters that minimize the negative likelihood. A custom MatLab code is written to find the parameters for the mixture distribution of precipitation volume for each location and season combination. Kolmogorov-Smrinov goodness of fit test was implemented to check if the final distributionv fit the precipitation volume data well.

2.2 Data Analysis

The parameters of Weibull mixture are estimated for precipitation volumes for each combination of location and season. We first look at the Spring season as it has the most number of observations in all the 3 locations. Figure 2.9 shows the Weibull mixture model for Lakewood Spring season. The top plot in Figure 2.9 shows the histogram of precipitation volume in blue while the red and green lines are the estimated Weibull distributions of stratiform rainfall and convective rainfall respectively. The red distribution shows that most of the rainfall with less than 1 billion liters comes from stratiform rainfall. Rainfalls with volumes larger than 1 billion liters all come from convective rainfall. Generally, rainfall with thunderstorms (convective rainfall) leads to longer duration heavy downpours and occasionally it leads to a short duration quick downpour. These occurrences have been captured perfectly by the Weibull distribution of precipitation data.

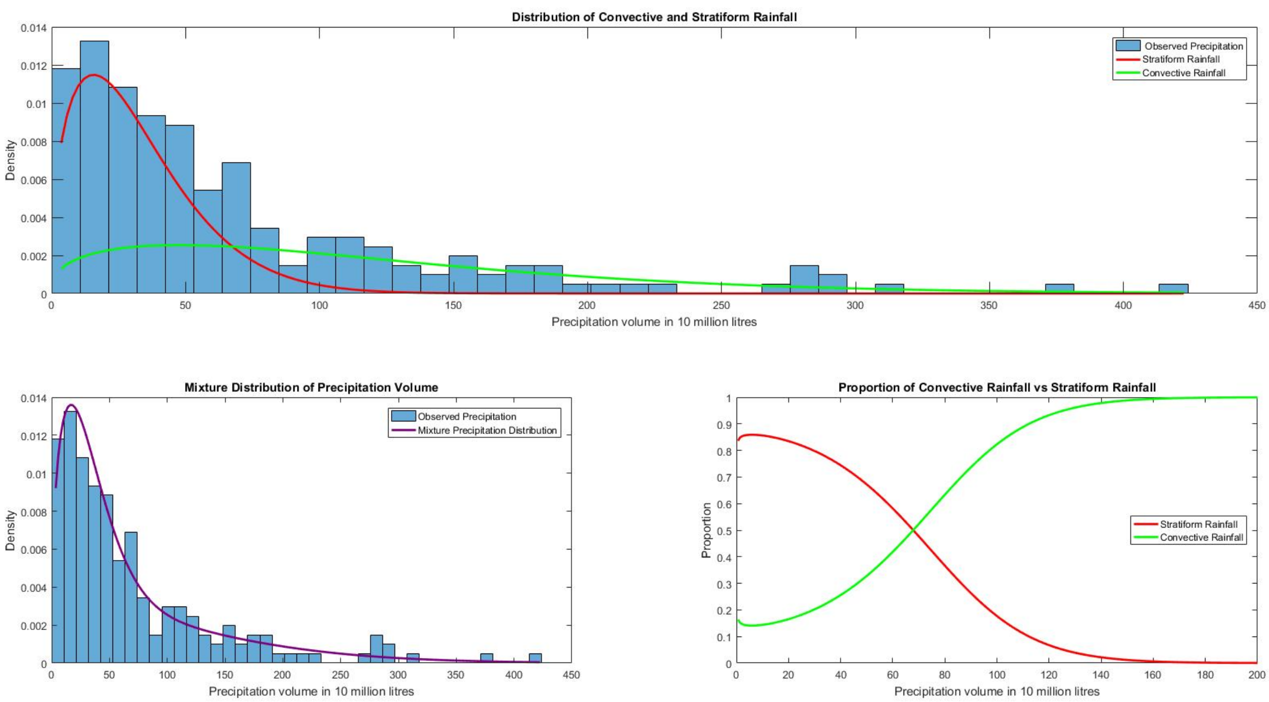

Figure 2.9: Lakewood Spring Mixture model for Precipitation Volumes

The bottom left plot in Figure 2.9 shows the overall fit of the mixture precipitation data. It can be seen that the Weibull mixture fits the data accurately except it underestimates volume at 2.5 billion litres. The bottom right plot shows the proportion of stratiform and convective rainfall for a given rainfall volume. For a rainfall of less than 670 million liters the odds of it being a stratiform rainfall are higher than convective rainfall and for a rainfall of more than 670 million liters the odds of it being a convective rainfall are higher than stratiform rainfall.

In order to test the goodness of fit of the mixture distribution, the Kolmogorov-Smirnov (KS) test is implemented. Since no weather stations keep records for a rainfall being convective or stratiform, we use the average MLCAPE (Mixed Layer Convective Available Potential Energy) measurement obtained as described in the introduction. MLCAPE measurements tells us if there is any precipitation found in a parcel of air. If the MLCAPE value is 0, then there is no precipitation present in the air. When the MLCAPE value is more than 3000 tornadoes are likely. To check the goodness of fit, we have used the most stringent constraint, MLCAPE values of more than 0 as convective vs MLCAPE values of 0 as stratiform rainfall.

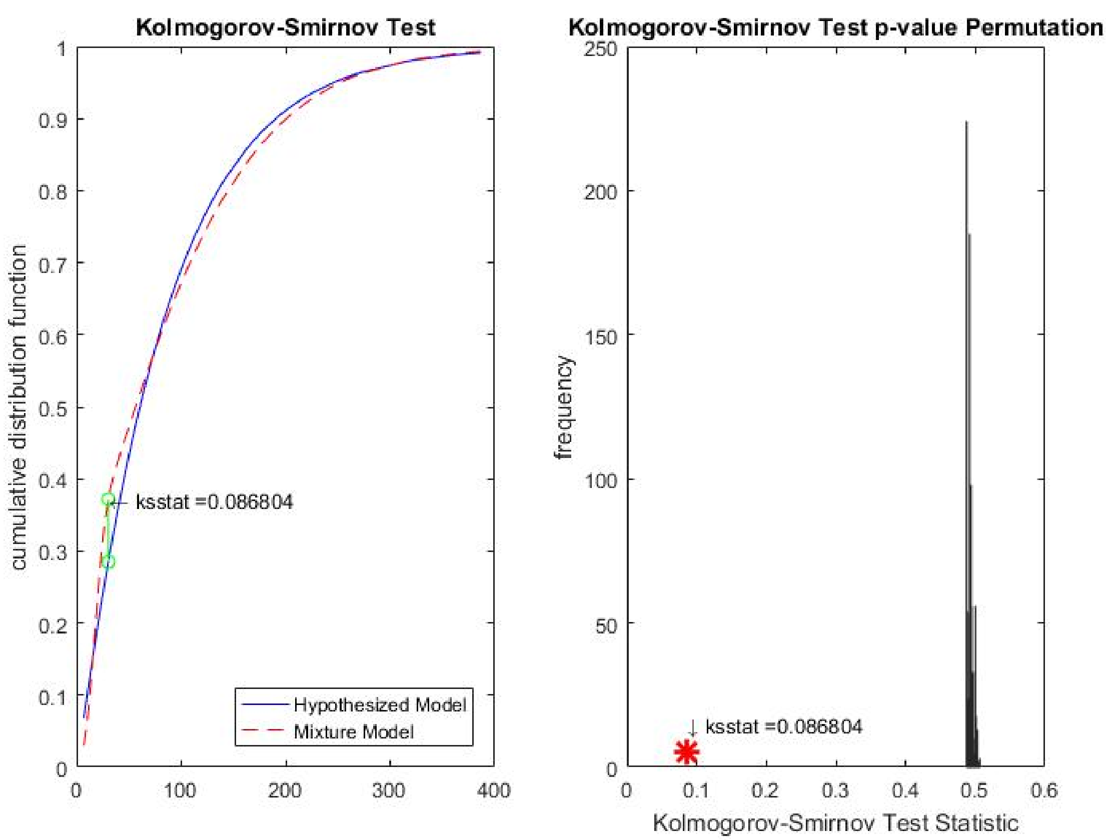

Figure 2.10: Kolmogorov Smirnov Test for Lakewood Spring Mixture

The null hypothesis for the KS test states that the distributions of MLCAPE measurements and the Weibull mixture model are the same versus the alternate hypothesis that the two distributions are different. The left plot on Figure 2.10 shows the KS test for Lakewood Spring season. The blue line is the hypothesized model derived from MLCAPE and red dotted line is the estimated Weibull mixture model. The two distribution lines are almost similar with the highest distance between the two lines at KS test value of 0.0526. The p-value is computed by building a sampling distribution of KS test statistic using exact (permutation) test. The right graph of Figure 2.10 shows the sampling distribution of KS statistic, which are computed by randomly assigning the observations to convective precipitation and stratified precipitation distributions without replacement. The red star is the original KS test statistic computed from the MLCAPE data. Since all of the sampling distributions KS statistics exceeds the red star, we obtain a p-value of 1.00. Therefore, we can conclude that the two distributions are not different from each other. Thus, mixture model correctly estimates the distribution of convective precipitation and stratiform precipitation.

Similar are obtained using the Weibull mixture model for all the other combination of location and season as seen in Figure 2.11 to Figure 2.32. Based on these Figures, one can see that Fort Collins has the highest proportion of convective precipitation compared to other locations. Lakewood and Boulder have similar precipitation proportions for all seasons. The most important finding of this exercise is that the seasonal effects are similarly for all locations. For Winter seasons one expects the number of precipitation events to be lowest of all seasons, as the air is cold and cannot contain much water vapor. This pattern is captured by the Weibull mixture models and can be seen in Figures 2.11, 2.17 and 2.25. The highest proportion of convective precipitation is observed during the Spring season. The proportion of convective precipitation for Summer and Fall season is similar.

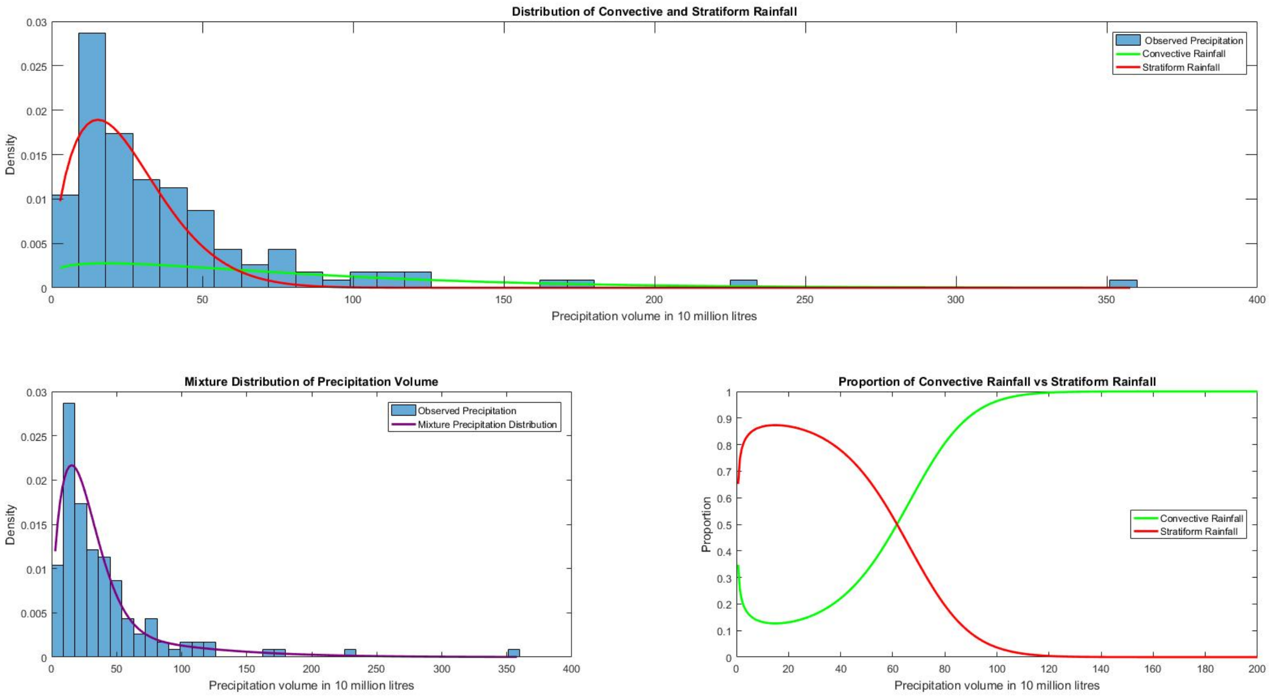

Figure 2.11: Lakewood Winter Mixture model for Precipitation Volumes

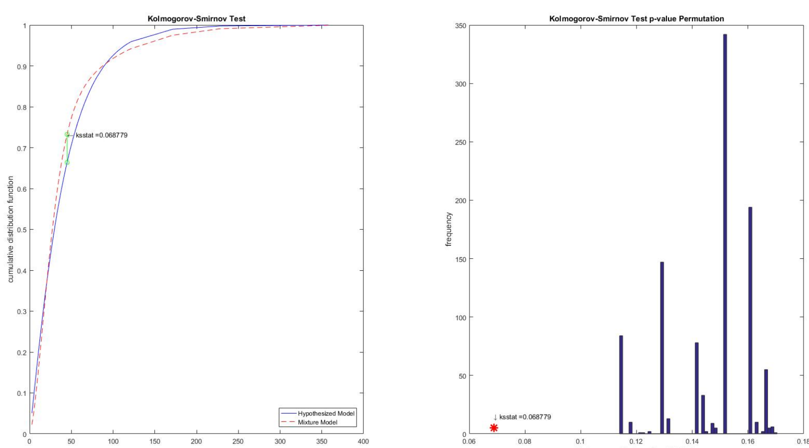

Figure 2.12: Kolmogorov Smirnov Test for Lakewood Winter Mixture

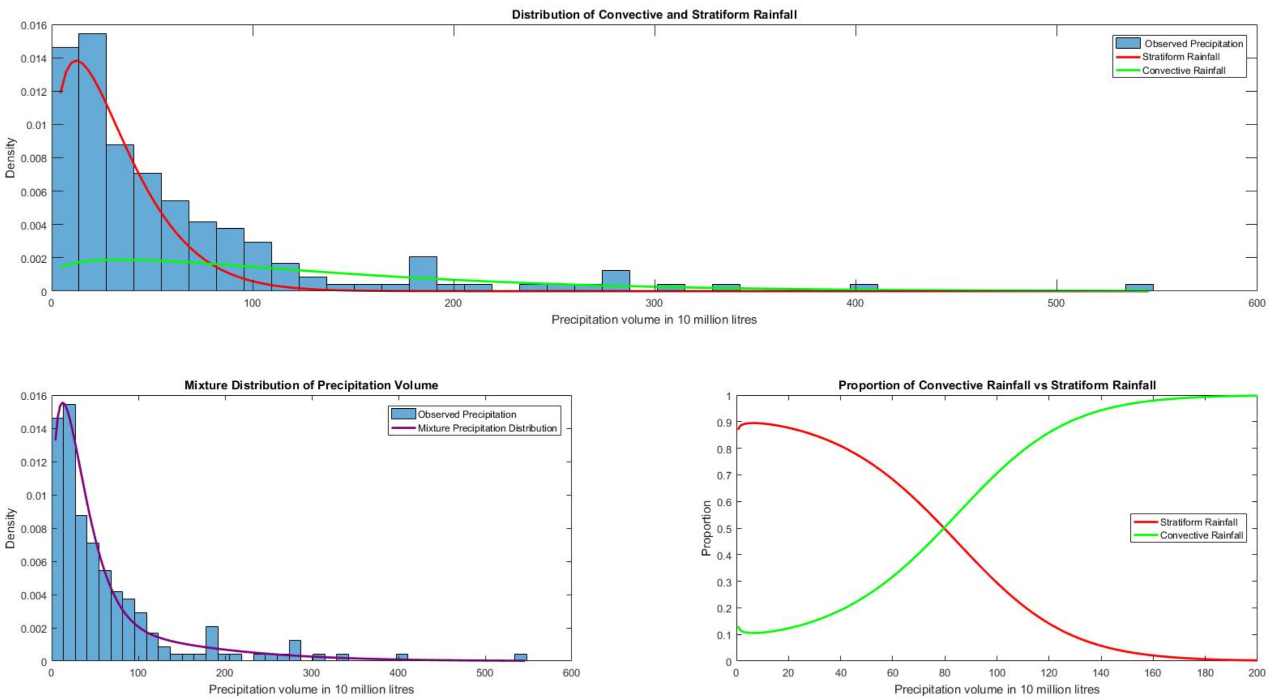

Figure 2.13: Lakewood Summer Mixture model for Precipitation Volumes

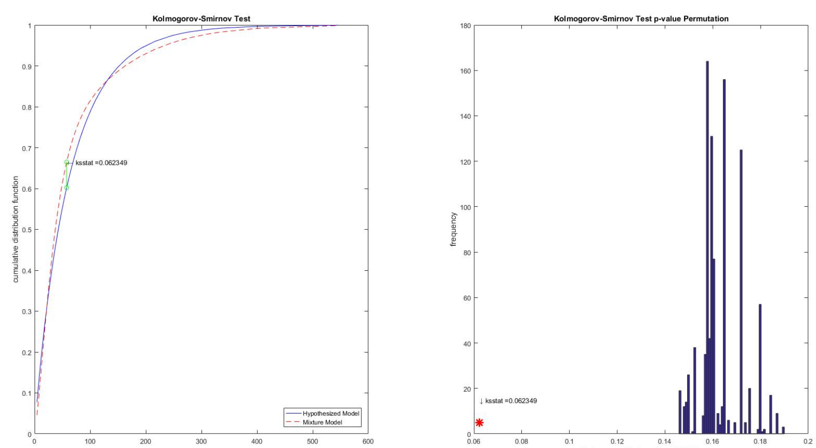

Figure 2.14: Kolmogorov Smirnov Test for Lakewood Summer Mixture

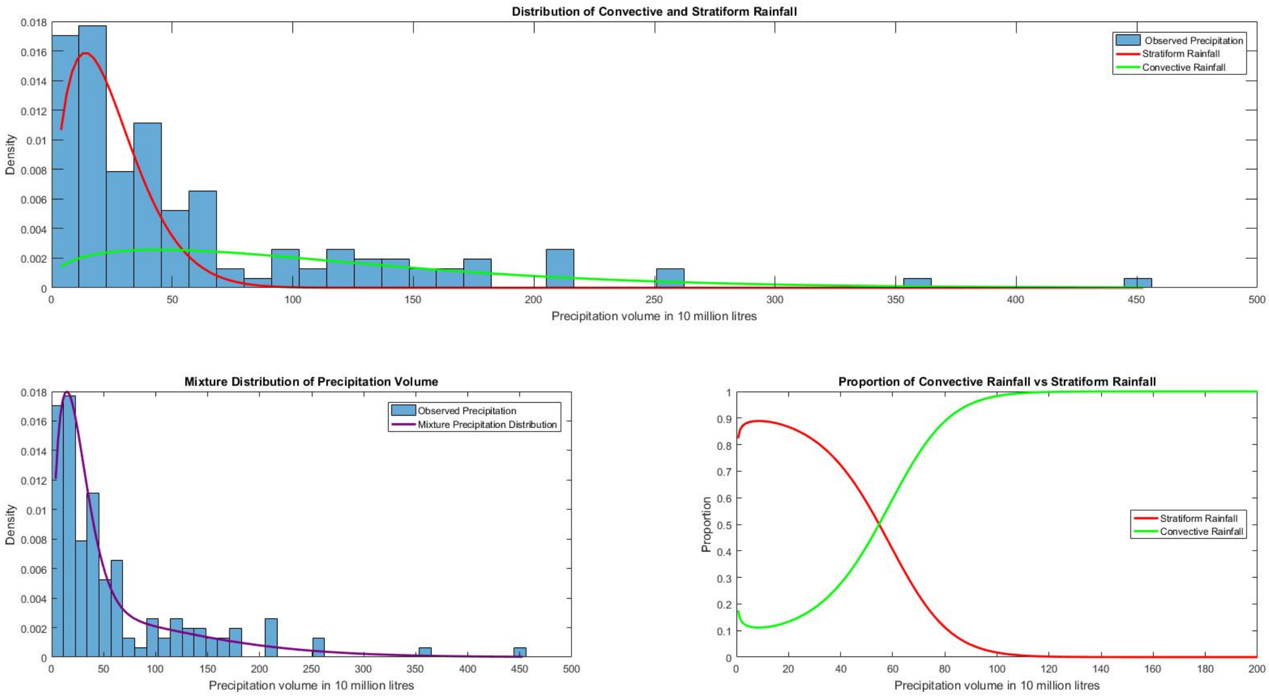

Figure 2.15: Lakewood Fall Mixture model for Precipitation Volumes

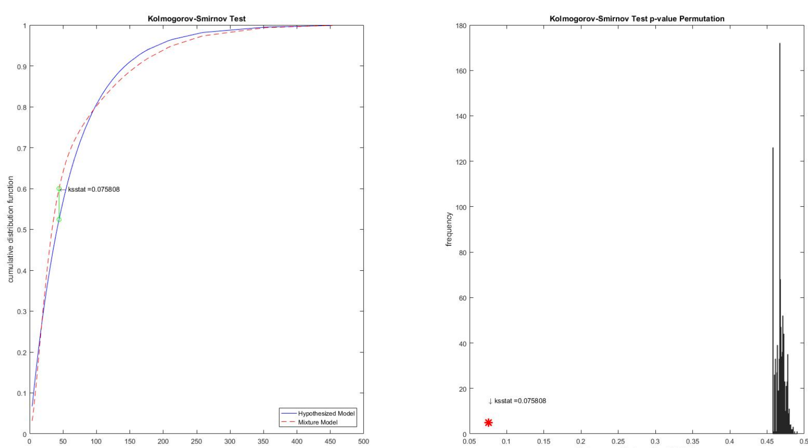

Figure 2.16: Kolmogorov Smirnov Test for Lakewood Fall Mixture

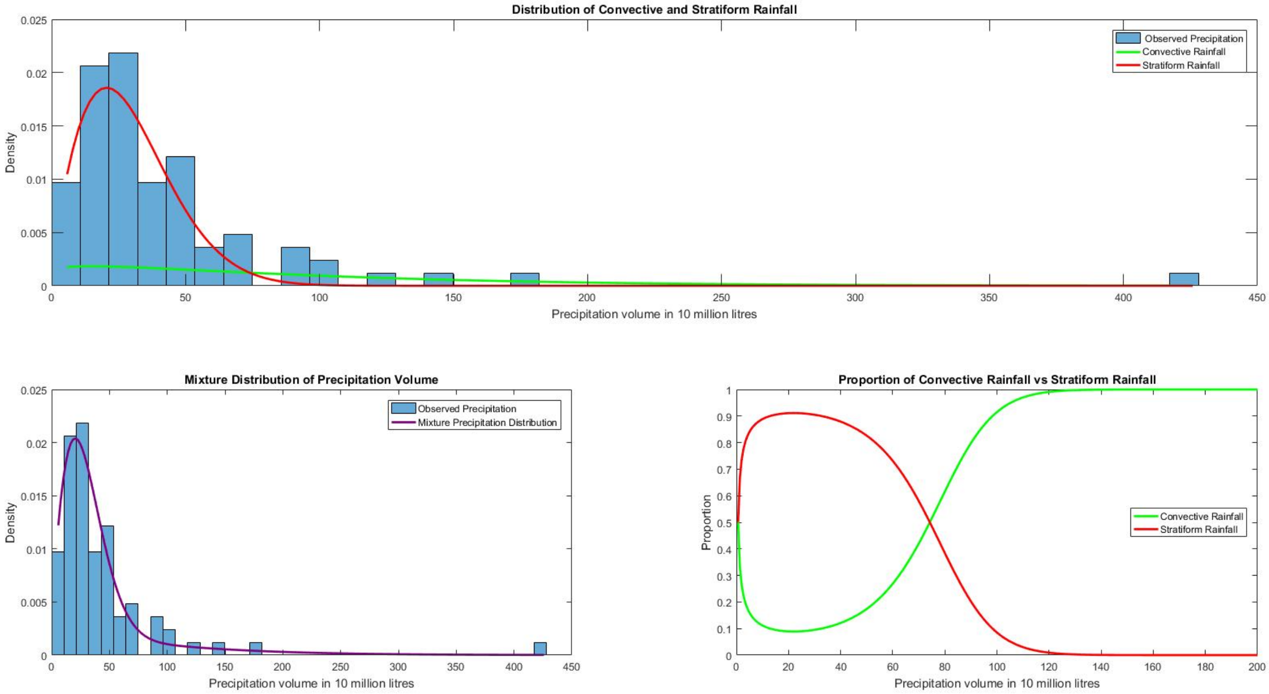

Figure 2.17: Fort Collins Winter Mixture model for Precipitation Volumes

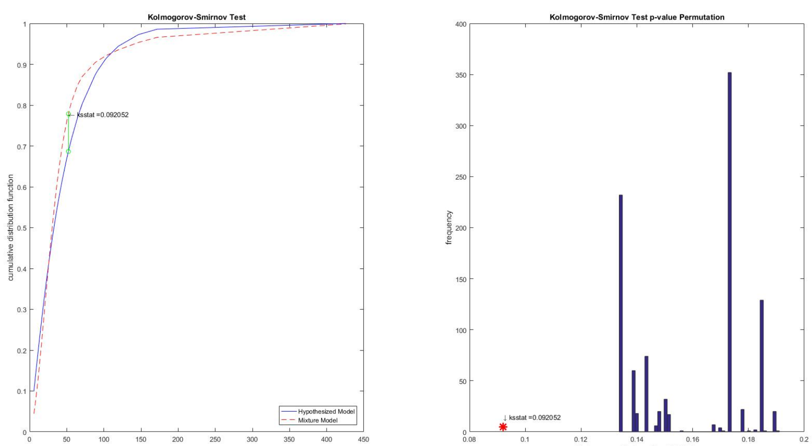

Figure 2.18: Kolmogorov Smirnov Test for Fort Collins Winter Mixture

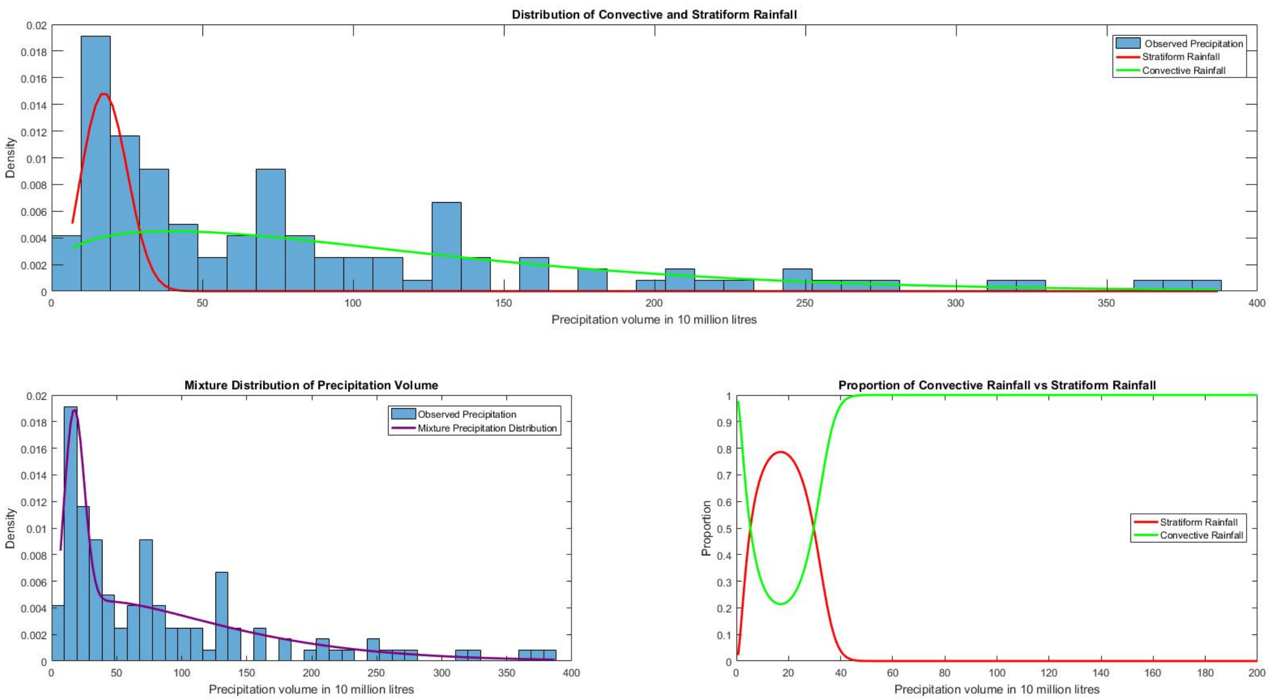

Figure 2.19: Fort Collins Spring Mixture model for Precipitation Volumes

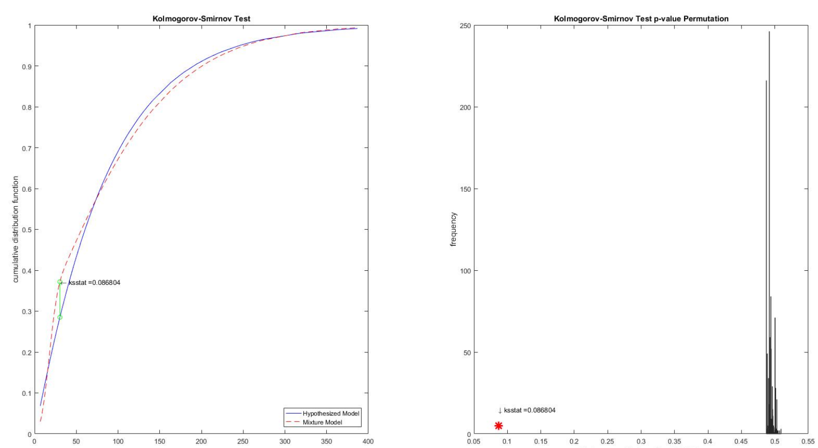

Figure 2.20: Kolmogorov Smirnov Test for Fort Collins Spring Mixture

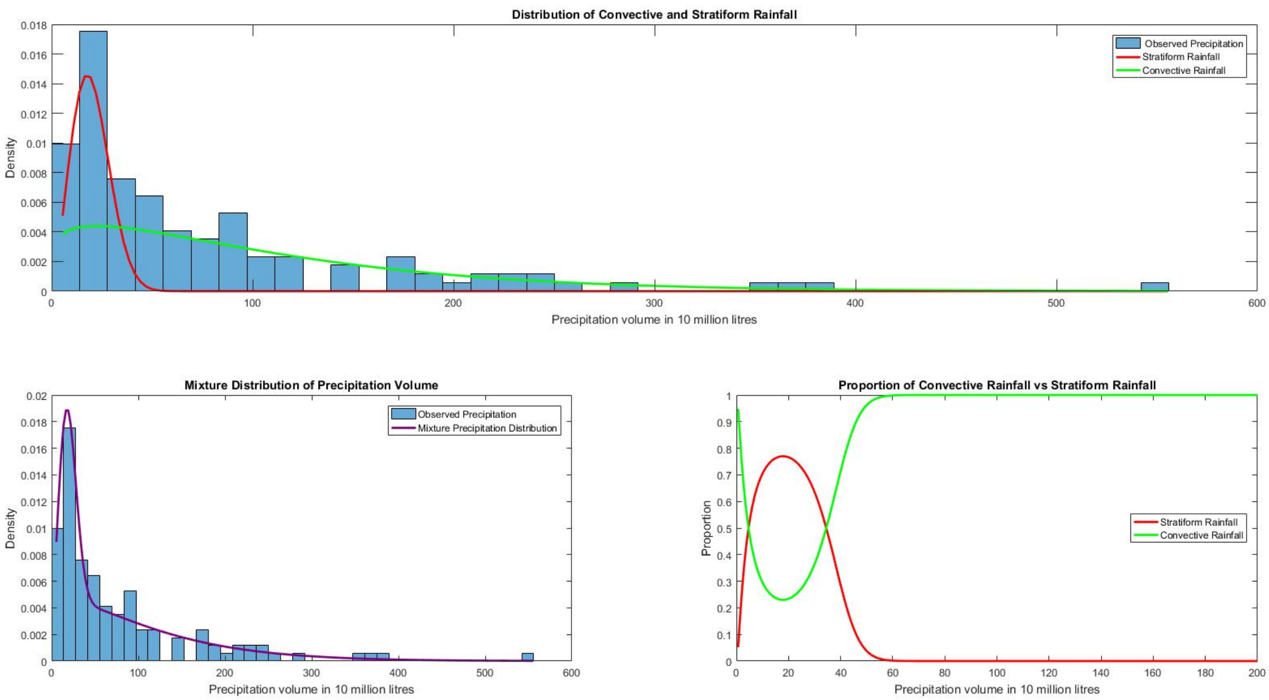

Figure 2.21: Fort Collins Summer Mixture model for Precipitation Volumes

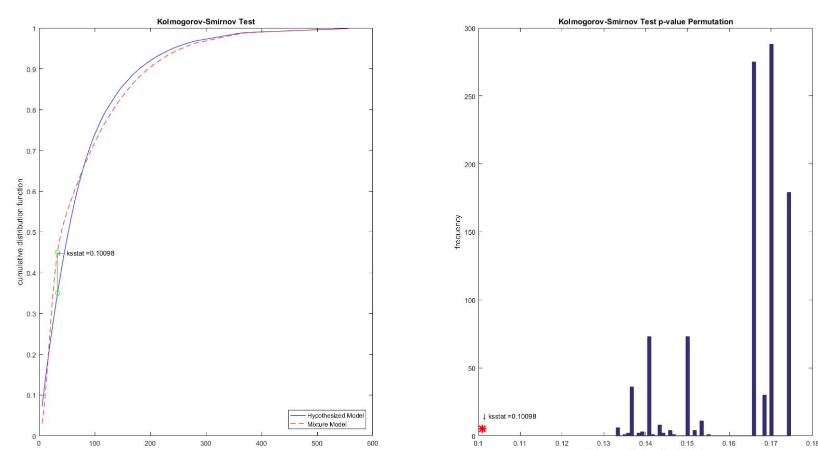

Figure 2.22: Kolmogorov Smirnov Test for Fort Collins Summer Mixture

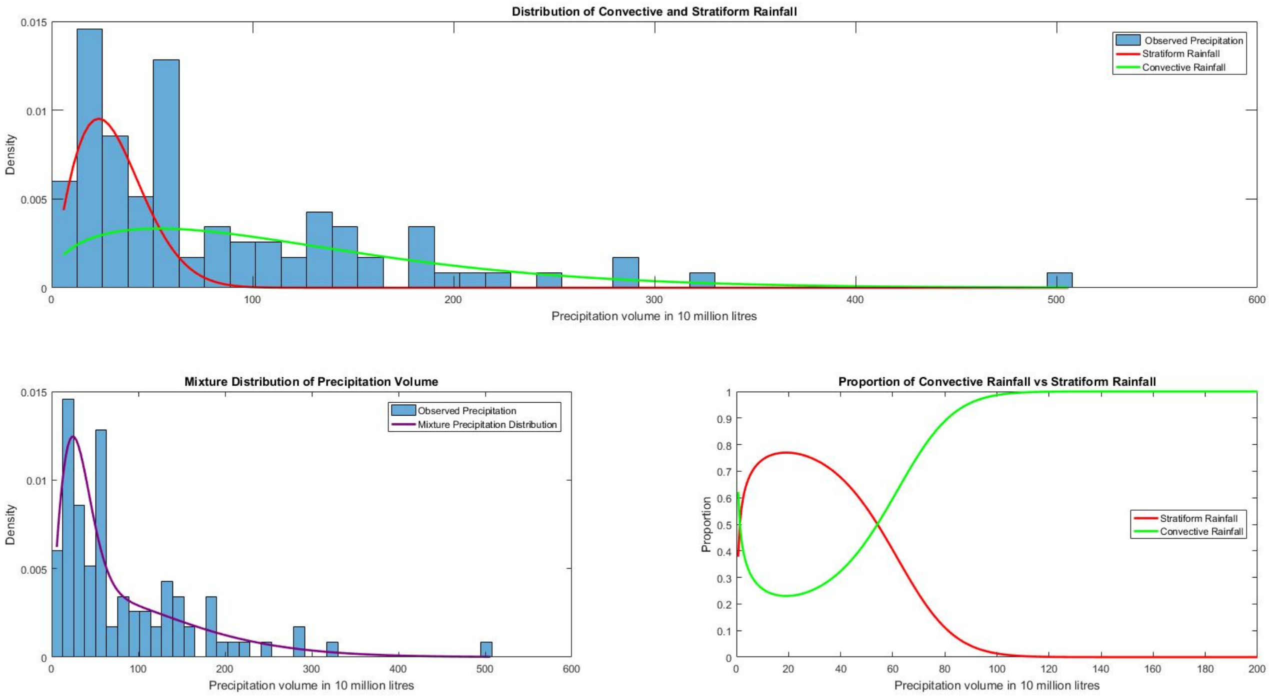

Figure 2.23: Fort Collins Fall Mixture model for Precipitation Volumes

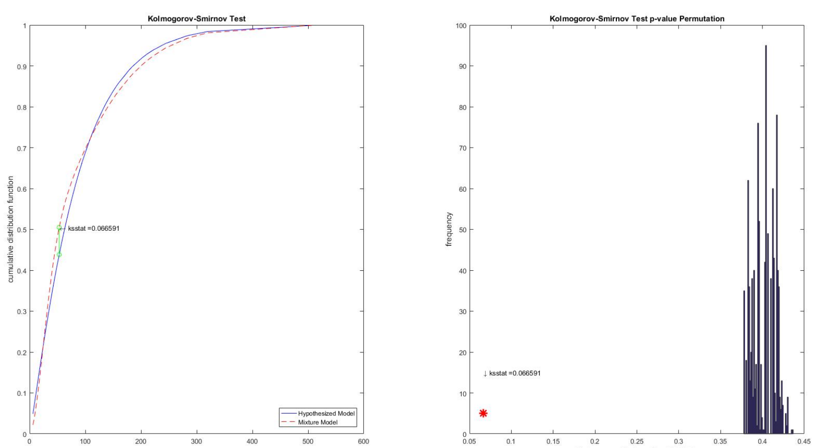

Figure 2.24: Kolmogorov Smirnov Test for Fort Collins Fall Mixture

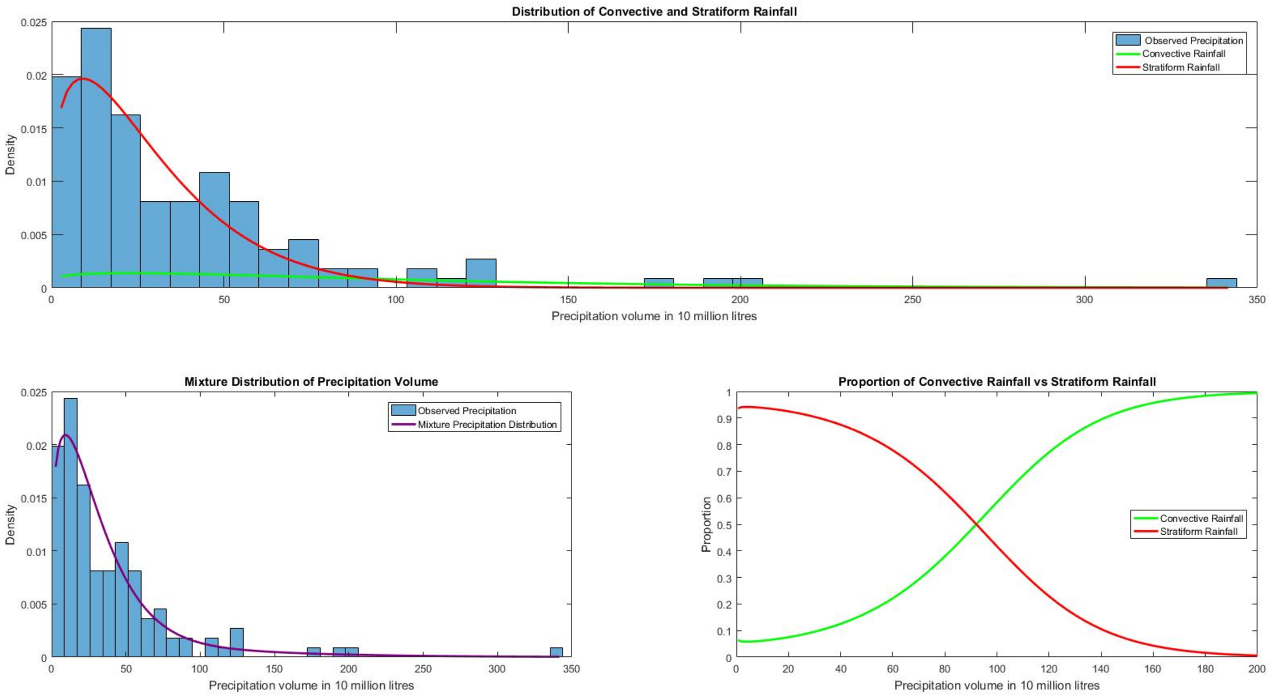

Figure 2.25: Boulder Winter Mixture model for Precipitation Volumes

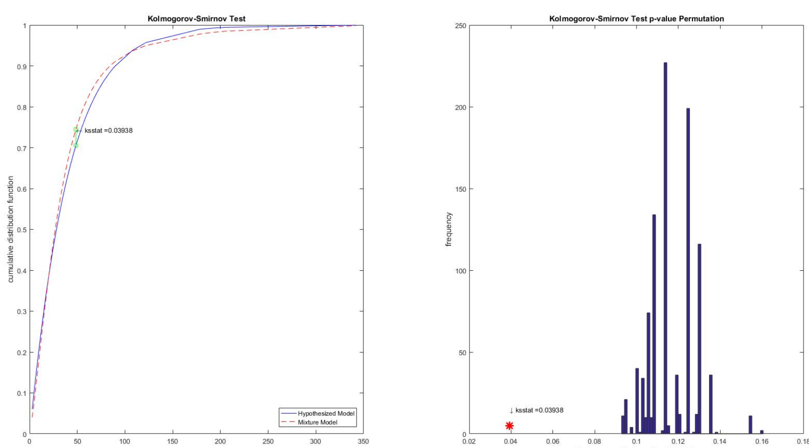

Figure 2.26: Kolmogorov Smirnov Test for Boulder Winter Mixture

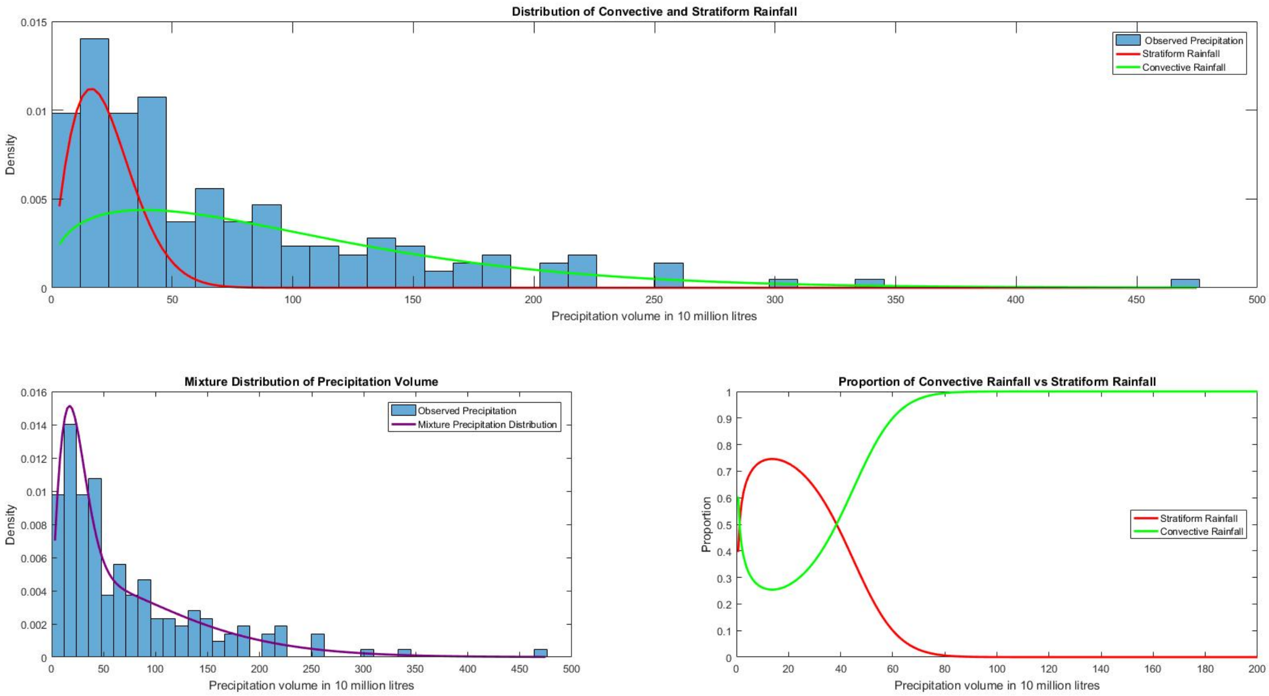

Figure 2.27: Boulder Spring Mixture model for Precipitation Volumes

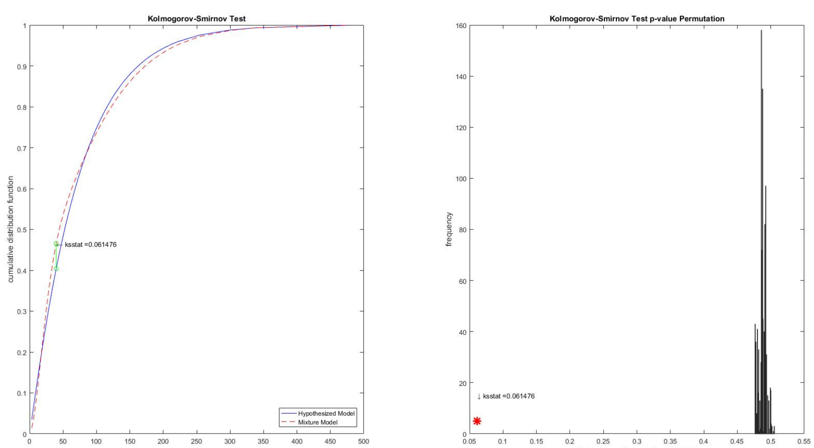

Figure 2.28: Kolmogorov Smirnov Test for Boulder Spring Mixture

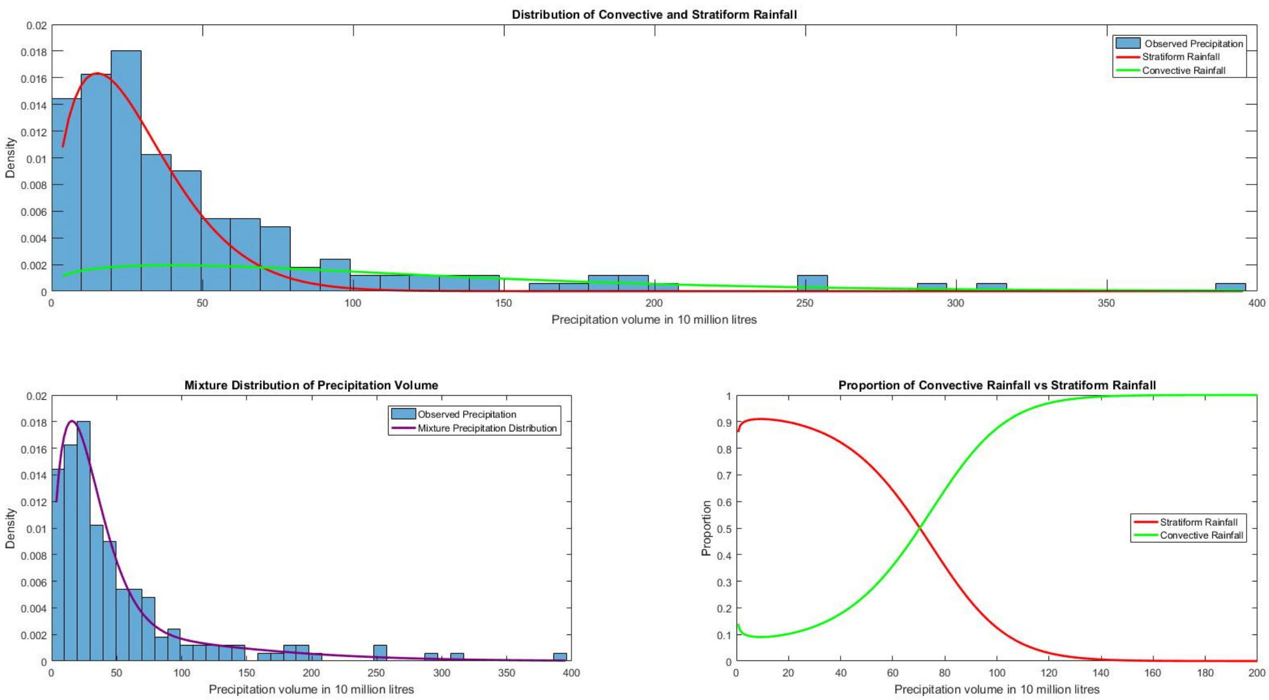

Figure 2.29: Boulder Summer Mixture model for Precipitation Volumes

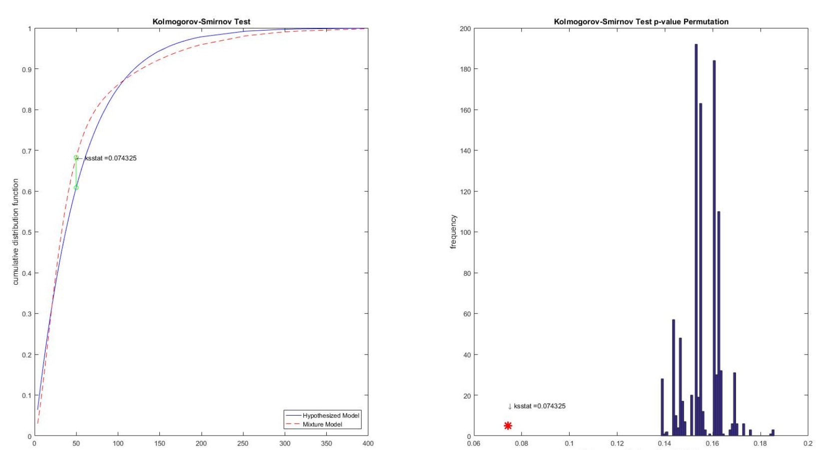

Figure 2.30: Kolmogorov Smirnov Test for Boulder Summer Mixture

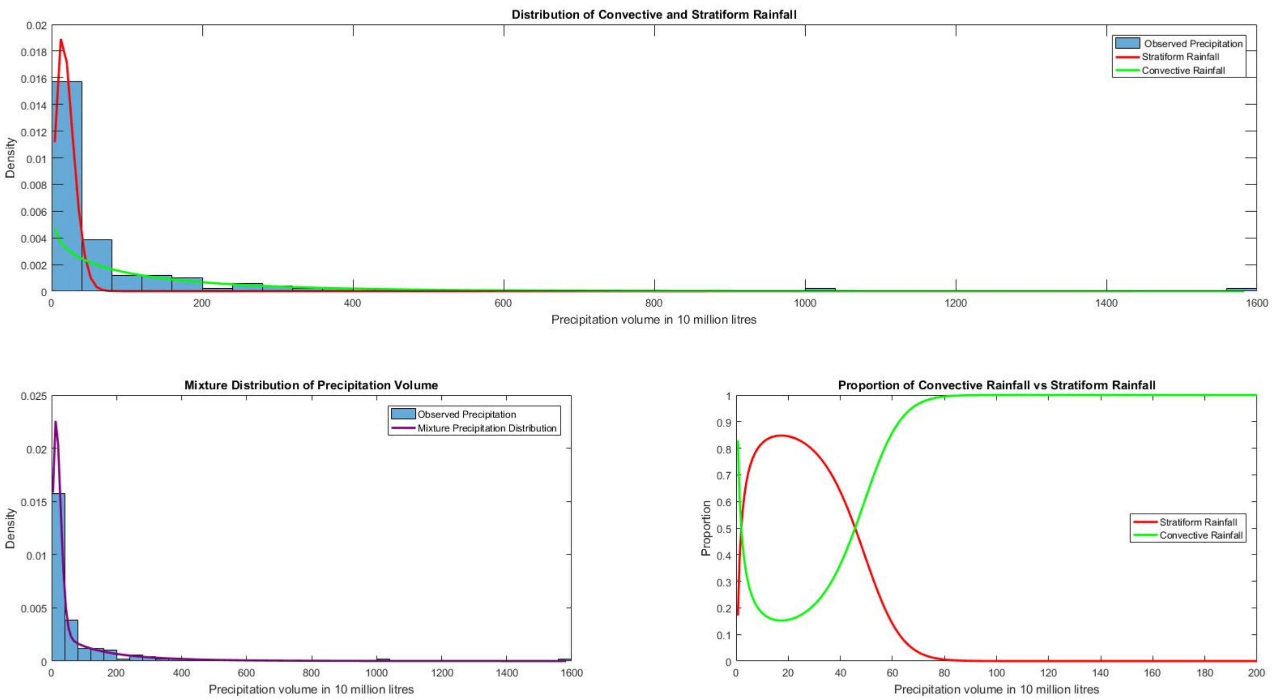

Figure 2.31: Boulder Fall Mixture model for Precipitation Volumes

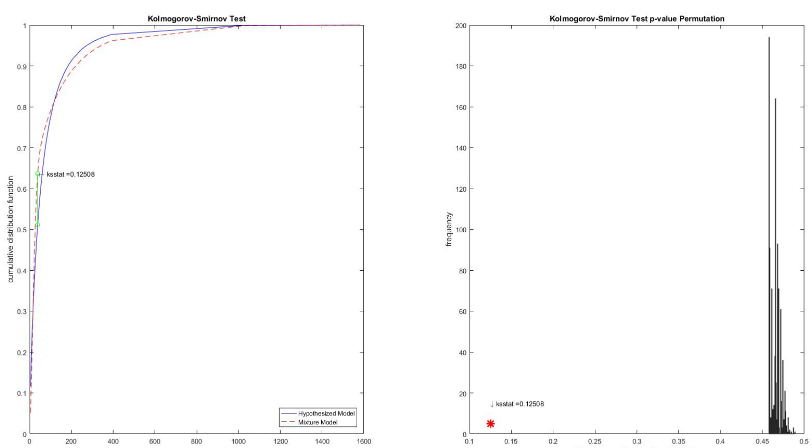

Figure 2.32: Kolmogorov Smirnov Test for Boulder Fall Mixture

Table 2.1 provides the minimized negative log likelihood, estimated convective and stratiform precipitation parameters. Table 2.2 provides with the stringent observed MLCAPE parameters, KS test statistic and the p-values. As mentioned earlier, Weibull mixtures of precipitation volumes behave similarly by seasons. This can be seen by looking at the range of proportions of convective precipitation (the parameter p). In Winter the proportion of convective precipitation ranges from 0.178 to 0.30 which is the lowest of all seasons. This is followed by Summer where proportions are 0.3091, 0.3453 and 0.6397 for Boulder, Lakewood and Fort Collins respectively. This means that there is a larger proportion of convective precipitation in Summer at Fort Collins. This is followed by Fall where proportions range from 0.427 to 0.598, with the higher proportion of convective precipitation of 0.598 observed in Fort Collins. Spring has the highest proportion of convective precipitation of all seasons with 0.4435, 0.6468 and 0.7278 in Lakewood, Boulder and Fort Collins respectively, thereby the likelihood of experiencing extreme storms is the highest in Spring.

The shape parameter for the convective precipitation in Winter for all three regions is around 1.1 to 1.2, along with the scale parameter ranging from 80 to 100, which is the lowest of all seasons indicating that there is less variability or extreme values/storms during Winter. The highest scale parameter is found in Fall season ranging from 122 to 134 indicating that there are more extreme values during this season, e.g. the extreme precipitation volume in liters from Colorado flooding in September 2013. One can see that Lakewood and Boulder have similar precipitation volume mixture distributions for all seasons.

Table 2.1: Parameter estimates for Weibull Mixture Distribution for Precipitation Volumes

Model	**Sample Size**	**(-Log Likelihood)**	**BIC**	**Convective Precipitation**	**Stratiform Precipitation**	**Proportion of Convective Precipitation**
Winter						
Lakewood	128	591.663	1207.587	β=80.78, α = 1.12,	β=28.323, α = 1.6	p = .2959
Fort Collins	77	358.288	740.835	β=97.38, α = 1.13,	β=33.05, α = 1.76	p = .2271
Boulder	129	599.631	1223.523	β=97.08, α = 1.21,	β=30.95, α = 1.27	p = .1779
Spring						
Lakewood	192	1001.067	2026.394	β=127.75, α = 1.36,	β=35.84, α = 1.44	p = .4435
Fort Collins	124	663.872	1352.004	β=119.00, α = 1.311,	β=20.32, α = 2.81	p = .7278
Boulder	180	943.012	1910.284	β=108.44, α = 1.33,	β=25.47, α = 1.82	p = .6468
Summer						
Lakewood	175	895.977	1816.214	β=136.48, α = 1.23,	β=34.76, α = 1.33	p = .3453
Fort Collins	123	648.411	1321.082	β=114.49, α = 1.17,	β=22.46, α = 2.36	p = .6397
Boulder	168	823.164	1670.588	β=116.80, α = 1.32,	β=31.50, α = 1.50	p = .3091
Fall						
Lakewood	134	677.122	1378.504	β=121.74, α = 1.35,	β=27.199, α = 1.55	p = .4266
Fort Collins	92	494.855	1013.97	β=131.75, α = 1.39,	β=34.95, α = 1.89	p = .5977
Boulder	124	633.226	1290.712	β=133.86, α = .85,	β=22.53, α = 1.77	p = .4597

Table 2.2: Parameter estimates for MLCAPE Weibull Distribution and Goodness of Fit Test

Model	**MLCAPE Parameters**	**KSSTAT**	**p-value**
Winter			
Lakewood	β_1=57.432, α_1 = 1.002, p = .0781, β_2=40.524,α_2 = 1.109	0.0688	**1.00**
Fort Collins	β_1=46.07, α_1 = 1.134, p = .0909, β_2=40.24,α_2 = .82	0.0921	**1.00**
Boulder	β_1=48.31, α_1 = .85, p = .0775, β_2=439.57,α_2 = 1.06	0.0394	**1.00**
Spring			
Lakewood	β_1=71.17, α_1 = 1.06, p = .5, β_2=70.20,α_2 = 1.002	0.0526	**1.00**
Fort Collins	β_1=82.32, α_1 = 1.094, p = .4839, β_2=90.13,α_2 = 1.006	0.0868	**1.00**
Boulder	β_1=71.22, α_1 = 1.082, p = .5278, β_2=77.37,α_2 = 1.04	0.0615	**1.00**
Summer			
Lakewood	β_1=62.145, α_1 = 1.226, p = .126, β_2=63.01,α_2 = .959	0.0623	**1.00**
Fort Collins	β_1=60.68, α_1 = 2.188, p = .0989, β_2=77.96,α_2 = .94	0.101	**1.00**
Boulder	β_1=54.72, α_1 = 1.39, p = .1012, β_2=52.84,α_2 = .99	0.0743	**1.00**
Fall			
Lakewood	β_1=50.820, α_1 = 1.007, p = .4553, β_2=69.0465,α_2 = .9553	0.0758	**1.00**
Fort Collins	β_1=94.03, α_1 = 1.095, p = .6304, β_2=77.06,α_2 =1.15	0.0666	**1.00**
Boulder	β_1=77.12, α_1 = .72, p = .5565, β_2=45.53,α_2 =.89	0.1251	**1.00**

Gamma Models for Precipitation as a Function of CAPE

3.1 Methodology

The present goal is to accurately predict the intensity of precipitation volume as a function of CAPE'. Precipitation volumes are positive and right skewed with extreme events. Thus we use gamma regression, a special case of generalized linear model (GLM) family. Gamma regression is generally used for positive skewed data and is used extensively in the area of quality control, insurance, weather extremes and clinical trials.

The general form of gamma regression is

$$y = \beta_0 + \beta_i x + \sigma z$$

$$\mu = E(y) = \beta_0 + \beta_i x \tag{3.7}$$

where the errors z follows the gamma distribution. The gamma distribution is

$$f(y) = \frac{1}{\delta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{y/\delta} \qquad \alpha, \delta > 0 \tag{3.8}$$

and

$$\text{E(y)}= \alpha\delta = \mu \qquad Var(y) = \alpha\delta^2 = \mu^2/\delta$$

where α is the shape parameter and δ is the scale parameter. There are several special cases of gamma distributions; for example, when $\alpha = 1$ it is an exponential distribution and when $\alpha \to \infty$ it is a normal distribution. In order to treat gamma as a part of a generalized linear model, we assume that α is known. Thus the parameter of interest is the scale, δ. The exponential form of a gamma distribution is

$$f(y) = exp\{-y_i/\delta + (\alpha - 1)ln(y_i) - ln(\delta^\alpha) - ln(\Gamma(\alpha))\} \tag{3.9}$$

where the canonical parameter: $\theta = -1/\delta$ and the dispersion parameter: $\phi = 1/\alpha$. Since we need $\mu > 0$ and $\theta < 0$ which gives restriction on δ. Thus the identity canonical link is not often used and the most common choices are log and inverse link.The use of inverse link can sometimes provide negative fitted values. As a result gamma glm with log link is used to predict precipitation volume. The form of gamma regression with log link is

$$\begin{aligned} &log(\mu) = log(E(y)) = \beta_0 + \beta_i x \\ &\text{or} \\ &\mu = \hat{y} = exp(\beta_0 + \beta_i x) \end{aligned} \tag{3.10}$$

To check the goodness of fit of gamma regression, the residual deviance test is implemented. The null hypothesis states that the model fits the data well, against the alternative that the model does not fit the data. The test statistic is computed as:

$$D = 2\sum_i [-log(y/\hat{\mu}) + (y - \hat{\mu})/\hat{\mu}]/\phi \tag{3.11}$$

where D is the scaled deviance statistic, which is asymptotically distributed as χ^2 given that the model is correct. Thus *p*-value for the test statistics is obtained by using the χ^2 distribution.

3.2 Data Analysis

Before fitting the gamma log link regression model,we look at the correlation between variables. From Table 3.3 one can see that there appear to be no correlation between day and night time CAPE measurements. All of CAPE0's variables are highly correlated with each other and the same is true for CAPE12's. This can be explained as all the CAPE0 and CAPE12 precipitation measurements are recorded at the same time of 6-7 PM and 6-7 AM respectively. As a result one can conclude that multicollinearity might be an issue in our dataset.

Table 3.3: Correlation Matrix of Precipitation Volumes and CAPE's

	Volume	SBCAPE0	MUCAPE0	MLCAPE0	SBCAPE12	MUCAPE12	MLCAPE12
Volume	1	0.08	0.06	0.05	0.07	0.05	0.04
SBCAPE0	0.08	1	0.86	0.76	0.4	0.48	0.38
MUCAPE0	0.06	0.86	1	0.76	0.34	0.4	0.33
MLCAPE0	0.05	0.76	0.76	1	0.3	0.36	0.33
SBCAPE12	0.07	0.4	0.34	0.3	1	0.77	0.56
MUCAPE12	0.05	0.48	0.4	0.36	0.77	1	0.75
MLCAPE12	0.04	0.38	0.33	0.33	0.56	0.75	1

3.2.1 Gamma Regression for Precipitation Volume for All Seasons

A gamma regression with log link is fitted to precipitation volumes as a function of CAPE', location and season. The best fit model is selected based on both Akaike Information Criterion (AIC) and p-value for each coefficient, and results are presented in Table 3.4. In this model, surface based convective precipitation measurements for day and night along with season and location are significant predictors for precipitation volumes. The positive coefficients on SBCAPE0 and SBCAPE12 indicate that surface based potential energy is positively associated with volume. Relative to Winter, the remaining three seasons have higher precipitation. Also, Boulder and Fort Collins are more likely to have higher precipitation than Lakewood. The model is of the form

$$\begin{aligned} \widehat{Precipitation} = \exp(&3.6637 + 0.0001 * \text{SBCAPE0} + 0.0003 * \text{SBCAPE12} \\ &+ 0.5852 * \text{Spring} + 0.2915 * \text{Summer} + 0.5237 * \text{Fall} \\ &+ .0067 * \text{Boulder} + .2009 * \text{Fort Collins}). \end{aligned} \tag{3.12}$$

Figure 3.33 shows the observed vs fitted plot of the best fit model for all seasons. The predicted volumes appear to be on a flat line with little variability and fails to capture extreme events. This might be due to the fact that season is a significant factor with a lot of heterogeneity. Therefore, there is a need to fit the precipitation volume model by season.

Table 3.4: Regression Estimates for Best Fit Model of Precipitation, All Seasons

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.6637	0.0822	44.56	0.0000
SBCAPE0	0.0001	0.0001	1.69	0.0907
SBCAPE12	0.0003	0.0001	1.90	0.0571
Spring	0.5852	0.0924	6.34	0.0000
Summer	0.2915	0.1038	2.81	0.0050
Fall	0.5237	0.0993	5.28	0.0000
Boulder	0.0067	0.0723	0.09	0.9259
Fort Collins	0.2009	0.0802	2.50	0.0124

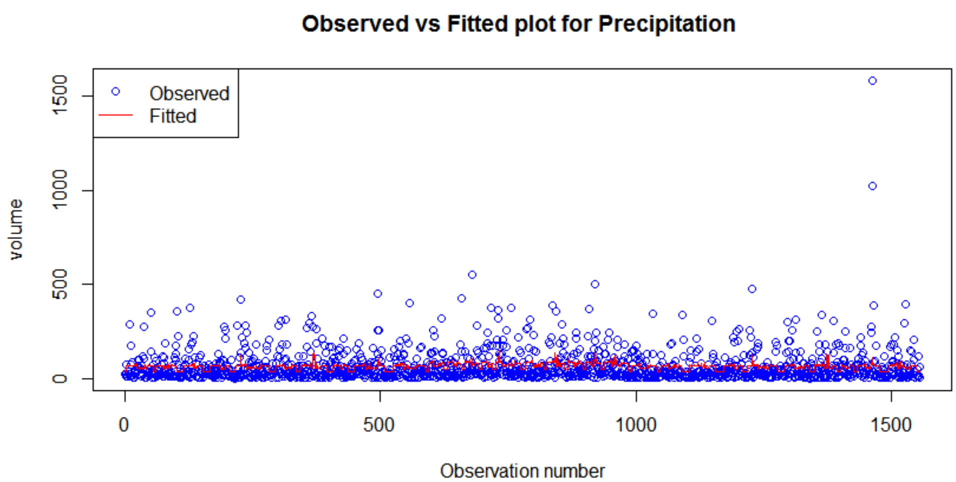

Figure 3.33: Diagnostic Plots for Best Fit Model, All Seasons

Table 3.5 shows the observation counts across season and location. Winter has the lowest number of observations for precipitation while Spring has the highest number of observations.

A deviance goodness of fit test is performed to check the validity of the model. The null hypothesis states that the model fits the data well, and the alternative hypothesis states that the model is lack of fit. The test statistic is derived by comparing the scaled residual deviance with its degrees of freedom. The p-value of deviance goodness of fit test ≈ 1.00,

Table 3.5: Observations across Season and Location

	Lw	Bo	Fc	Total
Winter	113	114	65	292
Spring	185	173	120	478
Summer	168	164	118	450
Fall	128	118	90	336

thereby, we fail to reject the null hypothesis and concluding that the model shows no lack of fit.

3.2.2 Gamma Regression for Winter Precipitation Volume

A gamma log link regression is fitted to precipitation volumes for Winter season as seen in Equation 3.13. Only the mixed layer precipitation measurement from the morning (MLCAPE12) is significant. The coefficient of MLCAPE12 is -0.03982, indicating that an increase in MLCAPE12 leads to a decrease in expected Winter precipitation. This is unexpected as precipitation volumes should increase with the rise in CAPE values.

$$Winter\widehat{Precipitation} = \exp\left(3.7188 - 0.03982 * \mathrm{MLCAPE12}\right) \tag{3.13}$$

We further examine the observed vs fitted plot in Figure 3.34. The left plot shows the observed vs fitted plot for Winter precipitation as a function of MLCAPE12. From the plot, it is seen that all the spikes are downward facing, that is MLCAPE12 only predict low precipitation. To make sure that the same downward pattern is not observed elsewhere, we fit a new regression model for Winter Precipitation as a function of SBCAPE12 (highly correlated with MLCAPE12), and the right plot shows the observed vs fitted plot for this model. The spikes are upward facing which means SBCAPE12 predicts high volumes well. SBCAPE, MUCAPE and MLCAPE provide us with precipitation measurements in a parcel at different altitudes in atmosphere at a given time. Hence, a function of all these CAPE'

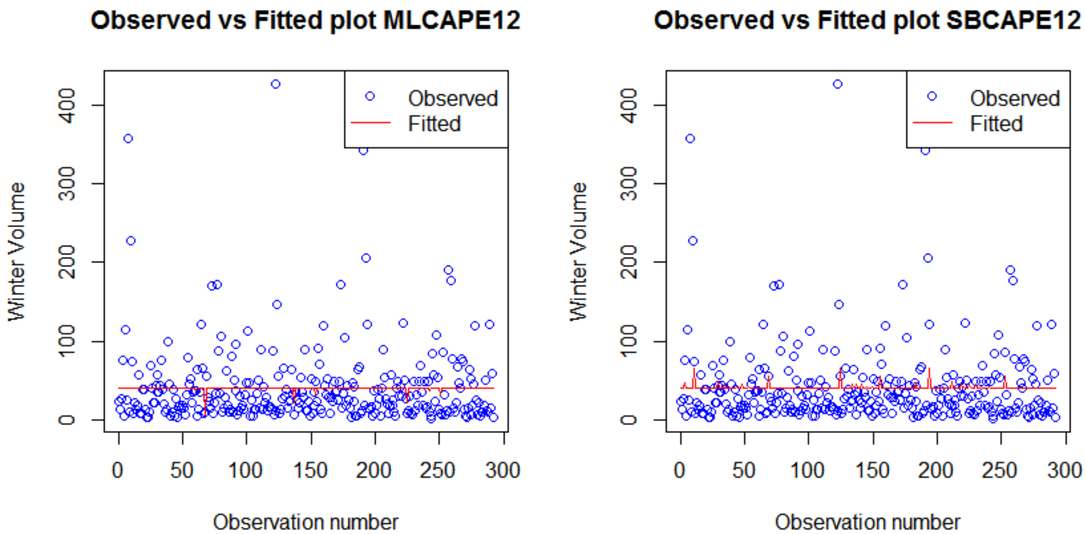

Figure 3.34: Observed vs Fitted Plot for Winter Precipitation

combined together can help accurately measure the precipitation. Therefore even though the CAPE12 variables are highly correlated and some of them are insignificant, we include them in our final model.The same is true for modeling precipitation for other seasons. If only one of the CAPE variable is significant, then we include remaining CAPE values recorded at the same time in the model.

From the output of Table 3.6, it is seen that none of the p-values are significant, this is likely due to multicollinearity. The observed vs the fitted plot in Figure 3.35 shows spikes in both direction with the predicted values closer to the y intercept line of $e^{3.6972} = 40$ (10 million liters). For the Winter season, the precipitation volumes are the lowest with rare extreme storm events. Thus, the model is a good fit as it predicts most of low volume precipitation compared to high precipitation volumes. The dispersion parameter for the model is 1.545, along with residual deviance of 265.55. Therefore, the deviance goodness of fit test statistic is $265.55/1.545 = 171.89$ with a corresponding p-value of ≈ 1. Thereby, we fail to reject the null hypothesis and conclude that the model shows no lack of fit.

$$\widehat{WinterPrecipitation} = \exp(3.6972 + 0.0417 * \text{SBCAPE12} + 0.0054 * \text{MUCAPE12} - 0.0612 * \text{MLCAPE12}) \tag{3.14}$$

Table 3.6: Regression Estimates for Best Fit Model of Winter Precipitation

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.6972	0.0756	48.87	0.0000
SBCAPE12	0.0417	0.0444	0.94	0.3485
MUCAPE12	0.0054	0.0111	0.48	0.6298
MLCAPE12	-0.0612	0.0453	-1.35	0.1778

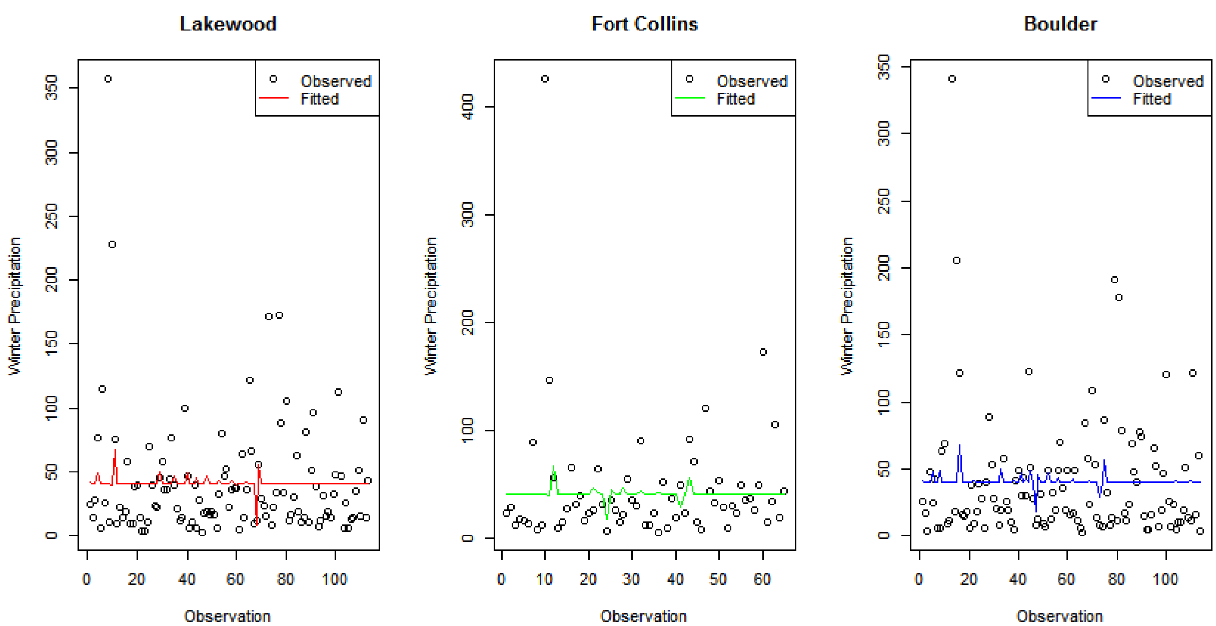

Figure 3.35: Diagnostic Plots for Winter Precipitation Model

3.2.3 Gamma Regression for Spring Precipitation Volume

A gamma regression with log link is fitted to the Spring precipitation volumes. The best fit model is a function of surface based CAPE measured at evening time and most unstable CAPE' measured at morning and evening times, as seen in Equation 3.15.

$$\widehat{SpringPrecipitation} = \exp(4.3465 + 0.0001 * \text{SBCAPE0} - 0.0011 * \text{MUCAPE0} + 0.0003 * \text{MUCAPE12}) \tag{3.15}$$

Table 3.7: Regression Estimates for Best Fit Model of Spring Precipitation

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.3465	0.0531	81.89	0.0000
SBCAPE0	0.0009	0.0004	2.59	0.0098
MUCAPE0	-0.0011	0.0003	-3.17	0.0016
MUCAPE12	0.0003	0.0002	1.51	0.1309

From Table 3.7, we can see that variables SBCAPE0 and MUCAPE0 are significant predictors of precipitation in Spring. Although MUCAPE12 is not significant, we still retain it in the model because it gives us the lowest AIC value, compared with that from the model without this variable. Besides, precipitation volume in Spring is highly volatile due to the many extreme storm events, therefore, it is reasonable to include most unstable potential energy readings (MUCAPE) in our final model. We also perform the deviance test and find the test statistic is 461.3258 with a p-value of 0.6531. Thus, we fail to reject our null hypothesis that the sample data is from a Gamma distribution.

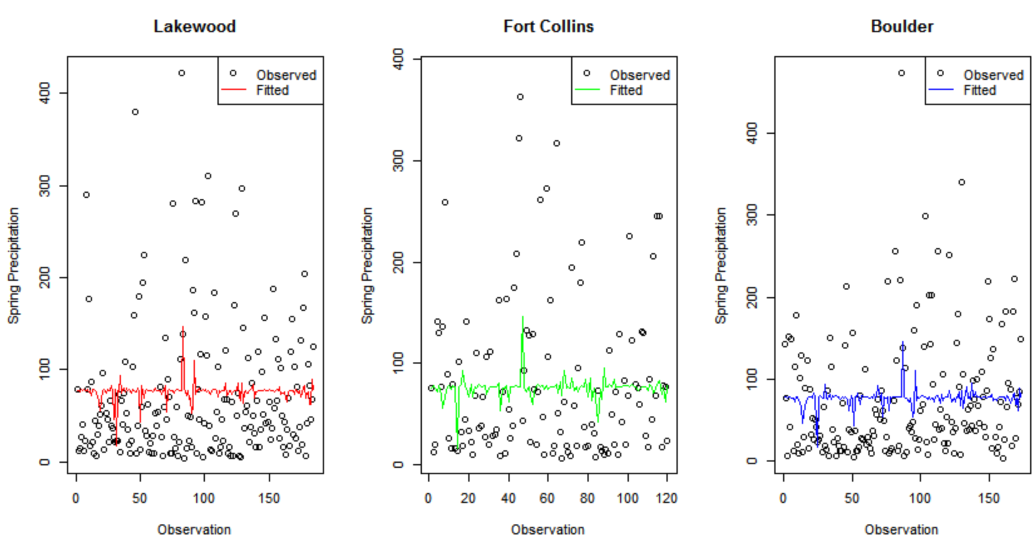

Figure 3.36: Diagnostic Plots for Spring Precipitation Model

Figure 3.36 presents the observed versus fitted plot for the Spring precipitation model. We can see from the scatter plot that our model correctly predicts the precipitation volumes when there is only light rainfall/snowfall. However, when there is heavy rain/snow or even catastrophic floods, our model does not forecast the precipitation volume well, which is indicated by the significant distance between the extreme values and the upper spikes.

3.2.4 Gamma Regression for Summer Precipitation Volume

A gamma log link regression model is fitted to Summer precipitation volumes. From Table 3.8, we can see that none of the variables are significant and the parameter estimates are tiny. A backward selection process was implemented and we find that the model with only location factor was the best according to AIC. Table 3.9 displays the gamma regression coefficients of the Summer precipitation model where only location is included. This means that the rainfall volumes in Summer only depend on location of city, which is inadequate. Upon further inspection of the Summer precipitation distribution in Figures 1.4 to 1.6 as shown in the Introduction, we see that Summer volumes for all CAPE' are right skewed with the highest variability across all seasons. Hence, there is a need to transform the CAPE values to better fit the Summer precipitation volumes.

A log transformation is applied to the CAPE values in order to account for the skewness and high variability. We reestimate the model with the transformed variables and present the results from the best fit model in Table 3.10. Before examining individual estimates, we should inspect the overall fit of the model. A deviance test of the model gives a p-value of 0.9746, indicating there is no lack of fit, and the analysis of deviance table also shows the variables included in Table 3.10 are overall significant predictors of Summer precipitation. Mixed layer convective available potential energy in the night (MLCAPE0) is significantly and negatively associated with Summer precipitation at 1% level, whereas surface-based potential energy in the morning (SBCAPE12) is positively related to precipitation volume in Summer. Besides, we find significant difference in the rainfall volume in Summer across city. On average, Fort Collins has a higher level of precipitation in Summer holding everything else constant. Precipitation in Boudler, however, is not significantly different from that in Lakewood.

Table 3.8: Regression Estimates for Full Model of Summer Precipitation

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.1135	0.1096	37.53	0.0000
SBCAPE0	-0.0000	0.0001	-0.01	0.9952
MUCAPE0	0.0001	0.0001	0.50	0.6139
MLCAPE0	-0.0001	0.0002	-0.26	0.7934
SBCAPE12	-0.0002	0.0002	-0.65	0.5143
MUCAPE12	-0.0000	0.0002	-0.04	0.9652
MLCAPE12	0.0001	0.0004	0.35	0.7238
Boulder	-0.1763	0.1299	-1.36	0.1755
FortCollins	0.2247	0.1423	1.58	0.1151

Table 3.9: Regression Estimates of Summer Precipitation Model by Location

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.1293	0.0900	45.90	0.0000
Boulder	-0.1737	0.1280	-1.36	0.1755
FortCollins	0.2293	0.1401	1.64	0.1023

$$\begin{aligned} Summer\widehat{Precipitation} = &\exp(4.172 - 0.1638 * \text{Boulder} + 0.2559 * \text{FortCollins} \\ &- 0.0622 * log(\text{MLCAPE0}) + .0574 * log(\text{SBCAPE12}) \end{aligned} \tag{3.16}$$

Table 3.10: Regression Estimates for Best Fit Model of Summer Precipitation

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.1720	0.0960	43.48	0.0000
log(MLCAPE0)	-0.0622	0.0156	-3.98	0.0001
log(SBCAPE12)	0.0574	0.0143	4.02	0.0001
Boulder	-0.1638	0.1172	-1.40	0.1629
FortCollins	0.2559	0.1282	2.00	0.0466

Similar to that for Winter and Spring, we also present the observed-fitted plot from model (3.16) in Figure 3.37. Since a large proportion of the observations have low precipitation volume, the scatter plot is concentrated on the lower part of the graph vertically. The red, green and blue lines are the predicted volume from the model fitted with Gamma distribution for Lakewood, Fort Collins and Boulder, respectively. We can see that they correctly predict both light and medium rainfall as the spikes extend both to the bottom and top of the fitted line. One important difference to note is that Fort Collins have a higher mean predicted rainfall volumes. The average predicted rainfall volumes in Summer for Lakewood and Boulder are close to each other. This is consistent with the results shown in Table 3.10.

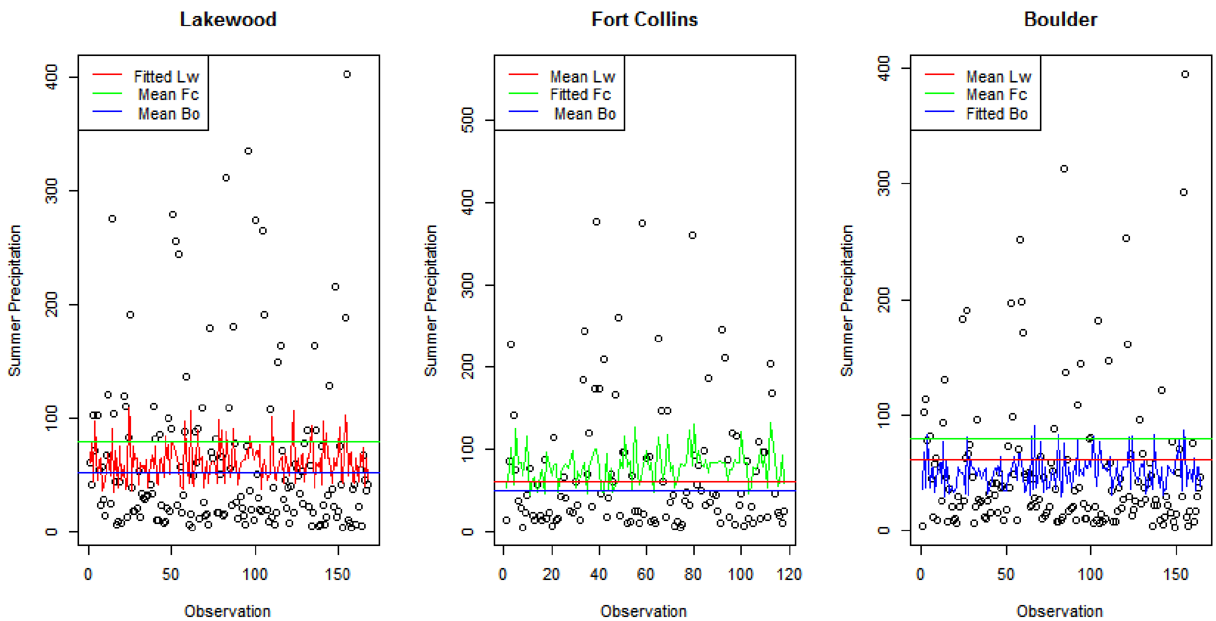

Figure 3.37: Observed vs Fitted Plot for Summer Precipitation Model

3.2.5 Gamma Regression for Fall Precipitation Volume

A gamma log link regression model is fit to Fall precipitation volumes. From Table 3.11, we can see that all variables are included in the best fit model with the exception of MLCAPE0. The variables of best fit model are all significant with a p-value of less than 5%. The surface based potential energy (SBCAPE) measurements recorded both in daytime and nighttime are positively associated with the Fall precipitation, whereas, the most unstable potential energy (MUCAPE) measurements for day and night are negatively associated with Fall precipitation. Also, we find significant difference in the rainfall volume across cities. Boulder and Lakewood have similar precipitation levels, however, Fort Collins is significantly different from the other two locations and has a higher level of precipitation in Fall. The deviance test is conducted to check the overall fit of the model. The deviance test statistic of 270.53

is obtained with a corresponding p-value of .99. Thereby, we can conclude that there is no lack of fit.

$$\begin{aligned} \widehat{FallPrecipitation} = \exp(&3.9901 + 0.0013 * \text{SBCAPE0} - 0.0006 * \text{MUCAPE0} \\ &+ 0.0025 * \text{SBCAPE12} - 0.0027 * \text{MUCAPE12} \\ &+ 0.0052 * \text{MLCAPE12} + 0.0700 * \text{Boulder} \\ &+ 0.3431 * \text{Fort Collins}) \end{aligned} \tag{3.17}$$

Table 3.11: Regression Estimates for Best Fit Model of Fall Precipitation

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.9901	0.1035	38.54	0.0000
SBCAPE0	0.0013	0.0003	3.93	0.0001
MUCAPE0	-0.0006	0.0002	-2.74	0.0064
SBCAPE12	0.0025	0.0010	2.55	0.0112
MUCAPE12	-0.0027	0.0008	-3.44	0.0007
MLCAPE12	0.0052	0.0023	2.29	0.0228
Boulder	0.0700	0.1439	0.49	0.6270
FortCollins	0.3431	0.1560	2.20	0.0286

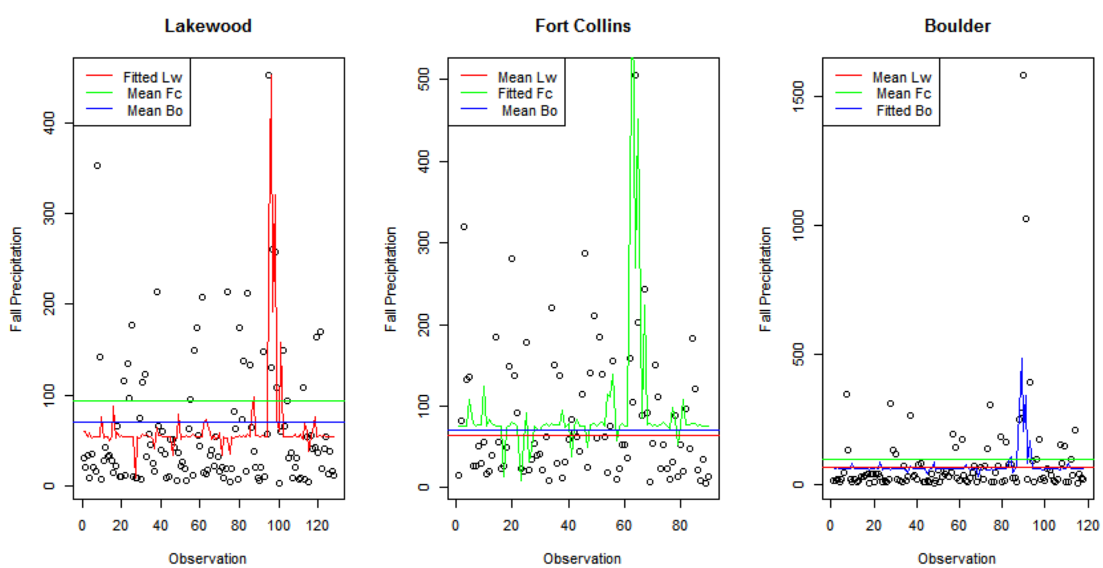

Figure 3.38: Observed vs Fitted Plot for Fall Precipitation Model

Figure 3.38 presents the observed vs fitted plot of Fall precipitation for the three cities. Different from the previous models, we find that model 3.17 fits the data well for all types of rainfall including extreme volumes, and this is true for all locations. The highest two observations in the plot for Boulder are from the Colorado flooding in September, 2013. The flooding precipitation volumes are at least three times that of other events on average. As a result, the model has not accurately predicted this event yet. However, the plots for Lakewood and Fort Collins, show that the model predicts the Fall precipitation well. In addition, we find that Fort Collins has the highest mean predicted precipitation volumes, followed by Boulder and Lakewood.

Conclusion

In this study, we examine the seasonal precipitation patterns in Lakewood, Fort Collins, and Boulder, Colorado. We first classify precipitation volumes into two categories, namely convective and stratiform precipitation. This was achieved by fitting a versatile Weibull mixture distribution for each location-season combination. Afterwards, Kolmogorov-Smirnov Test was implemented to check the goodness of fit of these mixture models. The Weibull mixture distributions accurately model the seasonal precipitation volumes for Lakewood, Fort Collins, and Boulder, Colorado. Results clearly show that seasons affect the proportion and the shape of the distribution of convective precipitation. Surprisingly, we find that seasonal mixtures for Lakewood and Boulder perform similarly, although Lakewood is located on a flat region compared with Boulder's mountainous landscape. The weather stations do not distinguish between convective and stratiform events, thereby, the mixture model is artificial to some extent.

The second part of this study aims to model the continuous precipitation volumes as a function of Convective Available Potential Energy (CAPE) measurements. Since the precipitation volumes are positive and right skewed with high extreme events, a gamma model was estimated. It was found that the log link gamma model for seasonal precipitation volumes fits the data well. For most of the seasons, the model accurately predicts only the lower and mid precipitation volumes except for Fall where the model accurately predicts even the extreme events.

There are some limitations of this study. First, observation for precipitation volumes are available only for times when all stations in the region record rainfall. Had we had daily observations, including the days with no rainfall, we could have fit a spatial temporal model to determine the trend and forecast future precipitation volumes and extreme storm events.

Second, the CAPE measurements are indicators of the convective energy present in the atmosphere. Future research could try to identify each precipitation event as convective or stratiform and then predict the extreme events with higher accuracy.

Bibliography

Byatt, D. (2000), 'Convergent variants of the nelder-mead algorithm', *Department of Mathematics, University of Canterbury* .

Gaal, L., Molnar, P. and Szolgay, J. (2014), 'Selection of intense rainfall events based on intensity thresholds and lightning data in switzerland', *Hydrology and Earth System Sciences* **18**(5), 1561–1573.

Gavin, H. P. (2013), 'The nelder-mead algorithm in two dimensions', *CEE 201L. Duke U* .

Houze Jr, R. A. (2014), *Cloud dynamics*, Vol. 104, Academic press.

Kriebel, P. J. (2016), 'Estimating precipitation volume distribution using data from the spatially dense cocorahs network', *Department of Statistics, University of Georgia* .

Qi, Y., Zhang, J. and Zhang, P. (2013), 'A real-time automated convective and stratiform precipitation segregation algorithm in native radar coordinates', *Quarterly Journal of the Royal Meteorological Society* **139**(677), 2233–2240.

Rulfová, Z. and Kyselỳ, J. (2014), 'Trends of convective and stratiform precipitation in the czech republic, 1982–2010', *Advances in Meteorology* **2014**.

Wright, M. H. (2012), 'Nelder, mead, and the other simplex method', *Documenta MathematicaExtra Volume ISMP* pp. 271–276.

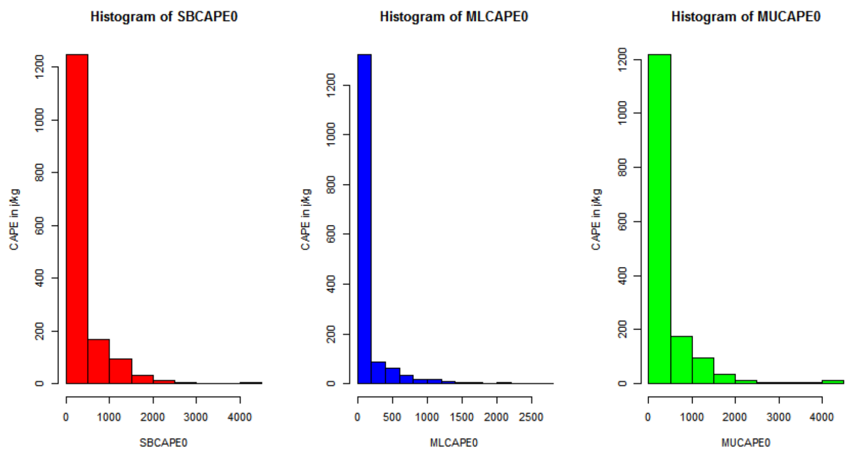

Figure 5.39: Histogram of CAPE0

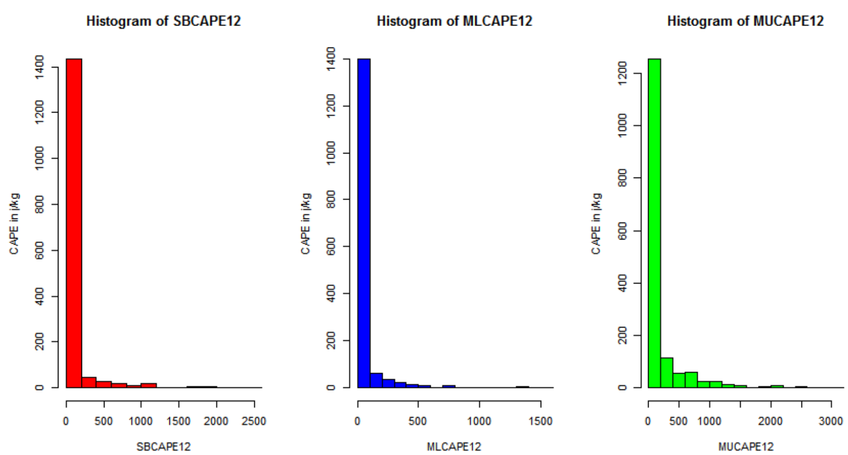

Figure 5.40: Histogram of CAPE12