

# Assignment\_1

*Vahid Salahi*

*February 3, 2017*

## Question 1:

Read the following papers:

- Milly et al., (2008); “*Stationarity is Dead - Wither Water Management?*”, Science, Vol 319; 573 - 574
- Milly, P. C. D., Julio Betancourt, Malin Falkenmark, Robert M. Hirsch, Zbigniew W. Kundzewicz, Dennis P. Lettenmaier, Ronald J. Stouffer, Michael D. Dettinger, and Valentina Krysanova. “On critiques of “*Stationarity is dead: whither water management?*”. Water Resources Research 51, no. 9 (2015): 7785-7789.

a) What is stationarity assumption? What are some of the reasons why stationarity was used in hydrologic design? What causes non-stationarity? What you are unclear about stationarity and non-stationarity?

A stationarity data set is one whose statistical properties such as mean, variance and autocorrelation remain constant over time. Nonstationarity can result from many anthropogenic processes. These include change in land and water use and also climate change. In stationarity models, planners assume that natural change and variability is small enough which makes the model reasonable to use. But generally stationarity is rejected by planner due to magnitude of hydroclimate change that is under way.

b) What are the main arguments for using non-stationarity assumption in future hydrologic design (paper A and B).

The main argument for using non-stationarity assumption in the hydrologic design is the anthropogenic climate change (ACC) due to temperature driven changes of the Earth. another argument to keep using non-stationarity is the climate change that is leading to the change in precipitation patterns, drought and flood periods, soil moisture, frequency of extreme weather conditions. Due to effects of one or some of the mentioned, it is believed that non-stationarity should be used for future hydrologic designs.

c) What are the main arguments for continuing to use stationarity assumptions in hydrologic design (You should refer to critiques referred in Paper B)?

On Critiques of “Stationarity is Dead: Whither Water Management?”, it was argued that non-stationarity is not necessarily the only or the best option to be considered specially when it comes to being accurate, reliable and when optimal model are not available. It was concluded that if a non-stationary method is used it might lead to some inaccurate results, in other words it is better to use a familiar method (stationarity) to reduce the errors and inaccuracy in the results.

---

### Citation:

I have came up with the answer of this question by collaboration with Farhang, Gladys and Yosselyn.

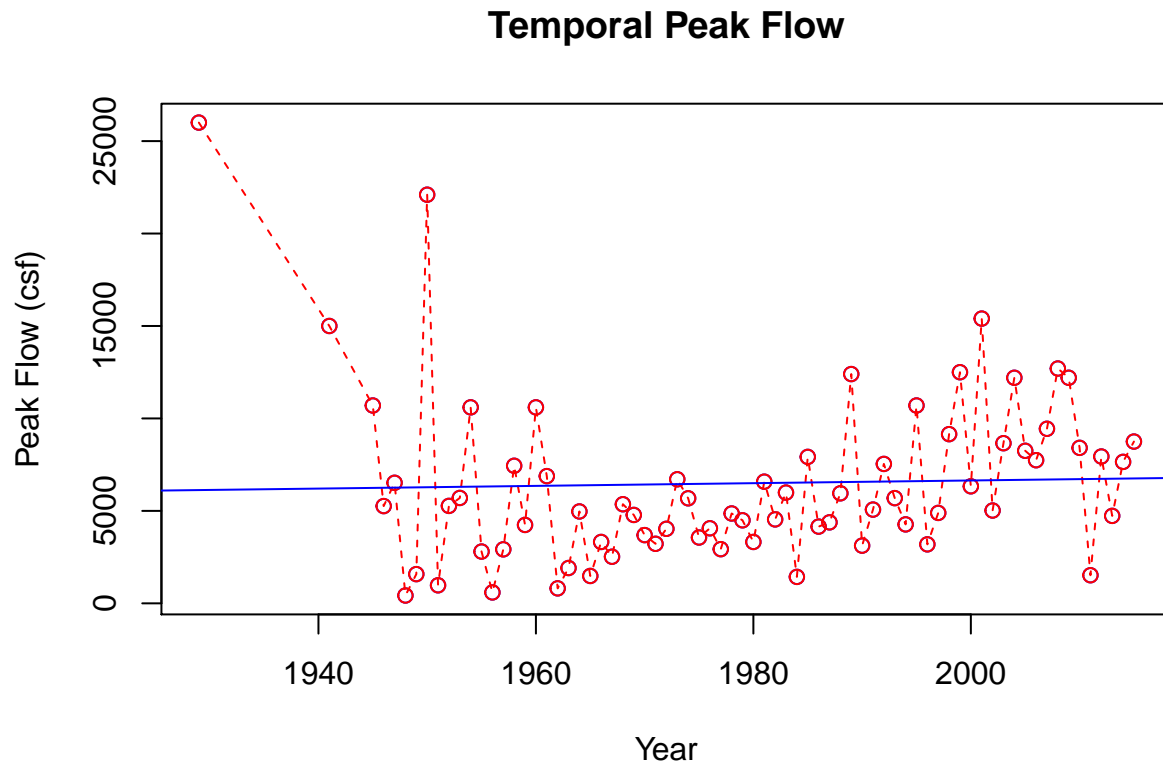
## Question 2:

Download the peak annual streamflow and daily streamflow data for USGS 08069000 Cypress Ck nr Westfield, TX and perform necessary analysis to answer the following questions: Note you need to justify your answers with appropriate analysis, charts, figures, etc.

a) Is there a temporal trend in peak flows?

To see if there is any temporal trend for the peak flow data, I first plot the data and obtain the regression line for it. The plot is shown below:

```
plot(hydyear,peakflow,ylab="Peak Flow (csf)",xlab="Year",col="blue")
lines(hydyear,peakflow,type="o",lty=2,col="red")
title("Temporal Peak Flow")
lm <- lm(peakflow ~ hydyear)
abline(lm,col="blue")
```



The regression line appears to have a very small slope which means there is no temporal trend in the data. The summary of the regression line shows a p-value of 0.7707 which indicates no correlation for the time series.

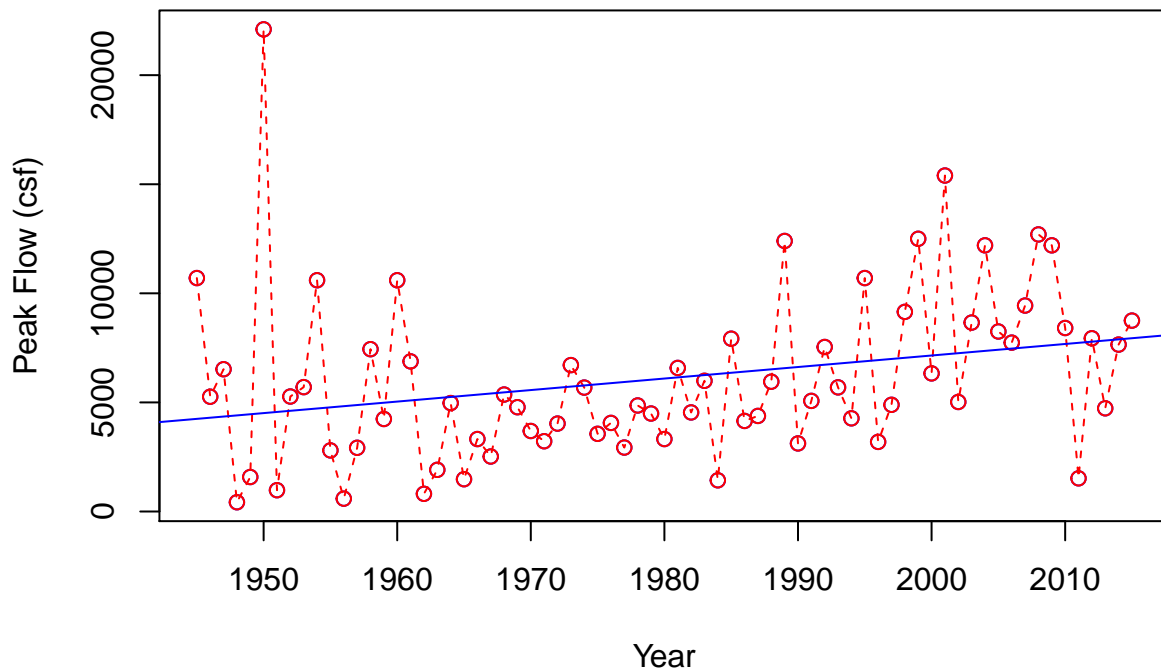
```
summary(lm)
```

```
##
## Call:
## lm(formula = peakflow ~ hydyear)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5840  -2902  -1023   1569  19875
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7996.21   49504.35  -0.162   0.872
## hydyear       7.32      25.02   0.293   0.771
##
## Residual standard error: 4601 on 71 degrees of freedom
## Multiple R-squared:  0.001205,    Adjusted R-squared:  -0.01286
## F-statistic: 0.08563 on 1 and 71 DF,  p-value: 0.7707
```

By looking at the data we notice that the first point of the data has very high value and after that for several years there is no data reported. I have excluded the first two points of the data to see how the regression line looks like. Regression can be affected very much by the outliers.

```
plot(hydyear[3:73],peakflow[3:73],ylab="Peak Flow (csf)",xlab="Year",col="blue")
lines(hydyear[3:73],peakflow[3:73],type="o",lty=2,col="red")
title("Temporal Peak Flow")
lm <- lm(peakflow[3:73] ~ hydyear[3:73])
abline(lm,col="blue")
```

## Temporal Peak Flow



The straight line with a significant positive slope can be observed in above plot which implies positive temporal trend in the time series. Summary of the regression line shows p-value of 0.01693 which indicates a strong correlation in the time series.

```
summary(lm)
```

```
##
## Call:
## lm(formula = peakflow[3:73] ~ hydyear[3:73])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6211.3 -2265.8  -733.5  1191.3 17589.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -98459.34   42719.37  -2.305   0.0242 *
## hydyear[3:73]    52.80     21.57    2.448   0.0169 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3726 on 69 degrees of freedom
## Multiple R-squared:  0.07989,    Adjusted R-squared:  0.06655
```

```
## F-statistic: 5.991 on 1 and 69 DF, p-value: 0.01693
```

The Mann-Kendall trend test is another tool that can be used for monotonic trend in a time series. A p-value of 0.006353 is obtained from this test which means a significant correlation in time series.

```
MannKendall(peakflow)
```

```
## tau = 0.219, 2-sided pvalue =0.0063537
```

b) Is the time-series non-stationary? if so what do you think are the ca

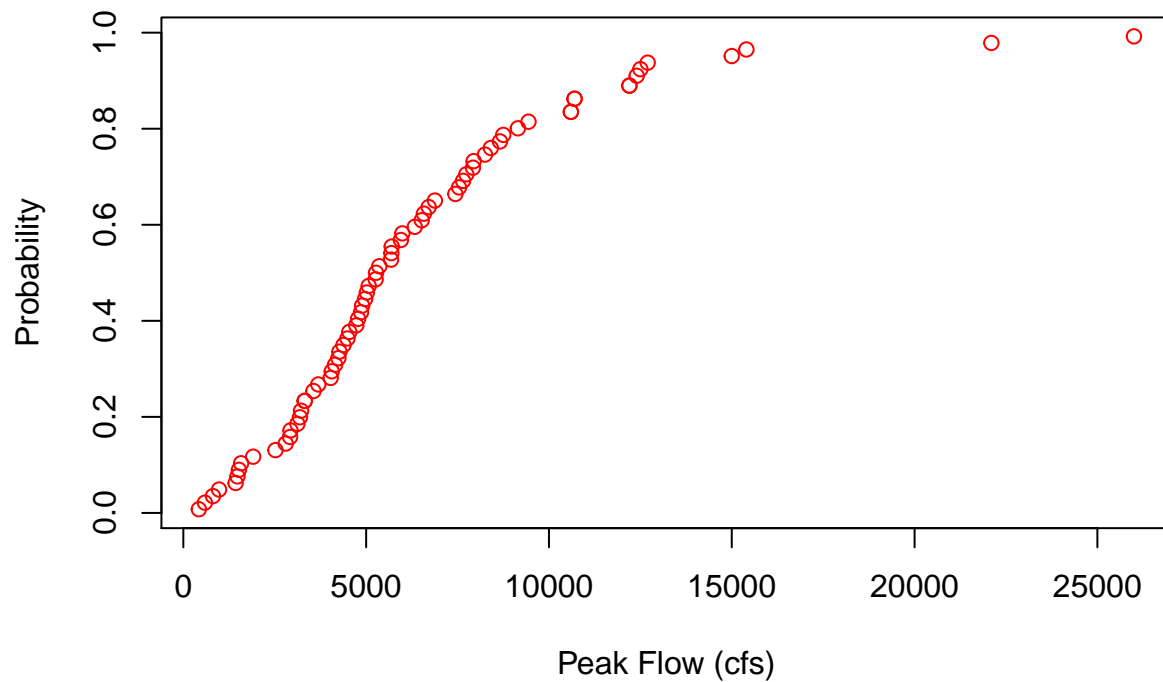
Based on the analysis have been done in previous part, we can conclude that the time series is nonstationary. This is due to a significant trend in temporal data which means change in statistics properties such as mean and variance. The time series has a positive trend meaning increase in peak flow data.

c) Assuming stationarity Draw a cumulative distribution function for the annual peak discharges assuming Gringorten Plotting position.

```
gringorten <- function(data)
{
  grig <- (rank(data)-0.44)/(length(data)+0.12)
  return(grig)
}

grig <- gringorten(peakflow.sort)
plot(peakflow.sort,grig,xlab="Peak Flow (cfs)",ylab="Probability",col="red")
title("Prob Dis. of Annual Discharge Using Gringorten")
```

## Prob Dis. of Annual Discharge Using Gringorten

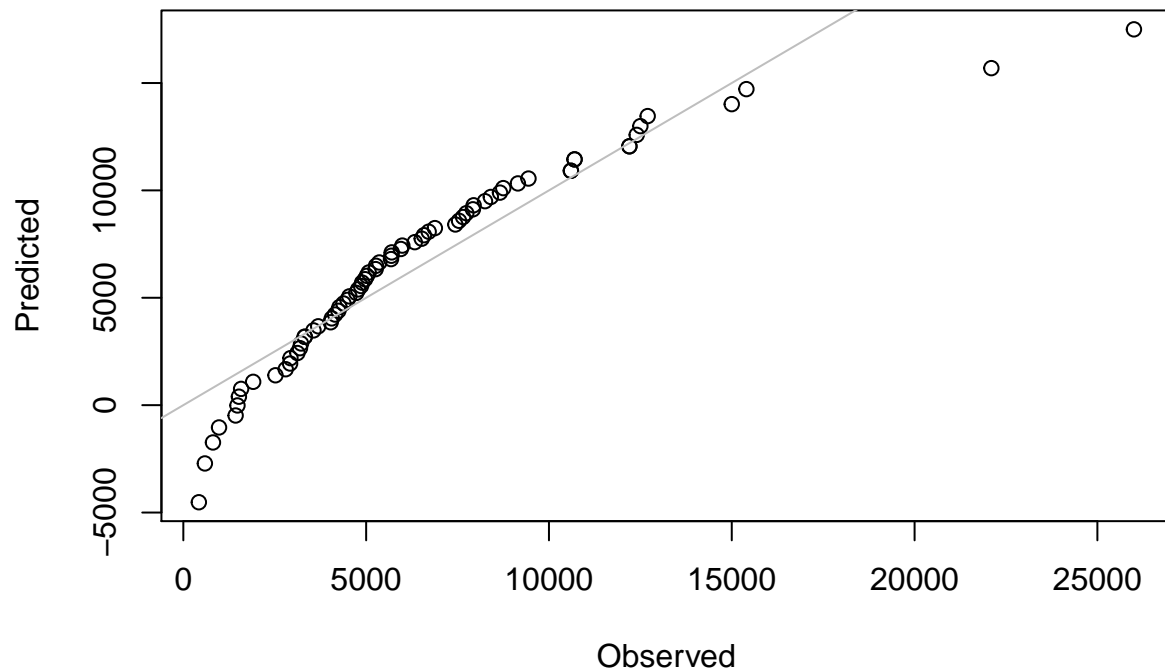


d) Assuming stationarity, develop theoretical cumulative distribution functions for the annual peak flows assuming Normal, Log-normal, Gumbel and Generalized Extreme Value (GEV) distributions

```
# Normal Distrubution
```

```
sf.nd<-fitdistr(peakflow.sort,"normal")  
q.nd<-qnorm(grig,sf.nd$estimate[1],sf.nd$estimate[2])  
plot(peakflow.sort,q.nd,xlab="Observed",ylab="Predicted")  
abline(0,1,col="gray")  
title("Peak Flow (cfs); Observed vs. Predicted by Normal Distribution")
```

## Peak Flow (cfs); Observed vs. Predicted by Normal Distribution

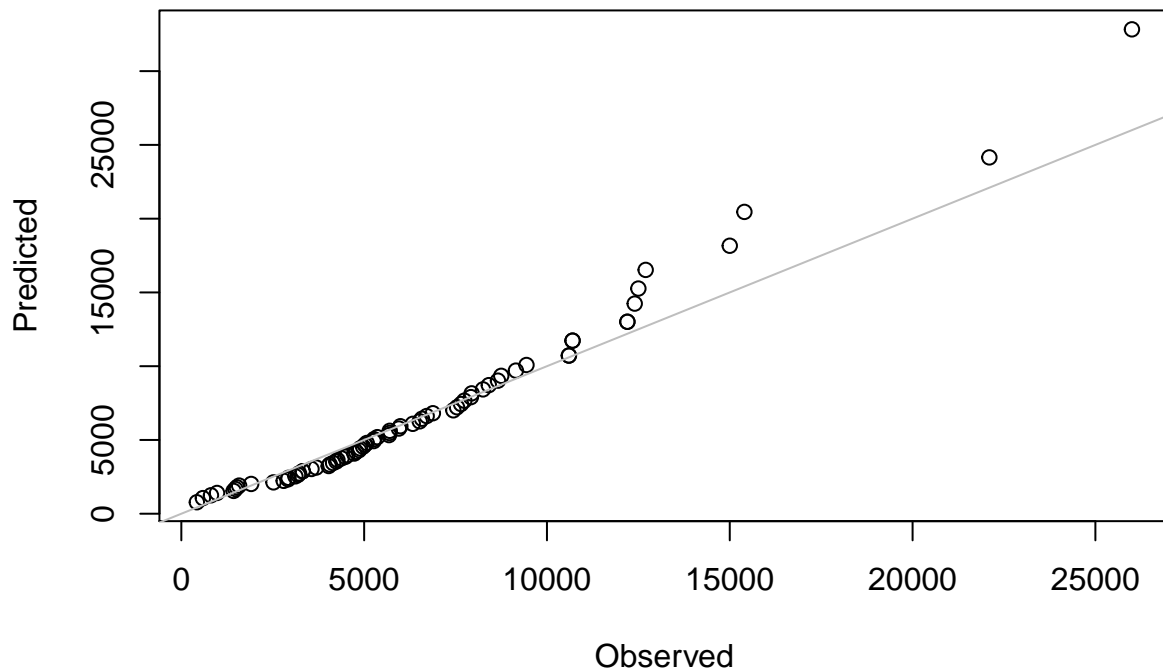


```
# Log-Normal Distribution
```

```
sf.ln<-fitdistr(peakflow.sort,"Log-normal")  
q.ln<-qlnorm(grig,sf.ln$estimate[1],sf.ln$estimate[2])  
plot(peakflow.sort,q.ln,xlab="Observed",ylab="Predicted")  
abline(0,1,col="gray")  
title("Peak Flow (cfs); Observed vs. Predicted by Log-Normal Distribution")
```



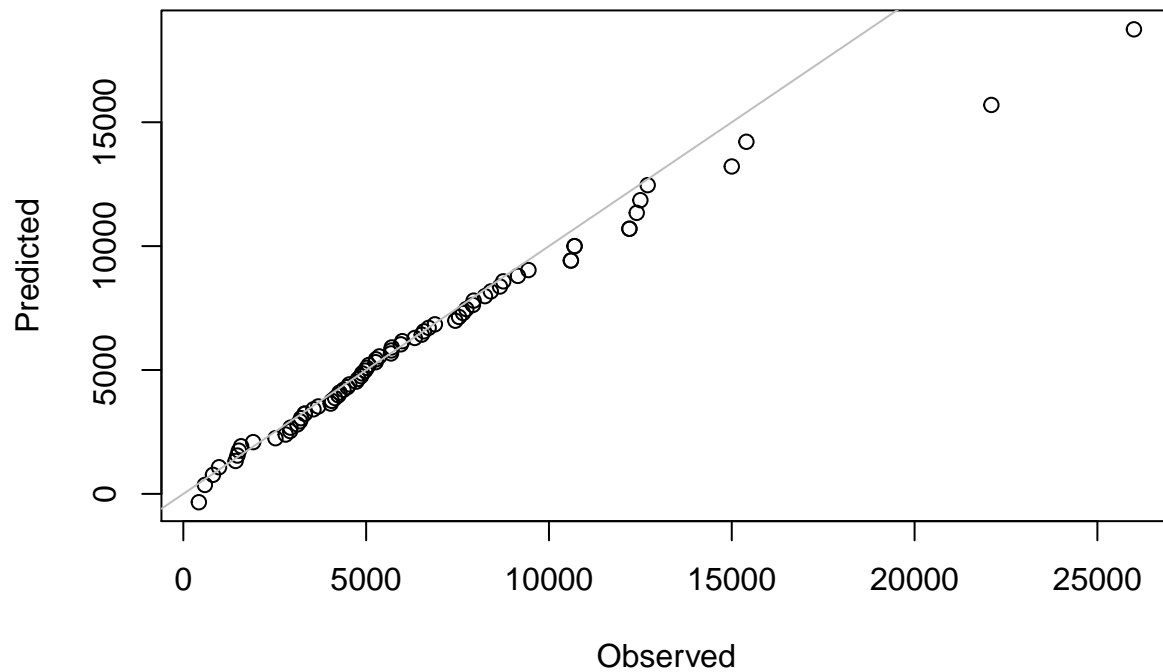
## Peak Flow (cfs); Observed vs. Predicted by Log-Normal Distribution



```
# Gumbel
```

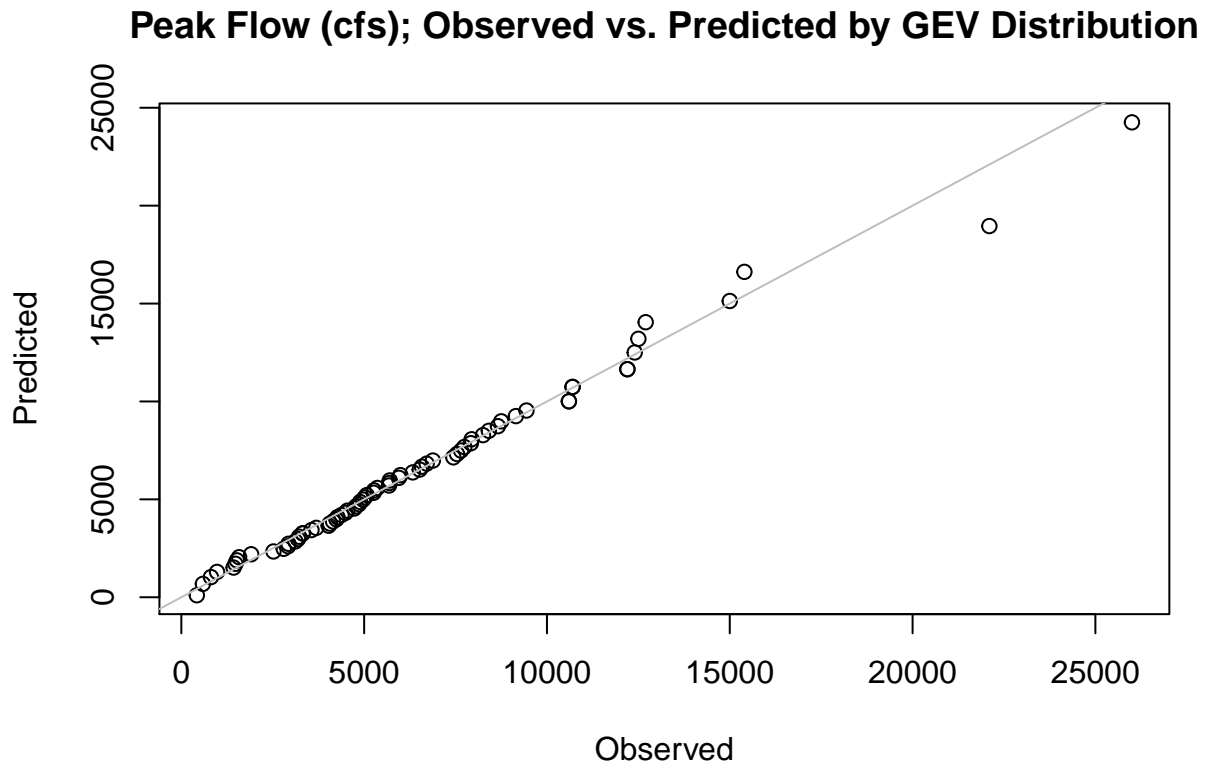
```
sf.gum <- fevd(peakflow,location.fun=~1,scale.fun=~1,shape.fun=0,type='Gumbel')
q.gum<-qevd(grig,sf.gum$results$par[1],sf.gum$results$par[2],type='Gumbel')
plot(peakflow.sort,q.gum,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by Gumbel Distribution")
```

## Peak Flow (cfs); Observed vs. Predicted by Gumbel Distribution



```
# GEV

sf.evd <- fevd(peakflow,location.fun=~1,scale.fun=~1,shape.fun=0,type='GEV')
q.gev <- qevd(grig,sf.evd$results$par[1],sf.evd$results$par[2],
              sf.evd$results$par[3],type='GEV')
plot(peakflow.sort,q.gev,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by GEV Distribution")
```



According to the presented plots corresponding to each distribution model, GEV distribution model seems to be better fit to the observed data rather than the other three models. The GEV model has a good fit for late data as well.

e) Assuming stationarity what are the discharges corresponding to 10, 50 and 100 year floods

The results for all the flood years and different distribution models has been presented as a data frame as following:

```
floods_cfs
```

```
##          10Year   50Year  100Year
## q_nd  12308.00  15814.21  17051.97
## q_ln  13587.90  24657.51  30430.49
## q_gum 11004.84  15890.77  17956.33
## q_gev 12049.49  19273.72  22812.05
```

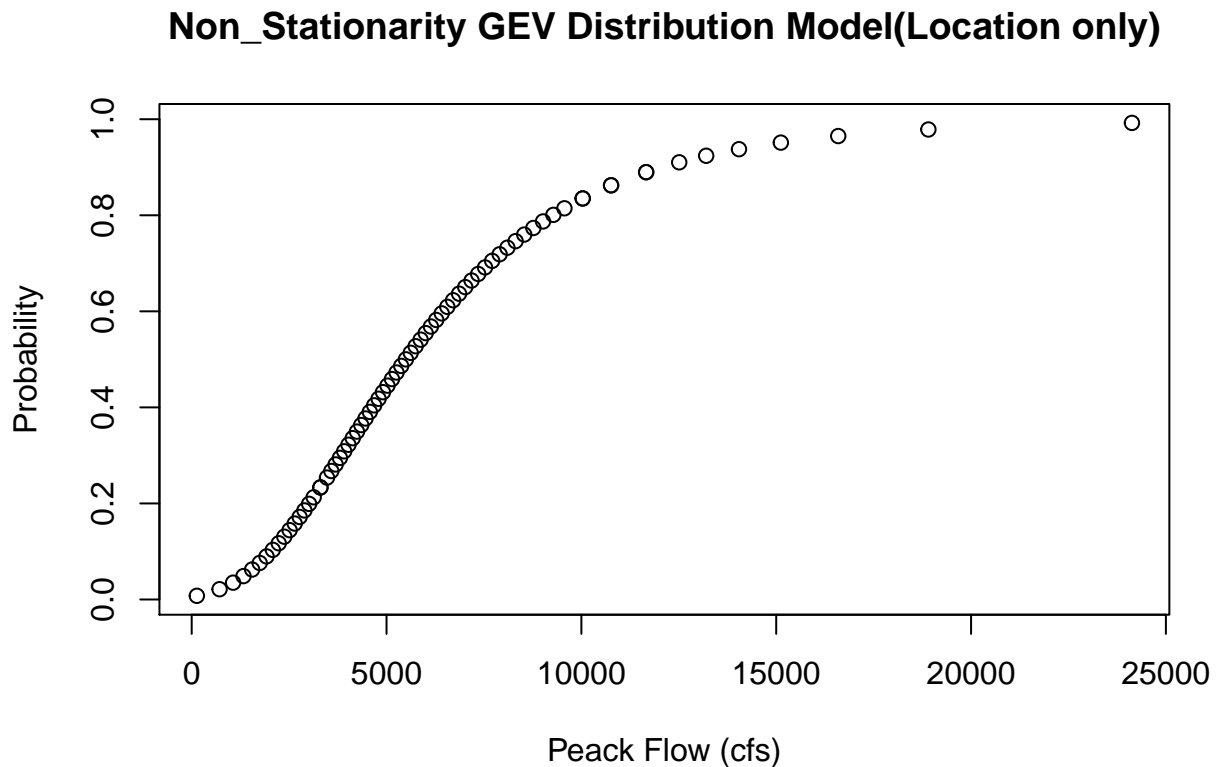
f) Fit a non-stationary GEV distribution assuming a linear temporal variation in

- Location alone
- Location and scale parameters

```

evd.fit.ns1 <- fevd(peakflow,dat,location.fun = ~hydyear)
loc1 <- evd.fit.ns1$results$par[1]+evd.fit.ns1$results$par[2]*hydyear
quan1 <- vector()
for (i in seq(1,length(peakflow),1)){
  quan1[i] <- qevd(grig[i],loc=loc1[i],evd.fit.ns1$results$par[3],
    evd.fit.ns1$results$par[4],type='GEV')
}
plot(quan1,grig,xlab="Peack Flow (cfs)",ylab="Probability")
title("Non_Stationarity GEV Distribution Model(Location only)")

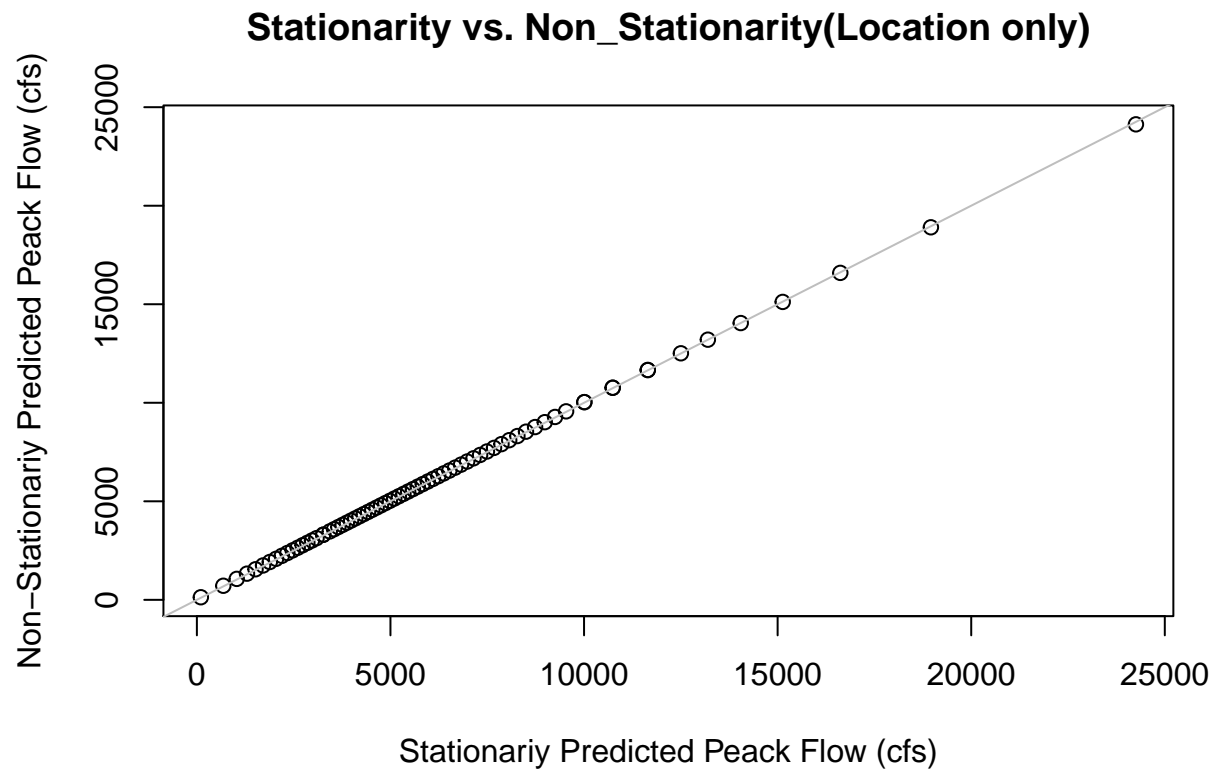
```



```

plot(q.gev,quan1,xlab="Stationariy Predicted Peack Flow (cfs)"
  ,ylab="Non-Stationariy Predicted Peack Flow (cfs)")
abline(0,1,col="gray")
title("Stationarity vs. Non_Stationarity(Location only)")

```



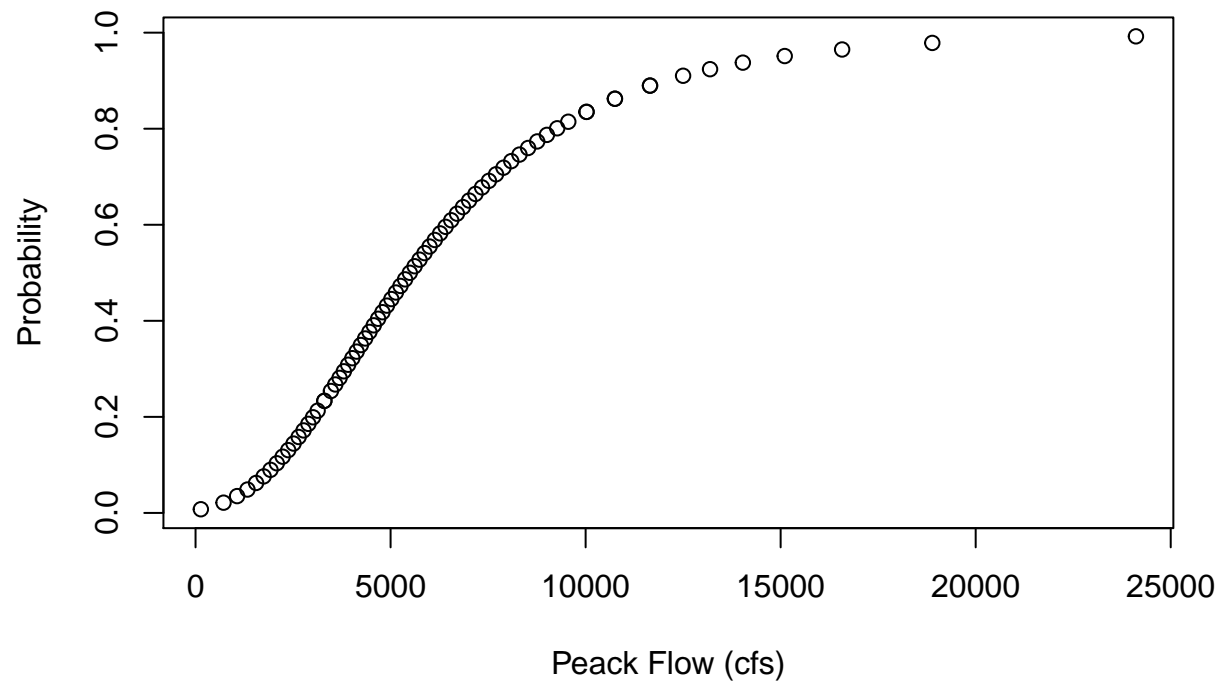
The comparison between nonstationary and stationary GEV models shows similar trend for both. Visually the fit on the 45 degree line which means they are close to each other. However, they have differences which cannot be seen visually.

The next part is the model of nonstationarity when location and scale parameters are a linear temporal functions:

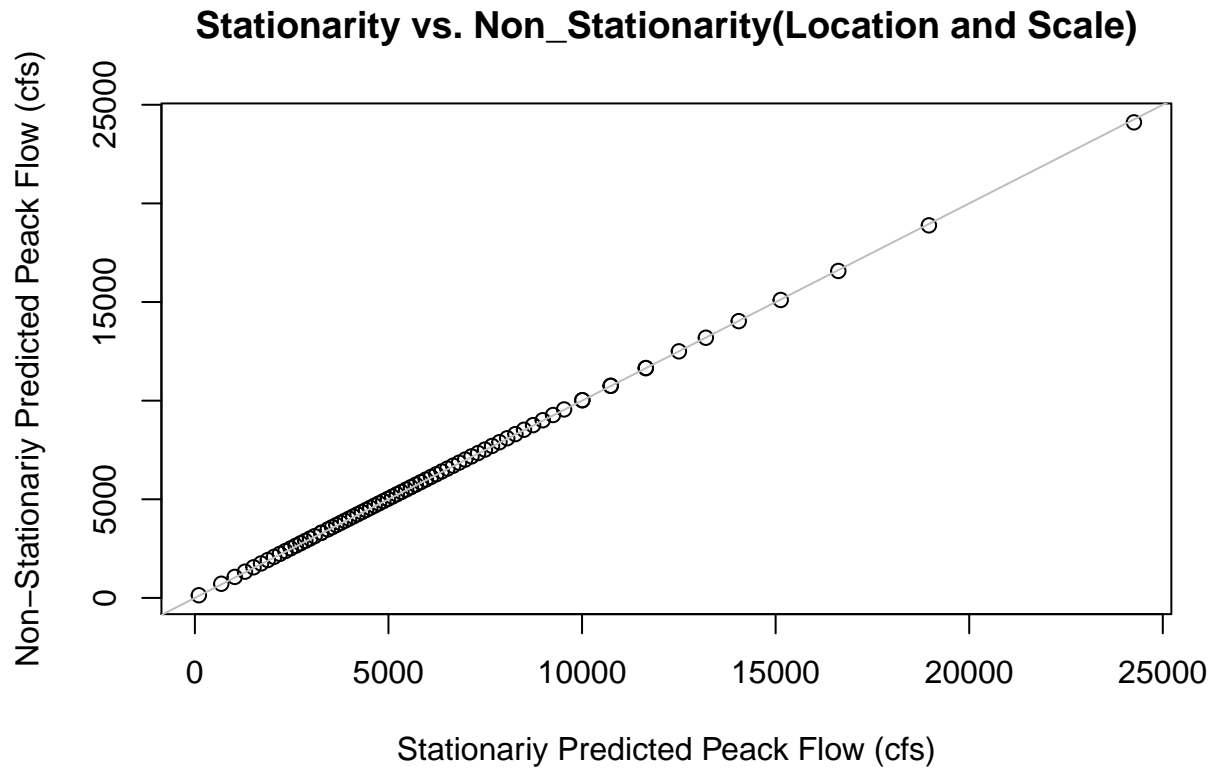
```
evd.fit.ns2 <- fevd(peakflow,dat,location.fun = ~hydyear,scale.fun = ~hydyear)
loc2 <- evd.fit.ns2$results$par[1]+evd.fit.ns2$results$par[2]*hydyear
sca <- evd.fit.ns2$results$par[3]+evd.fit.ns2$results$par[4]*hydyear
quan2 <- vector()
for (i in seq(1,length(peakflow),1)){
  quan2[i] <- qevd(grig[i],loc=loc2[i],scale=sca[i],
    evd.fit.ns2$results$par[5],type='GEV')
}

plot(quan2,grig,xlab="Peack Flow (cfs)",ylab="Probability")
title("Non_Stationarity GEV Distribution Model(Location and scale)")
```

## Non\_Stationarity GEV Distribution Model(Location and scale)



```
plot(q.gev,quan2,xlab="Stationariy Predicted Peack Flow (cfs)",  
     ylab="Non-Stationariy Predicted Peack Flow (cfs)")  
abline(0,1,col="gray")  
title("Stationarity vs. Non_Stationarity(Location and Scale)")
```



Same as previous case, the nonstationary GEV model when location and scale parameters are changing by time is also very close to the stationary model. The following plot shows all the models and empirical data models:

```

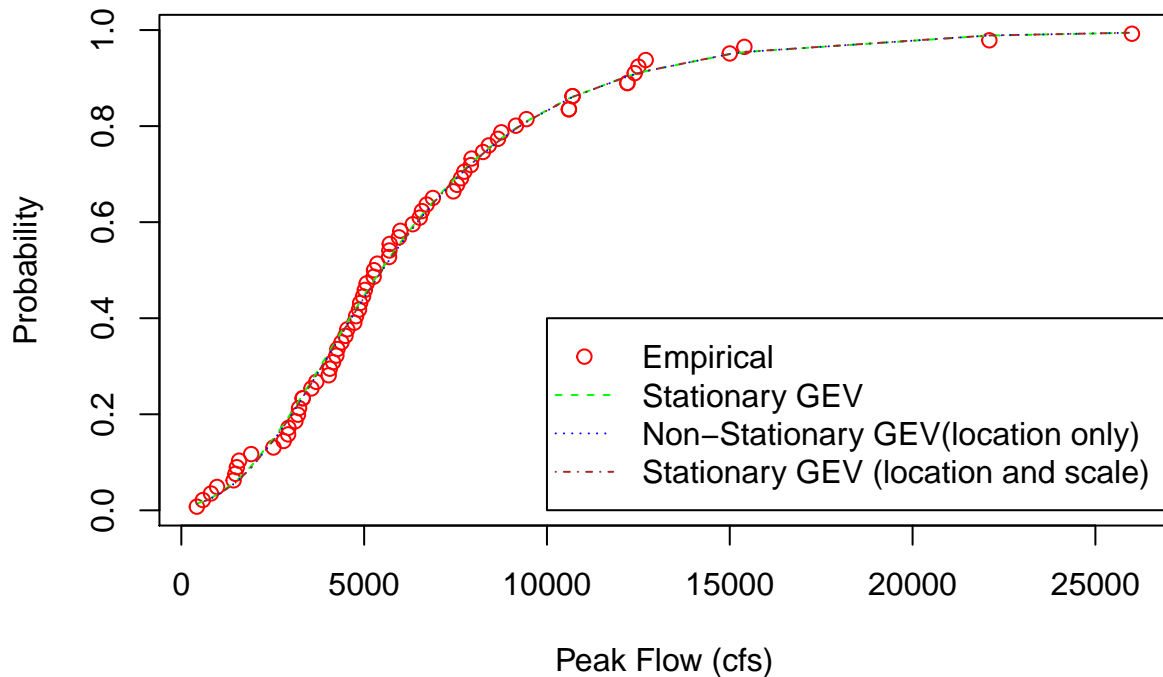
prob1 <- vector()
prob2 <- vector()
prob3 <- vector()

for (i in seq(1,length(peakflow.sort),1)){
  prob1[i] <- pevd(peakflow.sort[i],sf.evd$results$par[1],
                  sf.evd$results$par[2],sf.evd$results$par[3],type='GEV')
  prob2[i] <- pevd(peakflow.sort[i],loc=loc1[i],
                  evd.fit.ns1$results$par[3],evd.fit.ns1$results$par[4],type='GEV')
  prob3[i] <- pevd(peakflow.sort[i],loc=loc2[i],
                  scale=sca[i],evd.fit.ns2$results$par[5],type='GEV')
}

plot(peakflow.sort,grig,xlab="Peak Flow (cfs)",ylab="Probability",col="red",pch=1)
lines(peakflow.sort,prob1,xlab="Peak Flow (cfs)",ylab="Probability",col="green",lty=2)
lines(peakflow.sort,prob2,xlab="Peak Flow (cfs)",ylab="Probability",col="blue",lty=3)
lines(peakflow.sort,prob3,xlab="Peak Flow (cfs)",ylab="Probability",col="brown",lty=4)
legend(1e4,0.4,legend=c("Empirical","Stationary GEV","Non-Stationary GEV(location only)",
                        "Stationary GEV (location and scale)"),

```

```
lty=c(NA,2,3,4),col=c('red','green','blue','brown'),
pch=c(1,NA,NA,NA))
```



g) What is the probability that a flood exceeding a stationary 50 year return period occurs once during a 5 year design period? Compute the probabilities under stationary and nonstationary conditions described in (f)

```
Probability <- dbinom(1, size=5, prob=0.02)
```

```
Probability
```

```
## [1] 0.09223682
```

The probability of that 50 year return period flood occurs once during a 5 year design period is about 9%. I have used a built in function for binomial distribution which can be useful in obtaining this probability. The probability of 2 % has been obtained from 50 years flood and *size=5* is the design period.

---

### Citation:

Code for question#2 was developed from R-script provided by Dr. Uddameri (Blackboard) and collaborative efforts of Gladys, Farhang, Jordan, Kushal and Yosselyn.



### Question 3:

a) Construct a flow duration curve and calculate

- The maximum flow rate that is not exceeded 66 % of the time
- The exceedance corresponding to the peak annual flood in the year 2011 (the driest single year in the recorded history of Texas)

The Flow Duration Curve has been plotted as following:

```
a <- read.csv('Dailydata.csv')
sf.data <- a[,4]
date.x <- a[,3]
date <- as.Date(date.x,"%m/%d/%Y")

# Create a flow duration curve
summary(sf.data) # check to see if there are any zeros

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0     14.0     37.0   206.7   109.0 15600.0

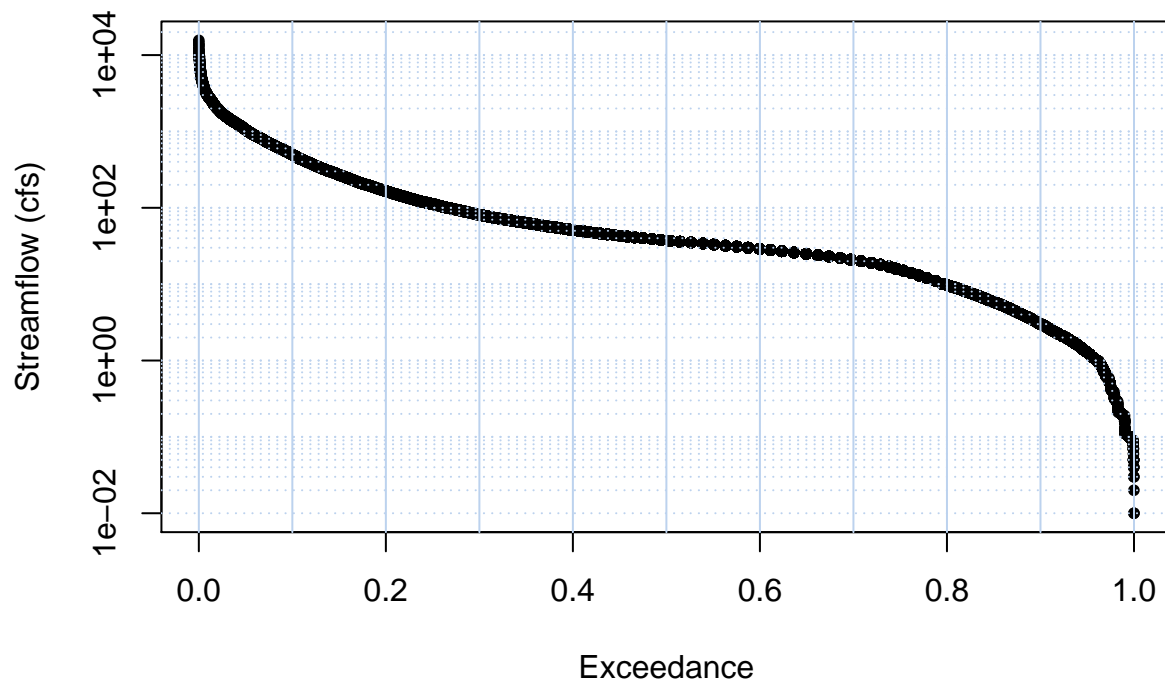
# Remove zeros
sf.data1 <- sf.data[ sf.data != 0 ]

# Calculate exceedance using Gringorten Formula
N <- length(sf.data1)
sf.order <- rank(sf.data1)
exceed <- 1 -(sf.order-0.44)/(N+0.12)

#Make a semi-log plot
plot(exceed,sf.data1,log="y",ylab='Streamflow (cfs)',xlab='Exceedance',pch=20,
     main="Flow Curve Duration")

## Put grid lines on the plot, using a light blue color ("lightsteelblue2")
# Put grid lines on the Y-axis
ypts <- seq(-2,5,1)
for(i in seq(1,length(ypts)-1,1))
{
    seqa <- seq(10^ypts[i],10^ypts[i+1],10^ypts[i])
    abline( h = seqa, lty = 3, col = colors()[ 440 ] )
}
# Put grid lines on X-axis
xpts <- seq(0.0,1.0,0.1)
for(j in seq(1,length(xpts),1))
{
    abline(v=xpts[j],col=colors()[440])
}
```

## Flow Curve Duration



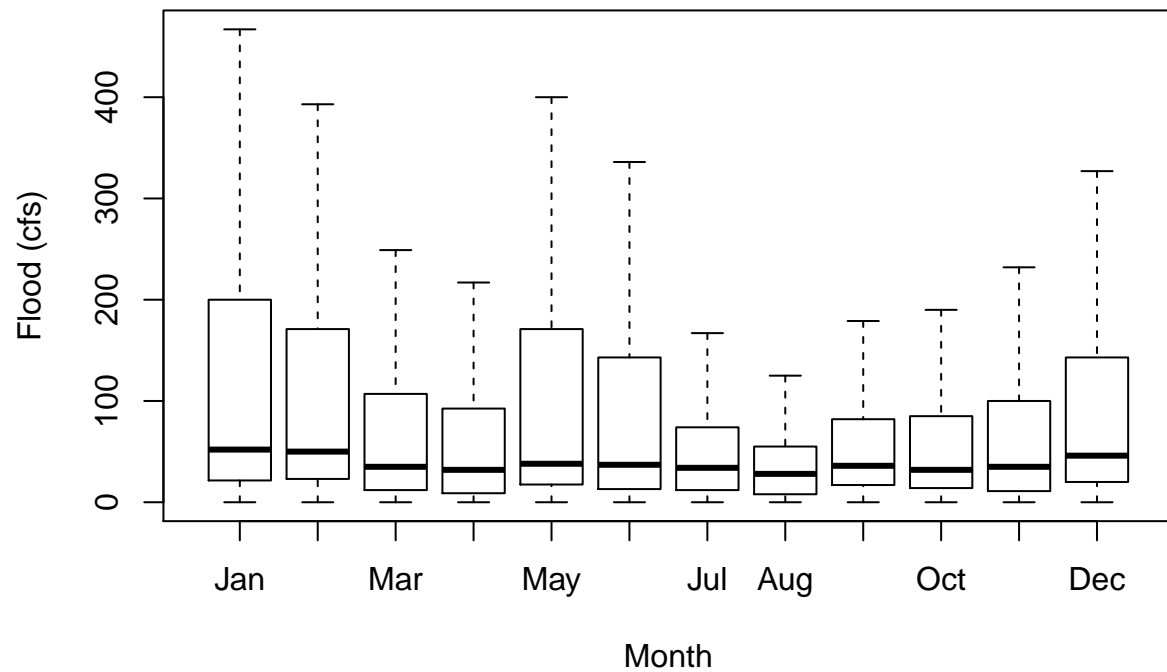
Based on the flow duration curve, the maximum flow rate that is not exceeding 66 % of the time is approximately 20 cfs

The maximum flow that occurs in 2011 is 1780 cfs. The corresponding exceedance to this flow is approximately 3 % or 0.03 based on flow duration curve.

b) Make a box-plot depicting the intra-annual and intra-monthly variations in streamflows. Which months have the highest variability and why?

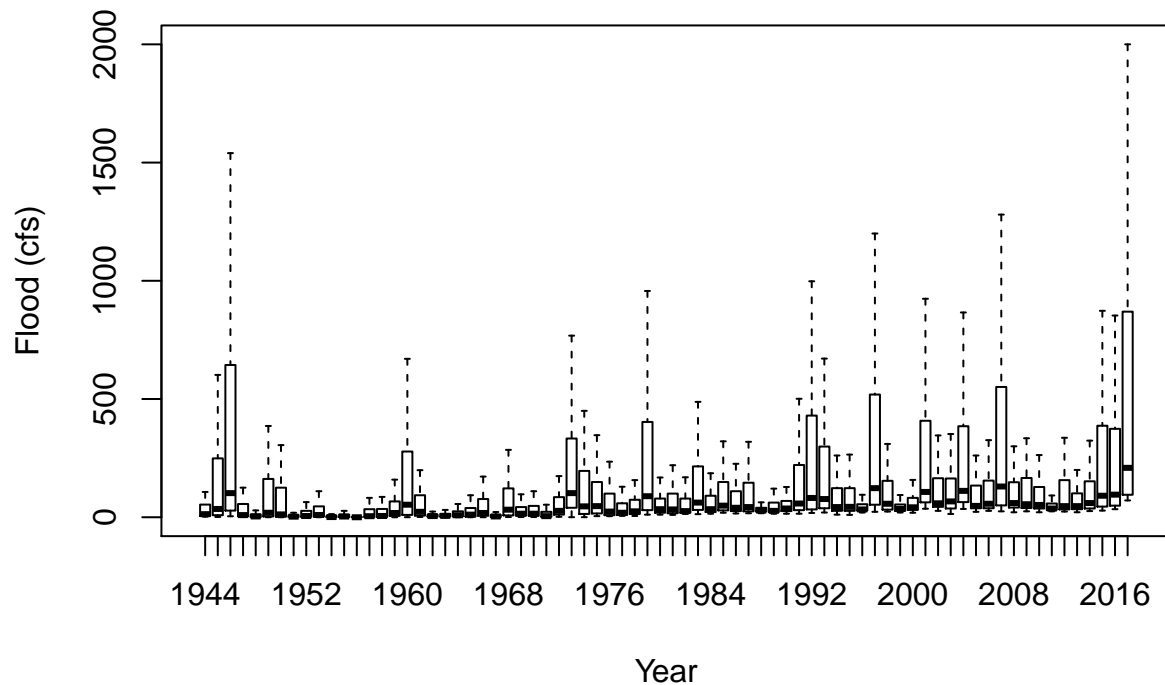
```
mon <- format(date,"%b")
month <-factor(mon,levels=c('Jan','Feb','Mar','Apr','May','Jun',
                           'Jul','Aug','Sep','Oct','Nov','Dec'))
boxplot(as.numeric(sf.data)~month,outline=F,main="intra-monthly variation boxplot",
        xlab="Month",ylab="Flood (cfs)")
```

### intra-monthly variation boxplot



```
Year <- format(date,"%Y")
boxplot(as.numeric(sf.data)~Year,outline=F,main="intra-mannual variation boxplot",
        xlab="Year",ylab="Flood (cfs)")
```

### intra-mannual variation boxplot

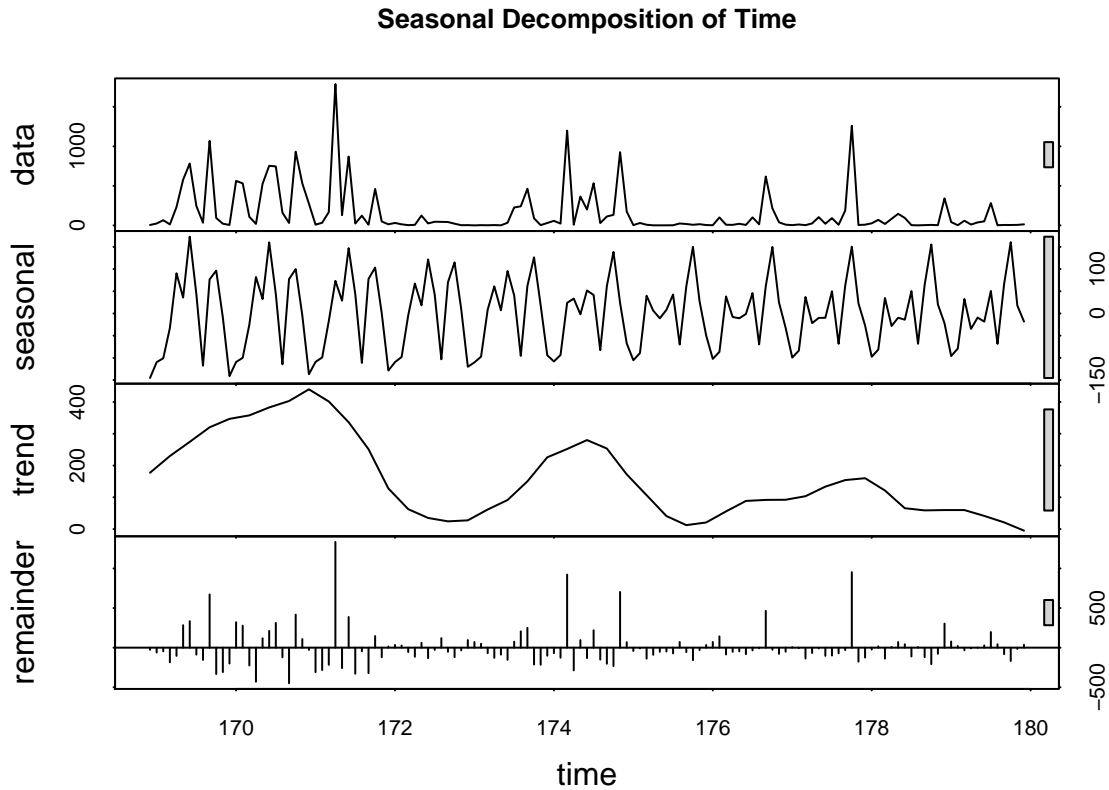


The boxplot for intra monthly variations shows high variability for January. January is in winter, it is possible that some years experience very high flood and on the other hand some years may drought happens, causing high variation in the pick flow data.

c) Decompose the data into seasonal, trend and noise components. Are the seasonal and trend components able to explain most of the variations in data?

```
tss <- xts(sf.data,date)
ts_m = as.vector(apply.monthly(tss, mean))
Q.ts <- ts(ts_m,start=c(07,1944),end=c(12,2016),frequency=12)
SDT <- stl(Q.ts,s.window = 12)

plot(SDT,main="Seasonal Decomposition of Time")
```



Based on the plot of Seasonal Decomposition of Time, trend can show variation in flood better than seasonal plot. The trend plot is a pretty good representative variation in the data and it follows the same pattern which data has. But, seasonal plot does not show the variation in the flood data because it's pattern does not have similarity to the data plot and shows a repetitive set of fluctuations. For example at recent time, the annual trend is showing decrease in flood magnitude while seasonal trend is unable to represent this drought. There are some other plots and analysis for this case that has been coded and is presented in the Appendix

---

### Citation:

Code for question#3 was developed from R-script provided by Dr. Uddameri (Blakboard) and collaborative efforts of Gladys, Farhang, Jordan, Kushal and Yosselyn.

## Question 4

Read at least the paper assigned to you from below:

Vogel, Richard M., Chad Yaindl, and Meghan Walter. *“Nonstationarity: flood magnification and recurrence reduction factors in the United States1.”* (2011): 464-474. (GROUP B)

Develop a 2 page report to answer the following questions:

a) Develop 5 bullets each per paper summarizing the main points of these papers.

- Most of the recent work in water resource consider climate change as the most important and the only factor effecting planning of trends. This is while this article’s study takes a different approach and consider the anthropogenic influences as contributing parameters in water resource planning and evaluate the impact of trends in flood series on flood frequency analysis.
- A statistical model has been suggested by this article which which consider nonstationarity condition and can be used to model flood time series and frequency.
- Nonstationarity in floods can result from a variety of anthropogenic processes including changes in land use, climate, and water use, with likely interactions among those processes making it very difficult to attribute trends to a particular cause
- Based on the report has been done in this study, two parameter Log-normal probability distribution function is amoung the best approximations to annual maximum flood seris. The reason is this pdf doen not exhibit significant bias in future flood frequencies. Another reason is when the two parameter Log-normal combined with a log-linear trend model produces a simple and elegant nonstationarity model of flood frequency which generates insights into the impact of flood trends on various flood statistics.
- Based on this paper, it can be realized that other more complex flood frequency models can results in change in determination of the impact of flood trends on design flood estimate. This paper did not investigate the role of flood frequency distribution model.

b) What is the main take-away message of the paper?

The most important outcome of reading this paper was differentiation between stationarity and nonstationarity models. Many previous studies consider simple stationarity models while this assumption is not correct any more. By increase in humman influences on the land, water and climate, nonstationary models should be applied in water resource planning.

c) What techniques were used in the paper?

1. Are you familiar with the techniques presented in the paper?
2. Will you be able to perform similar research in data/resources were made available?

To describe the relationship between flood magnitude and its frequency, a probabilistic model has been suggested in this article. This model has utilized a two parameter Log-normal probability distribution fucntion model and based on that a trend model for annual maximum flow has been generated. A nonstationary flood frequency model was approached by the article to describe the relationship between flood frequency and flood magnitude.

Some of the techniques were similar to what has been done in this assignment, such as fitting distribution model. It also included some advanced techniques in nonstationary model generation that was new to me.

Similar analysis to what has been done in this article was performed in this assignment. We tried to find the best fit for the distribution model and to develop also a nonstationary distribution function model. The method that was presented in this paper also can be utilized to do analysis on flood magnitude and frequency trend.

d) Are the techniques used suitable for the study?

As article declared, no sensitivity analysis was performed on the flood frequency distribution model to analyze the role of it on flood trends. But generally the idea of nonstationarity that has been applied in developing the model is suitable for flood trend analysis on the contrary of the most previous model which they assume stationarity.

e) Are you comfortable with the level of mathematics and techniques presented in the paper? What do you think you need to do to improve your level of comfort?

The level of mathematics presented in this article was not too high. This level is what we had worked with and seen before. Although, discussing paper in the group made it much more easier to understand. I need to focus more on the concept of hydrology to be more comfortable with the mathematics since I am new to this field of research. Some advanced skills in statistics also needed to be able to fully understand, analyze and evaluate the mathematical concept in this area. Working on the similar data sets and writing programs for doing analysis by our own can be a valuable practice helps us improve our skills.

f) Are the tables and Figures informative? Why? Why not?

The article included one table and a couple of plots (some boxplot and some spatial maps). The maps were depicting the location of stations (regulated, non-regulated and HCDN stations) over United States. These maps can give insights of how coverage of the stations is over the country. The other boxplots were very useful to evaluate the variation in different categories of stations.

g) Did the paper provide supplemental information? Did you follow what was in the supplemental information?

To test if the exponential trend model residuals were normally distributed and if logarithms of the annual maximum floods were approximately normally distributed, an automated normality test was presented in appendix of this article. I was not familiar with the statistical tools this article used in the appendix and seems it needs studying some other papers to be able to have good understanding of the concept.

h) Was the abstract self-contained? - Can you suggest a way to improve it.

Reading the abstract only cannot provide the details of mathematics concepts and approaches were used in the article. But generally, the abstract of this article was able to provide reader a general idea about the major approaches utilized in the article. It can be improved by providing a final conclusion of the paper at the end of the abstract. By doing this, reader can get the main outcome of the article.

i) What skills/information do you think you are missing that is making it difficult to read these papers?

In my case I need to put more effort to learn the concepts in this area and also I need to improve my knowledge of statistics since this subject is involving in statistical concepts a lot.

j) What are you going to do to improve your skills?

The biggest step I had to improve my skills was taking the Surface Water Hydrology course. This course puts coding skills and statistical concepts in practice which can help a lot. I would suggest this course continue engaging students with coding since it is a very practical tool for data analysis. I also would like to take some more courses in statistics. It would help me improve my knowledge of statistical skills and tools.

k) Do you think engineers should read this journal article or such as these? Why? Why Not?

Yes, as a hydrologic student we supposed to read such kind of article which are proposing new approaches to water resource planning.

---

### **Citation:**

Code for question#2 was developed from R-script provided by Dr. Uddameri (Blackboard) and collaborative efforts of Gladys, Farhang, Jordan, Kushal and Yosselyn.



## Appendix

### The written code for question 2

```
### Q2

# Load Libraries

library(MASS)
library(extRemes)
library(fitdistrplus)
library(evd)
library(Kendall)
library(GEVcdn)

## Part (a,b) #####

setwd("C:\\Users\\vahid\\Google Drive\\GraduateStudies\\Semester 6\\
      SurfaceWaterHydrology\\Module1\\RCode")

pf <- read.csv("PeakWest.csv")
hydyear <- pf[,5]
peakflow <- pf[,4]
dat <- as.data.frame(cbind(hydyear,peakflow))
peakflow.sort <- sort(peakflow)
plot(hydyear,peakflow,ylab="Peak Flow (cfs)",xlab="Year",col="blue")
lines(hydyear,peakflow,type="o",lty=2,col="red")
title("Temporal Peak Flow")
lm <- lm(peakflow ~ hydyear)
abline(lm,col="blue")
summary(lm)
Mk.test <- MannKendall(peakflow)

#Plot the boxplot
boxplot(peakflow.sort,ylab='Peakflow (cfs)')
grid()

## Part (c) #####

gringorten <- function(data)
{
  grig <- (rank(data)-0.44)/(length(data)+0.12)
  return(grig)
}

grig <- gringorten(peakflow.sort)
plot(peakflow.sort,grig,xlab="Peak Flow (cfs)",ylab="Probability",col="red")
title("Prob Dis. of Annual Discharge Using Gringorten")
```

```

## Part (d) #####

par(mfrow=c(2,2))
# Normal Distrubution

sf.nd<-fitdistr(peakflow.sort,"normal")
q.nd<-qnorm(grig,sf.nd$estimate[1],sf.nd$estimate[2])
plot(peakflow.sort,q.nd,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by Normal Distribution")

# Log-Normal Distribution

sf.ln<-fitdistr(peakflow.sort,"Log-normal")
q.ln<-qlnorm(grig,sf.ln$estimate[1],sf.ln$estimate[2])
plot(peakflow.sort,q.ln,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by Log-Normal Distribution")

# Gumbel

sf.gum <- fevd(peakflow,location.fun=~1,scale.fun=~1,shape.fun=0,type='Gumbel')
q.gum<-qevd(grig,sf.gum$results$par[1],sf.gum$results$par[2],type='Gumbel')
plot(peakflow.sort,q.gum,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by Gumbel Distribution")

# GEV

sf.evd <- fevd(peakflow,location.fun=~1,scale.fun=~1,shape.fun=0,type='GEV')
q.gev <- qevd(grig,sf.evd$results$par[1],sf.evd$results$par[2],
              sf.evd$results$par[3],type='GEV')
plot(peakflow.sort,q.gev,xlab="Observed",ylab="Predicted")
abline(0,1,col="gray")
title("Peak Flow (cfs); Observed vs. Predicted by GEV Distribution")

## Part (e) #####

Fx <- function(YEAR){
  rp <- 1-1/YEAR
  return(rp)
}

Y <- c(10,50,100)
q_nd <- vector()
q_ln <- vector()
q_gum <- vector()
q_gev <- vector()

for (i in Y){

```

```

q_nd <- c(q_nd,qnorm(Fx(i),sf.nd$estimate[1],sf.nd$estimate[2]))      # normal
q_ln <- c(q_ln,qlnorm(Fx(i),sf.ln$estimate[1],sf.ln$estimate[2]))    # log-norm
q_gum <- c(q_gum,qevd(Fx(i),sf.gum$results$par[1],
                      sf.gum$results$par[2],type='Gumbel'))          # Gumbel
q_gev <- c(q_gev,qevd(Fx(i),sf.evd$results$par[1],
                      sf.evd$results$par[2],
                      sf.evd$results$par[3],type='GEV'))             # GEV
}

floods_cfs <- rbind(q_nd,q_ln,q_gum,q_gev)
colnames(floods_cfs) <- c("10Year","50Year","100Year")

## Part (f) #####

evd.fit.ns1 <- fevd(peakflow,dat,location.fun = ~hydyear)
evd.fit.ns2 <- fevd(peakflow,dat,location.fun = ~hydyear,scale.fun = ~hydyear)

loc1 <- evd.fit.ns1$results$par[1]+evd.fit.ns1$results$par[2]*hydyear
loc2 <- evd.fit.ns2$results$par[1]+evd.fit.ns2$results$par[2]*hydyear
sca <- evd.fit.ns2$results$par[3]+evd.fit.ns2$results$par[4]*hydyear

quan1 <- vector()
quan2 <- vector()

for (i in seq(1,length(peakflow),1)){
  quan1[i] <- qevd(grig[i],loc=loc1[i],evd.fit.ns1$results$par[3],
                  evd.fit.ns1$results$par[4],type='GEV')
  quan2[i] <- qevd(grig[i],loc=loc2[i],scale=sca[i],
                  evd.fit.ns2$results$par[5],type='GEV')
}

par(mfrow=c(2,2))
plot(q.gev,quan1,xlab="Stationariy Predicted Peack Flow (cfs)",
     ylab="Non-Stationariy Predicted Peack Flow (cfs)")
abline(0,1,col="gray")
title("Stationarity vs. Non_Stationarity(Location only)")
plot(q.gev,quan2,xlab="Stationariy Predicted Peack Flow (cfs)",
     ylab="Non-Stationariy Predicted Peack Flow (cfs)")
abline(0,1,col="gray")
title("Stationarity vs. Non_Stationarity(Location and Scale)")

## Part (g) #####

###Determine Magnitude of 50 year flood

Probability <- dbinom(1, size=5, prob=0.02)

conP <- 1-(1-1/50)^5

```

```

prob1 <- vector()
prob2 <- vector()
prob3 <- vector()

for (i in seq(1,length(peakflow.sort),1)){
  prob1[i] <- pevd(peakflow.sort[i],sf.evd$results$par[1],
                  sf.evd$results$par[2],sf.evd$results$par[3],type='GEV')
  prob2[i] <- pevd(peakflow.sort[i],loc=loc1[i],
                  evd.fit.ns1$results$par[3],evd.fit.ns1$results$par[4],type='GEV')
  prob3[i] <- pevd(peakflow.sort[i],loc=loc2[i],
                  scale=sca[i],evd.fit.ns2$results$par[5],type='GEV')
}

plot(peakflow.sort,grig,xlab="Peak Flow (cfs)",ylab="Probability",col="red",pch=1)
lines(peakflow.sort,prob1,xlab="Peak Flow (cfs)",ylab="Probability",col="green",lty=2)
lines(peakflow.sort,prob2,xlab="Peak Flow (cfs)",ylab="Probability",col="blue",lty=3)
lines(peakflow.sort,prob3,xlab="Peak Flow (cfs)",ylab="Probability",col="brown",lty=4)
legend(1e4,0.4,legend=c("Empirical","Stationary GEV","Non-Stationary GEV(location only)",
                        "Stationary GEV (location and scale)"),
      lty=c(NA,2,3,4),col=c('red','green','blue','brown'),
      pch=c(1,NA,NA,NA))

```

### The written code for question 3

```
# Q3

# Load libraries
library(zoo)
library(forecast)
library(xts)

# Set working Directory
setwd("C:\\Users\\vahid\\Google Drive\\GraduateStudies\\Semester 6\\
      SurfaceWaterHydrology\\Module1\\RCode")

# Read data and extract relevant portions
a <- read.csv('Dailydata.csv')
sf.data <- a[,4]
date.x <- a[,3]
date <- as.Date(date.x,"%m/%d/%Y")

# Create a zoo object and plot the time-series
sf.zoo <- zoo(sf.data,date)
plot(sf.zoo,xlab="Time", ylab = 'Discharge (cfs)')
grid()

# Look at monthly variation
# extract month and bind it to the discharge data
mon <- format(date,"%b")
sf.mon <- cbind(mon,sf.data)
boxplot(as.numeric(sf.data)~mon,outline=F,ylab='Discharge (cfs)')
# Make sure the months plot as per the calender
month <-factor(mon,levels=c('Jan','Feb','Mar','Apr','May',
                           'Jun','Jul','Aug','Sep','Oct','Nov','Dec'))
boxplot(as.numeric(sf.data)~month,outline=F)

#plot acf and pacf function
par(mfrow=c(2,1))
Acf(sf.zoo,lag=15, main='ACF Daily Streamflow Data')
Pacf(sf.zoo,lag=15, main='PACF Daily Streamflow Data')
par(mfrow=c(1,1))

# aggregate data on a monthly, quarterly and yearly scales
tss <- xts(sf.data,date)
ts_m = as.vector(apply.monthly(tss, mean))
ts_y = as.vector(apply.yearly(tss, mean))
ts_q = as.vector(apply.quarterly(tss, mean))

Q.ts <- ts(ts_m,start=c(07,1944),end=c(12,2016),frequency=12)
SDT <- stl(Q.ts,s.window = 12)

plot(SDT,main="Seasonal Decomposition of Time")
par(mfrow=c(1,1))
acf(Q.ts)
```

```

plot(ts_m)
plot(ts_q)
plot(ts_y)

par(mfcol=c(3,2))
Acf(ts_m,main='Monthly ACF')
Acf(ts_q,main='Quarterly ACF')
Acf(ts_y,main='Yearly ACF')
Pacf(ts_m,main='Monthly PACF')
Pacf(ts_q,main='Quarterly PACF')
Pacf(ts_y,main='Yearly PACF')
par(mfrow=c(1,1))

# Create a flow duration curve
summary(sf.data) # check to see if there are any zeros
# Remove zeros
sf.data1 <- sf.data[ sf.data != 0 ]

# Calculate exceedance using Gringorten Formula
N <- length(sf.data1)
sf.order <- rank(sf.data1)
exceed <- 1 -(sf.order-0.44)/(N+0.12)

#Make a semi-log plot
plot(exceed,sf.data1,log="y",ylab='Streamflow (cfs)',xlab='Exceedance',pch=20)

## Put grid lines on the plot, using a light blue color ("lightsteelblue2")
# Put grid lines on the Y-axis
ypts <- seq(-2,5,1)
for(i in seq(1,length(ypts)-1,1))
{
  seqa <- seq(10^ypts[i],10^ypts[i+1],10^ypts[i])
  abline( h = seqa, lty = 3, col = colors()[ 440 ] )
}
# Put grid lines on X-axis
xpts <- seq(0.0,1.0,0.1)
for(j in seq(1,length(xpts),1))
{
  abline(v=xpts[j],col=colors()[440])
}

```