RWater Module 7
Flood Frequency Analysis
to Analyze the Effect of Climate and Land Use Change on Streamflow
Adnan Rajib and Venkatesh Merwade

Lyles School of Civil Engineering, Purdue University

Learning Goals

Urban development directly effects streamflow by increasing the peak discharge. These changes are becoming more significant with the ongoing process of climatic change. The climate being highly variable in recent times, we have been facing cyclones, tornados, droughts and high intensity rainfalls more frequently. Although climate change is a much slower process compared to urbanization, together they can lead to more frequent high flow events in the streams. The image below shows such an example, being published by USGS. Number of times per year that daily discharge exceeded 1000 cfs on the Northeast Branch of the Anacostia River in Maryland (USGS 01649500) increased from once or twice per year in the 1940s and 1950s to as much as six times per year in the 1990s.
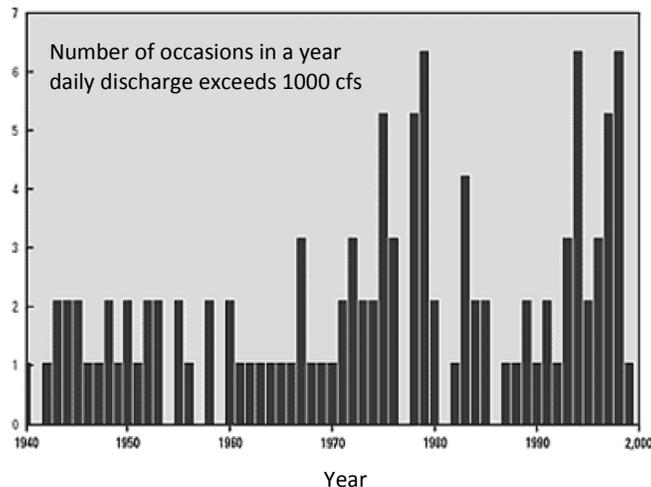

Figure 1. Source: http://pubs.usgs.gov/fs/fs07603/

Under these circumstances, it is important to quantify how frequent an extreme flow condition would be over a designated period of time. The Flood Frequency Analysis is a helpful tool in this regard. This analysis involves fair amount of calculation and data processing, requiring significant level of understanding of the process and expertise. However, this RWater module is designed in a way that after completing this module, students will be able to:

   i.   perform flood frequency analysis using USGS peak flow records by writing RWater scripts without any sort of data pre-processing,

  ii.   understand the concepts of return period and probability of occurrence at which a flow of specific magnitude may occur over a certain period of time, and calculate them for an actual stream location,

 iii.   identify how does the return period of a specific magnitude flood at the same location get altered in course of time as a result of climate and landuse change.

Flood Frequency, Return Period and Probability of Occurrence

The method for Flood Frequency Analysis uses long-term annual peak discharge data to construct a plot, which quantitatively tells the expectancy or likelihood of various discharge values as a function of return period.

The general meaning of return period (also called the recurrence interval) is the average number of years between the occurrences of an extreme event having certain magnitude. Very often we encounter terms like "this city will withstand a 100-year flood". This obviously refers to a specific discharge magnitude where 100 years being its return period. It does not necessarily mean that such a flood is likely to occur only once in every 100 years. Because, return period (T) is by-definition designated with the "probability of occurrence" or the exceedance probability (p) in one year, which is simply the inverse of T. Thus, from the relationship of p and T, there will be a 1-in-100 (1/100 = 0.01 = 1%) chance that a flood this size will happen <u>in any year</u> (USGS Fact Sheet 229-96, http://pubs.usgs.gov/fs/FS-229-96/). Accordingly, in a span of 100 years, such an event may occur once, twice, more, or not at all. Just for an example, consider we are interested to know the chances for a 100-year flood to occur not just in a single year, rather in a period of whole 25 year time.

- Let, '1' be all the possible flows that might occur, over the whole time period (n).
- p is the chance of the particular peak discharge occurring, in any year.
- $(1–p)$ is the chance of that particular flow <u>not</u> occurring, or the non–exceedance probability, in any year.
- $(1–p)^n$ is the non–exceedance probability for that particular flow, over the whole time period. This virtually refers to all the flows that are less than the peak discharge in concern.
- Finally, '1', meaning all possible flows, minus $(1–p)^n$, all flows during the time period that are lower than our flood of interest, leaves us with $1– (1–p)^n$, the probability of those flows of interest occurring over the whole stated time period (n).

So, if we want to calculate the chances for a 100-year flood (p = 0.01) over a 25 year time period (n = 25), we can use the formula: $1– (1–p)^n$. Hence, there is $1– (1–0.01)^{25}$ = 0.22 or 22% chance for that 100-year event to occur once or more in the 25 year span, with (100-22) = 78% chance of not occurring at all.

There are four widely used methods for flood frequency analysis. In this module, we will be using one of those methods called the "Gumbel Distribution".

Flood Frequency Analysis for Blackberry Creek near Chicago City

So far in the previous modules we have only used the daily streamflow data. But, to perform Flood Frequency Analysis, we are going to need the annual peak discharge values. This is basically the maximum discharge that might have taken place in any day within a year. Figure 2 below shows a daily streamflow hydrograph. Letter A on the figure denotes the largest value per year. Only these values (A) comprise a data set of annual peak discharges for subsequent use in flood frequency analysis.

RWater can directly download the annual peak discharge values for any USGS gage station within USA. In this module, we are going to perform flood frequency analysis for the watershed in Chicago area only (USGS 05551700: Blackberry Creek).
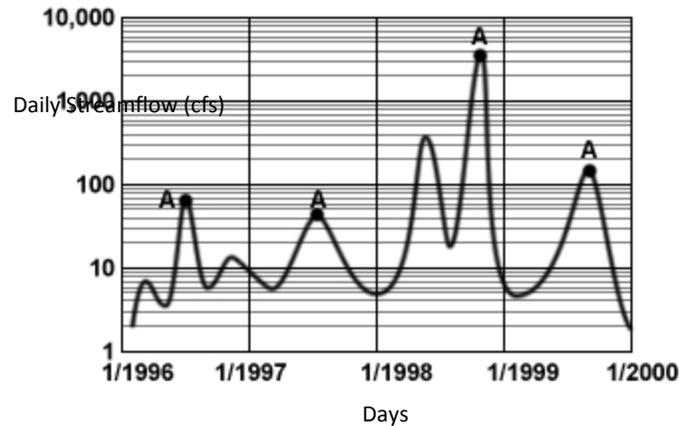
Figure 2. Source: http://comet.ucar.edu/

- Load the script for this module from your working directory.
- Select certain segments of the script and run in steps as shown below.
- **Make changes only in the portion with "XXXX" or as directed**
- Relevant explanations associated with each step are also provided here. These explanatory notes are only for building user's perception over the code.

```
### STEP 1
### Removing previously used scripts from RWater
### Removing all previously generated datasets and plots
cat("\014")
rm(list = ls())
dev.off()
```

A package is a collection of R functions to serve specific analysis purpose. Package names are case sensitive (**dataRetrieval**: **R** is capital letter)

```
### STEP 2
### Loading two specific packages into RWater
library(dataRetrieval)
library(xts)
```

```
### STEP3
### Get the Peak Annual Discharge
mysite<-'XXXX'
annualpeak<-readNWISpeak(mysite)
```

In this example, replace **XXXX** with **05551700**.

readNWISpeak is the function name which pulls the annual peak discharge data from USGS website (case sensitive: **NWIS** in capital letters)

```
### STEP4
### Split the downloaded data into two periods
bb71_90<-subset(annualpeak,
        peak_dt>="YYYY-MM-DD"&peak_dt<="YYYY-MM-DD")
bb91_10<-subset(annualpeak,
        peak_dt>="YYYY-MM-DD"&peak_dt<="YYYY-MM-DD")
```

```
### STEP5
### Split the plot window in two columns
par(mfrow=c(1,2))
```

Provide a range of date in YYYY-MM-DD format. In this example, use 1971-01-01 to 1990-12-31 for 'bb71_90' and 1991-01-01 to 2010-12-31 for 'bb91_10' respectively

peak_dt is the column name that contains the dates in the downloaded file (annualpeak)

```
### STEP6
### Perform Flood Freqency Analysis
### Locate the column of your data set that has the peak discharges
### Click the 'bb71_90' from your workspace (upper right)
### You can see that peak discharges are stored in the 6th column (peak_va)

Q = XXXX
graphlab = "Your Duration"

#Generate plotting positions
n = length(Q)
r = n + 1 - rank(Q)  # highest Q has rank r = 1
T = (n + 1)/r

# Set up x axis tick positions and labels
Ttick =
c(1.001,1.01,1.1,1.5,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,25,30,3
5,40,45,50,60,70,80,90,100)
xtlab =
c(1.001,1.01,1.1,1.5,2,NA,NA,5,NA,NA,NA,NA,10,NA,NA,NA,NA,15,NA,NA,NA,NA,20,N
A,30,NA,NA,NA,50,NA,NA,NA,NA,100)
y = -log(-log(1 - 1/T))
ytick = -log(-log(1 - 1/Ttick))
xmin = min(min(y),min(ytick))
xmax = max(ytick)

# Fit a line by method of moments, along with 95% confidence intervals
KTtick = -(sqrt(6)/pi)*(0.5772 + log(log(Ttick/(Ttick-1))))
QTtick = mean(Q) + KTtick*sd(Q)
nQ = length(Q)
se = (sd(Q)*sqrt((1+1.14*KTtick + 1.1*KTtick^2)))/sqrt(nQ)
LB = QTtick - qt(0.975, nQ - 1)*se
UB = QTtick + qt(0.975, nQ - 1)*se
max = max(UB)
Qmax = max(QTtick)

# Plot peak flow series with Gumbel axis
plot(y, Q,
     ylab = expression( "Annual Peak Flow (cfs)" ) ,
     xaxt = "n", xlab = "Return Period, T (year)",
     ylim = c(0, Qmax),
     xlim = c(xmin, xmax),
     pch = 21, bg = "red",
     main = paste( "XXXX,",graphlab )
)
par(cex = 0.65)
axis(1, at = ytick, labels = as.character(xtlab))

# Add fitted line and confidence limits
lines(ytick, QTtick, col = "black", lty=1, lwd=2)
lines(ytick, LB, col = "blue", lty = 1, lwd=1.5)
lines(ytick, UB, col = "red", lty = 1, lwd=1.5)

# Draw grid lines
abline(v = ytick, lty = 3, col="light gray")
abline(h = seq(500, floor(Qmax), 500), lty = 3,col="light gray")
par(cex = 1)
```

> Replace **XXXX** with **bb71_90$peak_va**. This will command RWater to plot the annual peak discharge values from the **peak_va** column of the 'bb71_90' dataframe
>
> Put 1971-1990 as **graphlab**

> Replace **XXXX** with **Blackberry Creek** (name of your watershed)

```
### STEP7
### Similarly, perform Flood Freqency Analysis for the second time period

Q = XXXX
graphlab = "Your Duration"

#Generate plotting positions
n = length(Q)
r = n + 1 - rank(Q)   # highest Q has rank r = 1
T = (n + 1)/r

# Set up x axis tick positions and labels
Ttick =
c(1.001,1.01,1.1,1.5,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,25,30,3
5,40,45,50,60,70,80,90,100)
xtlab =
c(1.001,1.01,1.1,1.5,2,NA,NA,5,NA,NA,NA,NA,10,NA,NA,NA,NA,15,NA,NA,NA,NA,20,N
A,30,NA,NA,NA,50,NA,NA,NA,NA,100)
y = -log(-log(1 - 1/T))
ytick = -log(-log(1 - 1/Ttick))
xmin = min(min(y),min(ytick))
xmax = max(ytick)

# Fit a line by method of moments, along with 95% confidence intervals
KTtick = -(sqrt(6)/pi)*(0.5772 + log(log(Ttick/(Ttick-1))))
QTtick = mean(Q) + KTtick*sd(Q)
nQ = length(Q)
se = (sd(Q)*sqrt((1+1.14*KTtick + 1.1*KTtick^2)))/sqrt(nQ)
LB = QTtick - qt(0.975, nQ - 1)*se
UB = QTtick + qt(0.975, nQ - 1)*se
max = max(UB)
Qmax = max(QTtick)

# Plot peak flow series with Gumbel axis
plot(y, Q,
     ylab = expression( "Annual Peak Flow (cfs)" ) ,
     xaxt = "n", xlab = "Return Period, T (year)",
     ylim = c(0, Qmax),
     xlim = c(xmin, xmax),
     pch = 21, bg = "red",
     main = paste( "XXXX,",graphlab )
)
par(cex = 0.65)
axis(1, at = ytick, labels = as.character(xtlab))

# Add fitted line and confidence limits
lines(ytick, QTtick, col = "black", lty=1, lwd=2)
lines(ytick, LB, col = "blue", lty = 1, lwd=1.5)
lines(ytick, UB, col = "red", lty = 1, lwd=1.5)

# Draw grid lines
abline(v = ytick, lty = 3, col="light gray")
abline(h = seq(500, floor(Qmax), 500), lty = 3,col="light gray")
par(cex = 1)
```

> Replace **XXXX** with **bb91_10$peak_va**. This will command RWater to plot the annual peak discharge values from the **peak_va** column of the 'bb91_10' dataframe
>
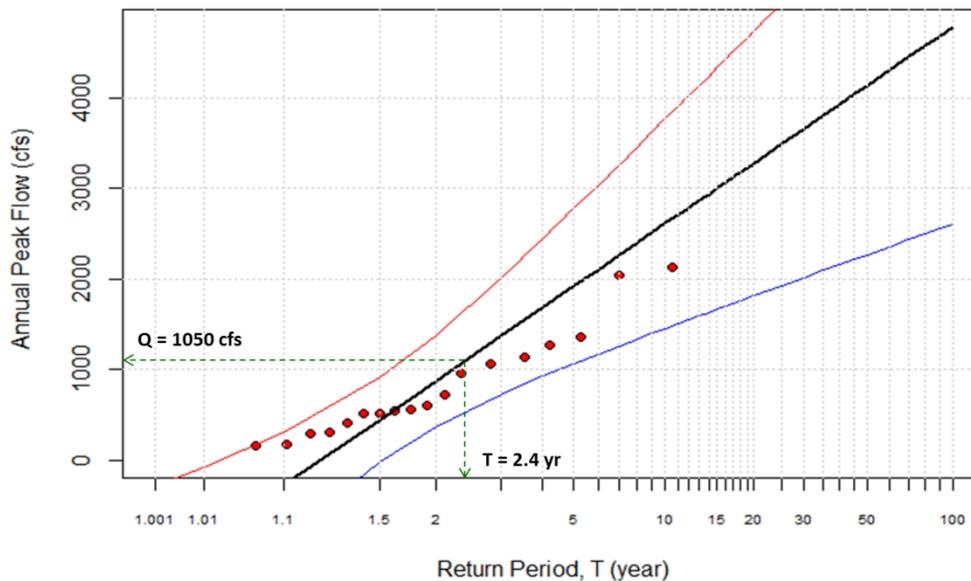> Put 1991-2010 as **graphlab**

> Replace **XXXX** with **Blackberry Creek** (name of your watershed)

You have just created flood frequency graphs for the two periods. Use the left/right arrows in the Plots window to see the desired graph.

Flood Frequency Analysis, Blackbeery Creek (1971-1990)
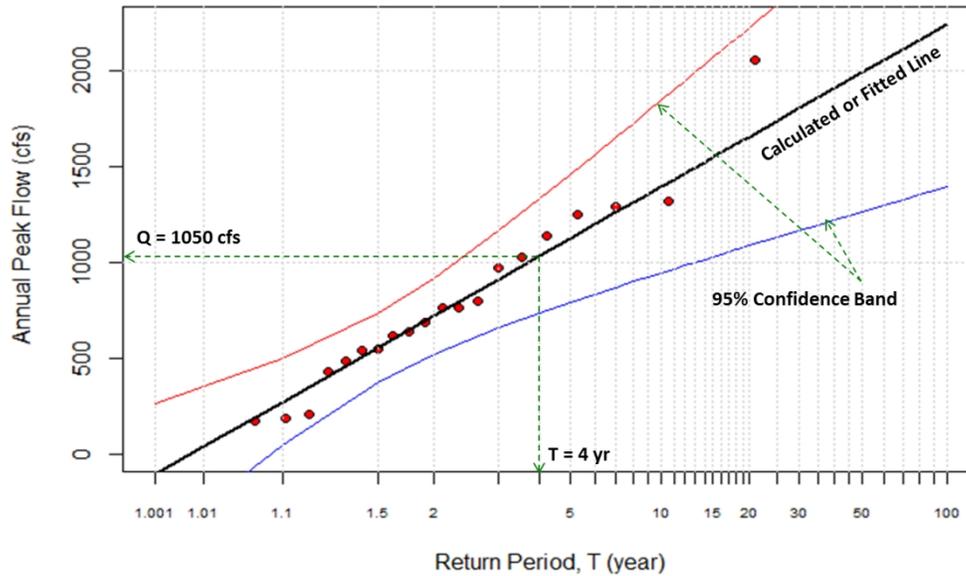
Flood Frequency Analysis, Blackbeery Creek (1991-2010)

The dots inside the plot are the actual observed peak discharge values. The thick straight line in the middle is the calculated probability distribution. To read the graph, pick the flood value of your concern from Y axis and trace a horizontal line up to where it intersects the line of calculated probability distribution. From that point of intersection, trace a vertical line towards X axis to find out the corresponding return period. You can simply do the opposite if you want to find out the flood magnitude for a given return period. Peak discharge/ return period read from the line of calculated

probability distribution can have errors due to the limited sample data used. Hence, a pair of lines (red and blue) are also plotted which denotes the confidence band, indicating 95% chance for the calculated values to lie within this limit.

Now, let's find out the return period for a flood event of 1050 cfs magnitude from the plots which you have generated. The results are as follows:

- During 1971-1990, a flood of 1050 cfs has a return period of 4 years.
- During the recent period of 1991-2010, the same magnitude flood shows a shorter return period of only 2.4 years.

Clearly, extreme flow events have become more frequent in recent time in this particular watershed.

Quiz

1. State whether the following statement is correct: "Probability of occurrence for a flood event is the inverse of its return period."   **TRUE/FALSE**

2. State whether the following statement is correct: "A flood of 10,000 cfs has a return period of 50 years. Such a flood will occur only once in every 50 years and twice in every 100 years"

                                                            TRUE/FALSE

3. A flood of 10,000 cfs has a return period of 50 years. What is the chance that such a flood will occur <u>in any single year</u>?
   a. 2%   b. 1%   c. 0.02%

4. What is the probability of occurrence of a 10-year flood in the current year if that same magnitude flood just happened last year?
   a. 0.2   b. 0.1   c. No chance of occuring

5. What is the chance for a 100-year flood to occur during a whole time period of 25 years?
   Hint: $p=1/100=0.01$, $n= 25$, use the formula: $1– (1–p)^n$
   a. 78%   b. 64%   c. 22%

6. Use the Flood Frequency Plots for the Blackberry Creek and complete the following table:

1971-1990		1991-2010	
Return Period (Year)	Peak Discharge (cfs)	Return Period (Year)	Peak Discharge (cfs)
3.75	1000	2.3	1000
	1500		1500
	2000		2000

Look into the values of the completed table. Calculate the chance (in percentage) for a 1500 cfs flood to occur in any year during 1971-1990 and 1991-2010 [Hint: approximate answers are 7% and 29%].

What can you conclude about the frequency of extreme events in this watershed? What are the possible reasons for such changes in frequency?

Exercise 1

Peak discharge datasets are provided for the following two watersheds:

(i)   McAlpine Creek (USGS 02146600) in the **Charlotte city area, North Carolina**

(ii)   Las Vegas Wash at Pabco Road (USGS 09419700) in the **Las Vegas city area, Nevada**

For each case, there are two datasets containing the peak discharge values for two different periods (1971-1990 and 1991-2010).

Follow the Blackberry Creek's example and perform Flood Frequency Analysis for these two watersheds. Comment on the change in frequency of extreme events in these locations.

Notes:
* Provide appropriate names while loading the datasets in RWater. For example, use 'mc71_90', 'mc91-10' for McAlpine Creek and 'vegas71_90', 'vegas91_10' for Las Vegas Wash.

Exercise 2

Using the RWater tool, download annual peak discharge for the Skunk Creek watershed near **Sioux Falls, South Dakota** (USGS 06481500) and compare the change in return period of a 10,000 cfs flood event between 1981-2000 and 1995-2014.

Do you think extreme events are becoming more frequent in recent years in this particular watershed?