Flood Frequency Analysis Guide

Marc Los Huertos

March 5, 2020

1 Introduction

1.1 What is a Flood Frequency Analysis?

1.2 Using R to Analyze Flood Frequencies

1.2.1 Using Packages to Import Data

We will use two packages (libraries), 'dataRetrieval' and 'xts'. To do this, navigage to the Packages tab in the lower right window of Rstudio and click on Install. Type in the names of packages to download.

```
### STEP 1
### Removing previously used scripts from RWater
### Removing all previously generated datasets and plots
cat("\014")
```

```
rm(list = ls())
dev.off()

## null device
##           1
```

```
### STEP 2
### Loading two specific packages into RWater  -- not sure what Rwater is...
library(dataRetrieval)
library(xts)

## Loading required package:  zoo
##
## Attaching package:  'zoo'
## The following objects are masked from 'package:base':
##
##    as.Date, as.Date.numeric
```

2 Selecting and Obtaining Gaging Station Data

2.1 Finding the Station ID

I recommend finding a station that has a long record, certianly more than 40 years if you can.

Using the USGS site, find the station ID and enter below as mysite. The package

```
### STEP 3
### Get the Peak Annual Discharge
mysite<-'11266500' # You want to change this code to match your USGS site code.

annualpeak<-readNWISpeak(mysite)
annualpeak_title <- "Merced River at XXX"
```

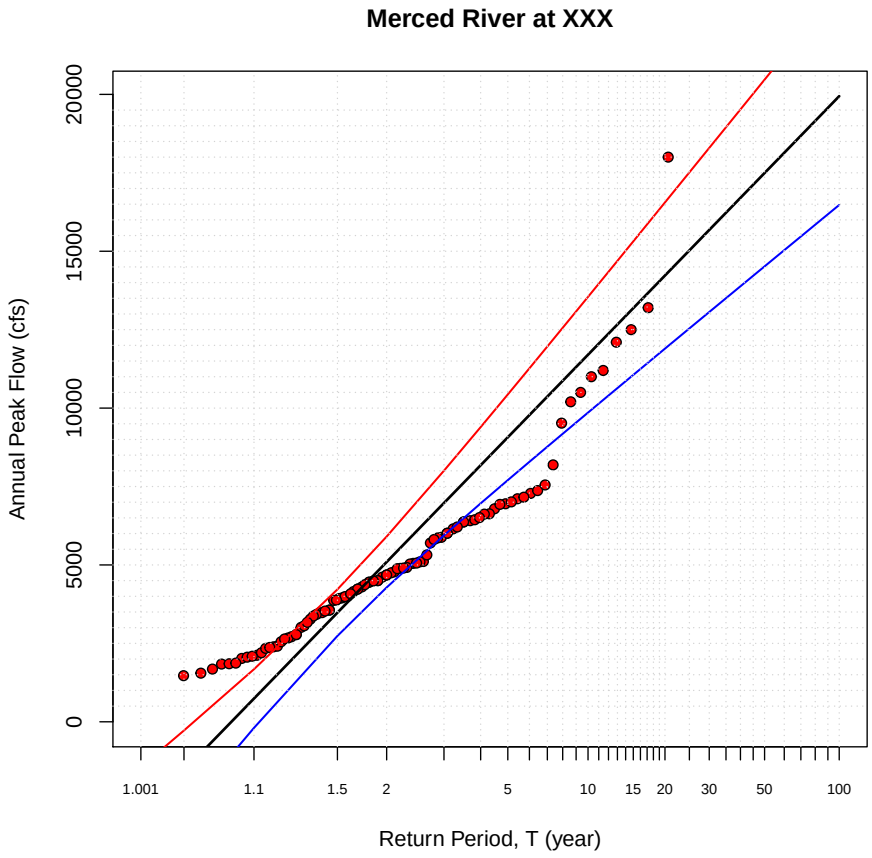
3 Flood Frequency Analysis

3.1 Flood Frequency for Entire Record

First, we'll analyze the data for the entire record – but you should make sure to do the second part below, where we split the data in half to see if the flood frequencies are consistent.

```
### Locate the column of your data set that has the peak discharges
### Click the 'annualpeak' from your 'Environment' (upper right)
### You can see that peak discharges are stored in the 6th column (peak_va)

Q <- annualpeak$peak_va

# 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,
    13,14,15,16,17,18,19,20,25,30,35,40,45,50,60,70,
    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,NA,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 = annualpeak_title
)
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)
```

4 Are Flood Frequencies Stationary?

4.1 Testing if the data are consistent over time

Look at the data and evaluate how to split the data in half – then we can see if the estimate for flood frequency has changed.

Remember, in California, the water year actually starts on the 1st of October each year. In the example, I have below, I have define the dates, name of the station and dates for the graphic labels in this section too.

```
### STEP 4
### Split the downloaded data into two 20 year periods
```

```
### Water year in CA, begins Oct 1 each year.

period1<-subset(annualpeak,
                peak_dt>="1980-10-01"
                &peak_dt<="1999-09-30")
period1_title = "Merced River at XXX, (1980-1999)"
period2<-subset(annualpeak,
                 peak_dt>="1999-10-01"
                 &peak_dt<="2019-09-30")
period2_title = "Merced River at XXX, (2000-2019)"
ymax = round(max(annualpeak$peak_va, na.rm=T)*1.1, -3)
```

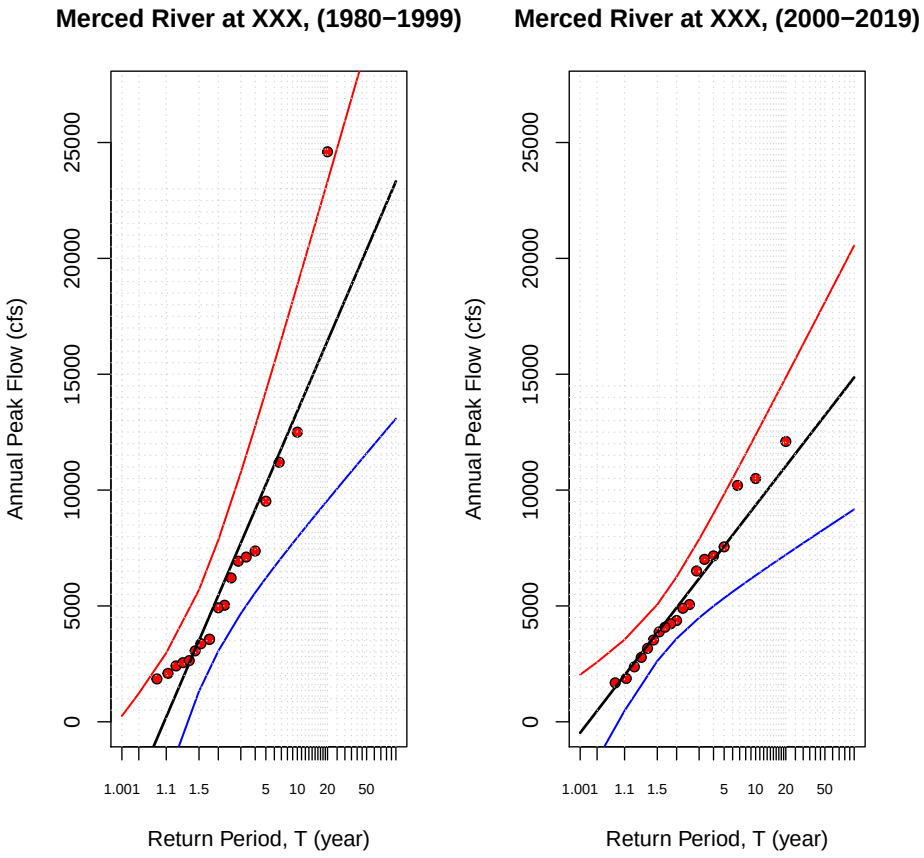
```
### STEP 5
```

5 Flood Frequency Analysis for Two Periods

```
### STEP 5
### Perform Flood Freqency Analysis
### Locate the column of your data set that has the peak discharges
### Click the 'period1' from your 'Environment' (upper right)
### You can see that peak discharges are stored in the 6th column (peak_va)

Q <- period1$peak_va

#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,
    13,14,15,16,17,18,19,20,25,30,35,40,45,50,60,70,
    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,NA,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)

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

# 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, ymax),
     xlim = c(xmin, xmax),
     pch = 21, bg = "red",
     main = period1_title
)
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)


### Perform Flood Freqency Analysis for the second time period

Q = period2$peak_va

#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
```

```r
#Ttick = c(1.001,1.01,1.1,1.5,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,25,30,35,40,
#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,NA,30,
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, ymax),
     xlim = c(xmin, xmax),
     pch = 21, bg = "red",
     main = period2_title
)
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)
```

5.1 Next Steps

make scales on y-axis the same!

6 Creating a function