USGS Discharge Analysis

Location: Delaware River at Trenton NJ
(USGS Site No. 01463500)
cc

Ahsan Aziz
Ahsan Karim
Moshe Fixler
Syed Lakitullah
Muhammad Anwar

Delaware River map

Purpose for recording and analyzing discharge data

- Keeping discharge records is used to study natural water resources
- Model information about the past, present, and future of the discharge at specific location
- Creating structures like dams, bridges, oil rigs, we need to know about how much discharge the body of water will have in a certain period of time
- Understand effects of natural hazards and like storms, floods, and rain on bodies of water
- Help scientists learn about the natural landscape and topography and how it is shaped by discharge
- Regulate water resources in states for public use
- Analyze on a global scale the conditions of water level and water flow as global warming and climate change is taking place and how discharge is affected

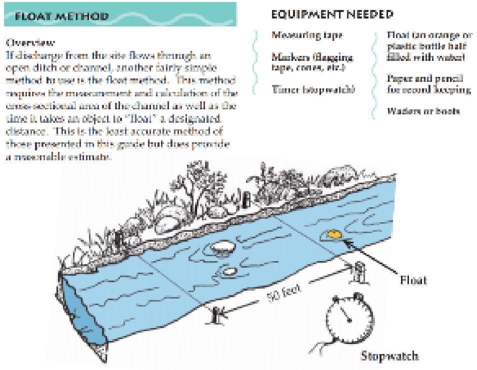
What is discharge?

- In hydrology, discharge is the volume rate of water flow that is transported through a given cross-sectional area.

- In fluid dynamics and hydrometry, the volumetric flow rate, (also known as volume flow rate), is the volume of fluid which passes per unit time.

Methods of collecting discharge data

Float method:



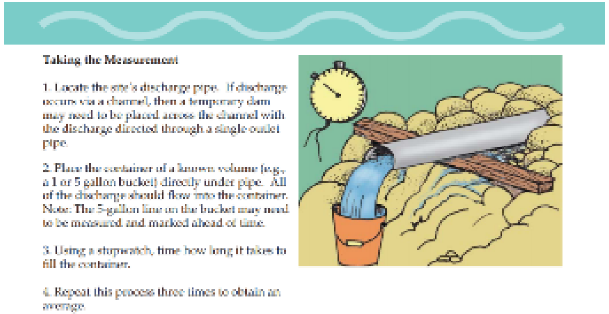
FLOAT METHOD

EQUIPMENT NEEDED

- Measuring tape
- Markers (flagging tape, cones, etc.)
- Timer (stopwatch)
- Float (an orange or plastic bottle half filled with water)
- Paper and pencil for record keeping
- Waders or boots

Overview
If discharge from the site flows through an open ditch or channel, another fairly simple method to use is the float method. This method requires the measurement and calculation of the cross-sectional area of the channel as well as the time it takes an object to "float" a designated distance. This is the least accurate method of those presented in this guide but does provide a reasonable estimate.

Bucket and stopwatch method:



Taking the Measurement

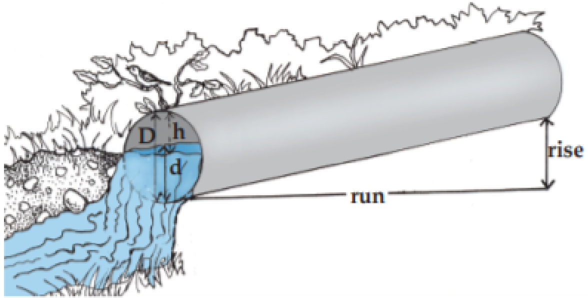
1. Locate the site's discharge pipe. If discharge occurs via a channel, then a temporary dam may need to be placed across the channel with the discharge directed through a single outlet pipe.

2. Place the container of a known volume (e.g., a 1 or 5 gallon bucket) directly under pipe. All of the discharge should flow into the container. Note: The 5-gallon line on the bucket may need to be measured and marked ahead of time.

3. Using a stopwatch, time how long it takes to fill the container.

4. Repeat this process three times to obtain an average.

Manning's equation method:

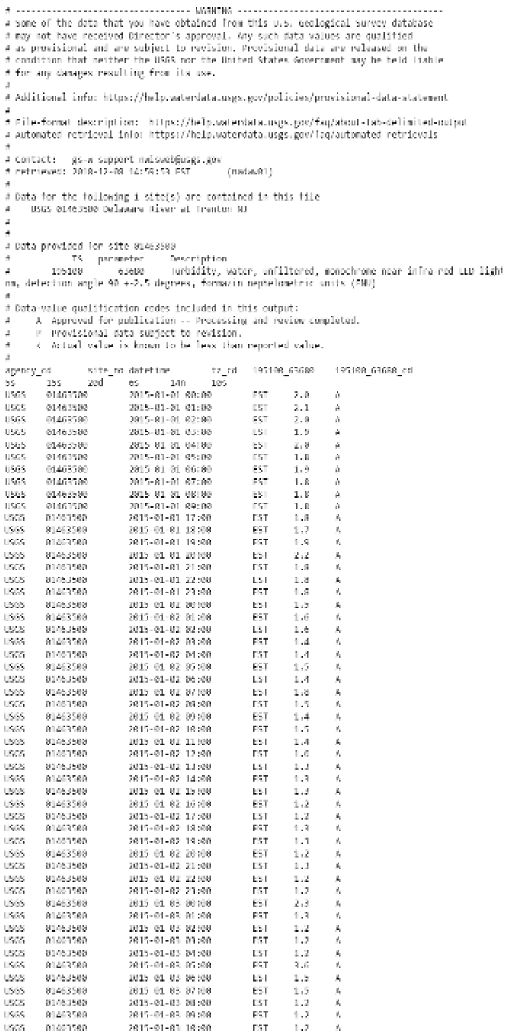


$$Q = \frac{1.49\, A\, R^{2/3} S^{1/2}}{n}$$

A = cross-sectional area
R = hydraulic radius
S = slope
n = Manning's roughness coefficient
Q = discharge

Procedure for our project

1. Google search “usgs water data”
2. On the USGS site, click on “Surface Water”
3. On the next page, click on “Daily Data”
4. Choose site selection criteria
    a. Site location: State/Territory
    b. Site Identifier: Site Number “01463500”
5. When found the desired site, select dates
6. When selecting type, select Tab-Separated format
7. Save this [url](url) as text file named “USGS_01463500_Delaware_River_at_Trenton_NJ”
8. Read instructions in project description for creating graphs
9. Many of the function names are given in project description
10. Plot the necessary graphs and analyze the data



```
# ---------------------------------- WARNING ----------------------------------------
# Some of the data that you have obtained from this U.S. Geological Survey database
# may not have received Director's approval. Any such data values are qualified
# as provisional and are subject to revision. Provisional data are released on the
# condition that neither the USGS nor the United States Government may be held liable
# for any damages resulting from its use.
#
# Additional info: https://help.waterdata.usgs.gov/policies/provisional-data-statement
#
# File-format description:  https://help.waterdata.usgs.gov/faq/about-tab-delimited-output
# Automated retrieval info: https://help.waterdata.usgs.gov/faq/automated-retrievals
#
# Contact:   gs-w support nwisweb@usgs.gov
#
# Data for the following 1 site(s) are contained in this file
#    USGS 01463500 Delaware River at Trenton NJ
#
#
# Data provided for site 01463500
#            TS   parameter     Description
#
# Data-value qualification codes included in this output:
#     A  Approved for publication -- Processing and review completed.
#     P  Provisional data subject to revision.
#     <  Actual value is known to be less than reported value.
#
agency_cd	site_no	datetime	tz_cd
```

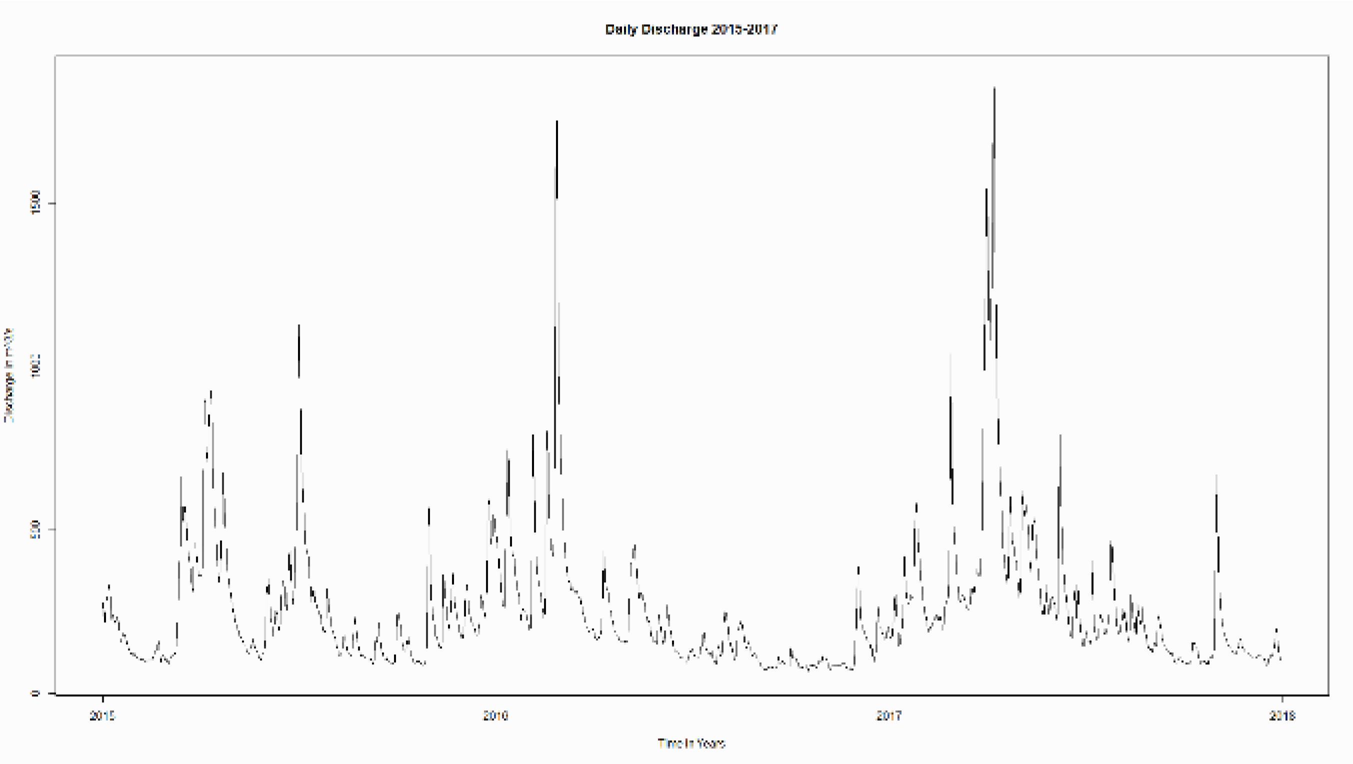
R programming

```
#2.1================================================================================================================
setwd("C:/Users/Ahsan Karim/Desktop/Data_Analysis_Project")
USGS_01463500=read.table("USGS_01463500_Delaware_River_at_Trenton_NJ.txt",header=T)
USGS_01463500=USGS_01463500[-c(1),]
filtered_data=USGS_01463500[,c(3,4)]
colnames(filtered_data)[2]="discharge"
filtered_data[,2] = as.numeric(as.character(filtered_data$discharge))
filtered_data[,2] = filtered_data[,2]*0.0283168
YYYY_MM_DD = as.Date(USGS_01463500$datetime,"%Y-%m-%d")
par(bg = 'gray99')
plot(YYYY_MM_DD,filtered_data$discharge,type="l",xlab = "Time in Years",ylab = "Discharge in m^3/s", main = "Daily
Discharge 2015-2017")
#2.2================================================================================================================
Y=substr(filtered_data$datetime,1,4)
M=substr(filtered_data$datetime,6,7)
D=substr(filtered_data$datetime,9,10)
combined_data=cbind(filtered_data,Y,M,D,YYYY_MM_DD)
library(dplyr)
Y_discharge=group_by(combined_data,Y)
Y_tbl=summarise(Y_discharge,
                     mean=mean(discharge),
                     sd=sd(discharge),
                     min=min(discharge),
                     max=max(discharge)
                     )
par(bg = 'gray99')
plot(Y_tbl$Y,Y_tbl$discharge,xlab="Time [years]",ylab="Discharge [m^3/s]",main="Annual Discharge",ylim=c
(0,2000),type="l")
lines(Y_tbl$Y,Y_tbl$mean,col="black",lwd=2)
lines(Y_tbl$Y,Y_tbl$sd,col="royalblue",lwd=2)
lines(Y_tbl$Y,Y_tbl$min,col="green",lwd=2)
lines(Y_tbl$Y,Y_tbl$max,col="red3",lwd=2)
legend("right",c("Mean","Standard Deviation","Minimum","Maximum"),
       col=c("black","royalblue","green","red3"),lty=1,lwd=2,cex=.75)
days_index=strftime(strptime(combined_data$YYYY_MM_DD,format="%Y-%m-%d"),format="%j")
combined_data=cbind(combined_data,days_index)
D_discharge=group_by(combined_data,days_index)
D_tbl=summarise(D_discharge,
                 mean=mean(discharge),
                 sd=sd(discharge),
                 min=min(discharge),
                 max=max(discharge)
                 )
par(bg = 'gray99')
plot(D_tbl$days_index,D_tbl$mean,type="l",ylim=c(0,2000),xlab="days_index",ylab="Discharge [m^3/s]",main="Daily
Discharge Climatology")
lines(D_tbl$days_index,D_tbl$sd,col="blue",lw=2)
lines(D_tbl$days_index,D_tbl$min,col="green",lw=2)
lines(D_tbl$days_index,D_tbl$max,col="red3",lw=2)
legend("topright",c("Mean", "Standard Deviation", "Minimum", "Maximum"),col = c
("black","royalblue","green","red3"),lty = 1, lwd = 2, cex=.75)
```

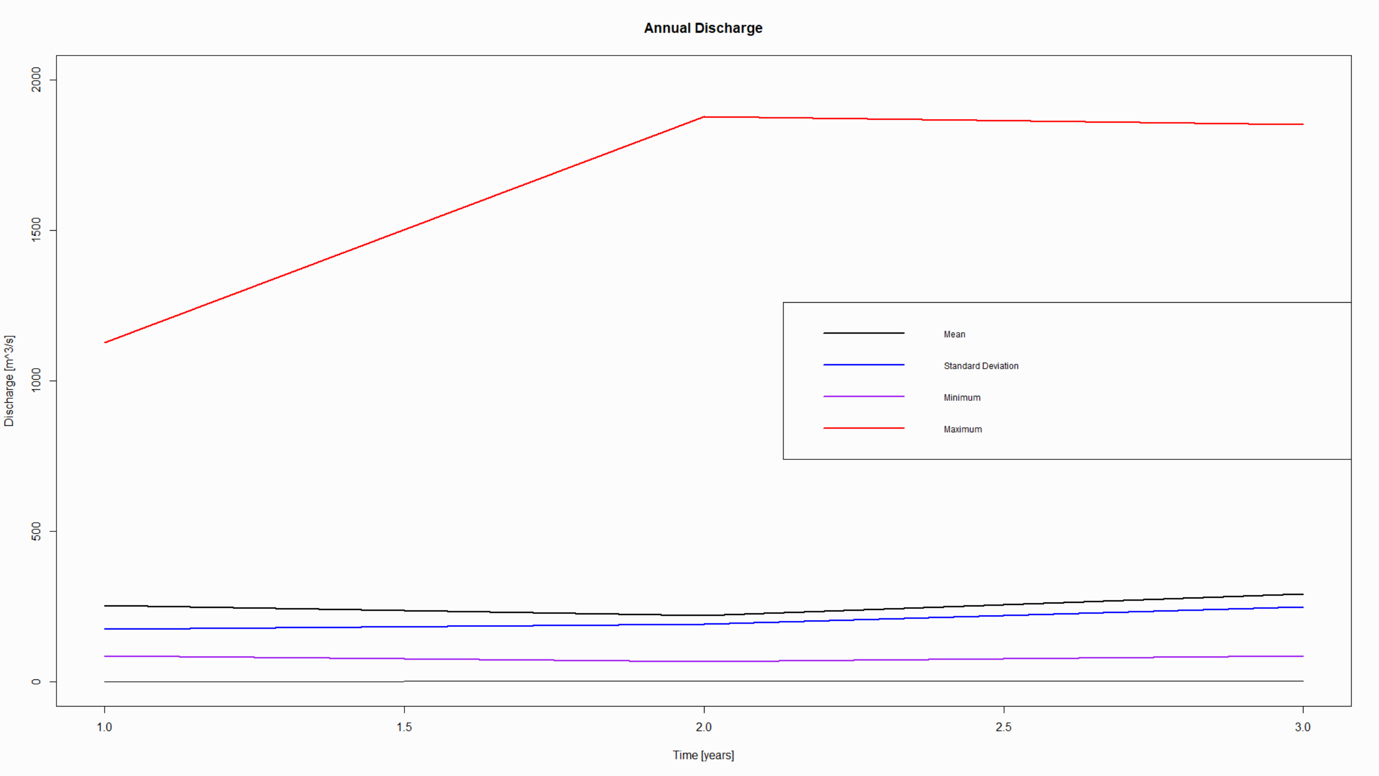
```
#2.3================================================================================================================
library(fitdistrplus)
library((fitdistrplus), MASS)
data_length = length(filtered_data$discharge)
normal = fitdistr(filtered_data$discharge,"normal")
normalestimate = rnorm(data_length,normal$estimate['mean'],normal$estimate['sd'])
lognormal = fitdistr(filtered_data$discharge,"lognormal")
lognormallestimate = rlnorm(data_length,lognormal$estimate['meanlog'],lognormal$estimate['sdlog'])
gamma = fitdistr(filtered_data$discharge,"gamma")
gammaestimate = rgamma(data_length,gamma$estimate['shape'],gamma$estimate['rate'])
exponential = fitdistr(filtered_data$discharge,"exponential")
exponentialestimate = rexp(data_length,exponential$estimate['rate'])
hist(filtered_data$discharge,xlim=c(0,1000),ylim=c(0,.005),xlab="Discharge [m^3/s]",ylab="Density",main="PDF of
Daily Discharge")
lines(density(filtered_data$discharge),col="black",lwd=2)
lines(density(normalestimate),col="yellow2", lwd=1)
lines(density(lognormallestimate),col="purple",lwd=1)
lines(density(gammaestimate),col="orchid1",lwd=2)
lines(density(exponentialestimate),col="blue",lwd=1)
legend("topright", c("Empirical","Normal","Exponential","Lognormal","Gamma"),col = c
("black","seagreen3","purple","orchid1","blue"), lty = 1)
par(bg = 'gray99')
plot.ecdf(filtered_data$discharge,col="red3",xlab="Discharge m^3/s",ylab="Probability",main="Empirical vs.
Theoretical CDFs",pch="o")
lines(ecdf(normalestimate),col="slateblue4",lwd=2)
lines(ecdf(exponentialestimate),col="blue",lwd=2)
lines(ecdf(lognormallestimate),col="purple",lwd=2)
lines(ecdf(gammaestimate),col="orchid1",lwd=2)
legend("bottomright", c("Empirical","Normal","Exponential","Lognormal","Gamma"),
       col = c("red3","yellow2","blue","purple","orchid1"), lty = 1)
install.packages("extRemes")
library(extRemes)
par(bg = 'gray99')
qqplot(normalestimate,filtered_data$discharge,xlab="Standard Normal Distribution Quantities m^3/s",ylab="Observed
Discharge Quantities m^3/s",main="Q-Q Plot, Normal Distribution")
qqline(normalestimate,probs=c(0.1,0.6))
qqplot(lognormallestimate,filtered_data$discharge,xlab="Standard Lognormal Distribution Quantities
m^3/s",ylab="Observed Discharge Quantities m^3/s",main="Q-Q Plot, Lognormal Distribution")
qqline(lognormallestimate,prob=c(0.1,0.6))
qqplot(gammaestimate,filtered_data$discharge,xlab="Standard Gamma Distribution Quantities m^3/s",ylab="Observed
Discharge Quantities m^3/s",main="Q-Q Plot, Gamma Distribution")
qqline(gammaestimate,prob=c(0.1,0.6))
#2.4================================================================================================================
install.packages("stats")
library(stats)
normalks=ks.test(normalestimate,ecdf(filtered_data$discharge))
lognormalks=ks.test(lognormallestimate,ecdf(filtered_data$discharge))
gammaks=ks.test(gammaestimate,ecdf(filtered_data$discharge))
expontialks=ks.test(exponentialestimate,ecdf(filtered_data$discharge))
fevd=fevd(filtered_data$discharge,filtered_data,type="GEV")
Table_5 = ci(fevd,alpha=0.05,type="parameter")
Table_6 = return.level(fevd,return.period=c(5,10,50,100,500),do.ci=TRUE)
plot(fevd)
```

Daily discharge over 3 years

- Showing daily statistics of water discharge at
- Peak occurences of water discharge in 2016 and 2017 with discharge levels being greater than 1500 [m^3/s]
- Inconsistent discharge records throughout these 3 years, 2015 is most inconsistent
- Severe storm on February 24th, 2016 which affected water discharge nearby
- Discharge values stayed low during the middle of the year in 2016 and 2017



Statistics for each year

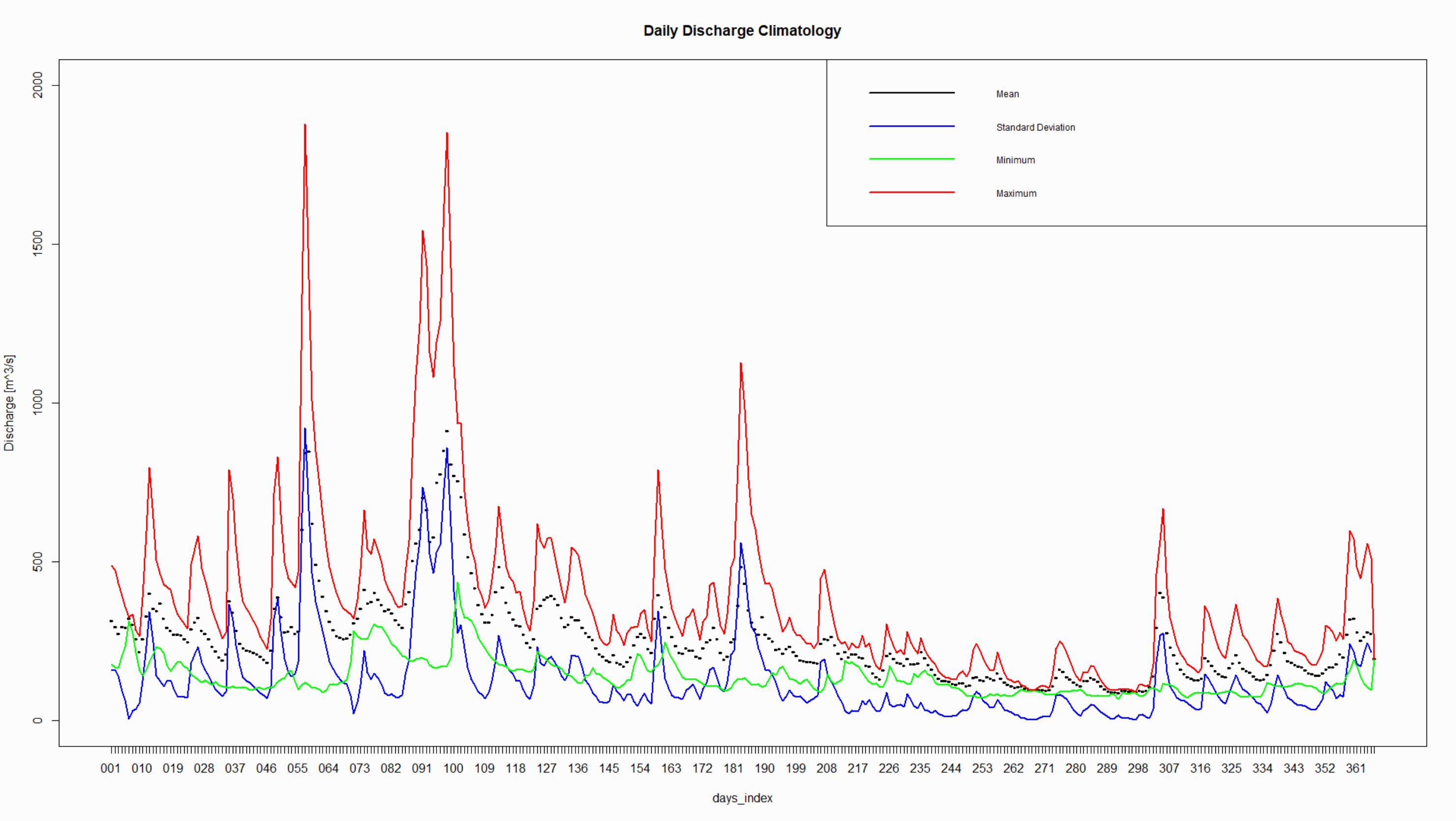


Y	mean	sd	min	max
2015	252.7294	175.1562	87.49891	1127.009
2016	220.4130	190.6686	67.11082	1877.404
2017	291.5529	248.6656	86.36624	1851.919

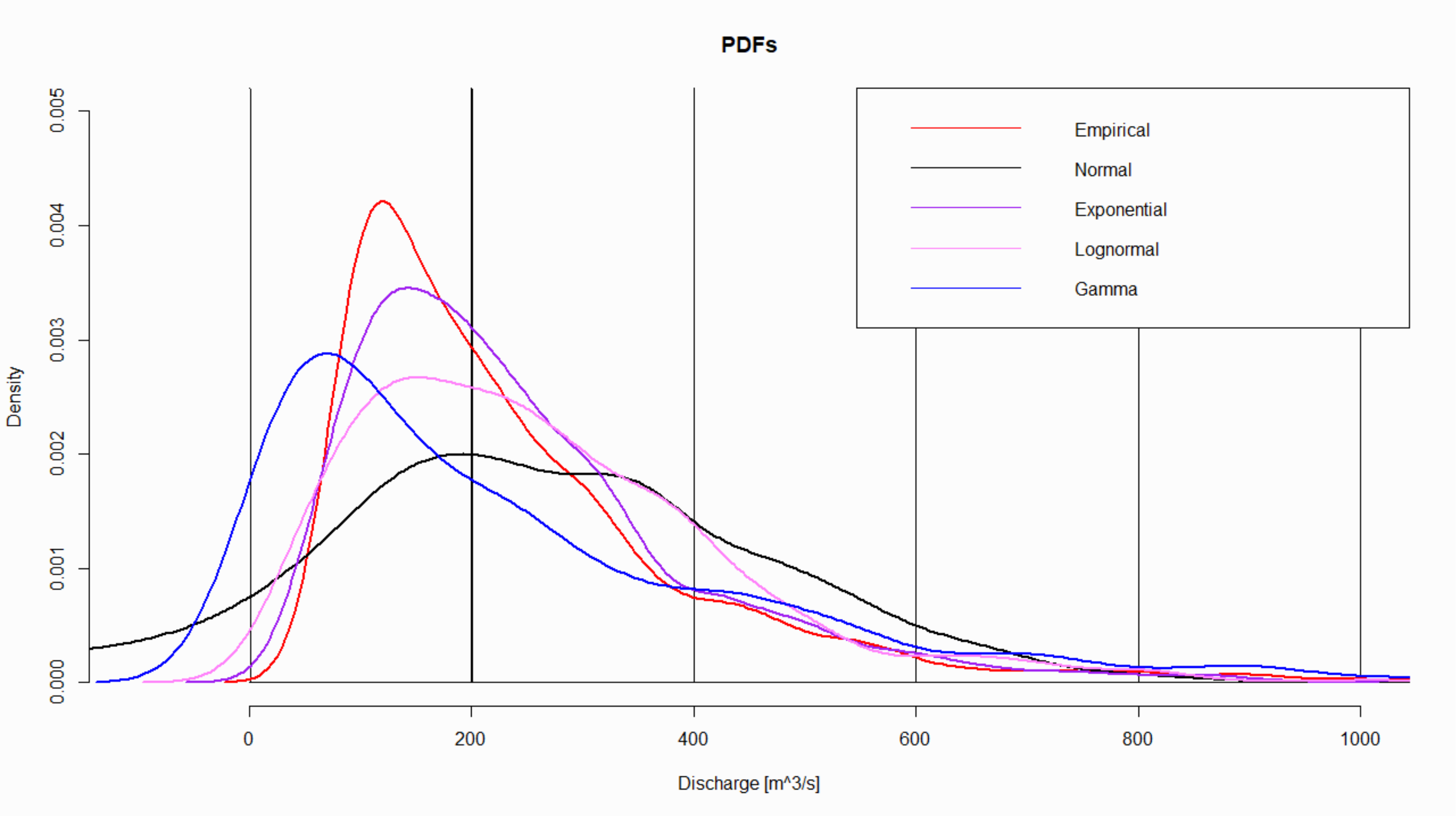
- Mean, standard deviation and minimum discharges were all similar for the three years
- Maximum discharges increased drastically by over 750 [m^3/s] from 2015-2016
- Interesting to see the maximum did not affect the average on graph, implies heavy rainfall once or twice a year do not significantly impact the river's discharge average

Statistics for each day

- From this graph we can see that the mean values indicated by black representing the data correctly
- The maximum of each day is shown to be greater in the range of days 60-110, from March to April
- Standard deviation is showing similar trend to the maximum, which makes sense because as discharge values are extensively away from the mean, the standard deviation is also extended



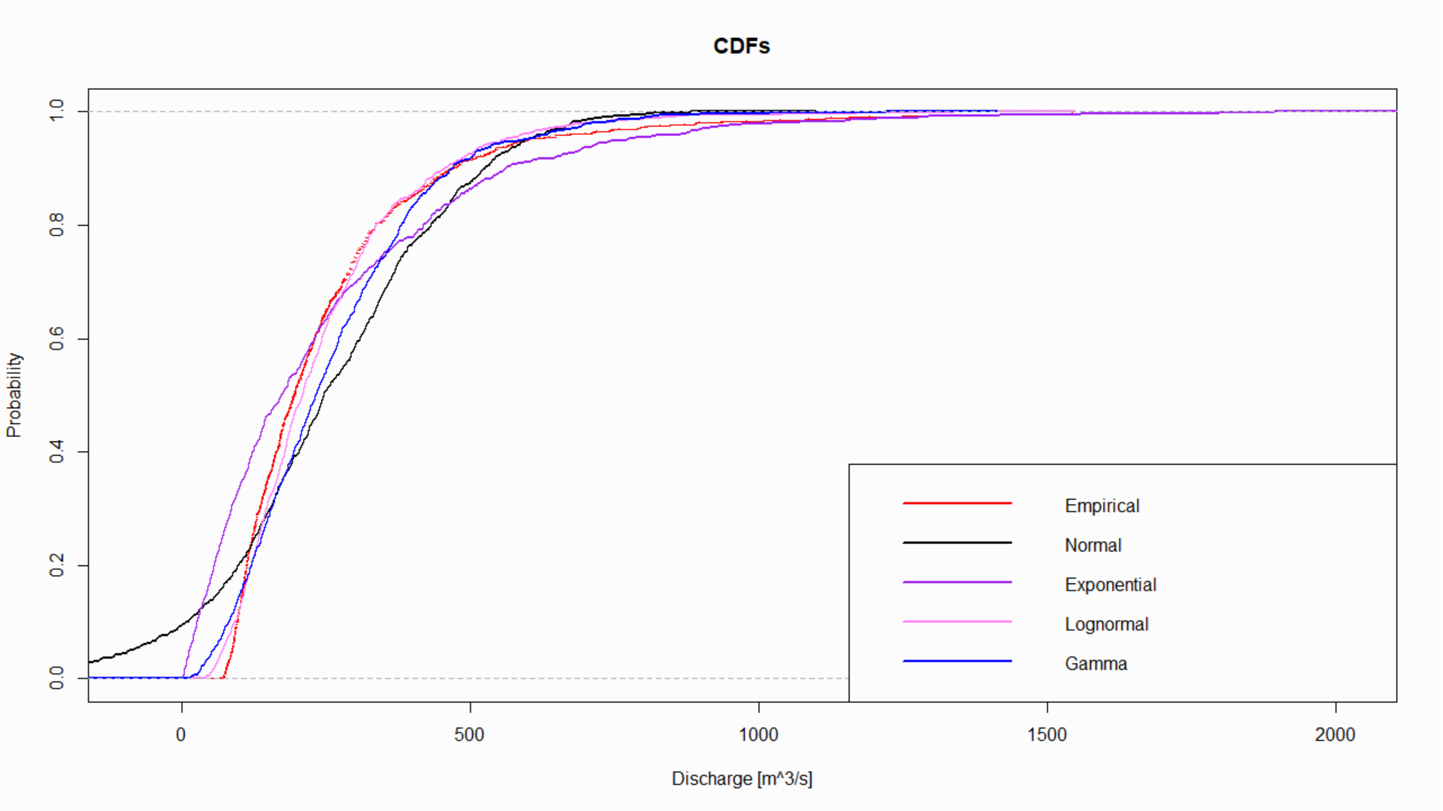
PDFs



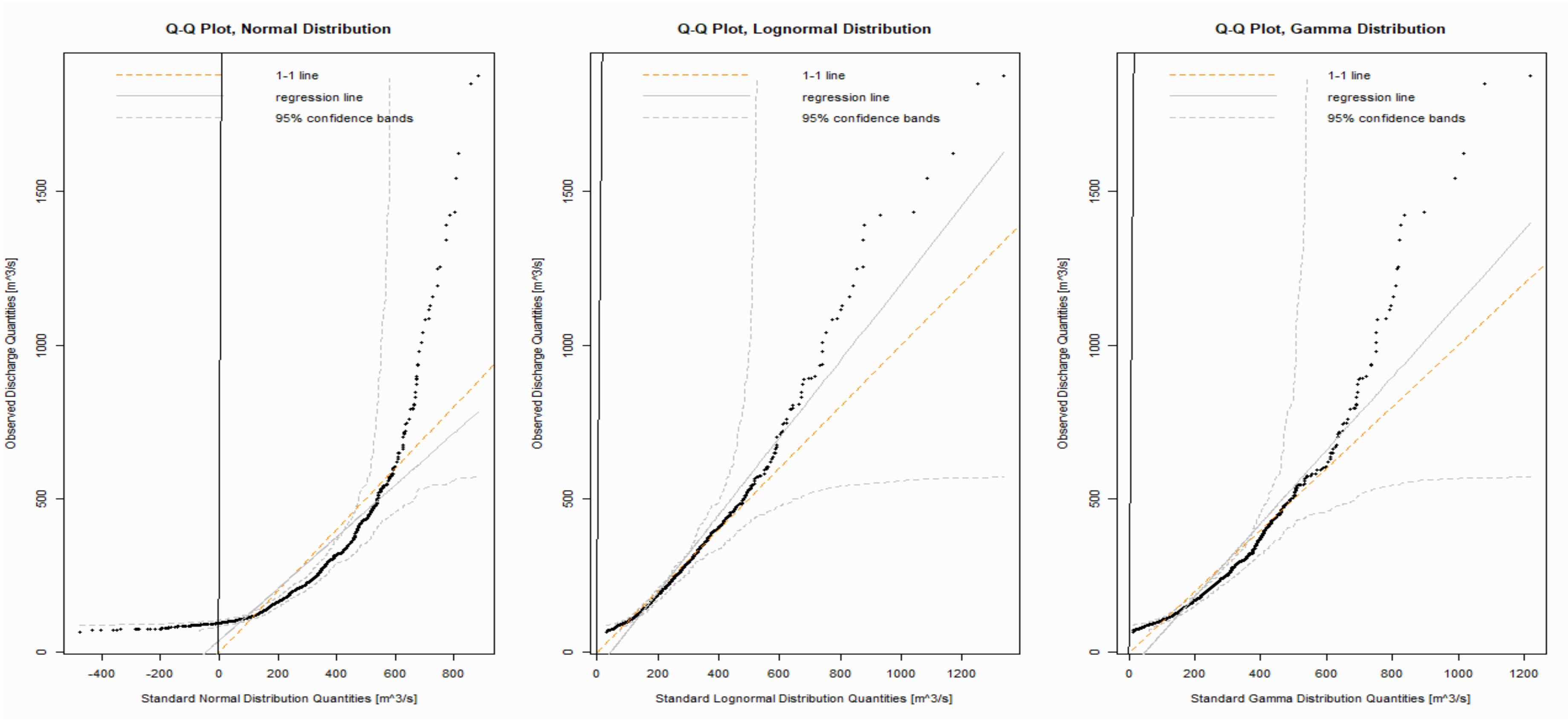
- Empirical distribution is from USGS data
- Normal, Exponential, Lognormal and Gamma are distributions are from model-simulated data based on their parameters
- From this plot we can see that exponential distribution best fits the empirical distribution

CDFs

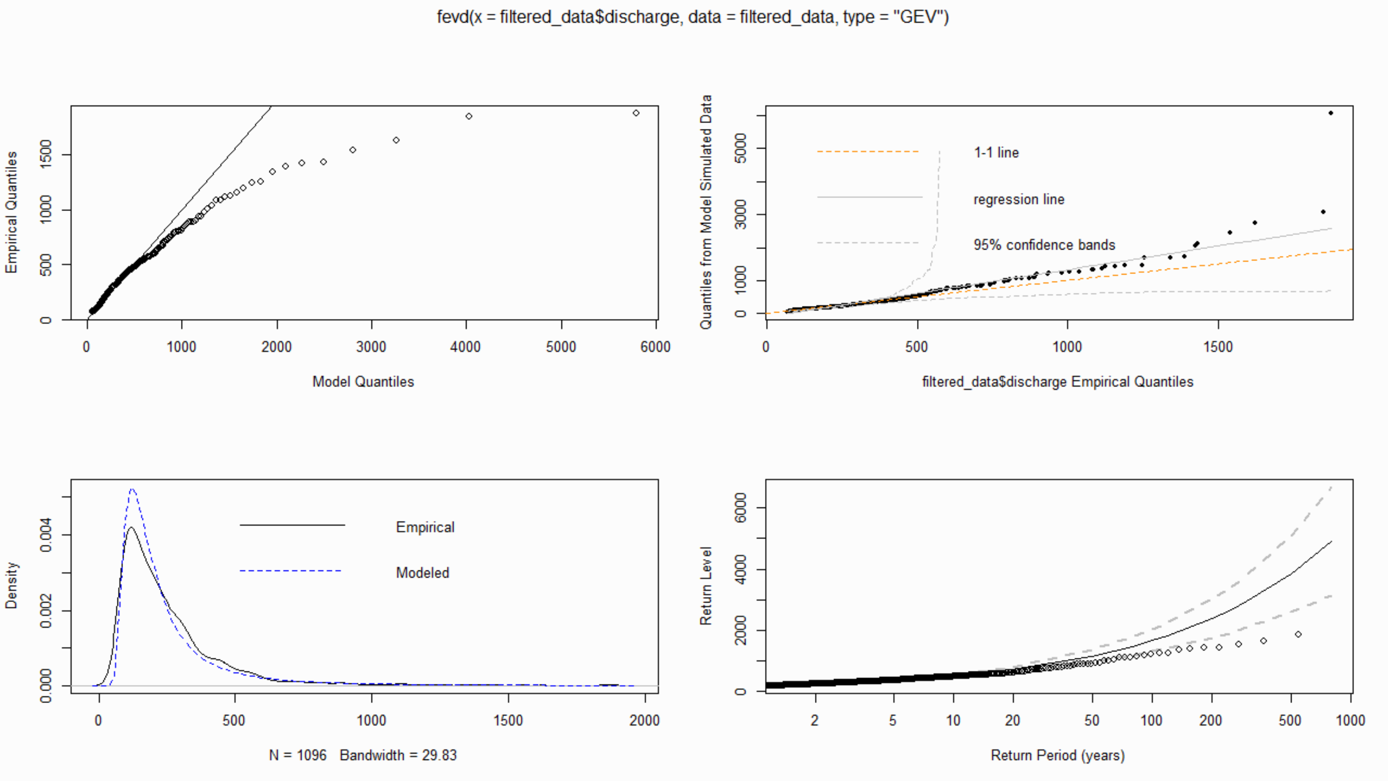
- Cumulative Density Function as the name suggests shows the sum of the individual probabilities for each type of distributions
- Sum of each of the probabilities equal to 1.0
- We used the plot.ecdf() function in RStudio to generate the CDF plots of the discharge for each distribution.
- The lognormal CDF best relates to the empirical CDF - as is evident from the plot.



Quantile-Quantile Plots



Extreme value




```
> Table_5
fevd(x = filtered_data$discharge, data = filtered_data, type = "GEV")

[1] "Normal Approx."

          95% lower CI    Estimate 95% upper CI
location  147.0583037 152.6739335  158.2895634
scale      72.6807508  78.0933408   83.5059307
shape       0.4476913   0.5223031    0.5969149
```

```
> Table_6
fevd(x = filtered_data$discharge, data = filtered_data, type = "GEV")

[1] "Normal Approx."

                      95% lower CI  Estimate 95% upper CI
5-year return level       311.7517  330.4429     349.1342
10-year return level      446.9685  487.4960     528.0236
50-year return level      952.1473 1150.7321    1349.3168
100-year return level    1295.9739 1655.7221    2015.4703
500-year return level    2581.9524 3841.5084    5101.0644
```

Conclusion

What we learned from working on this project:

- Retrieve data from USGS database
- Significance of recording discharge data
- Program in R
- Analyze the discharge data
- Apply different distributions of data and compare them

How we can apply this knowledge in the future:

- Work with large data in various fields
- Make predictable inferences based on statistical models
- Create valuable graphs and data representations for general public

References

Evenson, E.J., Orndorff, R.C., Blome, C.D., Böhlke, J.K., Hershberger, P.K., Langenheim, V.E., McCabe, G.J., Morlock, S.E., Reeves, H.W., Verdin, J.P., Weyers, H.S., and Wood, T.M., 2013, U.S. Geological Survey water science strategy—Observing, understanding, predicting, and delivering water science to the Nation: U.S. Geological Survey Circular 1383–G, 49 p.

“Manning's Equation.” *Continuity Equation - Manning's Equation*, <www.fsl.orst.edu/geowater/FX3/help/8_Hydraulic_Reference/Manning_s_Equation.htm.>

“USGS Water Data for the Nation” U.S. Department of the Interior. U.S. Geological Survey. Data 2018-12-09. <https://waterdata.usgs.gov/nwis>