System hydrology term project (2010)

The purpose of this term project is to use moment generating function to estimate parameters on real world hydrological data. The river discharge data were picked from the USGS National Hydrology Dataset (http://nhd.usgs.gov/data.html), and the distribution model fitted were decided based on the descriptive plots.

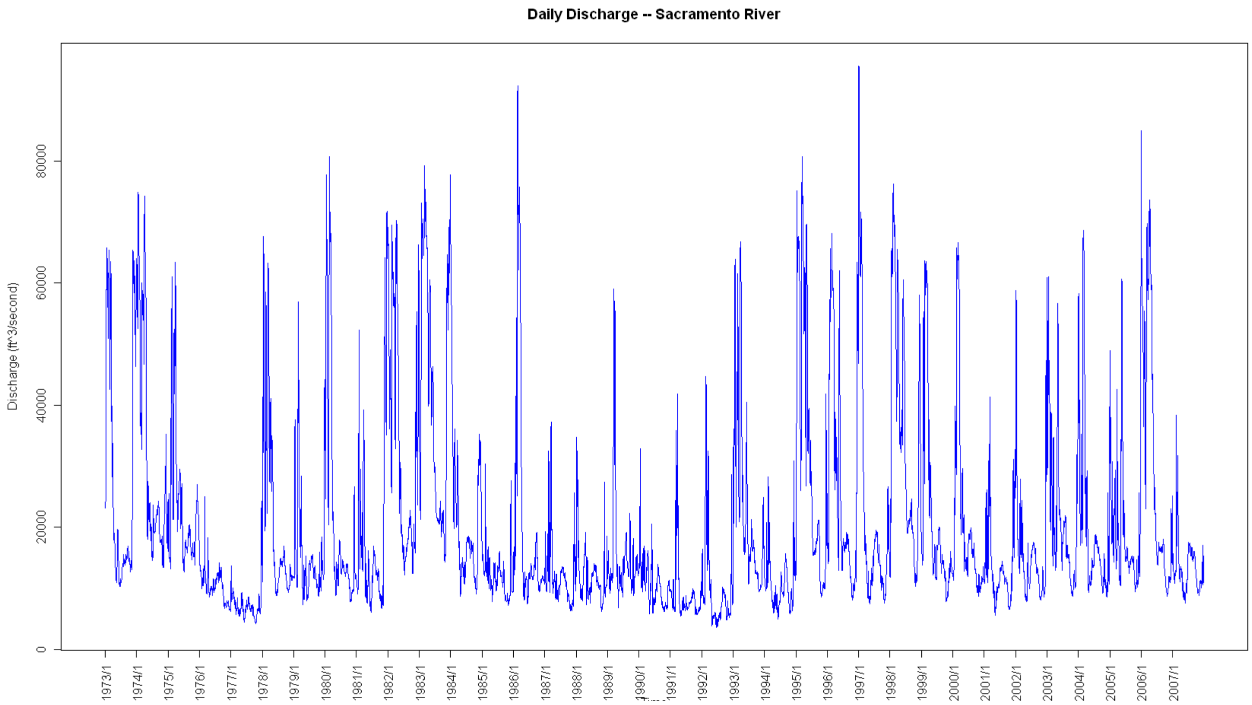
- Wet Climate: Sacramento River, CA, 1973-2007
- Dry Climate: San Pedro River, AZ, 1973-2007

Daily Discharge

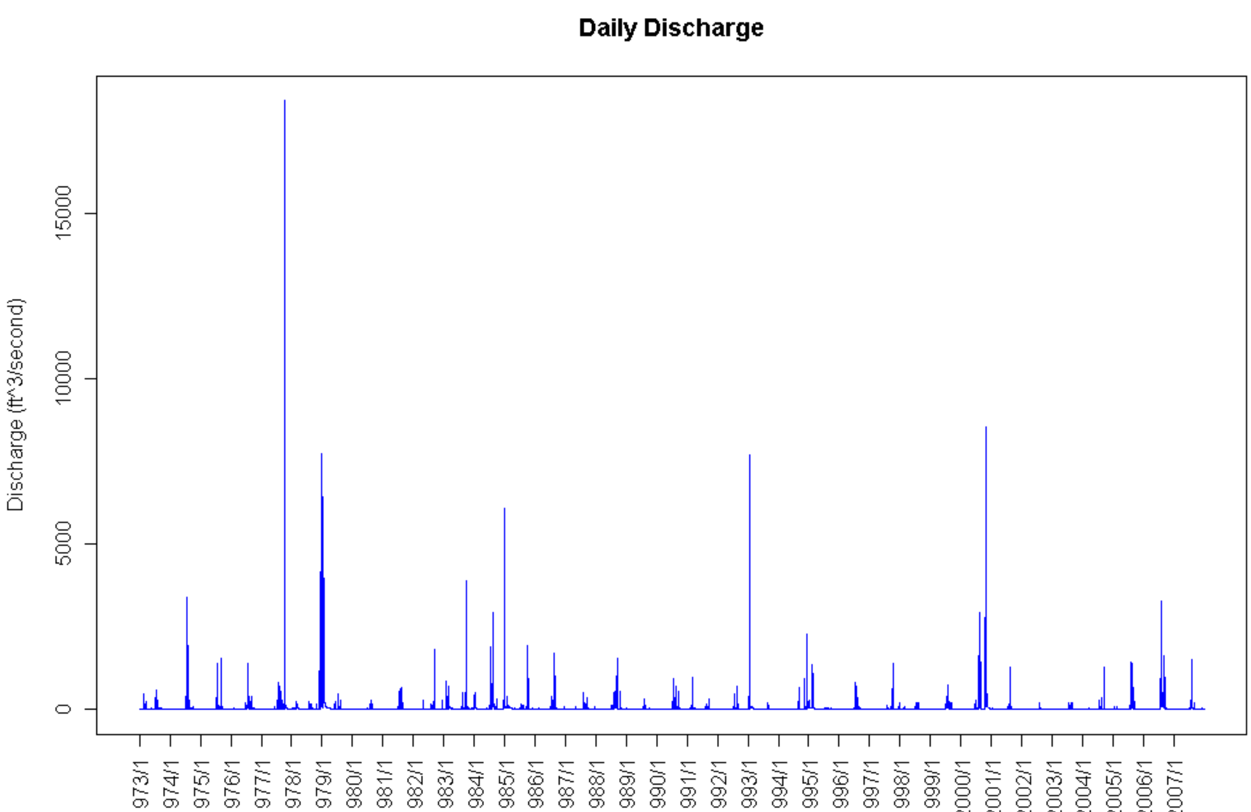
(a) Plot the data sets, find their histograms, and also find the first four moments and the coefficient of variation.

	1^{st} mo.	2^{nd}mo.	3^{rd}mo.	4^{th}mo	c.v.
Sacramento	2.033860e+04	6.747210e+08	3.205110e+13	1.819517e+18	0.794453
San Pedro	4.030381e+01	7.487907e+04	7.187577e+08	1.046287e+13	6.715658

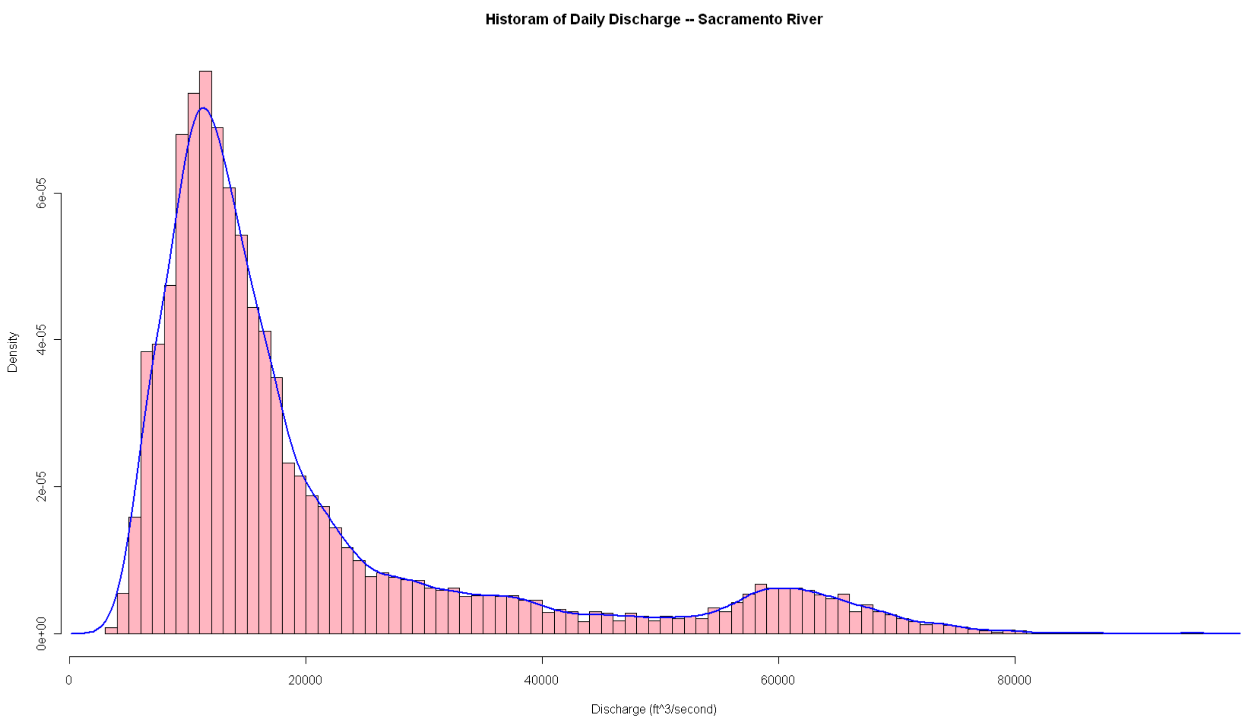
– Sacramento River (wet site)



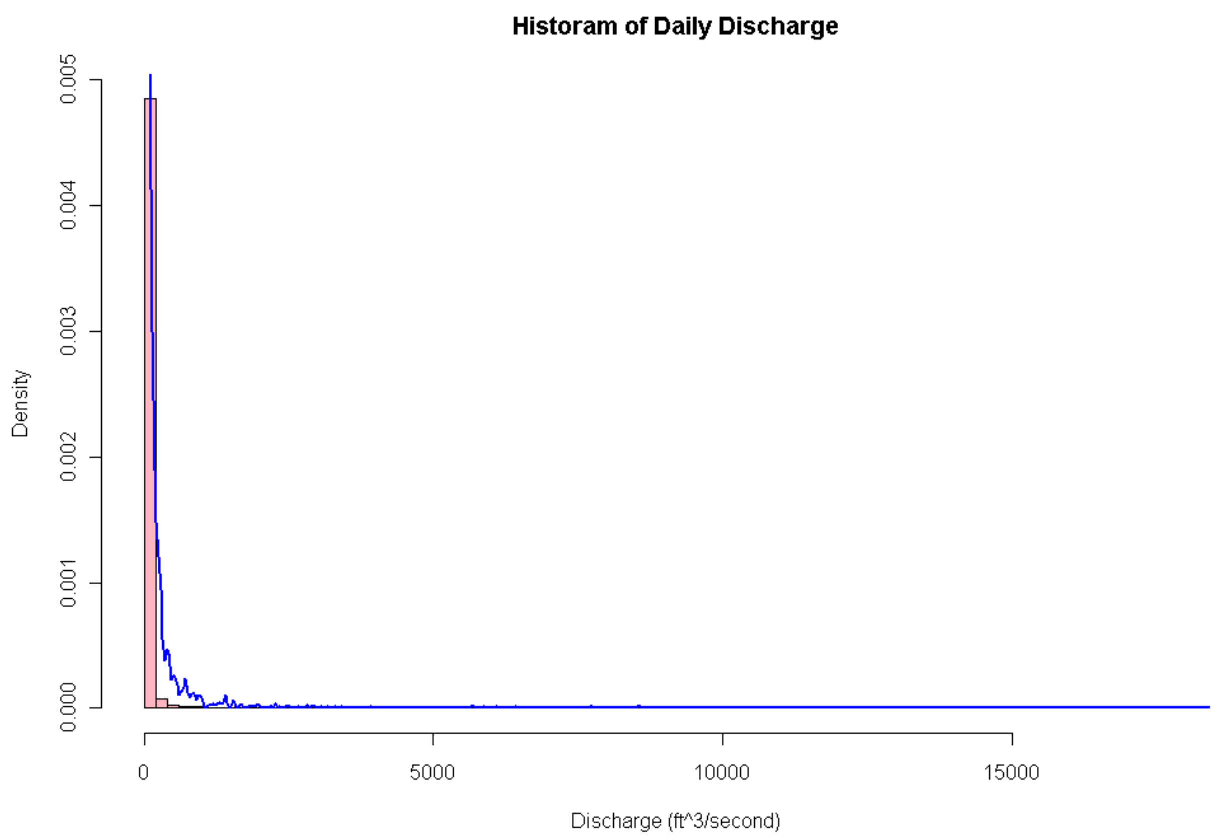
– San Pedro River (dry site)



– Sacramento River (wet site)



– San Pedro River (dry site)



(b) Using the method of moments, suggest two plausible models for such histograms. Plot your results and comment on whether one of the models turns out to be better than the other.

Both lognormal and exponential distribution fit the data well (ks tests), and the plot below indicates that the data is better described by lognormal distribution.

– Sacramento River(wet site)

A) lognormal distribution

1) Parameter estimate by methods of moments: lognormal distribution

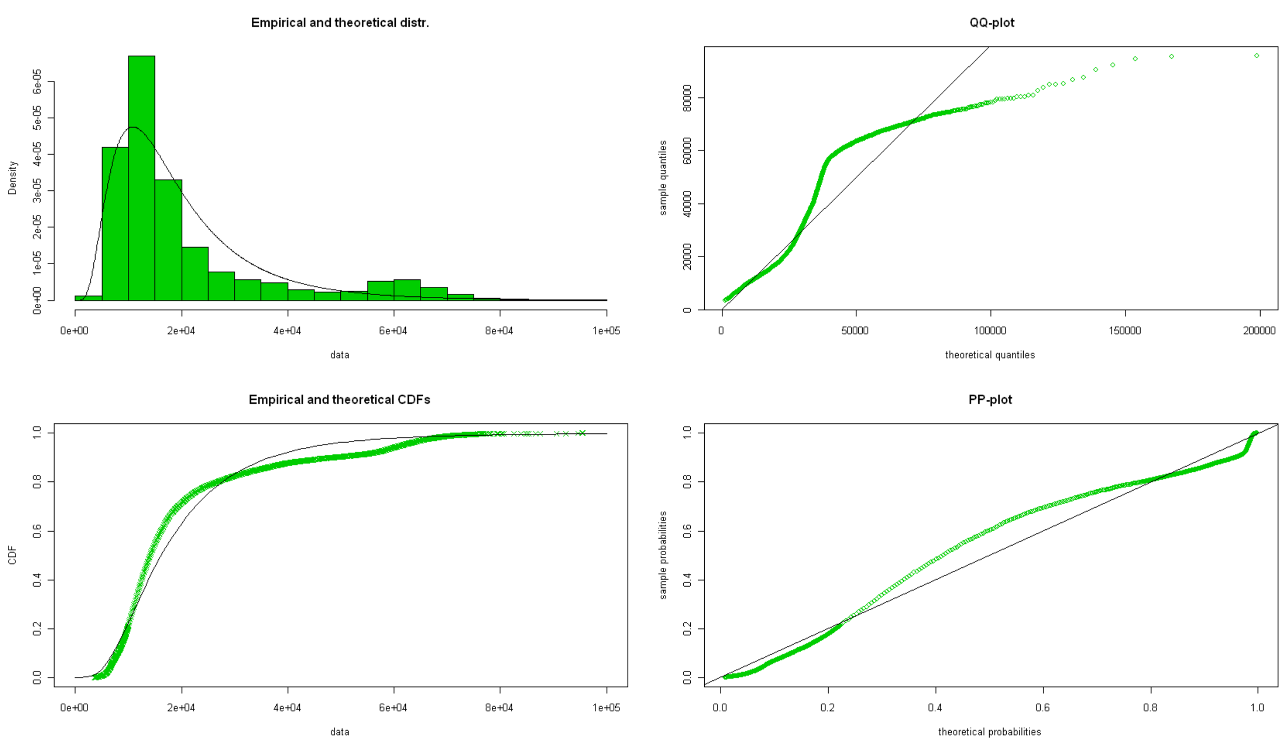
meanlog	9.6916833
sdlog	0.6350401

2) Fitting lognormal distribution:

One-sample Kolmogorov-Smirnov test

D = 1, p-value < 2.2e-16 ➔ Reject the H_o

H_A: two-sided



B) exponential distribution

1) Parameter estimate by methods of moments: exponential distribution

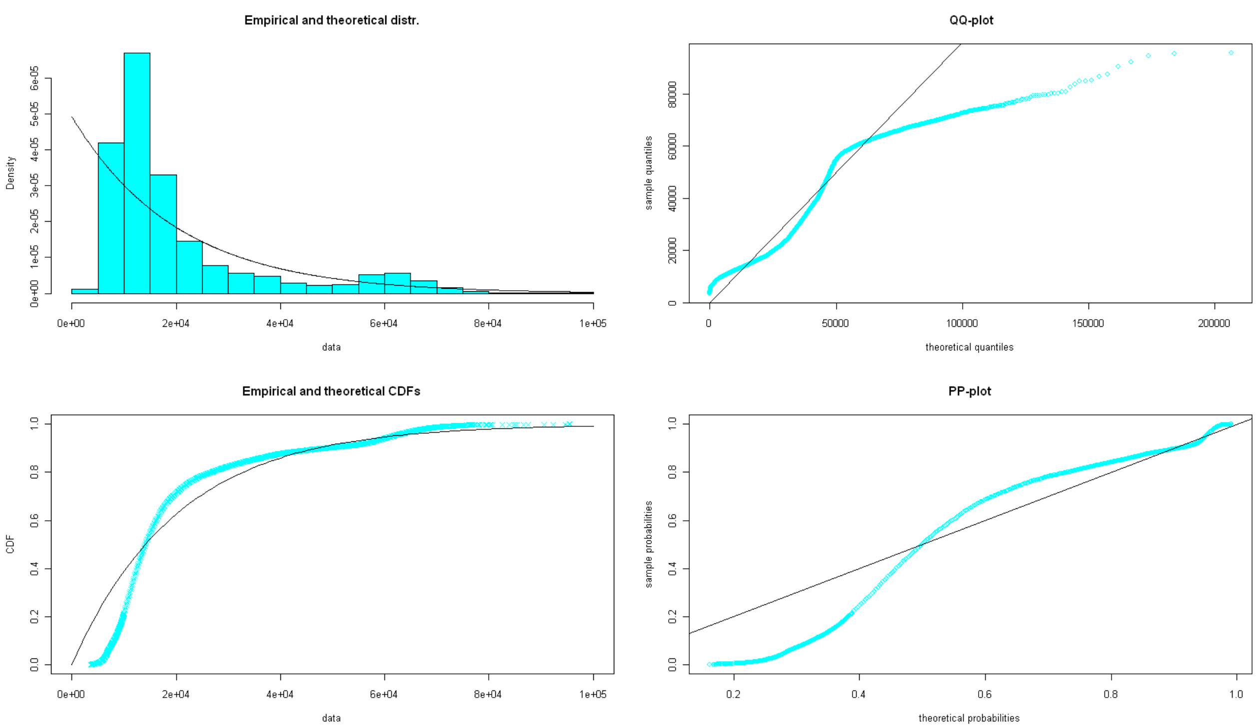
rate
4.91676e-05

2) Fitting exponential distribution:

One-sample Kolmogorov-Smirnov test

D = 1, p-value < 2.2e-16 ➔ Reject the H_o

H_A: two-sided



– San Pedro River (dry)

Both logistic distribution and exponential distribution fit the data well (ks tests), and the plot below indicates that the data is better described by logistic distribution.

A) logistic distribution

1) Parameter estimate by methods of moments: exponential distribution

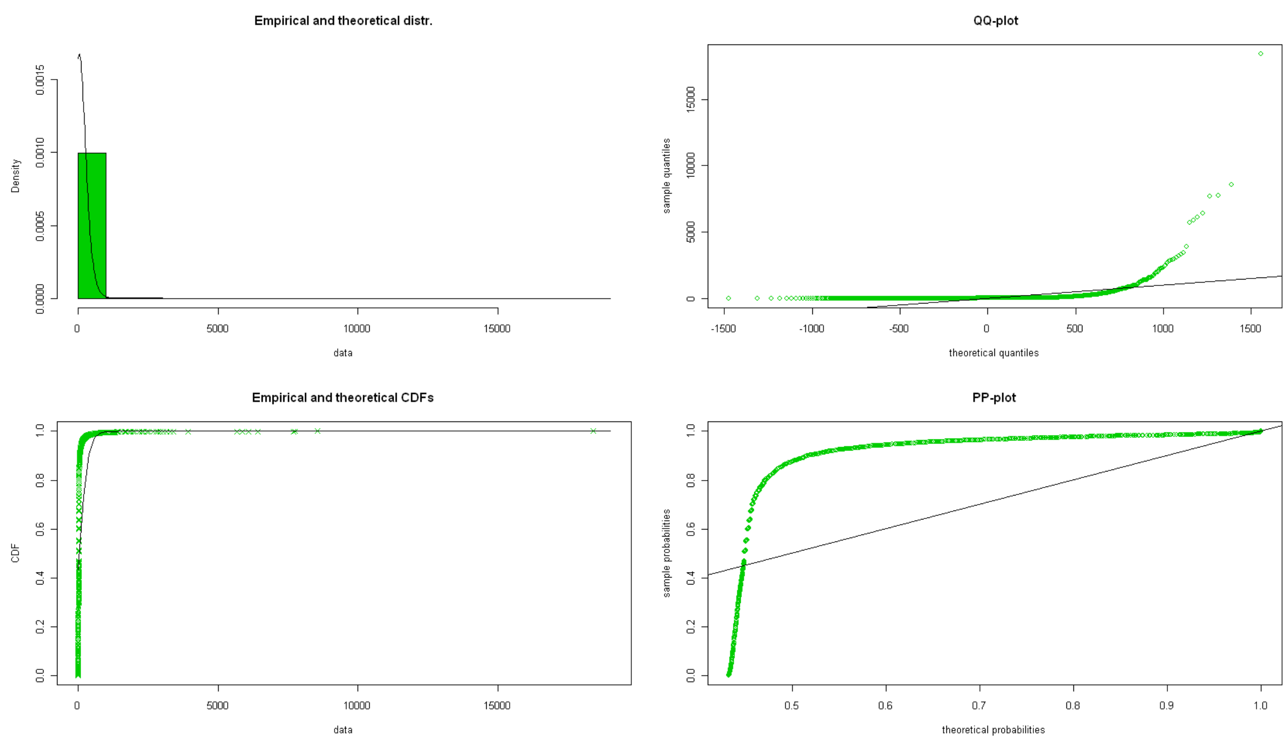
Location	40.30381
scale	149.22048

2) Fitting logistic distribution:

One-sample Kolmogorov-Smirnov test

D = 0.8406, p-value < 2.2e-1616 ➔ Reject the H_o

H_A: two-sided



B) exponential distribution

1) Parameter estimate by methods of moments: exponential distribution

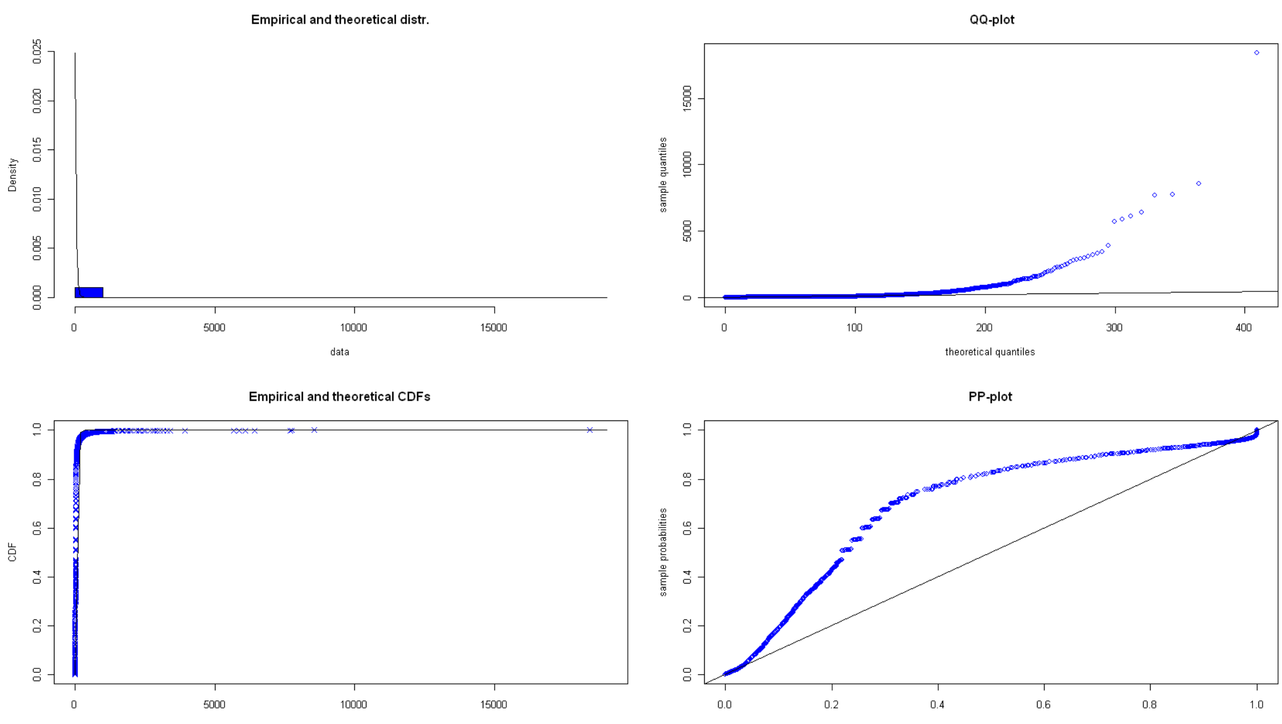
rate
0.02481155

2) Fitting exponential distribution:

One-sample Kolmogorov-Smirnov test

D = 0.8383, p-value < 2.2e-16➔ Reject the H_o

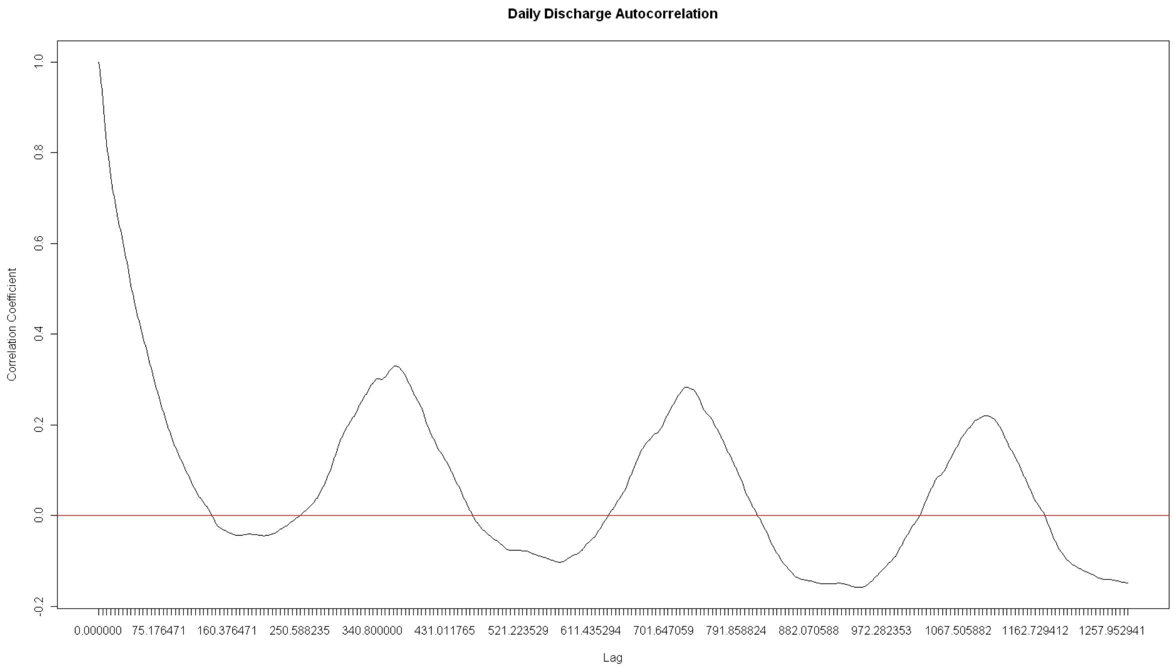
H_A: two-sided



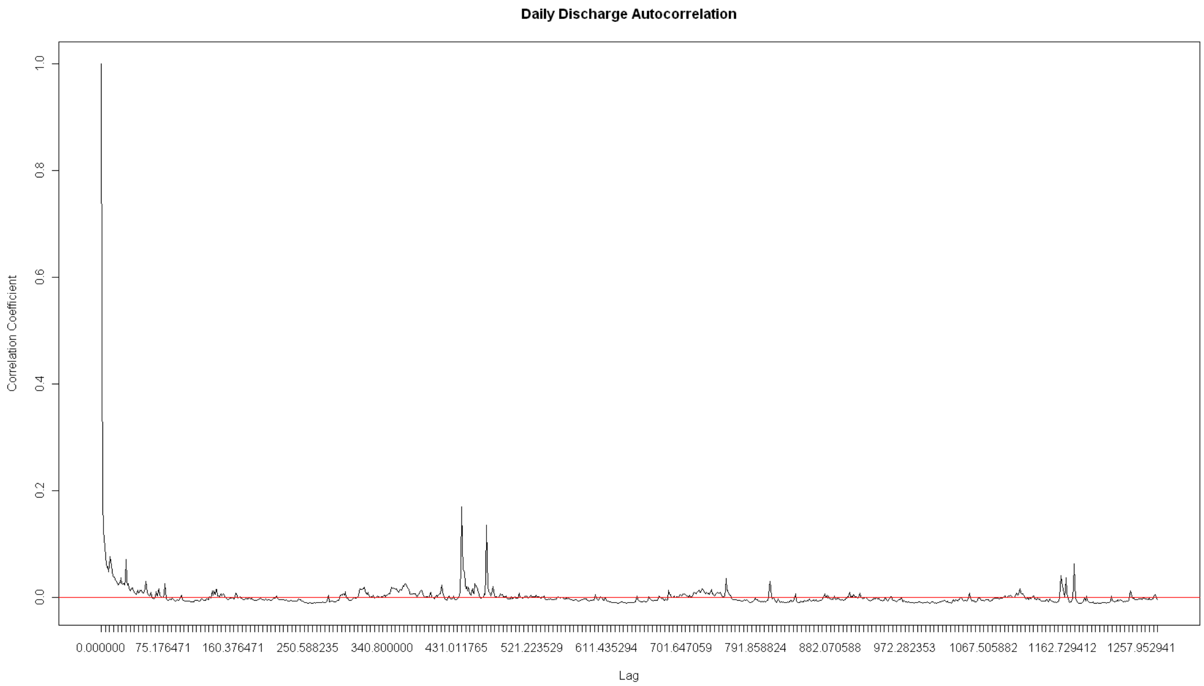
(c) Find the autocorrelation function of the records, extending your calculations to a lag equal to a tenth of the data length. Plot your results and comment on the relative decays on the two data sets.

Autocorrelation coefficient decays faster and fluctuates less in dry site sample.

– **Sacramento River(wet site)**



– **San Pedro River(dry site)**



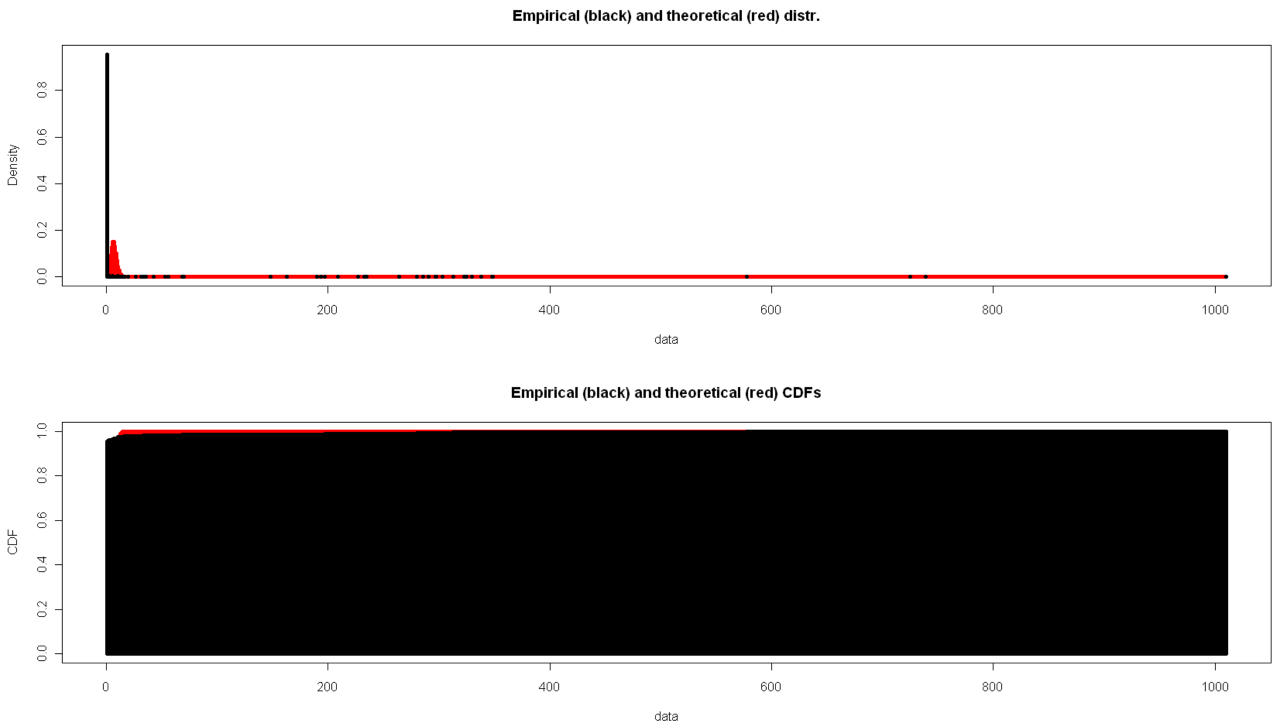
(d) Consider the upward crossings of the mean plus one standard deviation. Are the arrivals of such events consistent with a Poisson process?

The results of K-S tests for both samples are significant. According to the p-value and the density plots, both dataset are consistent with Poisson processes.

– **Sacramento River (wet site)**

One-sample Kolmogorov-Smirnov test, H_A: two-sided

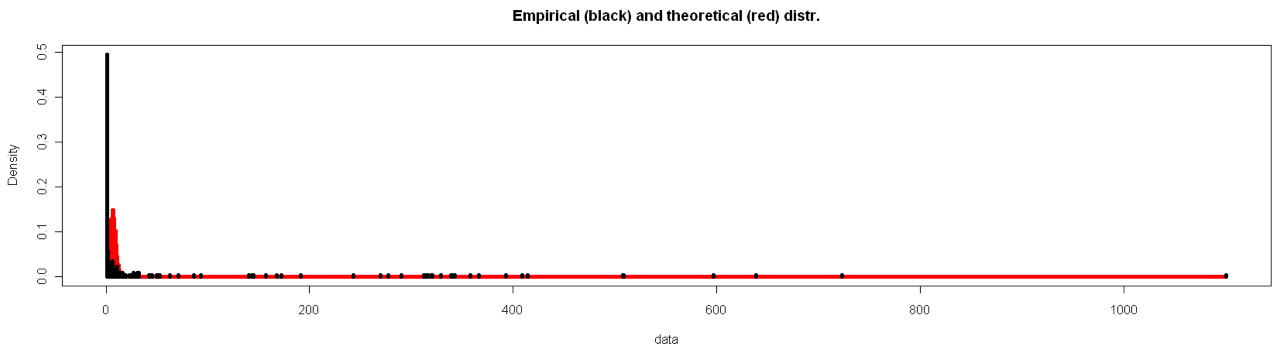
D = 1, p-value < 2.2e-16➔ Reject the H_o

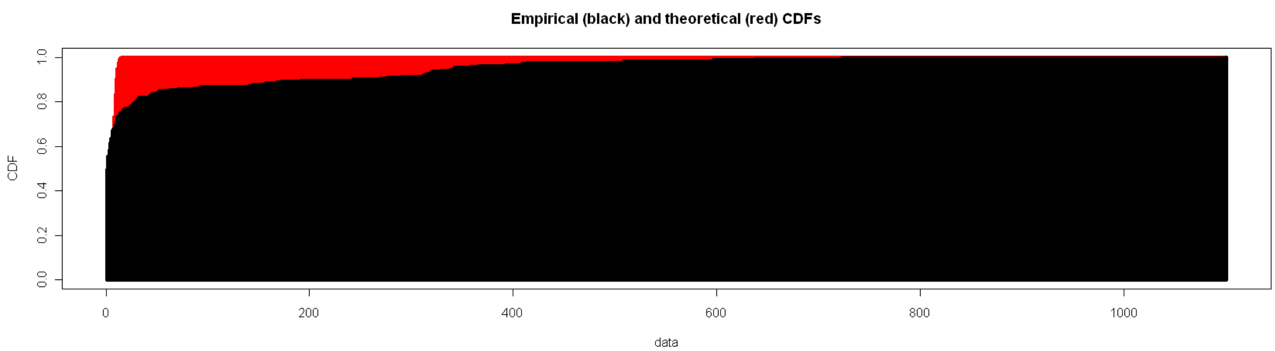

Empirical (black) and theoretical (red) distr.
Density
0.0
0.2
0.4
0.6
0.8
0
200
400
600
800
1000
data
Empirical (black) and theoretical (red) CDFs
CDF
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
data

- **San Pedro River(dry site)**

  One-sample Kolmogorov-Smirnov test, H_A: two-sided

  D = 1, p-value < 2.2e-16➔ Reject the H_o

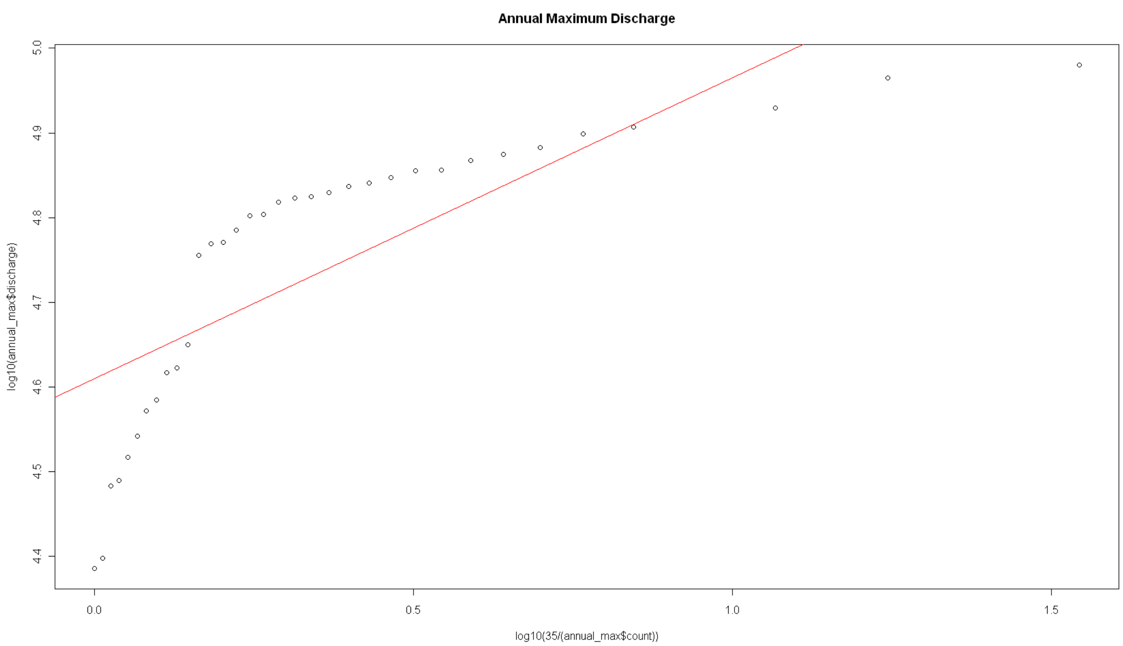


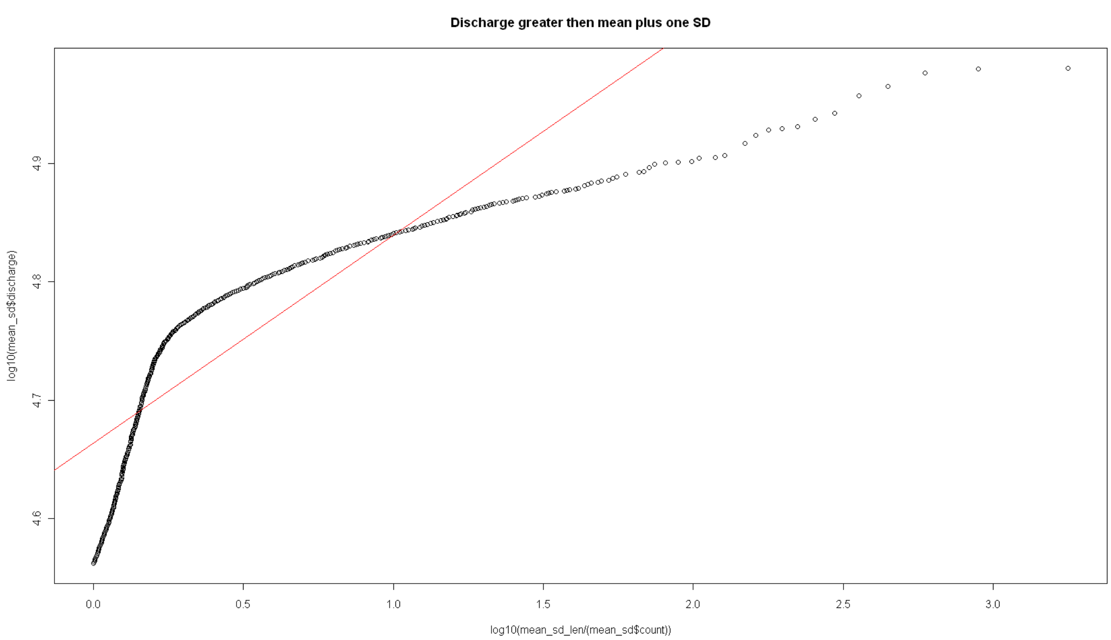


(e) Consider the series of maximal annual flows and also the series of daily flows greater than the mean plus one standard deviation. Are the distributions of such series nicely described by power-laws? As needed refer to the class presentations on the topic.

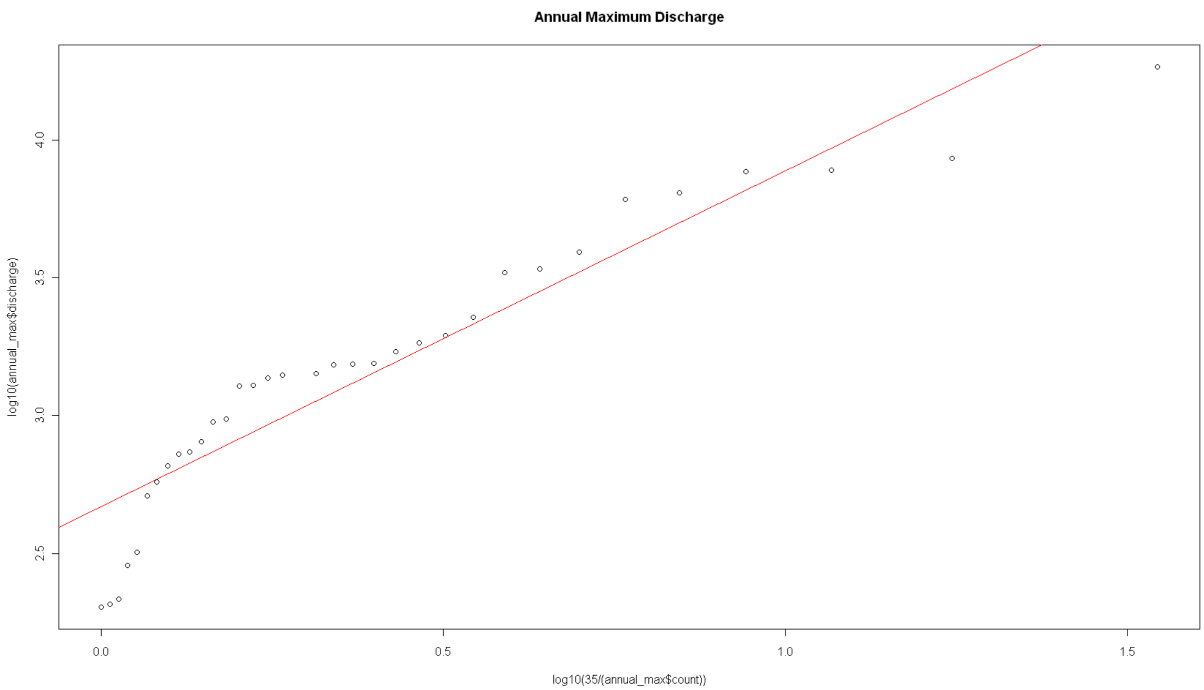
Both samples from the dry site, which is “dammed”, are nicely described by power-law.

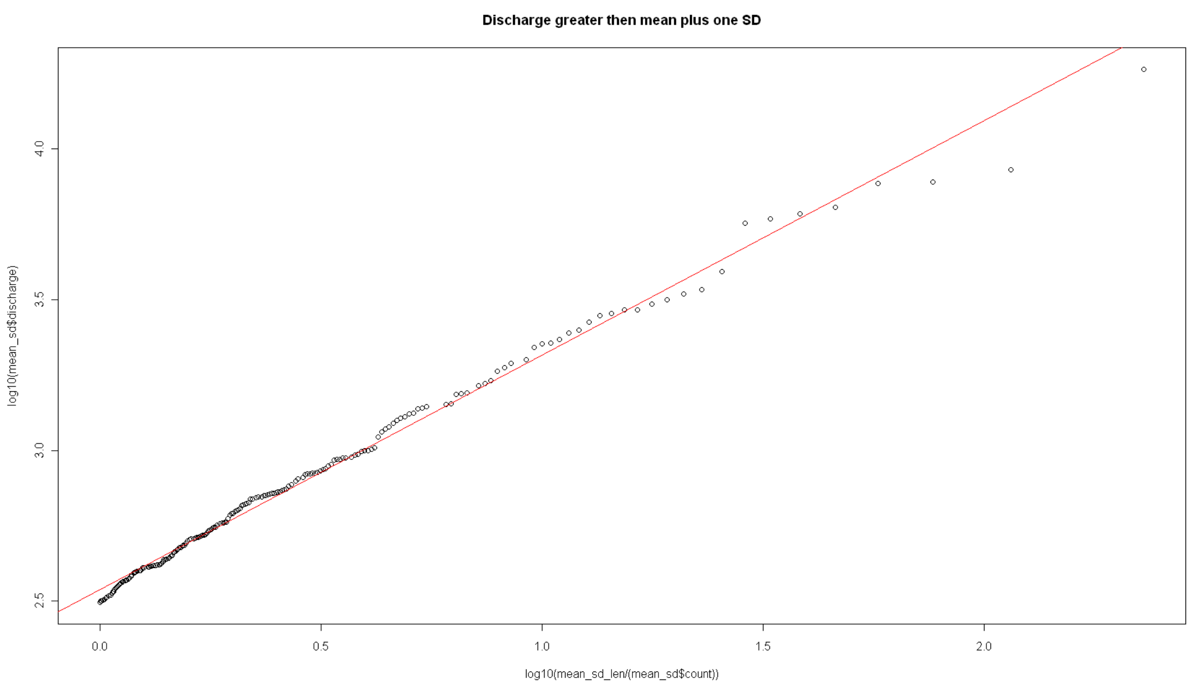
– **Sacramento River(wet site)**





- San Pedro River(dry site)



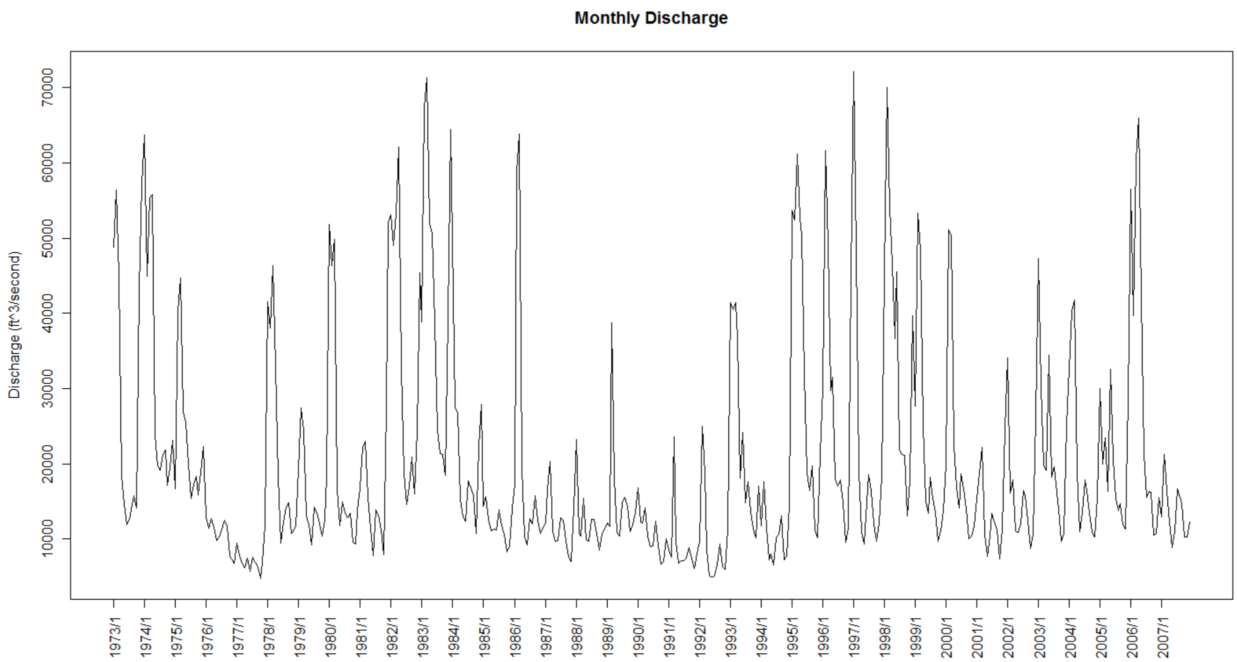


Monthly Discharge

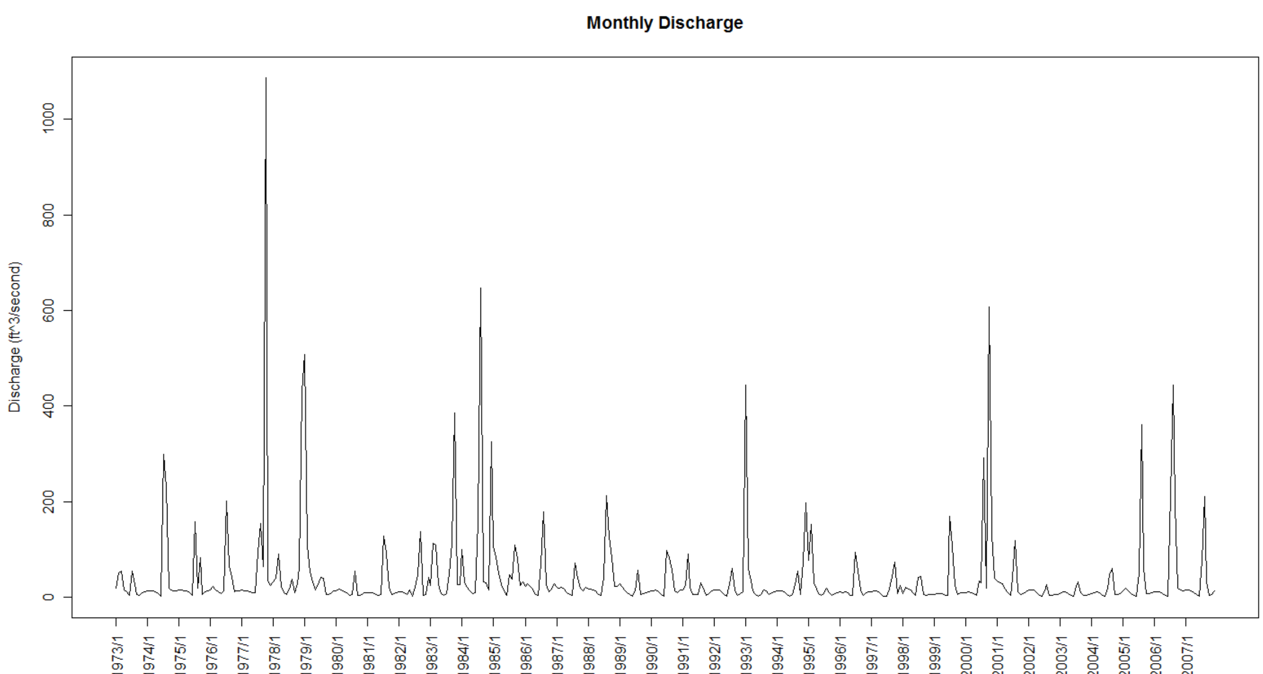
(f) Plot the new records, find their histograms, and also find the first four moments and coefficient of variation.

	1st mo.	2nd mo.	3rd mo.	4th mo	c.v.
Sacramento	2.04 X 104	6.32 X 108	2.67 X 1013	1.34 X 1018	0.7209
San Pedro	40	1.02 X 104	5.90 X 106	4.69 X 109	2.3168

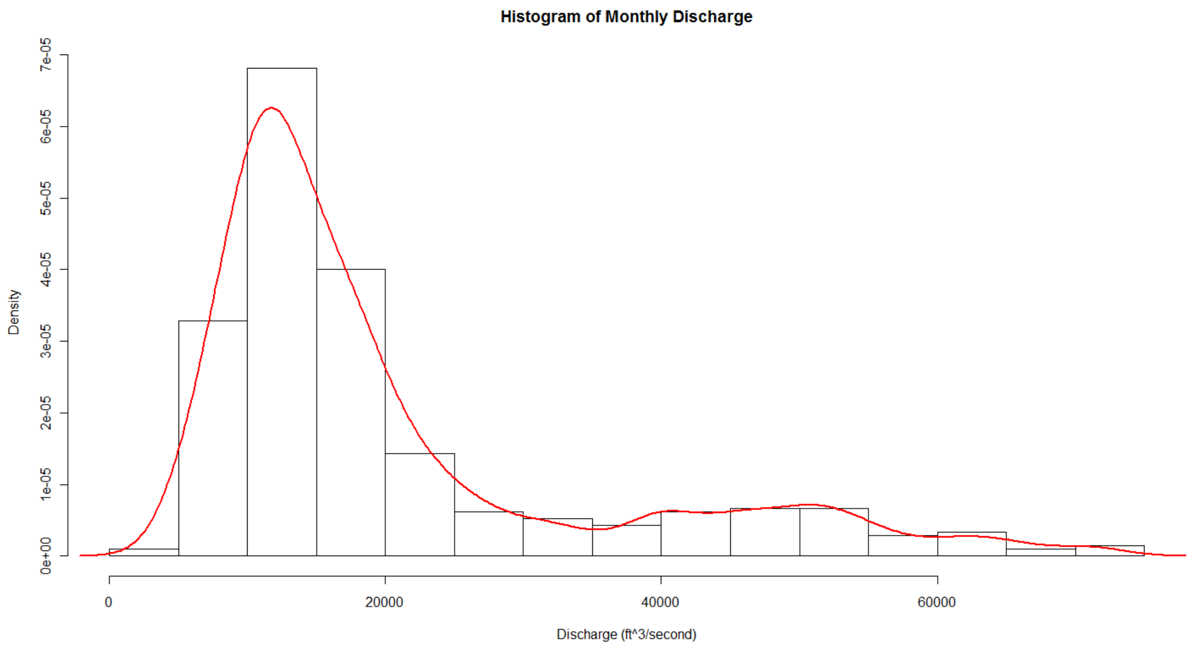
– Sacramento River(wet site)



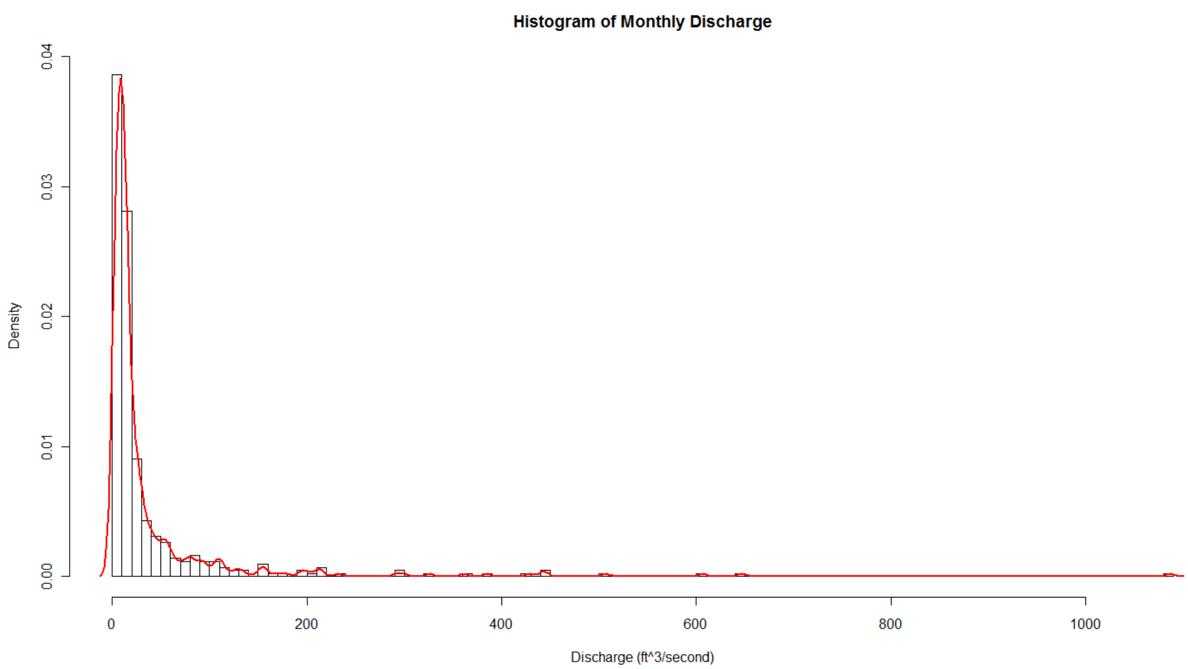
– San Pedro River(dry site)



- Sacramento River(wet site)



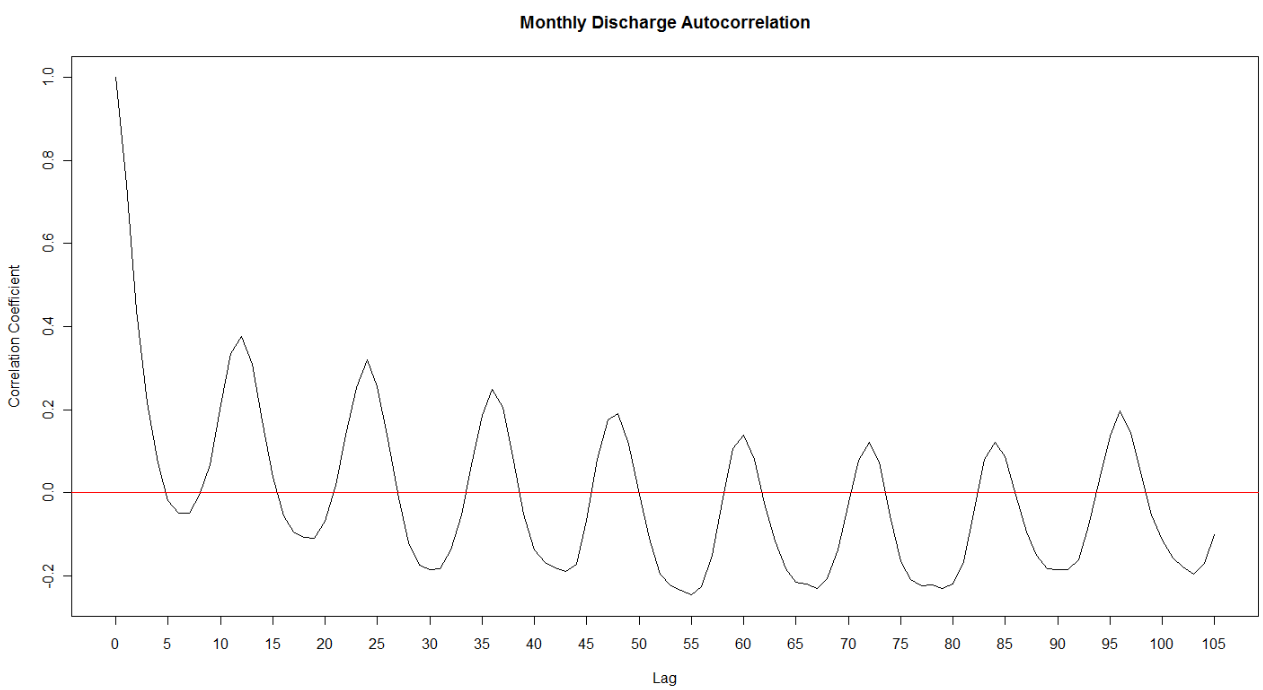
- San Pedro River(dry site)



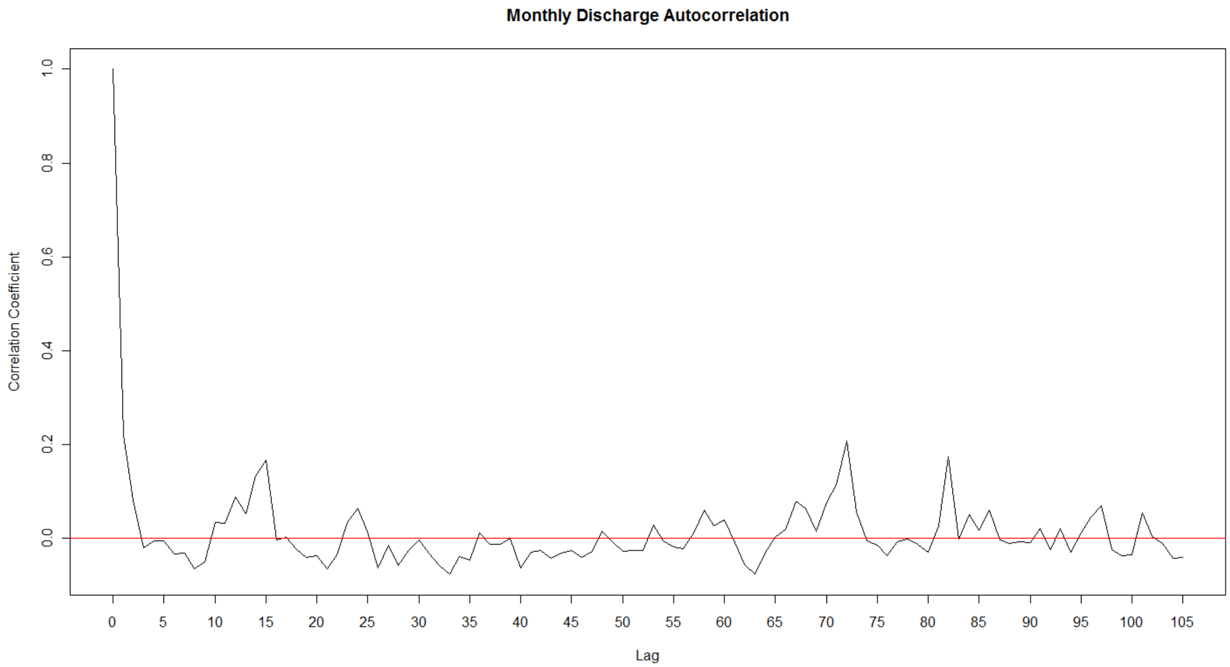
(g) Find the autocorrelation function of the new records, extending your calculations to a lag equal to a quarter of the data length. Plot your results and comment on the trends and relative decays on thetwo (three) data sets.

Dry site decays faster

- **Sacramento River(wet site)**



- **San Pedro River(dry site)**



(h) Compute the monthly means, variances and month to month correlations for the monthly sets and fit a seasonal autoregressive model to the records. For each of the two monthly series, simulate three synthetic series based on such a seasonal model. Plot your results together with the original series and comment on their overall features. Hint: Use the book "Numerical Recipes" to generate standard Gaussian values.

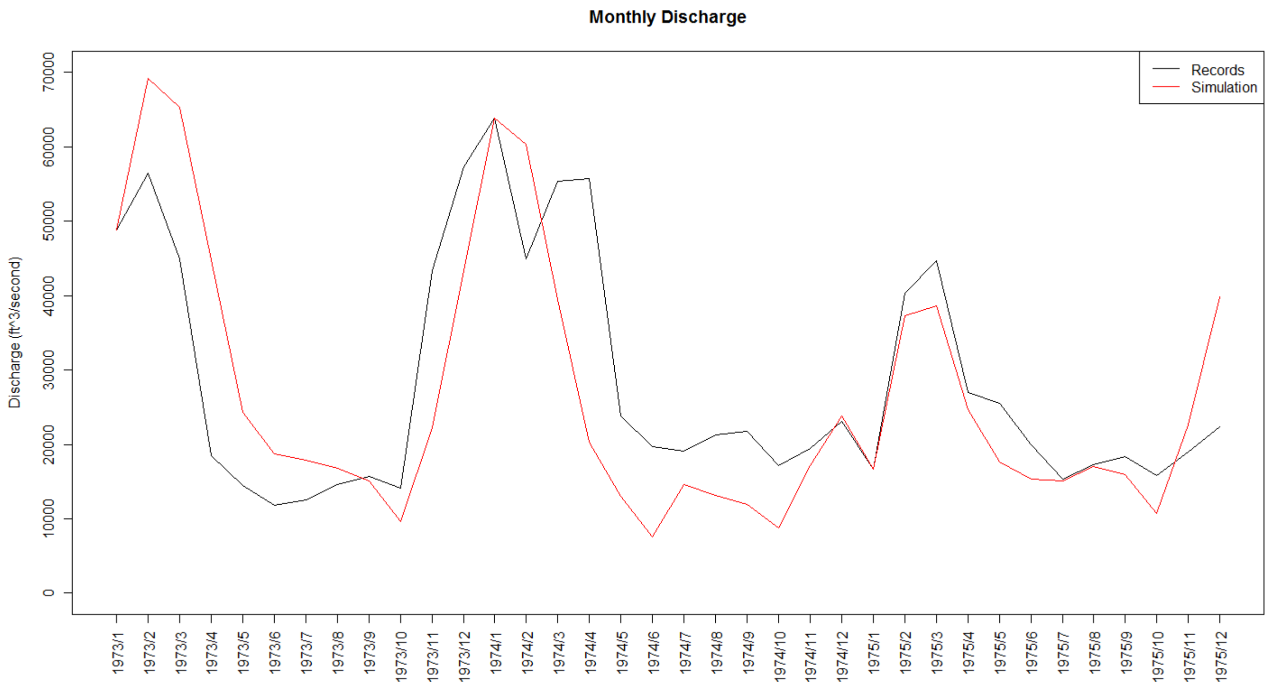
Lower CV, higher correlation coefficient => better predication

	Wet Site				Dry Site		
Month	Mean	Variance	Correlation Coefficient*		Mean	Variance	Correlation Coefficient*
1	30,514	325,874,901	N/A		48.08	12015.50	N/A
2	33,679	344,213,610	0.5964		29.88	1133.05	0.5182
3	34,198	342,798,848	0.8523		25.30	634.73	0.5856
4	23,441	279,284,924	0.7888		12.72	37.33	0.7939
5	18,251	160,435,367	0.8174		6.65	10.68	0.5967
6	14,979	79,886,917	0.8465		4.96	45.64	0.1771
7	14,799	17,034,952	0.7342		64.72	5259.40	0.0105
8	14,633	12,743,272	0.9108		114.20	19199.72	0.4735
9	14,024	14,211,353	0.8201		41.94	2041.43	0.2815
10	10,369	9,568,706	0.9006		72.95	45134.21	0.1505
11	13,638	73,839,251	0.6803		17.43	621.63	0.4998
12	22,235	214,398,461	0.8669		41.29	8228.79	0.4069

* Month i means the correlation between month i and month i-1

Three years

– Sacramento River(wet site)



– San Pedro River(dry site)