GEOMORPHIC PROCESSES
15-040-504

Laboratory #5: Flood Frequency Analysis

Purpose:

1. Introduction to flood frequency analysis based on a log-normal and Log-Pearson Type III discharge frequency distribution models.
2. Familiarization with the U.S. Geological Survey Water Resources Division website and data products.
3. Familiarization with some of the more advanced statistical functions available on spreadsheet programs.

Reading:

Ritter, D.F.; Kochel, R.C.; and Miller, J.R., 1995, *Process Geomorphology, 3rd ed*.: Dubuque, Wm. C. Brown Publishers, p. 168-171.

References:

Chow, V.T.; Maidment, D.,R.; and Mays, L.W., 1988, *Applied Hydrology*: New York, McGraw-Hill, Inc., p. 380-415.

U.S. Water Resources Council, 1981, *Guidelines for determining flood flow frequency*: Bulletin 17B of the Hydrology Committee, 183p.

Discussion:

Most projects built within potential reach of flood waters are engineered to withstand a flood of a given average recurrence interval, the *design flood*. A design flood with a recurrence interval of 200 years means that a project is designed to withstand a flood of a magnitude that is equaled or exceeded *on average*, once in 200 years. I emphasize "on average" because this should not be interpreted as meaning the flood will occur every 200 years. In fact, a 200 year flood could occur twice in one year. It is probably less confusing to think of it as the flood discharge that could be expected to be equaled or exceeded 50 times in a 10,000 year period. Many municipalities limit development of areas within reach of the 100 or 200 year flood. The mapping of the limits of the 100 year flood is, needless to say, extremely contentious because it determines property value and insurance rate. The process for delineating the 100 year flood limits is, unfortunately, not as straightforward as one would hope (unless one is a lawyer).

In this exercise you will determine the limits of the 200 year flood in an area of interest to you. You will make this determination based on the historical record of flooding and using a log-normal discharge frequency distribution. Although this was once a fairly daunting undertaking, the data necessary is easily available from the USGS, WRD website and most spreadsheet programs have the statistical procedures necessary for the analysis built into them.

It should be noted that we will be using a dataset that Ritter *et al.* (1995) refers to as the *annual series*, the highest discharge encountered during a given water year (running from October 1 to September 31) and is referred to as the *annual maximum series* by Chow et al., 1988. This is not the same thing as a *partial duration series* which is all discharges above a certain *base value*. There may be many discharges above this value during some years and none during other years.

It should also be noted that the log-normal frequency model is no longer the model of choice (although still widely used). The log-normal distribution assumes that when a histogram is constructed from the log of each discharge in the annual series it shows a normal (or "bell-shaped" distribution) and is therefore log-normal. The log-normal frequency distribution is based on two parameters, the mean of the log of discharges, α,

$$\alpha = \frac{\sum_{i=1}^{n} \log Q_i}{n} \tag{1}$$

and the standard deviation of the log of discharges, β

$$\beta = \sqrt{\frac{\sum_{i=1}^{n} (\log Q_i - \alpha)^2}{n-1}}\ . \tag{2}$$

The preferred frequency distribution, log-Pearson Type III, uses a three parameter fit, the skew, C_s, of the log of discharges in addition to α and β

$$C_s = \frac{n \sum_{i=1}^{n} (\log Q_i - \alpha)^3}{(n-1)(n-2)\,\beta^3}\ . \tag{3}$$

This distribution is described at length (and FORTRAN code is provided for performing a flood frequency analysis) in U.S. Water Resources Council (1981). This publication outlines the flood prediction procedure used by all Federal agencies as well by many other state, municipal, and private engineering groups. The procedure is presented quite clearly by Chow et al (1988) (an excellent book that I believe is used in one of Engineering's hydrology courses).

Frequency Models:

As we discussed in our investigation of effective discharge, we must estimate the characteristics of an extreme event from the recorded history of events. Before making an estimate we must first collect the data and see how it fits models that have been proposed for the frequency distribution of that system.

Two models are commonly used to fit the discharge frequency of the annual flood: lognormal and log-Pearson Type III. The lognormal distribution is called a two parameter model because it uses two parameters (mean and standard deviation) in the model. The log-Pearson Type III model uses three parameters (mean, standard deviation and skew). Models using as many as nine parameters have been used.

The lognormal distribution assumes that the frequency distribution of the log of discharge is normally distributed (bell-shaped curve). Before dealing with a lognormal distribution, let's look at normal distribution. Let's assume that the weights of male college seniors is normally distributed with a mean of 153 pounds and a standard deviation of 17 pounds (determined by using the Excel functions ***AVERAGE()*** and ***STDEV()*** respectively. We might ask if we had 1,000 male college seniors, what would be the weight of the heaviest. In order to do this, we must first determine the ***z-score, z,*** of this extreme weight. We can do this using tables available in any statistics text or by using the Excel function ***NORMSINV().*** The probability of the weight of this one in one thousand big boy being exceeded is

$\frac{1}{1000}$ or 0.001. Another way of expressing this is that the probability of an individual with a less than or equal weight is 0.999. ***NORMSINV(0.999)*** tells us that the z-score of 0.999 is 3.09. We can determine the weight corresponding to this weight by multiplying by the standard deviation and adding the mean (10)(17)+(153) or 323 pounds.

Log-normal distribution:

Now lets take a look at the lognormal distribution of annual floods. The first step is to take the logarithm (either base 10 or Naparian) of each flood magnitude then determine the mean, standard deviation, and skew of these logs. Although (1), (2), and (3) could be used, it's much easier to use the Excel built-in functions ***AVERAGE()***, ***STDEV()***, and ***SKEW()*** respectively. The z-score of a particular magnitude of discharge Q_i is then

$$z = \frac{\ln Q_i - \alpha}{\beta} \qquad (4)$$

The probability of this discharge being equaled or exceeded, $P(Q>Q_i)$, is determined by 1 - ***NORMSDIST(z)***. You may similarly find the discharge corresponding to a particular magnitude of flood using the Excel function NORMINV(1-probability of exceedance, α, β).

Log-Pearson Type III:

Log-Pearson Type III distribution is similar to the log-normal distribution; in fact, if skew equals zero, log-Pearson Type III distribution becomes a log-normal distribution. We will use an approximation for the log-Pearson Type III presented in Chow et al., 1988. Instead of using the *z* score *z* in calculations of discharge from probability, an adjusted variable is used, K_t

$$k=\frac{C_s}{6} \qquad (5)$$

$$K_T=z+(z^2-1)k+\frac{1}{3}(z^3-6z)k^2-(z^2-1)k^3+zk^4+\frac{1}{3}k^5 \qquad (6)$$

so to calculate a corresponding discharge for the z-score of a particular probability use

$$\ln Q = \beta K_t + \alpha \qquad (7)$$

The Ohio River Flood Example

All right, let's look at some real data. Annual flood peaks for any rivers, including Ohio River at Cincinnati, may be downloaded from the Survey's data retrieval site, http://waterdata.usgs.gov/nwis-w/US/. A map of the U.S. is displayed where each state is "clickable". Clicking on the state of interest will take you to the home page the state has prepared. You will be asked if you want to call up the data either by entering the gage ID (which you probably don't know), or by selecting a county from a list of counties or from a map of counties, or by querying their data base. The gage ID for Ohio River at Cincinnati is 03255000. Several data options may be displayed but you want peak flow data (not offered for all rivers). Clicking on peak flow will display the range of years of data available (remember you need at least thirty years of data). There will be a number of "radio buttons" controlling the form of the output data. Make sure click the button labeled "Only annual peaks" otherwise you will not have an annual series. Also click the button labeled "Tab-delimited text data file - MM/DD/YYYY". All of the data for Ohio River at Cincinnati is presented in Appendix A.

Perform the following steps:

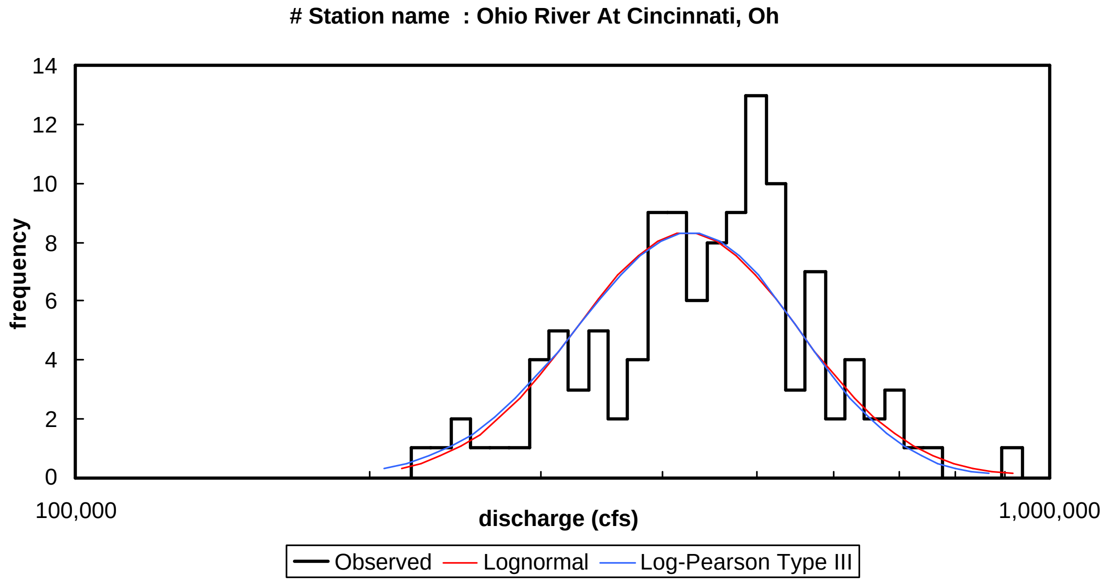
1. Add a fourth column of the log discharge, *=ln([cell address])* for the natural log.

2. Calculate the mean, standard deviation, minimum, and maximum value of the log of discharge, *=average([cell range]), =stdev([cell range]), =min([cell range]), =max([cell range])*

3. Calculate the bin width of a 50 bin histogram of discharge (*=([maximum discharge] - [minimum discharge])/49*) (note this is discharge not log of discharge).

4. Create the histogram bins by putting the minimum discharge in a cell and adding it and the bin width (calculated in the previous step) to the cell below it. Add the bin width to that value and enter it in the cell below. Keep doing this until you have 50 bins. The value in the fiftieth bin should equal the maximum discharge.

5. Histogram the log-discharges using the fairly complex (but neat) procedure discussed in class.



6. For each histogram bin, calulate the predicted frequency of occurrence according to the log normal distribution (*=NORMDIST([address of bin cell],[mean of ln discharge], [standard deviation of ln discharge], FALSE)*[# of observations] * [bin width of ln discharge histogram]*) and replace cummalative percent with predicted frequency.

7. Do the same thing for the log-Pearson Type III distribution

Now lets take a closure look at the disasterous flood of 1937 which had an estimated discharge of 894,000 cfs. The z-score is (13.70-12.96)/0.26 or 2.83. Assuming a log-normal distribution, this corresponds to an exceedance probability of 1-NORMSINV(2.83) or 0.002292 which corresponds to a recurrence interval of 436 years. Similarly with a skew of –0.20, Kt of 2.84 is determined corresponding to a recurrence interval of 940 years.

<u>Analysis and Questions</u>:

1. Unfortunately the closest station to where we gauged Little Miami River is a Fort Ancient (station # 03242500). Retrieve this data.

2. Download and carefully study the Ohio River analysis spreadsheet.

3. Enter the Little Miami data into the spreadsheet and perform the analysis.

4. Calculate the recurrence interval corresponding to the discharge you calculated in the first exercise using both the log-normal and log-Pearson Type III distribution.

5. Is the data best fit with a log-normal or log-Pearson Type III distribution.

6. What is the recurrence interval of the largest flood of record according to the log-normal and log-Pearson Type III distribution.

7. Neatly summarize your procedure and results.

```
# US GEOLOGICAL SURVEY
# PEAK FLOW DATA
#
# Station name  : Ohio River At Cincinnati, Oh
# Station number: 03255000
# latitude (ddmmss)............................ 390540
# longitude (dddmmss).......................... 0843038
# state code................................... 39
# county....................................... Hamilton
# hydrologic unit code......................... 05090203
# basin name................................... Middle Ohio-Laughery
# drainage area (square miles)................. 76580
# contributing drainage area (square miles).....
# gage datum (feet above NGVD)................. 428.93
# base discharge (cubic ft/sec)................
# Gage heights are given in feet above gage datum elevation.
# Discharge is listed in the table in cubic feet per second.
#
# Peak flow data were retrieved from the
# National Water Data Storage and Retrieval System (WATSTORE).
#
# Format of table is as follows.
# Lines starting with the # character are comment lines describing the data
# included in this file.  The next line is a row of tab-delimited column
# names.  The next line is a row of tab-delimited data type codes that
# describe the width and type of data in each column.  All following lines
# are rows of tab-delimited data values.
#
# ----Water Years Retrieved----
```

Date	**Annual Maximum Discharge**	**Gage at Peak**	**ln Discharge**
1773	821,000	76	*13.62*
1792	594,000	63	*13.29*
1793	498,000	57	*13.12*
02/18/1832	616,000	64.3	*13.33*
12/17/1847	604,000	63.6	*13.31*
06/16/1858	326,000	43.8	*12.69*
02/22/1859	472,000	55.4	*13.06*
04/17/1860	388,000	49.2	*12.87*
04/20/1861	391,000	49.4	*12.88*
01/24/1862	497,000	57.3	*13.12*
03/12/1863	316,000	42.8	*12.66*
05/20/1864	283,000	39.8	*12.55*
03/07/1865	484,000	56.3	*13.09*
09/26/1866	312,000	42.5	*12.65*
03/14/1867	477,000	55.8	*13.08*
03/30/1868	376,000	48.2	*12.84*
04/02/1869	384,000	48.8	*12.86*
01/19/1870	471,000	55.3	*13.06*
05/13/1871	290,000	40.5	*12.58*
04/13/1872	304,000	41.8	*12.62*
02/21/1873	300,000	41.5	*12.61*
01/13/1874	373,000	47.9	*12.83*

Date	Annual Maximum Discharge	Gage at Peak	ln Discharge
08/06/1875	472,000	55.4	*13.06*
01/29/1876	426,000	51.8	*12.96*
01/20/1877	452,000	53.8	*13.02*
03/17/1878	221,000	33.6	*12.31*
12/15/1878	297,000	41.2	*12.60*
02/17/1880	444,000	53.2	*13.00*
02/16/1881	414,000	50.8	*12.93*
02/21/1882	508,000	58.6	*13.14*
02/15/1883	650,000	66.3	*13.38*
02/14/1884	734,000	71.1	*13.51*
01/20/1885	350,000	46	*12.77*
04/09/1886	477,000	55.8	*13.08*
02/06/1887	488,000	56.3	*13.10*
04/01/1888	284,000	39.9	*12.56*
02/22/1889	268,000	38.3	*12.50*
03/26/1890	522,000	59.2	*13.17*
02/25/1891	498,000	57.4	*13.12*
04/25/1892	326,000	43.8	*12.69*
02/20/1893	460,000	54.9	*13.04*
02/15/1894	241,000	35.6	*12.39*
01/14/1895	379,000	48.4	*12.85*
04/04/1896	370,000	47.7	*12.82*
02/26/1897	544,000	61.2	*13.21*
03/29/1898	547,000	61.4	*13.21*
03/08/1899	495,000	57.4	*13.11*
02/17/1900	258,000	37.3	*12.46*
04/27/1901	524,000	59.7	*13.17*
03/05/1902	410,000	50.9	*12.92*
03/05/1903	439,000	53.2	*12.99*
03/09/1904	349,000	45.9	*12.76*
03/13/1905	378,000	48.3	*12.84*
04/02/1906	404,000	50.4	*12.91*
01/21/1907	631,000	65.2	*13.36*
04/04/1908	474,000	55.9	*13.07*
02/28/1909	457,000	54.6	*13.03*
03/07/1910	421,000	51.8	*12.95*
02/03/1911	387,000	49.11	*12.87*
03/27/1912	441,000	53.4	*13.00*
04/01/1913	662,000	69.9	*13.40*
04/04/1914	364,000	47.2	*12.80*
02/07/1915	482,000	55.9	*13.09*
04/01/1916	445,000	53.5	*13.01*
03/17/1917	484,000	56.1	*13.09*
02/12/1918	500,000	61.8	*13.12*
01/06/1919	424,000	52	*12.96*
03/22/1920	462,000	54.6	*13.04*
03/11/1921	305,000	41.5	*12.63*
12/27/1921	485,000	56.1	*13.09*
02/05/1923	369,000	47.6	*12.82*
01/06/1924	480,000	55.8	*13.08*
02/18/1925	309,000	42.3	*12.64*
01/25/1926	354,000	46.3	*12.78*
01/27/1927	530,000	59.1	*13.18*
12/20/1927	403,000	50.4	*12.91*
03/04/1929	434,000	52.7	*12.98*

Date	Annual Maximum Discharge	Gage at Peak	ln Discharge
11/22/1929	334,000	44.5	*12.72*
04/08/1931	329,000	44.1	*12.70*
02/07/1932	403,000	50.4	*12.91*
03/21/1933	604,000	63.6	*13.31*
03/09/1934	357,000	46.6	*12.79*
03/16/1935	430,000	52.4	*12.97*
03/28/1936	554,000	60.6	*13.22*
01/26/1937	894,000	80	*13.70*
01/03/1938	334,000	44.2	*12.72*
02/07/1939	538,000	58.28	*13.20*
04/24/1940	568,000	60.04	*13.25*
06/09/1941	242,000	35.62	*12.40*
03/20/1942	364,000	45.51	*12.80*
01/04/1943	594,000	60.8	*13.29*
04/15/1944	400,000	48.6	*12.90*
03/07/1945	708,000	69.2	*13.47*
01/11/1946	391,000	47.6	*12.88*
01/24/1947	306,000	41	*12.63*
04/17/1948	637,000	64.8	*13.36*
01/30/1949	449,000	52.63	*13.01*
02/04/1950	547,000	58.57	*13.21*
12/10/1950	506,000	55.98	*13.13*
02/01/1952	510,000	56.92	*13.14*
03/06/1953	245,000	35.95	*12.41*
03/07/1954	230,000	34.01	*12.35*
03/09/1955	592,000	61.04	*13.29*
03/17/1956	458,000	53.18	*13.03*
04/10/1957	441,000	52.3	*13.00*
05/11/1958	544,000	57.98	*13.21*
01/26/1959	493,000	55.52	*13.11*
04/05/1960	370,000	45.86	*12.82*
05/09/1961	452,000	55.34	*13.02*
03/02/1962	595,000	61.3	*13.30*
03/10/1963	540,000	59.41	*13.20*
03/11/1964	650,000	66.2	*13.38*
03/29/1965	396,000	47.19	*12.89*
02/17/1966	468,000	53.04	*13.06*
03/11/1967	566,000	59.78	*13.25*
05/30/1968	510,000	56.77	*13.14*
02/04/1969	284,000	40.75	*12.56*
04/05/1970	407,000	50.32	*12.92*
02/23/1971	384,000	47.68	*12.86*
04/25/1972	436,000	51.44	*12.99*
12/13/1972	469,000	53.81	*13.06*
01/14/1974	464,000	53.45	*13.05*
03/24/1975	417,000	50.1	*12.94*

count:	118
minimum:	12.31
maximum:	13.70
mean:	12.96
std dev:	0.26
skew:	-0.20
step:	0.05