

**Project EDDIE: STREAM DISCHARGE**

**Student Handout**

This module was initially developed by Bader, N.E., T. Meixner, C.A. Gibson, C.M. O’Reilly, and D.N. Castendyk. 26 June 2015. Project EDDIE: Stream Discharge. Project EDDIE Module 5, Version 2. <http://cemast.illinoisstate.edu/data-for-students/modules/stream-discharge.shtml>. Module development was supported by NSF DEB 1245707.

# Learning objectives:

* You will understand how to download, organize and analyze streamflow data.
* You will learn about major climate impacts in their region.
* You will analyze streamflow data to detect and quantify climate change impacts on water quantity in their region.
* You will learn about flood events and how to predict the likelihood of big flood events.

Why this matters:Fresh water is a fundamental resource for our society. Discharge measures the volume of fresh water passing by a point on a riverbank per unit time. Discharge is also a way to measure the quantity of water that is available. For example, water rights are often measured in units of discharge. Fish might require a certain discharge in a reach of stream in order to thrive or move through the reach. Conversely, the highest discharges result in floods that may harm people and structures. Discharge is fundamentally connected to our hydrologic cycle. Streams are fed by water that originally fell as rain and snow. Changes in the quantity of rain and snow, their distribution, and their timing, can be expected to cause changes in stream discharge. Clearly, it is important to be able to understand the way discharge varies. How can we measure these changes?

We must start with data. In this activity we will use data from the United States Geologic Survey (or USGS) network of stream gaging stations. The USGS is a governmental organization established in 1879, as part of the Department of the Interior. Originally tasked with the classification and mapping of United States public lands (and assessment of their mineral resources), the USGS has since expanded their role as a provider of impartial information on the status of ecosystems in the United States. (See <http://www.usgs.gov> for more details.) The USGS intensively monitors a network of 37 streams in relatively undisturbed watersheds around the US. These monitoring sites are collectively called the Hydrologic Benchmark Network.

Outline:

1. Discussion of papers read for class and quick PowerPoint introduction
2. Activity A: Variability in stream flow
3. Activity B: Changes in discharge over time
4. Activity C: Peak discharge and flood hazard

# **Activity A**: Variability in real stream data

## Viewing and accessing data:

1. Navigate to the Hydrologic Benchmark Data website, at <http://ny.cf.er.usgs.gov/hbn/index.cfm>

1. Click on “Site list” to see a map showing the locations of the gaging stations in the network. These stations have data on both streamflow and chemistry. Today we will concentrate on the streamflow data.
2. Select the Neversink River in New York to see a photo of the river and the types of data available.
3. Click on the Site Description link. You can see that the water draining past this point has been collected from a drainage area of 66.6 square miles (172.5 square km).
4. Click on Current/Historical observations and scroll down to see how discharge and temperature have changed over the past week.

**Questions**:

1. Look at the temperature data. How variable was temperature over the past week (the maximum minus the minimum)? A good way to think about variability is to think about percent change. Estimate the mean value of the data. Approximately how much higher (as a percent) are the highest temperatures?
2. Based on the temperature graph, what probably drives temperature changes in the Neversink River? Do you think there is any relationship between temperature and discharge? If so, what do you think it might be?
3. When discharge is high enough, flooding occurs. Is the discharge you observe here unusually high? Unusually low? Typical for the region? Can we answer these questions with only a week’s worth of data?

## Seasonal variation in streamflow:

We might expect flow in a stream to change seasonally. After all, most (or all) of the streamflow that you observe originated as rain and snow falling in the watershed, and precipitation in most places is seasonally variable. Let’s take a look at how streamflow changes over an entire year to see what happens.

* Go back to the summary of all available data for the Neversink River, and choose Daily Data. Type in a start date of January 1 of last year, and an end date of December 31 the same year. Make sure the output format is Graph and click Go.
1. **Question**: Temperature is high in summer and low in winter, as you might expect. What about discharge? Is it the same all year? What months have the highest discharge? The lowest?

#

# Activity B: Changes in discharge over time

##

## B.1: Change through time on the Neversink River

Let’s go back to an important question we had at the beginning of this activity. Is stream discharge changing through time? How might we answer this question, given the seasonal variability of the data?

One solution is to reduce the variability in our data by considering summer and winter data separately. In order to do this, we will need to manipulate the data ourselves using Excel.

1. First, export the data (your instructor may have done this step for you). Go to Monthly Statistics for the Neversink River. Select discharge, and leave the date range blank to get the entire date range. Choose tab-separated data in YYYY-MM-DD format, and save to file. Once you click Submit, a text file called “monthly” will be saved to your computer.
2. Now import the data into Excel. Open Excel, and select File > Import > Text file. Navigate to your “monthly” file and open it. (If it is grayed out, you may need to change “Text Files” to “All files” in the drop-down menu under the dialog box.) The text file is delimited by tabs, so check the appropriate boxes to tell this to Excel, accepting the remaining defaults.
3. You only need three columns: year, month, and mean discharge. Take a close look at your columns to see if you can figure out which is which. When you are confident, type “Year”, “Month”, and “Discharge” in the cells above the appropriate columns. Now you can delete the remaining columns, and the rows above your labels.
4. Now is a good time to save your Excel spreadsheet. Call it something sensible such as “Neversink monthly discharge” and save it in a folder where you can find it again. You should periodically re-save your spreadsheet so that you don’t lose your work.

### Plotting discharge by month

1. Currently your data is organized by year, and within each year it is organized by month. In order to make a plot of a particular month, you should arrange your data by month instead.
* Click one of the cells in the month column and use Sort Ascending in Excel to sort by month - you will see all of the January data first, followed by February, etc.
* Select all of the data cells from February. (Hint: you can click once to select the top left cell, then scroll down and shift-click on the bottom right cell to select the data you want.)
* Graph the data with a scatter plot. You should see a cloud of points with discharge on the Y axis and year on the X axis.
* Excel will also plot all of the integers representing the month in another series on the plot. You can click on a point in the series to select it, then delete the series to get it out of the way.
1. Take a look at the plot you made. This shows the mean monthly discharge for February, over the entire period of record. How variable is it? In which year was the highest mean discharge? What was it?
2. Now let’s compare this data to mean discharges from a summer month. Let’s use August.
* Right-click on the chart and Choose Select Data from the drop down menu.
* In the “Name” box, rename your series “February data.”
* Add a new series, and name it “August data.”
* Click on the small button next to the “X values” field. You can scroll down using the bar on the right and select the corresponding years next to the August data. (Alternatively, you can note the cell names at the top and bottom of the range, and type them into the field.)
* Do the same thing with the “Y values” field, selecting the August discharge data.
1. Examine the plot you made. How is August mean discharge different?

### Change in time

1. Can you see any change in time? It may be difficult to see, against such a variable background. We can use a regression line to visualize and measure this change through time.
* Right-click on a point from one of your data series, and select “Add trendline.”
* In the dialog box that pops up, make sure that you have selected a linear trendline and that you show the equation and R-squared on the screen.
* NOTE: The “**R-squared**” is the fraction of the variance in the Y-axis that can be explained by the X variable. For example, in your plot of Neversink discharge against time, an R-squared of 0.20 means that 20% of the variance in discharge can be explained by the trend. The other 80% of the variance in discharge would be due to other factors.
1. Is there a trend? Are August and February the same, or are they different?

## B.2: Try the analysis on a new watershed

1. This analysis tells us something about the drainage basin in Neversink, New York. Now let’s try this again in another location. Select another watershed in the Hydrologic Benchmark Network and repeat the analysis. When you are finished, convene with the rest of your class to share your results.
2. It is very likely that the watersheds that you analyzed are not the same. These watershed are from a variety of geographic regions, with different patterns of precipitation, different precipitation regimes (e.g., rain vs. snow), and different topography and soils. Why do you think the discharge pattern in your area is different from other places? Come up with a hypothesis to explain some of the differences that you observe. *How would you test this hypothesis - what additional data would you need?*

#

# **Activity C**: Peak discharges and flood hazard

## C.1 Extreme discharge events

When we are discussing water as a resource, we are often concerned with scarcity. For example, when water resources are allocated to water users, often the units are in discharge (e.g., cfs). Conversely, when discharge is high, we begin to worry about floods. A flood can be defined in several ways, but a good rule of thumb is: a flood is a discharge that is large enough to overtop the banks of the river channel. If a flood occurs in a populated area, it can endanger lives and structures. Floods are one of the most common types of natural disasters, and can occur in rainforests, deserts, or any landscape in between.

We have already developed most of the skills we need to understand flooding. What is different about flooding? The main difference is that we must focus our attention on extreme events, not average events. Sure, you can calculate that your stream has an average discharge of 1000 cfs, but this information is of no interest to you when a 10,000 cfs peak discharge event is carrying away your house. Thus, in order to understand floods, we will need data on peak discharges. A peak discharge is simply the highest recorded discharge over a period of time, often over one year.

We can use peak data provided by the USGS to ask questions about the likelihood of flood events. Because we are thinking about the interactions between rivers and humans, it makes sense to think about rivers in inhabited areas. The USGS Hydrologic Benchmark locations were selected to be in areas that are minimally affected by humans and human activities. However, these benchmark sites are a subset of a larger stream monitoring network maintained by the USGS. We will use this larger network for the next activity.

1. Navigate to the USGS network of stream gaging stations at <http://waterdata.usgs.gov/nwis/rt>
2. Notice the colored dots, depicting real-time conditions at stream gages nationwide. The dots are colored to show you if the streamflow is unusually high or low for this time of year. Note that "unusual" is a matter of degree; while the 25-74th percentile category is certainly not unusual, the next higher and lower categories (between the 10th and 90th percentiles) are perhaps mildly unusual. Discharges greater than 90th or less than 10th are occur about 20% of the time, on average.
3. You can click on individual states to pull up larger maps of that state. Do this now with New York State. You can verify that the USGS benchmark stations are available here: see if you can find the Neversink River in the southeastern part of the state. Notice that you can mouse over the dots to see the name of the stations.

## C.2 Floods on the Mississippi

1. Add a new worksheet to your Excel spreadsheet, and rename it "Peak data."
2. Find the gaging station on the Mississippi River at St. Louis, in Missouri. (Note the warning about the pre-1933 data.) You can see that the web interface is similar to the interface you used when accessing benchmark network data.
3. Back at the USGS website, select "Peak Streamflow" from the dropdown menu.
4. Select "Tab-separated file." As before, copy the text and then use "Paste special" as text to get the data into your new Excel worksheet.

**Question 1:** Examine the data. What years are covered by this dataset? Find the maximum discharge on record during this time. When did it occur? What was the discharge in cfs? What was the gage height during this discharge?

Back to Excel...

1. Clean up the spreadsheet as you did with the daily data. You should keep the date, gage height, and discharge columns. Data from before 1933 is not reliable, so delete these rows also.
2. Using the Excel functions AVERAGE() and STDEV(), calculate the mean and standard deviation of discharge for this dataset. To use formulas in Excel, you must first type `=' and that you can refer to cell contents by their position. An example formula to calculate the mean of some numbers in column B between rows 3 and 7 might therefore be: =AVERAGE(B3:B7)

**Question 2:** What is the mean peak discharge across all of the years (always include units)? What is the standard deviation of peak discharge? Approximately how many standard deviations above the mean PEAK discharge was the highest peak discharge?

### Flood probability

We can use peak data such as this to calculate the probability of floods of a particular size. This is how we figure out that a particular discharge is e.g. "a 100-year flood" for a given stream.

* **Discharge probability** is the probability of experiencing a flood *of this size or larger* in any year.
* **Long-term recurrence interval** is a misleading term. It is equal to 1/*p* where *p* is the discharge probability. A flood frequency of 100, for example, means that a discharge this high or higher is a "100-year flood," which means roughly that over a the period of record, a flood of this size or larger occurred once per 100 years of record.

**Questions 3 and 4:** Think carefully about what causes floods.

1. If there was a 100-year flood two years ago, how long would you have to wait for the next 100-year flood? (Careful, this is a trick question!)
2. Since probability and recurrence interval are related, is it more realistic to predict a flood event next year with probability or with recurrence interval?

Make sure you understand these questions before you go on.

###

### Calculating flood probability with ranked data

The simplest method for estimating the probability of flood events uses peak discharges organized by rank - the largest peak discharge on record has rank "1", the second-largest has rank "2", etc. Use Excel to organize your peak flow data by rank like this:

1. Copy the all of the Peak Discharge data to a new worksheet. Name the worksheet "Floods."
2. Now use the "Sort Z to A" function to sort the data from the highest discharge at the top to the lowest at the bottom.
3. Add a new column, named "Rank." In the top row, with the largest flood discharge on record, put a "1" in the Rank column. Then rank the remaining peak discharges. In the Rank column, in the cell just below the "1," type a formula adding one to the cell above it, and hit return. There should now be a "2" in the cell. Fill Down to calculate rank in the remaining cells.
4. Now start a new column, called "Probability." For this column, we can calculate the probability of a discharge this large or larger, based on the data. We will calculate this as the total number of peak events this high or higher divided by the total number of events on record, or *p* = *R* / (*n*+1) where *R* is the rank and *n* is the total number of events on record.

**Note:** we adjust the denominator by one to correct a systematic bias in this method. The explanation for this adjustment is easiest to see in small datasets. Imagine that we had only one peak discharge on record; our calculated probability of exceeding that peak discharge next year should properly be 1/(1+1) = 50%, indicating complete uncertainty, not 1/1 = 100%. (Obviously, a dataset containing only one peak discharge is not likely to be very useful to us.)

1. Use the formula to calculate the discharge probability for each discharge.
2. Make a new column, and calculate the long-term recurrence interval, or 1/*p*.
3. Plot your results. Make a new X-Y scatterplot with Recurrence Interval on the X-axis and Discharge on the Y axis. Make both axes logarithmic.
4. To improve readability, right-click on the numbers along an axis and select "add minor gridlines."

**Question 5:** Based on your graph, what discharge is a 50-year flood for this basin? How about a 100-year flood? Which of these estimates are you more confident in?

## C.3: Flood probability, changing with time

Up until now, we have calculated flood probability based on the entire period of record. This is sensible because it uses all of the available data. However, this method implicitly assumes that the probability of a particular flood event is consistent over the entire period of record. What if the probability of a 100-year flood changes with time?

In order to address this question, we need to consider some of the things that affect discharge. If something causes discharge to increase through time, then the probability of large flood events may also increase accordingly.

###

### Urbanization: the problem with pavement

Urbanization (expansion of cities) may seem to have nothing whatsoever to do with rainfall and streamflow. However, urbanization is accompanied by increased *impervious surfaces* - surfaces like pavement and rooftops that restrict the flow of water. When rain falls on pavement, it cannot infiltrate into the ground; instead, it runs off directly into streams. This causes rain from storm events to enter rivers quickly, resulting in rapid changes in discharge. Streams that respond rapidly and dramatically to rain events are called “flashy” streams, and are more likely to flood.

Humans affect floods in another important way: we intentionally design and build flood control projects, such as dams, in order to reduce the impact of flooding on populated areas. We certainly hope that such structures reduce the likelihood of flooding!

For the next part of this activity, we will see how the likelihood of flooding has changed in three different watersheds near Seattle, Washington. To analyze this data, you will use exactly the same technique you used to assess flood probability on the Mississippi. The only difference is that you will need to divide the peak flow data into two smaller datasets: one dataset containing the early years of data, and a second dataset containing the recent data. By calculating the size of the 100-year flood (for example) from each dataset, you can compare the results to see how the watershed has changed through time.

Here is a bit of background information on each watershed. The easiest way to locate each watershed is to go to the page of Washington data, then select the Statewide Streamflow Table link to access the list of watersheds. Mercer Creek is in the Lake Washington Basin; the other two are in the Green/Duwamish River Basin. If you prefer to locate them from the map, all three are southeast of the Puget Sound area.

1. **Mercer Creek** (USGS 12120000, Mercer Creek near Bellevue, Washington): This is the closest of the three gaging stations to downtown Seattle. We have peak streamflow data beginning in 1957. Rapid urbanization began in this area in the late 1970s. Compare the 1957-1977 peak flow data to peak flow data after 1977 to detect the effects of urbanization.
2. **Green River** (USGS 12113000, Green River near Auburn, Washington): Gaging on this stream began in 1937. In 1961, the Howard Hanson Dam was built upstream as part of a flood control project on the Green River. Thus, the 1937-1961 peak flow data may be different from data after 1961.
3. **Newaukum Creek** (USGS 12108500, Newaukum Creek near Black Diamond, Washington): This drainage is in a forested semi-rural setting, which has not changed greatly since peak streamflow data begins in 1945. This data will be a useful point of comparison with the other two drainages.

Your task is to calculate the size of the ten-year flood event (the event with a 10% chance of occurring in any given year). You will make the following comparisons:

1. Calculate how the size of the 10-year flood event changed after 1977 at the Mercer Creek gaging station. As a non-urbanized “control,” also calculate the same flood event to see how it changed after 1977 at the Newaukum Creek station. (In other words, you will need to calculate flood frequency four times to answer this question!) A good way to answer this question is to calculate the “before” and “after” events, then calculate the percent change.
2. Do the same analysis, but this time calculate the change in the size of the 10-year flood event after 1961 at the Green River gage, compared to the Newaukum Creek station. (Don’t bother recalculating the Newaukum Creek data for this very similar time period; the dataset is good enough to use the values you obtained in the last question.)