Freshwater Ecology
Fall 2016 Homework 1
35 points
Due Tuesday, 25 September 2018

Flood frequency analysis

You are a fancy freshwater ecologist studying the ecology of wetlands next to the Waccamaw River. You know that an endangered tree species living in these wetlands needs to be flooded occasionally in order to reduce competition by other plant species. You have been hired to describe the frequency of floods on the Waccamaw and think about whether these estimates make any sense. Real-time (as in NOW!) and historical discharge data exists online for the Waccamaw River thanks to the United States Geologic Survey. We'll access and process these data using the following steps.

Directions:

1) To access all flow data for South Carolina, first go to https://waterdata.usgs.gov/SC/nwis/current/?type=flow.
2) We need at least 20 years of data to make a good attempt at flood frequency analysis so we'll use data from the Waccamaw near Longs, SC. Scroll down to station number #02110500 and click on the hyperlink (i.e., the station number) for this site to access these data.
3) Then, using the drop-down menu toward the top of the page labeled "Available data from this site" (it has a blue background), select 'Surface-water: Peak streamflow'.
4) On the new page that appears, click on the hyperlink at the top for "Tab-separated file" under Output formats to see the peak flows (a new page appears).
5) Save this tab-separated file of data for annual peak streamflows to your computer as a text file (it should be called "peak.txt"). You'll need to find the 'Save" command in your web browser to save the file—click on the three short horizontal lines at the top right in Firefox to do so. In Chrome you click the right mouse button and a menu appears with the option to "Save as".)
6) Next, start MS Excel and open the peak.txt file from wherever you saved it. You may have to change the file type to find the .txt file type (instead of just ".xls" file types by clicking on the box next to the file name and selecting "All Files").
7) Excel will start its Text Import Wizard. Click on the "Delimited" option toward the top of the dialog box, and then click "Next" at the bottom. Click "Next" again as long as "Tab" is highlighted at the top. Click "Finish" as long as "General" is highlighted at the top.
8) After scrolling down a bit (Row 67 in my file following the directions above), you'll find the actual peak streamflow data. Copy and paste the dates (column C in my file) and the annual peak discharge data (column E in my file) into two new columns *in a different sheet* of your Excel file so that you do not alter the original data.
9) Replace the text in the two cells toward the top containing the unhelpful "10d" and "8s" if they are there with Date and Discharge, respectively (or something similar). Next, highlight both the discharge data AND the dates in your new sheet at the same time. Sort these data based on the annual discharges from highest to lowest discharge using "custom sort" (look for the "Sort and Filter" option at the top right of the Excel screen). Select the discharge data column option, the values option, and the descending order or Z to A sorting option). I encourage you to regularly save the Excel file that you are now working on (and be sure to save as an Excel file).
10) In the first blank column to the right, add the rank for each year, with the largest discharge value ranked as 1, and the smallest discharge equal to how many years are in your data set (67 in this case). In other words, you need to enter ranked integers (from 1 to 67) into this column.
11) Add another column to the right to calculate the recurrence interval (T) for each data point, using the following equation. Note: you'll need to translate this equation and later equations using Excel's formula system that starts with an equal sign ("=") in the first data cell and refers to the

cells to the left with the data. After entering the first formula, copy and paste it to the remaining cells below in the recurrence interval column.

   a. Recurrence interval formula: $T = (n + 1) / m$
   b. Where n = the number of years of discharge data and m = the magnitude of the flood by its rank. And yes, m is the rank from step 6, not the actual peak discharge value.

12) Add another column to calculate the probability of exceedance (P), using the following equation:
   a. $P = 1 / T * 100$
   b. Where T = recurrence interval (and yes, P/100 is just the reciprocal of T)
13) Add one last column consisting of the log10-transformed peak discharge data for each year (the formula would look something like: =log10(data in peak discharge cell).
14) These calculations will now allow you to predict the probability of certain-sized floods occurring in a given year. Typically, you would plot P or T versus peak discharge using log-scaled graph paper or a computer program. However, you don't have the paper and Excel is finicky with plotting log scales on axes. Instead, we'll do a simple linear regression in Excel where we predict the log10-transformed peak discharge from P (so P is the independent data and log10(peak discharge) is the dependent data). Use the Data Analysis wizard in Excel to calculate this regression. You may have to install this wizard under File, Options, Add-Ins, Analysis ToolPak.
15) Once you have the equation (it's on the Excel regression output towards the bottom), you can now predict the probability of various peak discharges. Of course, you can also predict T because it is the reciprocal of P/100. An important point: your regression equation is predicting peak discharge in a log-10 scale. So, you'll need to transform your predicted discharge back into 'standard' units (e.g., if discharge = 2.5, it's really 10^2.5 or 316) for the questions below.

Each individual must hand in the following typed information as a hard copy:

1) A table with the peak flow discharge for each "water year" (Oct 1 to Sep 30) of record and all of your calculated columns from your work following the above directions.
2) Your regression output (just copy and paste what Excel's Data Analysis spits out).
3) Answers to the following questions. When you are asked to calculate answers, **show your work**. Sometimes you will have to re-arrange the regression equation for these calculations (the power of algebra!).

Questions:

1) How many years did the peak discharge of the Waccamaw River exceed 6000 cfs?
2) What is the recurrence interval (T) for a flow of 3040 cfs using the data directly vs. using the equation? (In other words, you are going to have to enter the log10-transformed value for 3040 into the Excel-produced regression equation to calculate what T equals.) Using the equation, what is T for a flow of 6300 cfs? Do the estimated T's seem correct (i.e., compare them to actual data)? Why or why not?
3) What is the probability (P) of a flow exceeding 4000 cfs using the equation? Does this estimate of P seem correct? Why or why not? What is P for a flow exceeding 10,000 cfs using the data directly vs. using the equation?
4) What was the maximum and minimum discharge (cfs) for the entire historical record available? Does there seem to be any trend over the years with respect to dry or wet years? I would make a graph to answer this question and include it in your answer.
5) Other studies have found your endangered plant requires a flood with a 15-yr recurrence interval. What is the estimated discharge associated with a flood that has a 15-yr recurrence interval using the regression equation estimate? Does this estimated peak discharge seem correct? Why or why not?
6) Would you be comfortable advising whoever hired you about the likelihood of a 50-yr flood based on the data set you downloaded? In other words do you think your advice would be accurate using these standard methods you have just followed? Why or why not?