

Tips for Data Manipulation

This section of the website covers the following topics:

- [How to retrieve data.](#)
 - [How to perform time conversions with daily data.](#)
 - [How to work with ungages sites.](#)
 - [The importance of the length of the period of record.](#)
 - [Selecting class boundaries for flow duration analysis.](#)
 - [Importance of using instantaneous peak flow data for flood analysis.](#)
 - [How to build confidence in results of analyses.](#)
 - [A tutorial with companion data in a downloadable spreadsheet.](#)
 - [An example with companion data in a downloadable spreadsheet.](#)
-

Obtaining Streamflow Data for Gaged Sites

Format of Available Data:

The main sources for streamflow data are the [US Geological Survey](#) (US Department of Interior) and state water resources departments (e.g. the [Oregon Water Resources Department](#)). The [Hydro Data Links](#) page contains links to hydrologic data on these websites and to other data sources.

The summary values that are most readily available to the public from these sources are the mean daily discharges (the average flow during each consecutive 24-hour period) for the period of days, months, or years of interest. The USGS web site lists the year, month, date, and [discharge](#) (usually in cfs) for the period of record.

Calculation of Monthly and Annual Values:

It is often necessary to calculate the monthly and annual data from the daily values. The following are some tips for calculating monthly and annual values:

Step 1: Download daily data for period of interest into an Excel spreadsheet. The data should occupy four columns: 1) water year, 2) month, 3) date, 4) discharge.

Step 2: In an adjacent column, at the end of each month, you should calculate the average for that month. This can be done using the average function in Excel. It is advisable to calculate these monthly averages manually for the first four water years of data. This will ensure that the values for the included leap year have been calculated correctly.

Step 3: After the first four years, it is possible to copy and paste the cells from the calculation for the four years worth of data to the rest of the data set.

Step 4: The annual values can be calculated using the same method. Averages at the end of each year for the first four years should be done manually and then these cells can be copied and pasted to the rest of the data set.

Tutorial | Example

(What is the difference between Tutorial and Example files? [Click here for an explanation...](#))

Obtaining Streamflow Data for Ungaged Sites

Project sites are usually located on small or moderate-size streams that do not have operating stream gages. Locating streamflow data applicable to these sites for the purpose of making hydrologic analyses can be difficult. Generally there are two options for finding appropriate data:

1. Use streamflow values from a nearby gaged site that has the same drainage area size.
2. Scale down the streamflow values from a gage on a larger nearby watershed to the size of the drainage area of the project stream.

Using streamflow values from a gaged site of the same size:

Step 1: Determine the drainage area of the project site.

Step 2: Locate a gaged site with a drainage area similar in size to that of the project site. The [Hydro Data Links](#) page has a table of all USGS gages for coastal Oregon. The USGS web site will allow you to search for gages based on drainage area.

Step 3: From the list of gaged sites with a similar drainage area identify the ones that may be located in similar topographic and climatic zones as the project site. For example, if the project site is located in a coastal basin, it would not be appropriate to use streamflow values from a gaged site in Eastern Oregon. Eastern Oregon is generally much drier than the coastal range and the terrain is generally not as steep.

When using a different stream to estimate streamflow values for a project site, it is important to consider the following:

1. Drainage areas are similar
2. Topography is similar (mountainous, desert, valley, etc.)
3. Climatic patterns should be similar (precipitation patterns, snow accumulation, etc.)
4. Soil characteristics should be similar (porous, impermeable, etc.)
5. Land use should be similar (forested, urban, agricultural, etc.)
6. Length of Period of Record

The more components that are similar between the two basins, the more likely it is that the streamflow values from the gaged site will accurately depict the streamflow characteristics of the project site. Sometimes it may be necessary to look for gaged sites in other states with similar conditions (Washington, California) to find an appropriate site.

Scaling down streamflow values for a large watershed:

The project site may be located on a small part of a larger basin or the only gaged stream with similar topographic characteristics may have a large drainage area. For example, the project site may be on a small creek that is a tributary to a large river. The large river may be gaged near its mouth where the drainage area is quite large.

Step 1: Determine the drainage area of the project site.

Step 2: Determine if the project stream is part of a larger gaged basin.

Step 3: Determine the drainage area of the large basin.

Step 4: Calculate the proportionality (Drainage Area Ratio) between the project site drainage area and the drainage area of the gaged watershed.

Step 5: Scale down streamflow values by this ratio.

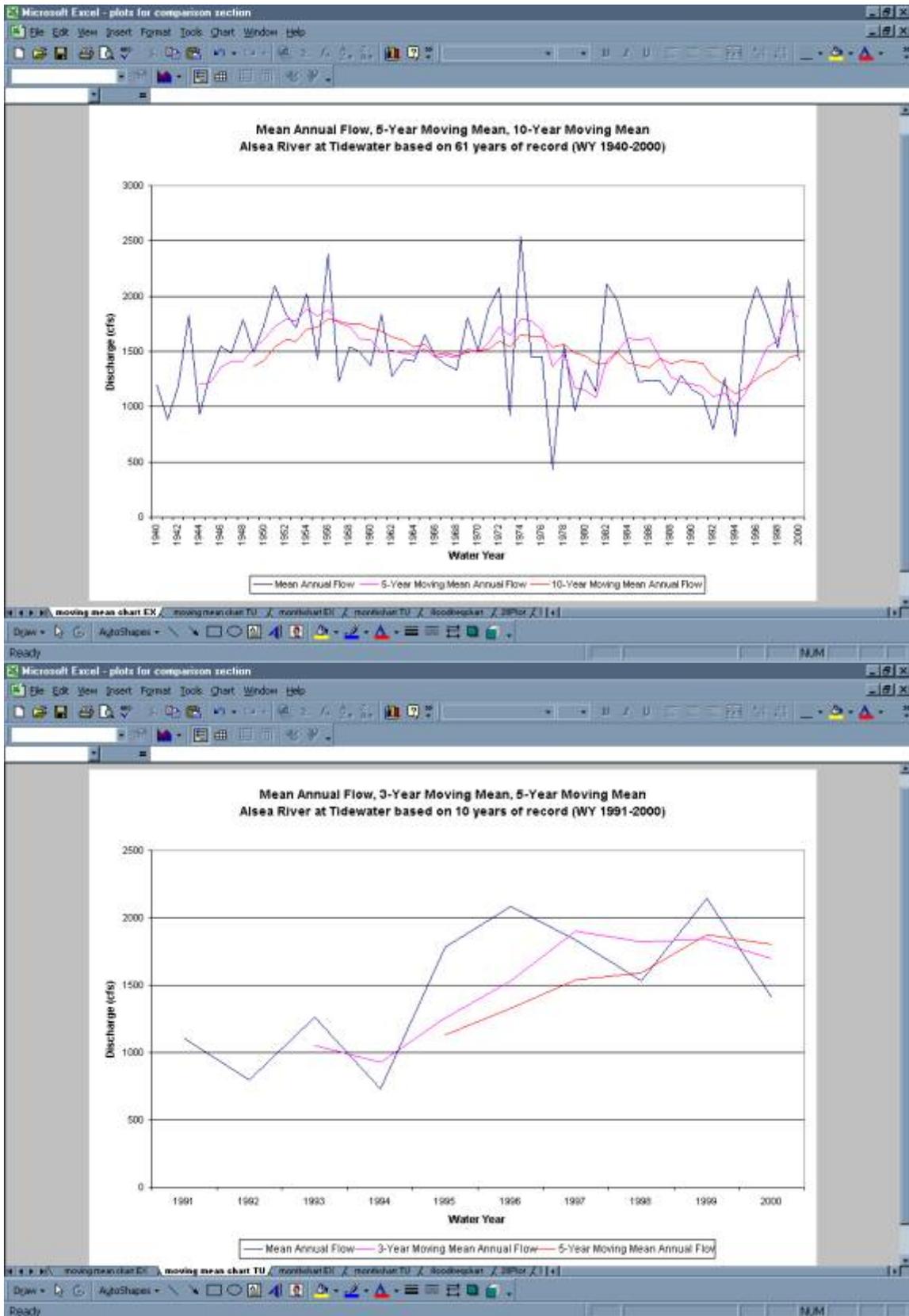
The size of a **watershed** affects the hydrologic characteristics of that drainage area. A small watershed responds to storms very differently from a large basin. Most of the precipitation on a small watershed tends to run off quickly because there is little time or area for **infiltration**. On a large basin, travel distances are much longer, channel storage may diminish flow peaks, precipitation has more time to infiltrate, and runoff reaches the basin outlet more slowly after peak rainfall than in a small watershed, such that the discharge per unit area is usually smaller. It is advisable to keep in mind that streamflow values from a large basin may not accurately represent the hydrologic characteristics of a smaller watershed for the reasons given.

The **Example Applications** presented in this web site are all located on ungaged coastal streams. The techniques described above were used to obtain streamflow data used in the analysis.

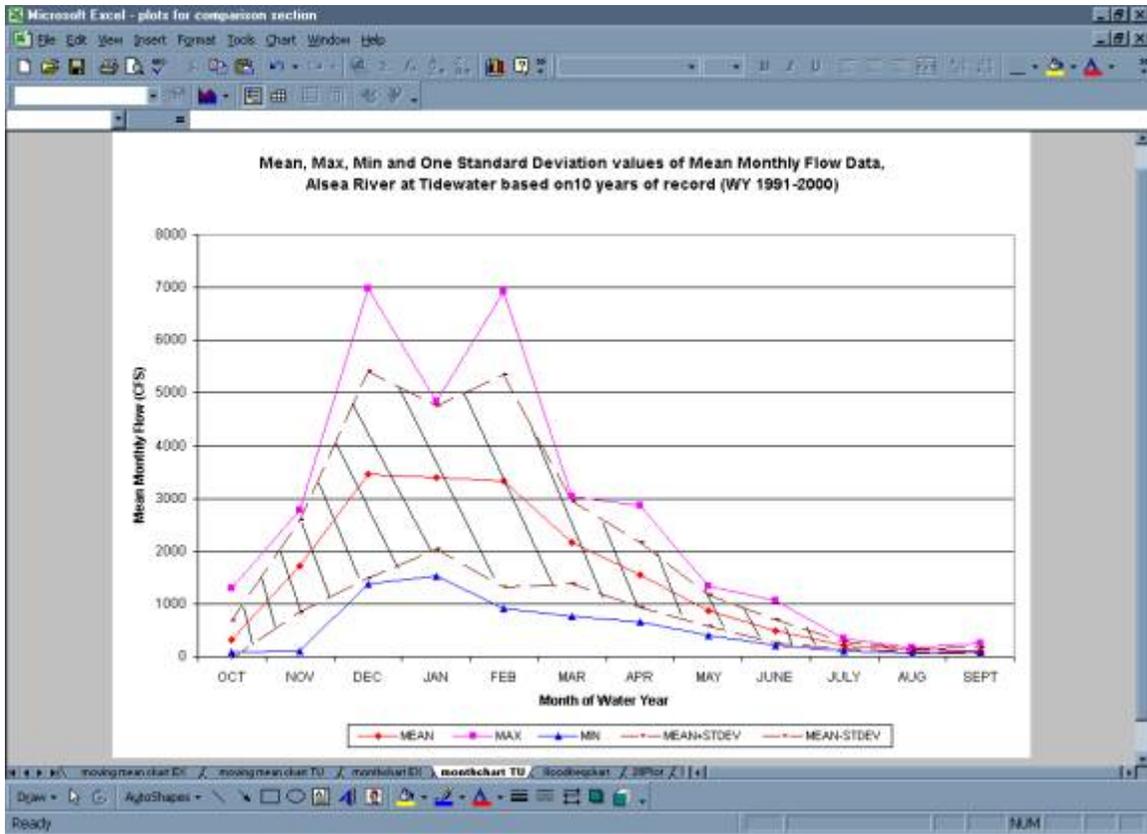
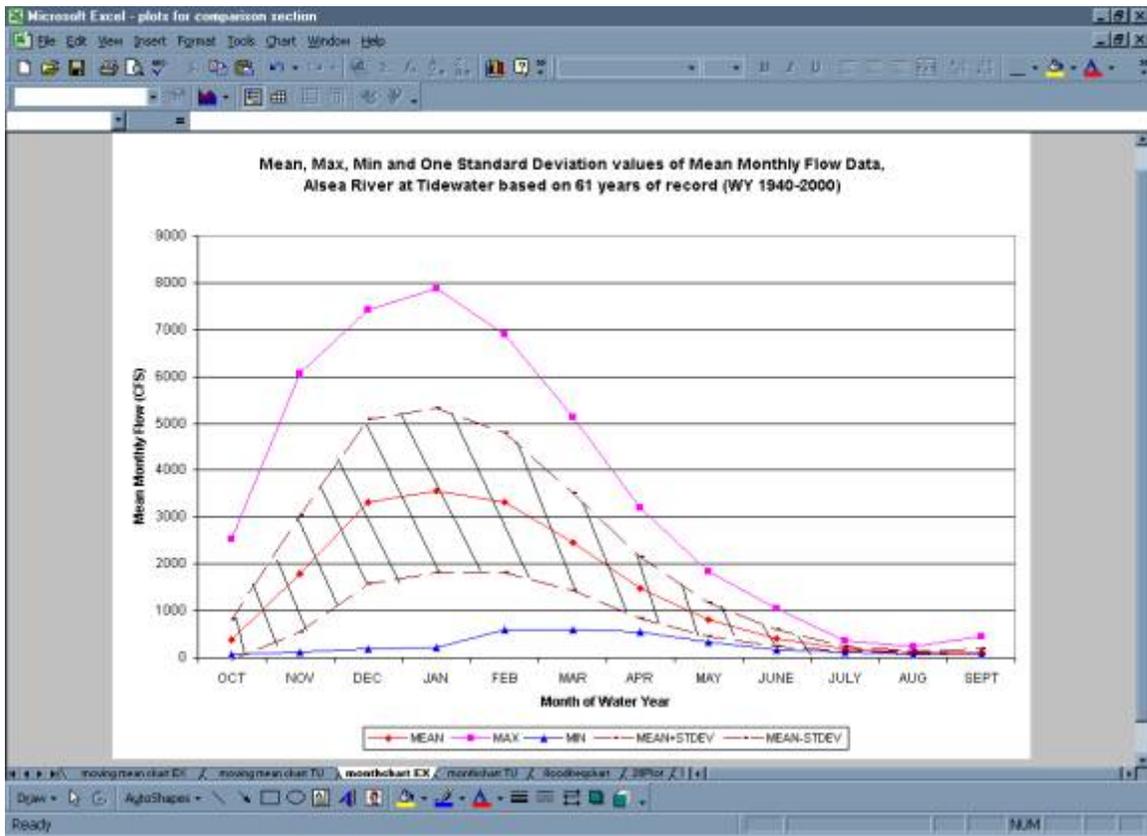
Importance of Length of Period of Record when Selecting a Gaged Site

While drainage area and topographic characteristics are important factors to address when selecting a gaged site to represent an ungaged project stream, the length of the period of record should also be considered. It is the nature of statistical analysis that a larger data set will produce more representative results. The longer the period of record, the more likely it is that the streamflow statistics generated will be representative of the long-term flow pattern of the stream in question. For example, values such as the mean flow for period of record and mean monthly flows estimated using 61 years of data (Aalsea example analysis) will allow more confidence than values estimated using 10 years of data (Aalsea tutorial analysis). When using a shorter period of record, it is more likely that flow values of interest may be overestimated or underestimated, which may lead to errors in determination of design flows. Also, discharge patterns are more easily discerned using a longer data set (see the following four figures). The shorter period of record shows the annual and monthly patterns to be more erratic than does the longer period of record. Thus, it is advisable to search for the longest period of record possible that also meets the other criteria listed in the previous section.

Annual and Monthly Analysis



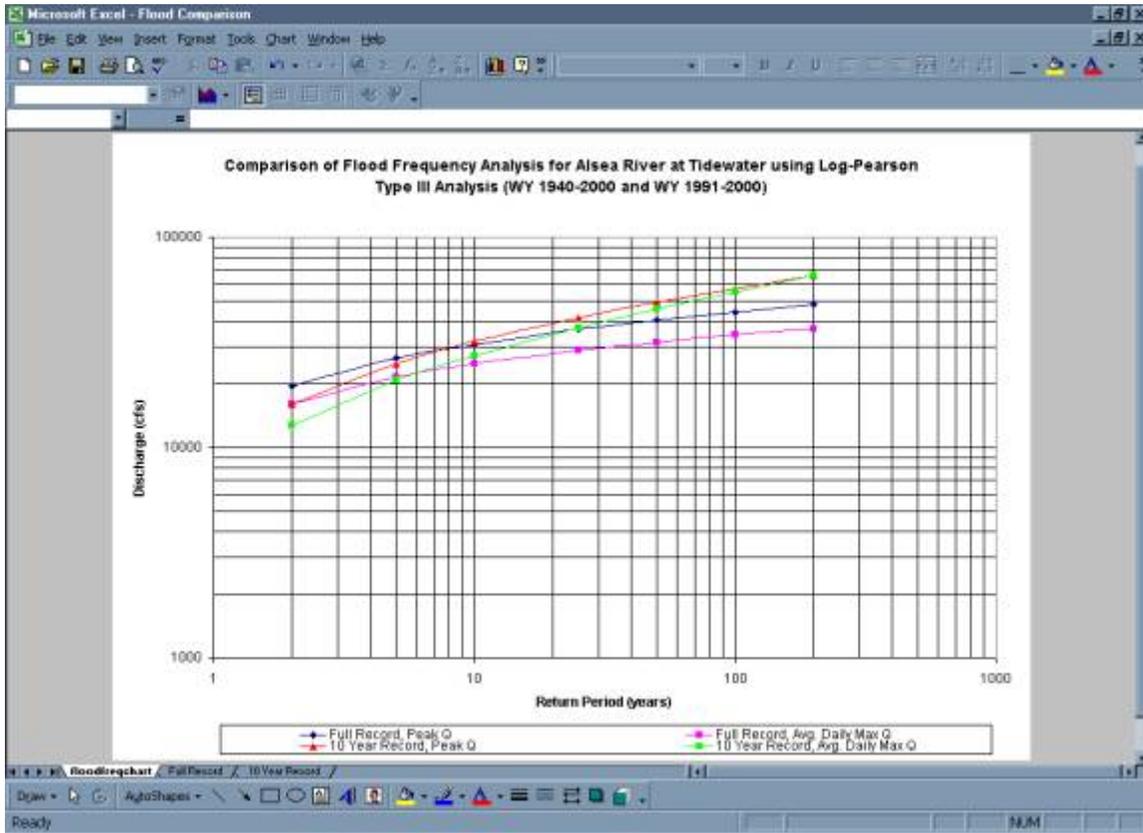
Tips for Data Manipulation from Streamflow Evaluations for Watershed Restoration Planning and Design, <http://water.oregonstate.edu/streamflow/>, Oregon State University, 2002-2005.



Tips for Data Manipulation from Streamflow Evaluations for Watershed Restoration Planning and Design, <http://water.oregonstate.edu/streamflow/>, Oregon State University, 2002-2005.

Flood Frequency Analysis

Using Instantaneous Peak Discharge values versus Average Daily Maximum Discharge values for Flood Frequency Calculations



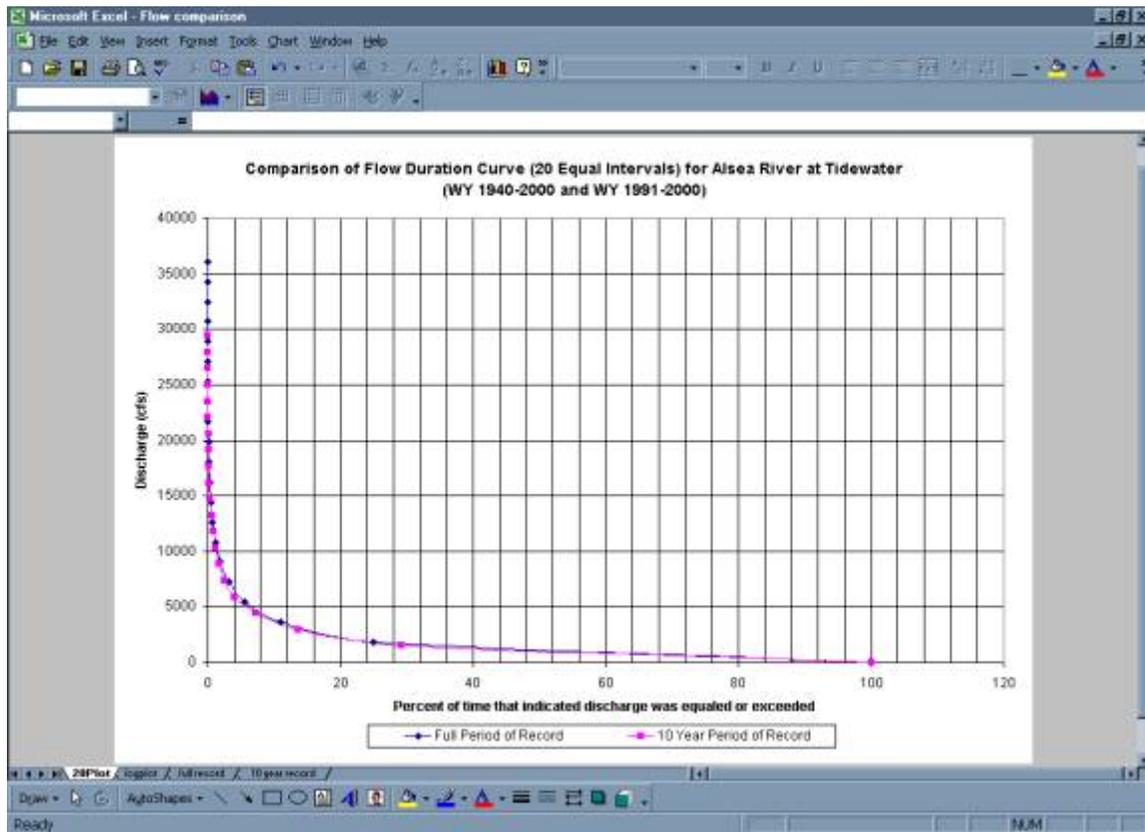
The length of the period of record is also important for flood frequency analysis, as shown by the Comparison of Flood Frequency Analysis plot. The graph shows flood frequency analysis performed using instantaneous peak flow values and average daily maximum values for both the 61-year period of record and the shorter 10-year period of record. In general, the use of average daily maximum discharge values instead of peak flow values in a flood frequency analysis is likely to result in underestimation of the flows associated with various return periods. This is because the daily data used to perform the flood frequency analysis are averaged over a 24-hour period and as such have smoothed out the extreme values. The instantaneous peak data are the largest values recorded at that moment and hence will provide larger values for flood discharges. When performing flood frequency analysis, instantaneous peak flow values should be used to determine design discharges. If peak values are not available, maximum mean daily discharges may be used (with caution).

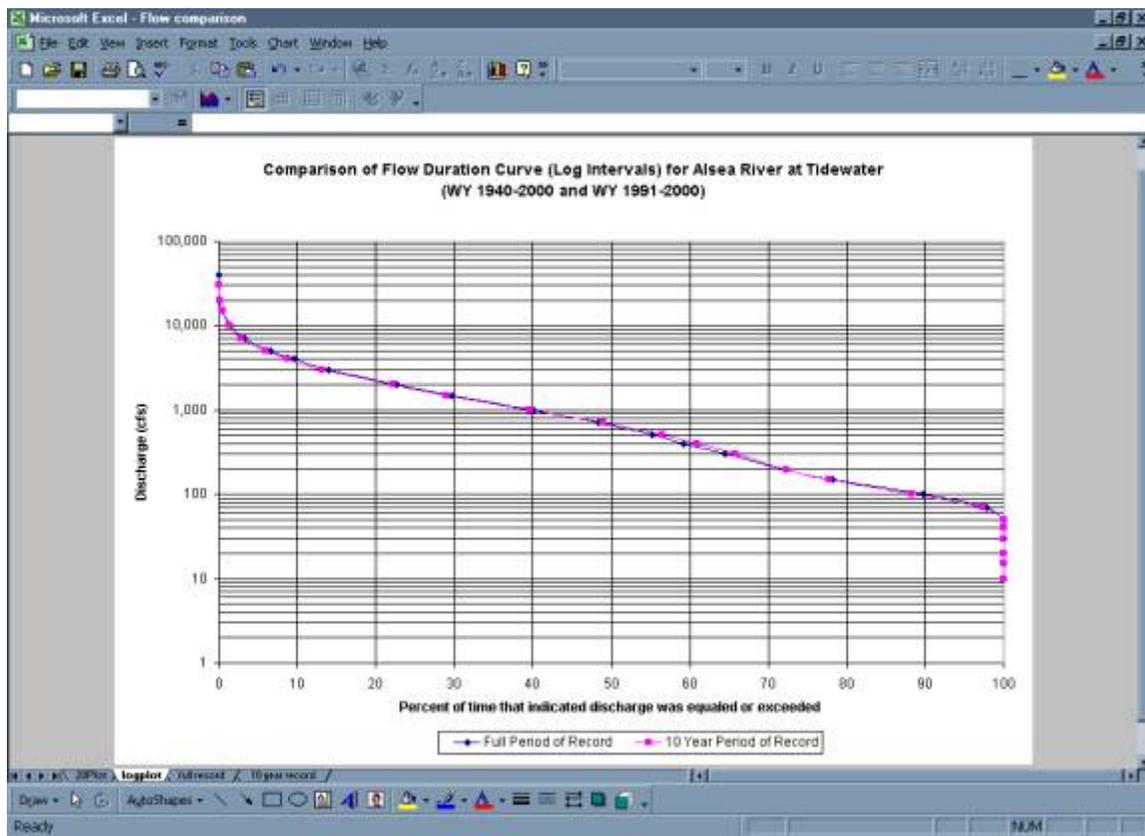
The graph also shows a discrepancy between the flow predictions based on the 61-year period of record and the 10-year period of record. The shorter period of record underestimates the flood discharge values for the smaller return periods and overestimates the flood discharge values for the larger return periods. If a structure were designed for the 100-year flood using the smaller data set, it would be over designed and may cost more than using the 100-year discharge based on the 61-year period of record.

Flow Duration Analysis

By viewing the Comparison of Flow Duration Curves graphs, the general shape of the flow duration curves does not seem to vary significantly between the 61-year period of record and the 10-year period of record. However, on a log scale, even small changes on the graph may indicate significant differences in discharge values.

When generating flow duration curves, one must choose a method for separating the data into intervals or classes. If the data is separated into an equal number of intervals, often the majority of flows will fall into the first few classes. The lower end of the flow duration curve is difficult to read. This problem can be avoided by separating the data into log intervals. The log intervals allow the data to be stretched out making it easier to read the flow duration curve at all points. (Note: The processing capabilities of modern computers makes it easy to generate flow duration curves using all individual daily values, so class limits are no longer as critical a consideration as they were a few years ago.)





Building confidence in flow estimates. How good are the results of the analysis?

It may not always be possible to find a nearby gage that meets all of the appropriate criteria. It may be necessary to use a gage that has a short period of record or is in a slightly different geographic region. In these situations, the user may question if the analysis results are representative of the long-term conditions of the stream in question. One method that can be used to build confidence in the results is to compare the data from the gage used for analysis with a gage in the project region that has a long period of record. The mean annual flows normalized to drainage area can be used for comparison.

Step 1: Identify a gage with a long period of record.

Step 2: Obtain mean annual (water years) streamflow data for full period of record.

Step 3: Normalize by drainage area.

Step 4: Obtain mean annual (water years) streamflow data for gage used in analysis.

Step 5: Normalize by drainage area.

Step 6: Plot Mean Annual Discharge/Unit Area vs. Discharge for both gages on same chart.

Step 7: Calculate the Average of all MAQ/DA values for both gages and plot as two horizontal lines on same chart as Step 6.

Consider the following questions:

- What information can be gained from this graph?
- Do the two gages follow a similar pattern?
- Does the period of record for the gage used for analysis occur at a particularly dry or wet cycle in the long-term record? What portion of the long-term wet/dry behavior does the gage used for analysis seem to represent? If the period of record occurs in a wetter than normal period, resulting flows may be overestimated. If the period of record occurs in a drier than normal period, resulting flows may be underestimated.

The data for the gage used for analysis can be scaled to the long-term data using a ratio of the average MAQ/DA values for both gages. However, it is important to consider all of the criteria listed in the Ungaged Sites section before scaling values.