STATISTICAL SOFTWARE PACKAGE

Jeff Harris david.j.harris@usace.army.mil, **Gary Brunner**gary.w.brunner@usace.army.mil**,**
Matt Fleming matthew.j.fleming@usace.army.mil**, Beth Faber Ph.D.**
beth.a.faber@usace.army.mil, Hydrologic Engineering Center, US Army Corps of Engineers,
Davis CA

Abstract Over a period of many years, the Hydrologic Engineering Center has supported a variety of statistical software that performs frequency analysis and other statistical computations. Four applications had been developed: HEC-FFA, HEC-STATS, REGFREQ and MLRP. This software was developed in the 80's and early 90's was DOS based and the time had come for its retirement. In 2007, HEC began an effort to combine and update this software into the Statistical Software package (HEC-SSP). Initially, version 1.0 performed the Bulletin 17B analysis. This incorporated the capabilities of HEC-FFA. Subsequently, volume-frequency and general frequency analyses capabilities in HEC-STATS were added. Currently, duration analysis and coincident frequency analysis are being designed and added to SSP. This presentation will detail the tools within the software, both existing and planned, and how they can help the hydrologist and engineer perform consistent statistical analyses critical to projects throughout the world.

INTRODUCTION

The current version of HEC-SSP, version 1.1, includes the ability to perform (1) frequency analysis based on Bulletin 17B, "Guidelines for Determining Flood Flow Frequency" (1982), (2) general frequency analysis that allows the user to adjust parameters and (3) volume-duration frequency analysis. The HEC-SSP software system was developed as a part of the Hydrologic Engineering Center's "Next Generation" (NexGen) of hydrologic engineering software. The NexGen project encompasses several aspects of hydrologic engineering, including: rainfall-runoff analysis; river hydraulics; reservoir system simulation; flood damage analysis; and real-time river forecasting for reservoir operations.

HEC-SSP is an integrated system of software, designed for interactive use in a multi-tasking environment. The system is comprised of a graphical user interface (GUI), separate statistical analysis components, data storage and management capabilities, mapping, graphics and reporting facilities.

The goal of HEC-SSP development is to ultimately combine all of the statistical analyses capabilities of HEC-FFA, STATS, REGFRQ and MLRP. New features and additional capabilities will be added in future releases. This paper will provide some detail on the tools and methods within the software, both existing and planned.

USER INTERACTION

HEC-SSP is designed to perform statistical analyses of hydrologic data. The user interacts with HEC-SSP through a graphical user interface (GUI). The main focus in the design of the interface was to make it intuitive, while still maintaining a high level of efficiency for the user. The interface provides for the following functions:

- File management

- Data entry, importing, and editing
- Statistical analyses
- Tabulation and graphical displays of results
- Reporting facilities
- On-line help

Statistical The underlying basis for all current and future analyses of unregulated, homogeneous annual peak flows is based on guidance recommended by Bulletin 17B "Guidelines for Determining Flood Flow Frequency (1982)," by the Interagency Advisory Committee on Water Data. HEC-SSP provides options to address all recommendation within 17B. HEC-SSP follows the 17B recommendation of using the log-Pearson Type III distribution and the method of moments to determine the statistical parameters of the station data. Methods are included in HEC-SSP to address: Broken Record; Incomplete Record; Zero Flood Years; Low and High Outliers; Historical Events. This guidance has been also applied to the volume-duration frequency analysis. The general frequency analysis, while allowing the user to follow the 17B Guidelines, also allows the user to adjust statistical parameters during an analysis. Bulletin 17B is published as recommended guidelines. However, if the analyst can show reason why the guidelines shouldn't or can't be followed, other procedures can be used. The general frequency analysis fills this need.

Data Data storage is accomplished through the use of "text" files (ASCII and XML), as well as the HEC Data Storage System (HEC-DSS). User input data are stored in flat files under separate categories of study, analyses, and a data storage list. Flows are stored in a project HEC-DSS file as time series data. Output data is predominantly stored in HEC-DSS, while a summary of the results is written to an XML file. Additionally, an analysis report file is generated whenever a computation is made. This report file is written to a standard ASCII text file. Data management is accomplished through the user interface. The modeler is requested to enter a Name and Description for each study being developed. Once the study name is entered, a directory with that name is created, as well as a study file. Additionally a set of subdirectories is created with the following names: Bulletin17bResults; Layouts; and Maps. As the user creates new analyses, an analysis file is created in the main project directory. The interface provides for renaming and deletion of files on a study-by-study basis.

Graphics Graphics include a map window and frequency curve plots. The map window can be used to display background map layers. Locations of the data being analyzed can be displayed on top of the map layers. The frequency curve plot shows the results of the frequency analysis, including: the analytically computed curve(s); the expected probability curve; confidence limits; and the raw data points plotted based on one of three available plotting position methods. Tabular output consists of tables showing the computed frequency curves, confidence limits, and summary statistics. All graphical and tabular output can be displayed on the screen, sent directly to a printer (or plotter), or passed through the Windows Clipboard to other software, such as a word-processor or spreadsheet.
A report file is available for each analysis. This report file includes the input data; preliminary results; all of the statistical tests (Low and High Outliers; Broken Record; Zero Flows Years;

Incomplete Record; Regional Skews; and Historic Information); and final results. For the 17B analysis this report file is similar to the HEC-FFA output file.

STEPS IN PERFORMING A BULLETIN 17B FREQUENCY ANALYSIS

The main steps in performing a flow frequency analysis with HEC-SSP include:

- Starting a new study
- Importing, Entering, and Editing Data
- Performing the Bulletin 17B Frequency Analysis
- Viewing and Printing Results

Starting a New Study The first step in performing a Bulletin 17B Flow Frequency analysis with HEC-SSP is to start a new study. The user is required to enter a name for the study; select a directory to work in (a default location is provided); and select the desired units system to work in. Adding a description of the study is optional. Once this information is provided a subdirectory will be created under the user chosen directory. The subdirectory will be labeled the same name as the user-entered study name. This study directory is where the project file, as well as other study files and directories will be located.

Importing, Entering, and Editing Data Before any analyses can be performed, the user must bring data into the HEC-SSP study. For a peak flow frequency analysis following Bulletin 17B, the data consists of peak annual flow data at gaged locations. Data is entered using the Data Editor shown in Figure 1.

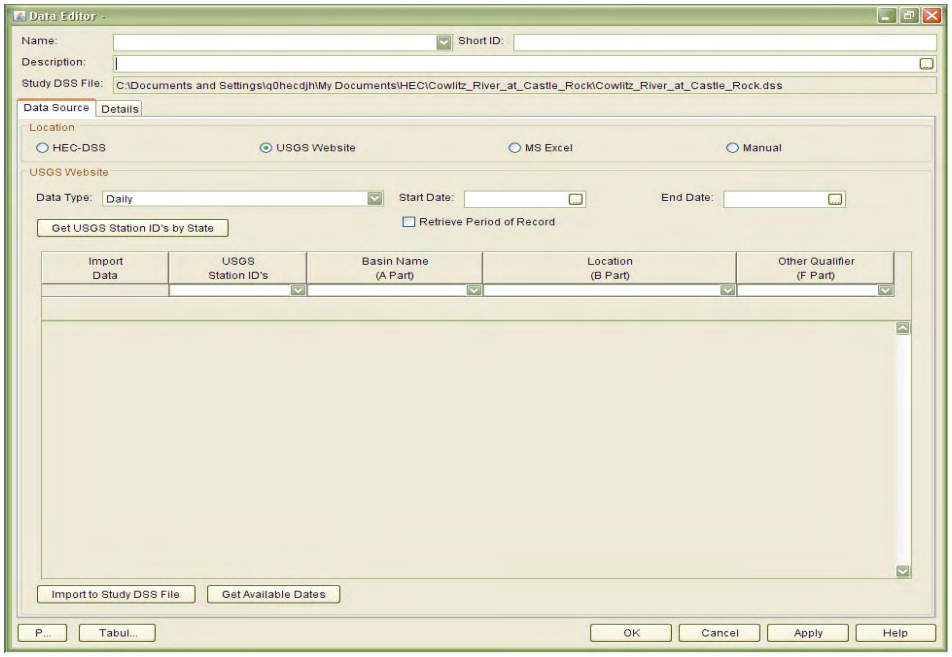


Figure 1 - Data Editor

As shown in Figure 1, the Data Editor has fields for the Name, Short Identifier, and the Description of the data at the top of the window. Additionally it lists the default DSS file name (Hydrologic Engineering Center-Data Storage System, HEC-DSS) that the data will be stored in once it is brought into the study. The default DSS file is always labeled the same name as the study with the .DSS file extension. The Data Editor contains two tabs, **Data Source** and **Details**. The **Data Source** tab is used for selecting and defining a source for bringing data into the HEC-SSP study. Currently, there are four ways to bring data into an HEC-SSP study: import from another HEC-DSS file; Import data from the USGS web site; import from a Microsoft Excel spreadsheet; and manually entering the data into a table. All of these methods will import data into the study DSS file.

When data is downloaded directly from the USGS website, in addition to the data itself, any metadata that is available will be downloaded and stored with the data. The metadata can be viewed from the **Details** Tab on the Data Editor. If the metadata includes coordinate location information, it will then be plotted on top of the background maps. HEC-SSP converts the coordinates of the point data to the default coordinate system of the base map. The user can interact with the plotted points by right clicking on the gage icon and a pop-up menu will appear. The user has the option to edit the data, plot the data, tabulate it, rename the label, or delete the data. If any of the other three methods are used to bring data into HEC-SSP, it is the responsibility of the user to populate the metadata.

It should be noted that when downloading data directly from the USGS website, no screening of the data is performed by HEC-SSP prior to analysis. It is the users responsibility to confirm that all data is appropriate for use. This can be done by scanning the USGS codes which are also downloaded with the data and are available by accessing the pathname in HEC-DSSVue. An example could be if a peak flow resulted from a failure of an upstream structure, it most likely shouldn't be used in a Bulletin 17B analysis since it would not be part of a homogeneous population of flows.

Performing the Bulletin 17B Flow-Frequency Analysis The Bulletin 17B Analysis Editor is shown in Figure 2. As shown in Figure 2, the user must enter a name for the analysis; a description (optional); select a flow data set (gage data stored in project DSS file); enter or select a name for the output DSS file; and enter or select a name for the report file. The analysis window contains three tabs: General; Options; and Results. The General tab contains settings for: Generalized Skew; Expected Probability Curve; Plotting Positions; and Confidence limits. Default settings are already established for each of the options on the General tab. However, the user can change any of these settings. The Options tab contains information on: Low Outlier Threshold; Historic Period Data; and User-Specified Frequency Ordinates. These options are not required for most analyses, but may be necessary depending upon the data. Once all of the settings and options have been selected, the user presses the Compute button to have the computations performed. When the computations have finished a message window indicating "Computations Complete" appears. Once the computations have finished the user can view the output.

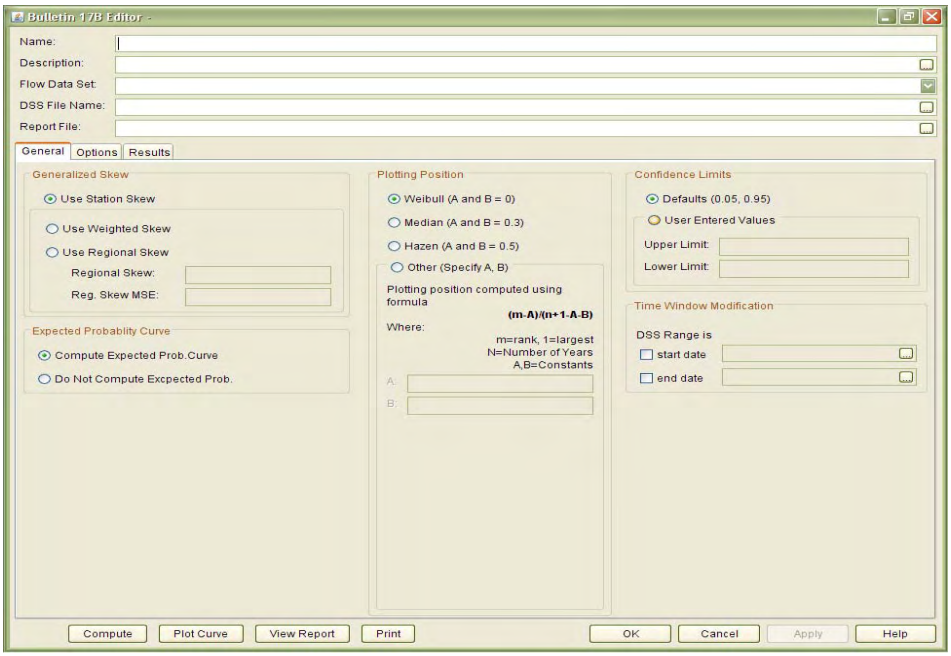


Figure 2 - Bulletin 17B Analysis Editor

Viewing and Printing Results Tabular output, Figure 3, is available on the Results Tab on the Bulletin 17B Editor. On the Results Tab the primary table consists of: Percent Chance Exceedance; Computed flow frequency curve; the Expected Probability adjusted curve; and the 5 and 95 percent Confidence Limits. The second table contains the general statistics about the data including: mean; standard deviation; station skew; regional skew; weighted skew; and the adopted skew of the analysis. The third table contains the number of: historic events; high outliers; low outliers; zero or missing values; systematic events in the data set; and the number of years in the historic period. The table can be printed by pressing the Print button at the bottom of the Analysis window. Graphical output, as shown in Figure 4, can be obtained by pressing the Plot Curve button at the bottom of the Analysis editor.

The final piece of output available is a text report file. The report file lists all of the input data and user settings; plotting positions of the data points; intermediate results; each of the various statistical tests performed (i.e. high and low outliers, historical data, etc…); and the final results. This file is often useful for understanding how the software arrived at the final frequency curve. To view the Report file press the **View Report** button at the bottom of the Analysis window.

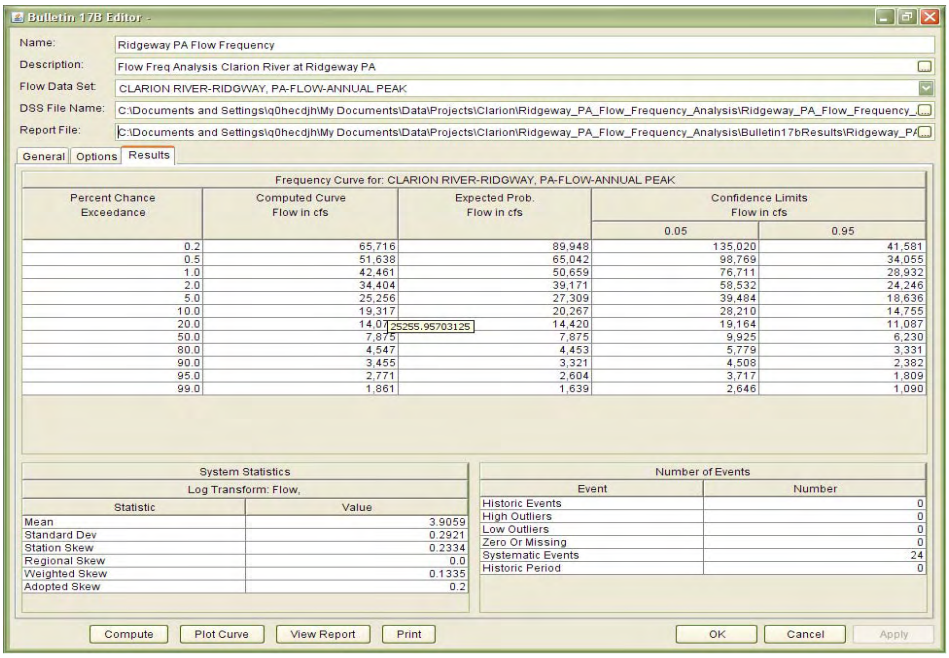


Figure 3 - Tabular Output

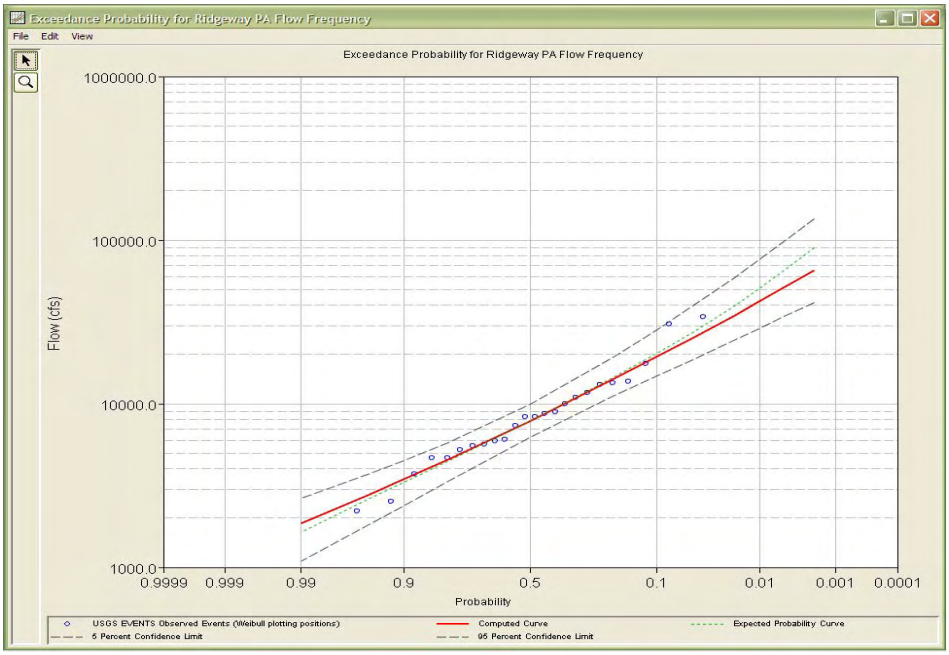


Figure 4 - Graphical Output

PERFORMING GENERAL FREQUENCY ANALYSIS

The method of performing a general frequency analysis is basically the same as the Bulletin 17B analysis. The user must make the appropriate data available to HEC-SSP. This can be through any of the four methods described earlier. In a General Frequency analysis, the user can perform either an analytical or graphical analysis. When a new analytical analysis is started, all options in

the analysis editor default to the Bulletin 17B recommendations. When performing a General Frequency Analysis, the user has the option to: choose if log transform is used and if so set the distribution to Log Normal or Log Pearson III; set confidence limits; select time window for the analysis; select plotting position method; use historic data and set outlier thresholds; turn off expected probability computation; set skew to station, weighted or regional.

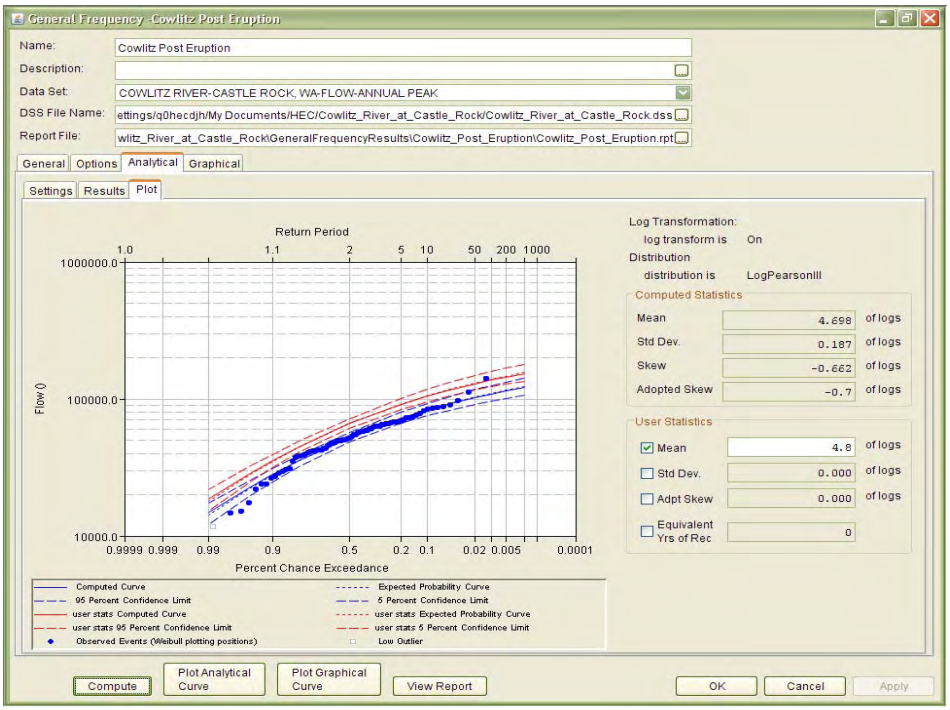


Figure 5 - Analytical Plot Tool

Figure 5 shows a helpful tool available when using the analytical analysis within the General Frequency Analysis. Once results of the analytic analysis are available the user has the opportunity to adjust the statistics of the computed curve and immediately see the results. This is accomplished by checking the appropriate box in the User Statistics area, entering the desired value then pressing the Compute button. Figure 5 shows, in blue, the actual data points and computed curves based on the actual data. The red curves depict the results if the mean is changed from 4.698 to 4.8. This curve is for illustration purposes only. Additionally, if the user has changed input options but not yet computed results based on these changes, a warning appears which informs the user that current results do not reflect input changes.

For a graphical analysis, once the appropriate data set has been selected, the desired options set and a Compute is completed, the user selects the Graphical Tab to see results and build the graphical curve. Here the user can enter values at desired exceedance intervals and build a curve that fits the data points.

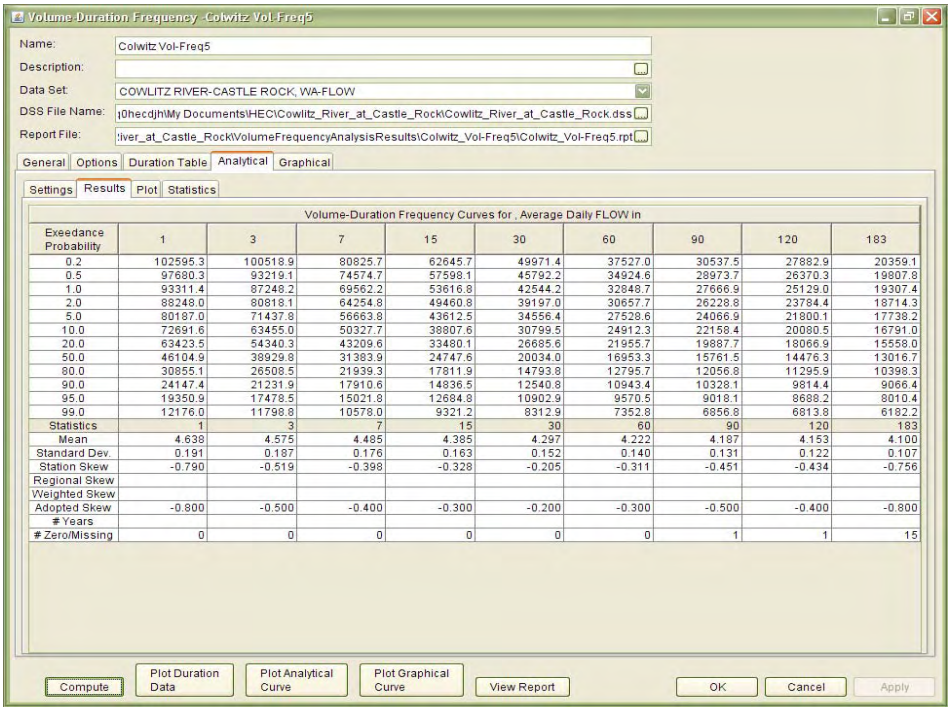
PERFORMING A VOLUME-DURATION FREQUENCY ANALYSIS

The Volume-Duration Frequency analysis is important to many aspects of the engineering and environmental communities. For an engineering analysis of a reservoir, it may be necessary to

generate synthetic inflow hydrographs based on a volume-duration frequency analysis of the gaged daily flows. Within the environmental community, many times it is necessary to compute 7Q10 value. The Volume-Duration Frequency module in HEC-SSP can satisfy both of these needs. As with the other types of analyses in HEC-SSP, the user must make the appropriate data available. This can be through any of the four methods described earlier. In a volume-duration analysis, the analysis is computed using mean daily flows. The user can perform either an analytical or graphical analysis. When performing a Volume-Duration Analysis, the user has the option to: choose if log transform is used and if so set the distribution to Log Normal or Log Pearson III; set the analysis to use maximum or minimum values; select water or calendar year; set time window and season of data to be analyzed; select plotting position method; select desired flow durations; select specific exceedance intervals for computation of values to be shown in results; turn off expected probability computation; set skew to station, weighted or regional.

Once a data set has been selected, the user has the option of performing a Compute or simply choosing to extract and view a table of the Volume-Duration data for all of the selected durations. This extract is automatically performed during a compute and the table populated. This table reflects the annual value for each of the selected durations. These values are used to generate the statistics of the data of each duration.

Once results are available, the user can view the resulting values at each selected exceedance interval, as shown in Figure 6, a plot of the data and a report of the computations. Additionally, two tools are available to the user on the Statistics Tab, when an analytic analysis has been computed. These tools allow the user to plot the computed statistics vs. duration and, as with General Frequency Analysis, enter statistics and do a compute based on those statistics. This is useful if, for example, the curves show a tendency to cross.



Exeedance Probability	1	3	7	15	30	60	90	120	183
0.2	102595.3	100518.9	80825.7	62645.7	49971.4	37527.0	30537.5	27882.9	20359.1
0.5	97680.3	93219.1	74574.7	57598.1	45792.2	34924.6	28973.7	26370.3	19807.8
1.0	93311.4	87248.2	69562.2	53616.8	42544.2	32848.7	27666.9	25129.0	19307.4
2.0	88248.0	80818.1	64254.8	49460.8	39197.0	30657.7	26228.8	23784.4	18714.3
5.0	80197.0	71437.9	56663.8	43612.5	34556.4	27528.6	24066.9	21800.1	17738.2
10.0	72691.6	63455.0	50327.7	38807.6	30799.5	24912.3	22158.4	20080.5	16791.0
20.0	63423.5	54340.3	43209.6	33480.1	26685.6	21955.7	19887.7	18066.9	15558.0
50.0	46104.9	38929.8	31383.9	24747.6	20034.0	16953.3	15761.5	14476.3	13016.7
80.0	30855.1	26508.5	21939.3	17811.9	14793.8	12795.7	12056.8	11295.9	10398.3
90.0	24147.4	21231.9	17910.6	14836.5	12540.8	10943.4	10328.1	9814.4	9066.4
95.0	19350.9	17478.5	15021.8	12684.8	10902.9	9570.5	9018.1	8688.2	8010.4
99.0	12176.0	11798.8	10578.0	9321.2	8312.9	7352.8	6856.8	6813.8	6182.2
Statistics	1	3	7	15	30	60	90	120	183
Mean	4.638	4.575	4.485	4.385	4.297	4.222	4.187	4.153	4.100
Standard Dev.	0.191	0.187	0.176	0.163	0.152	0.140	0.131	0.122	0.107
Station Skew	-0.790	-0.519	-0.398	-0.328	-0.205	-0.311	-0.451	-0.434	-0.756
Regional Skew									
Weighted Skew									
Adopted Skew	-0.800	-0.500	-0.400	-0.300	-0.200	-0.300	-0.500	-0.400	-0.800
# Years									
# Zero/Missing	0	0	0	0	0	0	1	1	15

Figure 6 - Volume-Duration Frequency Analysis Statistics for each Selected Duration

For a graphical analysis, once the appropriate data set has been selected, the desired options set and a Compute is completed, the user selects the Graphical Tab to see results and build the graphical curve. Here the user can enter values at desired exceedance intervals and build a curve that fits the data points.

PLANNED CAPABILITIES

Currently, additions to HEC-SSP are on going. New capabilities have been added and are currently in the testing phase. Three new capabilities methods are being tested. These include, Duration analysis, coincident frequency analysis and means to extrapolate frequency curves.

SOFTWARE AVAILABILITY

HEC-SSP is availability for anyone, free of charge, on the HEC website. Go to http://www.hec.usace.army.mil/ and select HEC-SSP 1.1 from the list of software. This will direct you to the download page where you can get the software and Users Manual.

REFERENCES

Interagency Advisory Committee on Water Data, Bulletin 17B, "*Guidelines for Determining Flood Flow Frequency*" (1982). http://water.usgs.gov/osw/bulletin17b/bulletin_17B.html

U.S. Army Corps of Engineers, EM 1110-2-1415, "*Hydrologic Frequency Analysis*" (1992)

U.S. Army Corps of Engineers, Hydrologic Engineering Center, "*HEC-SSP Users Manual*" (2007)

U.S. Army Corps of Engineers, Hydrologic Engineering Center, "*HEC-DSSVue Users Manual*" (2005)