Computer Programs for Obtaining and Analyzing Daily Mean Streamflow Data from the U.S. Geological Survey National Water Information System Web Site

Appendix 2. Make Plotting Position (MkPP Version 1.0)—A program for generating plotting positions, percentiles, and frequency factors from daily mean streamflow data

By Gregory E. Granato

Prepared in cooperation with the
U.S. Department of Transportation
Federal Highway Administration
Office of Natural and Human Environment

Open-File Report 2008–1362

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2009

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:
Granato, G.E., 2009, Computer programs for obtaining and analyzing daily mean streamflow data from the U.S. Geological Survey National Water Information System Web Site: U.S. Geological Survey Open-File Report 2008–1362, 123 p. on CD-ROM, appendix 2 of 5.

Contents

Abstract....31
Introduction....31
Graphical Data Analysis....32
Purpose and Scope....32
Governing Equations and Statistical Methods....32
Governing Equations....32
Statistical Methods....37
Use of the MkPP Program....37
Installation and Removal....37
Input-File Formats....38
Station-List Input-File Format....38
Streamflow Data Input-File Format....38
Output-File Format....39
Graphical User Interface....39
Summary....44
Acknowledgments....44
References Cited....44

Figures

2-1. Examples of statistical graphs derived from MkPP output that can be used for visual inspection and graphical data analysis of streamflow values using (*A*) a probability plot derived from daily mean streamflow data or (*B*) a flow-duration curve derived from daily mean streamflow data....33
2-2. Examples of differences among flow-duration curves that indicate potential effects of the surficial geology of a drainage basin on variations in streamflow....34
2-3 – 2-7. Screen captures showing—
2-3. Initial appearance of the MkPP program input form when the program opens....40
2-4. Example of the MkPP program file-specification dialog box....41
2-5. Example of the MkPP program input form when ready for data processing....42
2-6. Example of the MkPP program plotting-position equation-input form....42
2-7. Example of the MkPP program input form after completion of the data-processing activities for streamflow-gaging-station data....43

Tables

2-1. Summary of input and output files for the Make Plotting Position (MkPP Version 1.0) program....38

Conversion Factors and Abbreviations

Multiply	By	To obtain
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
cubic foot per second per square mile [$(ft^3/s)/mi^2$]	0.01093	cubic meter per second per square kilometer [$(m^3/s)/km^2$]

Abbreviations

1B3	1-day 3-year biological low flow
4B3	4-day 3-year biological low flow
7Q10	7-day 10-year low flow
ADAPS	USGS Automated Data Processing System
BCF	bias correction factor
BMP(s)	best management practice(s)
CD-ROM	computer disk-read only memory
dv	daily value
dd	data descriptor
EMC	event mean concentration
FHWA	Federal Highway Administration
GNWISQ	Get National Water Information System Streamflow (Q)
KTRLine	Kendall-Theil robust line
log10	common logarithm
MkDF	Make U.S. Environmental Protection Agency DFLOW3 batch input Files
MkPP	Make plotting position file
MOVE	maintenance of variance
NWIS	National Water Information System
NWISWeb	National Water Information System Web
PC	Personal Computer
Q	Streamflow
QSTATS	Streamflow (Q) Statistics
RDB	relational database
ROS	regression-on-order statistics
SELDM	stochastic empirical loading and dilution model
SQL	structured query language
SREF	Streamflow Record Extension Facilitator
TMDL(s)	Total Maximum Daily Load(s)
USGS	U.S. Geological Survey

Appendix 2. Make Plotting Position (MkPP Version 1.0)—A program for generating plotting positions, percentiles, and frequency factors from daily mean streamflow data

By Gregory E. Granato

Abstract

The Make Plotting Position (MkPP Version 1.0) program was developed by the U.S. Geological Survey (USGS) in cooperation with the Federal Highway Administration to facilitate visual analysis of daily mean streamflow data. Hydrologic analysis of streamflow data may be improved by visual inspection of a probability plot or flow-duration curve of the data values. The program was designed to sort the data and provide a condensed file of the plotting positions, percentiles, normal frequency factors, log-Pearson Type III frequency factors and associated data. The MkPP program uses daily value data files that have been obtained from the USGS National Water Information System Web site and reformatted to a common format by use of the program GNWISQ (appendix 1). The output file from the MkPP program can be used to facilitate rapid production of probability plots and flow-duration curves with simple spreadsheet or graphing software commonly available on most personal computers. This computer program has a graphical user interface that follows standard Microsoft Windows interface conventions.

Introduction

Streamflow information is important for many planning and design activities including water-supply analysis, habitat protection, bridge and culvert design, calibration of surface and ground-water models, and water-quality assessments. Streamflow information is especially critical for water-quality assessments (Warn and Brew, 1980; Di Toro, 1984; Driscoll and others, 1989; Driscoll and others 1990, a,b). Calculation of streamflow statistics for receiving waters is necessary to estimate the potential effects of point sources such as wastewater-treatment plants and nonpoint sources such as highway and urban-runoff discharges on receiving water. Streamflow statistics indicate the amount of flow that may be available for dilution and transport of contaminants (U.S. Environmental Protection Agency, 1986; Driscoll and others 1990, a,b). Streamflow statistics also may be used to indicate receiving-water quality because concentrations of water-quality constituents commonly vary naturally with streamflow. For example, concentrations of suspended sediment and sediment-associated constituents (such as nutrients, trace elements, and many organic compounds) commonly increase with increasing flows, and concentrations of many dissolved constituents commonly decrease with increasing flows in streams and rivers (O'Connor, 1976; Glysson, 1987; Vogel and others, 2003, 2005).

Reliable, efficient and repeatable methods are needed to access and process streamflow information and data. For example, the Nation's highway infrastructure includes an innumerable number of stream crossings and stormwater-outfall points for which estimates of stream-discharge statistics may be needed. The U.S. Geological Survey (USGS) streamflow data-collection program is designed to provide streamflow data at gaged sites and to provide information that can be used to estimate streamflows at almost any point along any stream in the United States (Benson and Carter, 1973; Wahl and others, 1995; National Research Council, 2004). The USGS maintains the National Water Information System (NWIS), a distributed network of computers and file servers used to store and retrieve hydrologic data (Mathey, 1998; U.S. Geological Survey, 2008). NWISWeb is an on-line version of this database that includes water data from more than 24,000 streamflow-gaging stations throughout the United States (U.S. Geological Survey, 2002, 2008). Information from NWISWeb is commonly used to characterize streamflows at gaged sites and to help predict streamflows at ungaged sites.

Long records (years to decades) are commonly needed to characterize variations in streamflow through time. One 40-year record may include about 14,610 daily mean streamflow values and require about 400,000 bytes of computer-disk space. Hundreds to thousands of such data files are necessary to characterize national variations in streamflow. The MkPP program was designed to reduce a large data set of daily mean streamflows into an output file of about 1,000 ranked streamflow values with their associated plotting positions, percentiles, and frequency factors. These MkPP program output files were designed to facilitate visual inspection and graphical analysis of the data.

Graphical Data Analysis

Graphical data analysis is important for assessing and understanding data (Helsel and Hirsch, 2002). Plotting positions, percentiles, and frequency factors are derived from a sorted data set to facilitate graphical data analysis. Plotting positions are the rank of each datum within the sorted data set that is normalized by the total number of measurements in the data set. The plotting positions are normalized to fractions between zero and one. The plotting positions also indicate the fraction of the data set above and below the value of interest. This fraction indicates the frequency of exceedance of each datum. The percentile of each datum is equal to the plotting position times 100. The frequency factors are used to scale the frequency of exceedance of each datum to a particular probability distribution. Graphing data with respect to the frequency factors is equivalent to plotting the data with respect to the percentiles without the need for software that can produce a graph with a probability axis (fig. 2-1).

Streamflow estimates are commonly made by selecting a statistical model that represents the relation between streamflow magnitude and the probability of occurrence. Plotting positions, percentiles, and frequency factors may be used with sample statistics to fit a theoretical distribution to the available data. Historically, different distributions have been used to estimate streamflow values, but the lognormal and log-Pearson type III distributions are most commonly used for hydrologic analysis of streamflow records (Marcovic, 1965; Haan, 1977; U.S. Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Hirsch and others, 1993; Stedinger and others, 1993; U.S. Army Corps of Engineers, 1993, 1994; Helsel and Hirsch, 2002; McCuen and others, 2002). Selection of a distribution may be facilitated by visual inspection of a probability plot of the data (Riggs, 1968; U.S. Interagency Advisory Committee on Water Data, 1982; Vogel and Kroll, 1989; Chow and others, 1988; Vogel and McMartin, 1991; Stedinger and others, 1993; McCuen and others, 2002). For example, visual inspection of figure 2-1 indicates that lognormal or log-Pearson type III distributions may provide a planning-level approximation for daily mean streamflow data from the Pawcatuck River at Wood River Junction, RI (USGS streamflow-gaging station 01117500) because the pattern of data may be approximated by a straight line when plotted on a logarithmic scale. It is clear, however, that daily mean streamflow data from Adamsville Brook at Adamsville, RI (USGS streamflow-gaging station 01106000) would not be well fit by these distributions.

Probability plots also may help the user identify the most representative streamflow-gaging station to represent flow characteristics. For example, Thomas (1966) developed flow-duration curves for 24 unregulated streams with 30 years of data in a gently rolling glaciated terrain in Connecticut (fig. 2-2). This graph indicates the potential effect of basin characteristics on streamflow statistics. In this area, the variability in normalized streamflow decreases with increasing percentage of stratified drift. Similar graphs may facilitate selection of representative hydrologically similar streamflow-gaging stations on the basis of basin characteristics to represent streamflow statistics at a site of interest.

Purpose and Scope

This manual describes the implementation, use, and interpretation of results from the Make Plotting Position MkPP (Version 1.0) program. The MkPP program was developed by the USGS in cooperation with the Federal Highway Administration for use in the analysis of local, regional, and national hydrologic data sets. This program was developed as part of a suite of tools to download and process streamflow information in support of a stochastic empirical loading and dilution model for planning-level estimates of the effects of highway runoff on receiving water. MkPP uses the relational database (RDB) daily value (dv) streamflow files retrieved from NWISWeb and reformatted to the original USGS RDB dv format by use of the Get NWIS streamflow (GNWISQ) program (appendix 1). Methods and governing equations for estimating streamflow statistics in the MkPP program are described. The formats of input data and output statistics are described. Step-by-step use of the program's graphical user interface is illustrated. The program code written in Microsoft Visual Basic 6.0 is documented as individual files in a Visual Basic project directory on the computer disk containing this manual.

Governing Equations and Statistical Methods

The MkPP program has a graphical user interface and calculates all necessary statistics with streamflow data from one or more streamflow-gaging stations. The section on governing equations describes the methods used for calculating plotting positions, percentiles, and frequency factors. The section on statistical methods provides information for interpreting and using the output from the MkPP program.

Governing Equations

Streamflow data commonly are ranked in either ascending or descending order to calculate the plotting positions (as fractions or percentiles of the data) that are either less than or greater than a particular data value, respectively (Riggs, 1968; Chow and others, 1988; Hirsch and others, 1993; Stedinger and others, 1993; U.S. Army Corps of Engineers, 1994; Helsel and Hirsch, 2002). In quantile (or cumulative-distribution) plots, data are commonly ranked from smallest to largest, and the fraction or percentage of data that are less than or equal to a given streamflow is indicated. In flow-duration curves (or probability plots), however, data are commonly ranked from the largest to the smallest flows to indicate the fraction or percentage of data that exceed a given flow. To reverse the

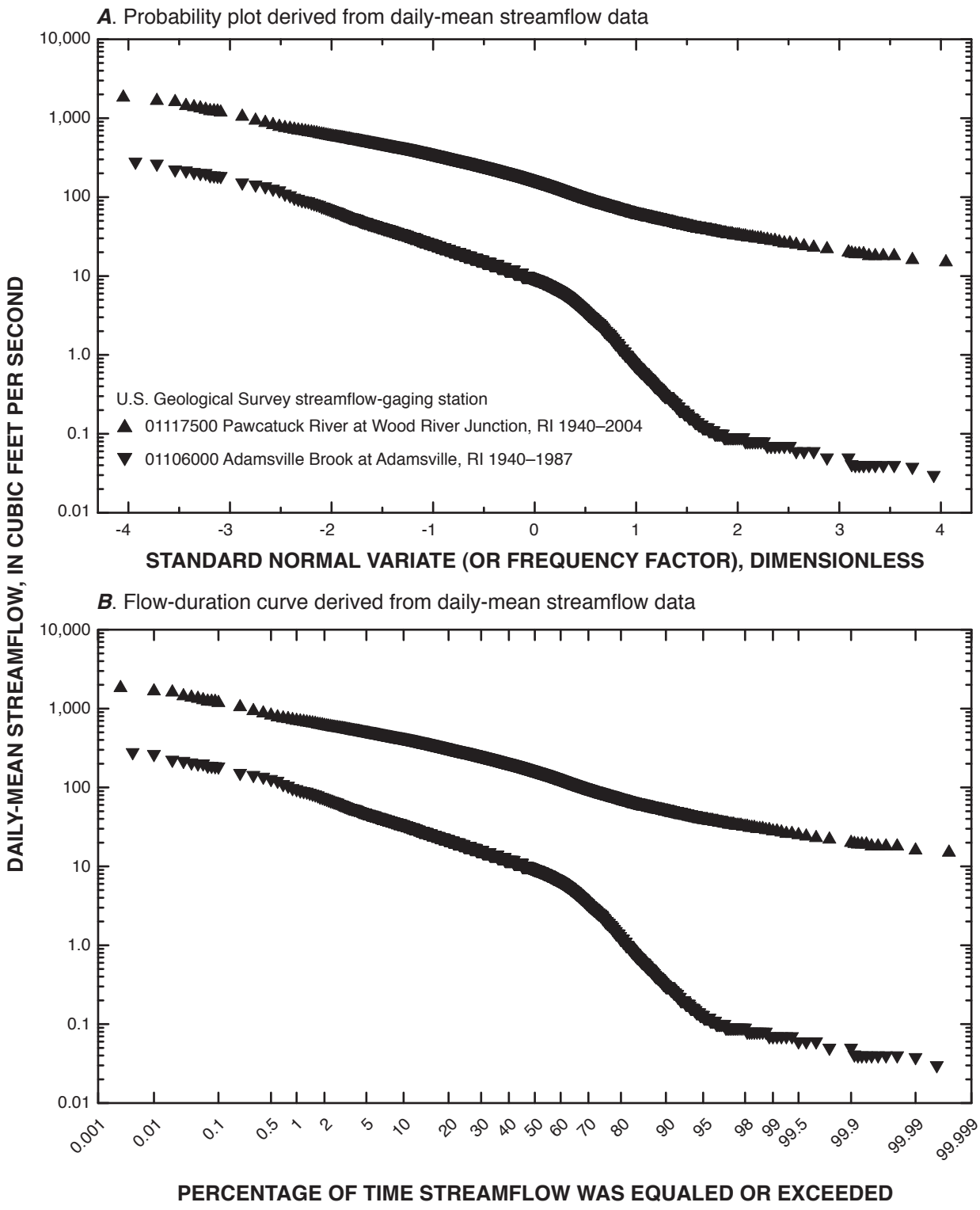

Figure 2-1. Examples of statistical graphs derived from MkPP output that can be used for visual inspection and graphical data analysis of streamflow values using (*A*) a probability plot derived from daily mean streamflow data or (*B*) a flow-duration curve derived from daily mean streamflow data.

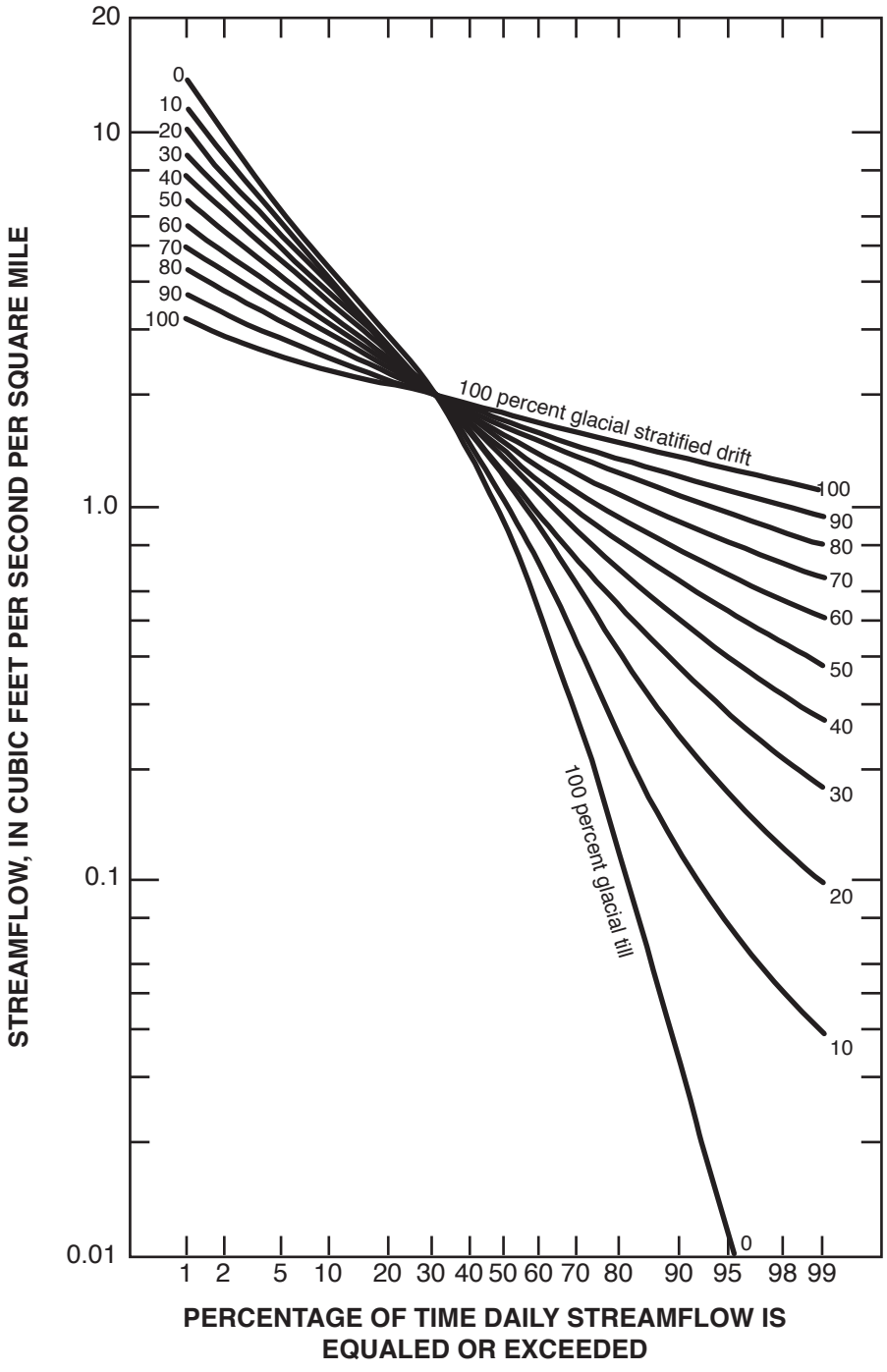

Figure 2-2. Examples of differences among flow-duration curves that indicate potential effects of the surficial geology of a drainage basin on variations in streamflow (from Thomas, 1966).

order of the rank, simply subtract the fractions to be converted from one. To convert the fractions to percentiles, multiply the fractions by 100. The general equation for plotting-position formulas that are used to calculate the proportion for each ranked data point is

$$p = \frac{(i-a)}{(n+1-2a)}, \quad (1)$$

where

p is the calculated plotting position;
i is the rank of the ith data point;
a is the plotting-position factor, which is different for each plotting-position formula; and
n is the number of data points in the population (Helsel and Hirsch, 2002).

The plotting-position factor (a) is 0.0 for the Weibull (1939), 0.375 for the Blom (1958), 0.4 for the Cunnane (1978), 0.44 for the Gringorten (1963), and 0.5 for the Hazen (1914) formula. Equation 1 produces a plotting position greater than zero and less than one for each data point if any of these plotting-position factors are used (Helsel and Hirsch, 2002). Differences in the plotting positions produced by the different formulas are not substantial for most of the data in large data sets, but the differences can be appreciable for the smallest and largest values, especially for small data sets (Stedinger and others, 1993). The MkPP program uses equation 1 to calculate the proportion of daily mean streamflows that are greater than or equal to selected ranked streamflow-data points. The associated proportion of streamflows less than or equal to these selected ranked streamflow data points ($1–p$) also is calculated. The MkPP program also converts these proportions to the respective percentile values by multiplying the proportions by a factor of 100.

To facilitate generation of normal-probability plots, the MkPP program calculates an approximate standard normal frequency factor (Z) for each plotting-position value. The standard normal frequency factor (or variate) is the x-axis value for each probability value in the probability-density function of a standard normal distribution (a normal distribution with a mean of zero and a standard deviation of one). The value of the standard normal frequency factor can be calculated from the exceedance probability by a three-step process (Chow and others 1988). First, the exceedance probability (p) is converted to an intermediate variable (W) in log space as follows:

for ($0< p \leq 0.5$),

$$W = \sqrt{\ln\left(\frac{1}{p^2}\right)}, \quad (2a)$$

and for ($0.5< p <1.0$),

$$W = \sqrt{\ln\left(\frac{1}{(1-p)^2}\right)}. \quad (2b)$$

The standard normal frequency factor (Z) is then approximated as

$$|Z| = W - \frac{2.515517 - 0.802853 \times W + 0.010328 \times W^2}{1.0 + 1.432788 \times W + 0.189269 \times W^2 + 0.001308 \times W^3} \quad (3)$$

For ($0< p \leq 0.5$),

$$Z = |Z|, \quad (4a)$$

and for ($0.5< p <1.0$),

$$Z = -1.0 \times |Z|. \quad (4b)$$

The expected error in Z is less than 0.00045 (Chow and others, 1988). The data will plot as a straight line in relation to the normal frequency factors if the data are normally distributed. If the data are lognormally distributed changing the streamflow axis to a log scale will improve linearity. The lognormal distribution is commonly used as a first-order approximation for streamflow data (DiToro, 1984; Driscoll and others, 1990; Nash, 1994; Van Buren and others, 1997; Limbrunner and others, 2000; Vogel and others, 2005). If the data are lognormally distributed, the skew of the data will be positive, and the skew of the logarithms of data will be approximately equal to zero. Many streamflow data sets, however, may exhibit a nonlinear probability plot and nonzero skew in logarithmic space. It should be noted, however, that only a relatively few measurements that are much larger or much smaller than the rest of a dataset are sufficient to produce a nonzero parametric skew (Helsel and Hirsch, 2002).

Helsel and Hirsch (2002) describe nonparametric methods to estimate skew coefficients, including the percentile skew, to calculate a robust measure of the symmetry of a data set above and below the median value. The MkPP program calculates and outputs a percentile skew by using the equation

$$Skew_p = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})}, \quad (5)$$

where

$Skew_p$ is the percentile skew,
P_{90} is the 90th percentile streamflow value,
P_{50} is the 50th percentile (or median) streamflow value, and
P_{10} is the 10th percentile streamflow value (Helsel and Hirsch, 2002).

The MkPP program calculates and outputs a percentile skew value for the streamflow data and the common (base 10) logarithms of streamflow data to help the user evaluate the lognormal approximation. If the lognormal plot is convex or concave and the skew is substantially different from zero, frequency factors for a three-parameter distribution such as the log-Pearson type III may linearize the probability plot, indicating a better approximation for the data (U.S. Interagency Advisory Committee on Water Data, 1982; Vogel and Kroll, 1989; Helsel and Hirsch, 2002).

The MkPP program uses the percentile skew (eq. 5) and the Wilson-Hilferty approximation (Kirby, 1972; U.S. Interagency Advisory Committee on Water Data, 1982; Vogel and Kroll, 1989; Helsel and Hirsch, 2002) to calculate an estimated frequency factor (ZLPIII) for a log-Pearson type III distribution from the approximate standard normal frequency factor (Z). The Wilson-Hilferty approximation is expressed as

$$Z_{LPIII} = \frac{2}{G} \times \left\{ \left[1 - \left(\frac{G}{6} \right)^2 + \left(\frac{G}{6} \right) \times Z \right]^3 - 1 \right\} \quad , \qquad (6)$$

where

Z_{LPIII} is the frequency factor for a log-Pearson type III distribution,
Z is the standard normal frequency factor, and
G is the estimated skew coefficient of the streamflows in logarithmic space; this coefficient is approximated by the MkPP program with the percentile skew ($Skew_P$).

The Wilson-Hilferty approximation is adequate for use if the skew is less than about plus or minus 3; otherwise, the modified Wilson-Hilferty approximation described by Kirby (1972) provides a better approximation. The MkPP program, however, uses a truncated skew value of plus or minus 2 to calculate the log-Pearson type III frequency factor. The percentile skew may be less than the parametric skew; therefore, skew values outside the range of plus or minus 2 are truncated so that the estimated log-Pearson type III frequency factors will be well within the accepted range of the unmodified Wilson-Hilferty approximation. If skew values are beyond the range of plus or minus 2, the user may substitute values from published tables of log-Pearson type III frequency factors for the values calculated by MkPP (Harter, 1969; 1971; U.S. Interagency Advisory Committee on Water Data, 1982; U.S. Natural Resources Conservation Service, 1998). Use of Z_{LPIII} as the independent variable on a probability plot of the streamflow values should increase linearity of the plotted streamflow values if the log-Pearson type III distribution is a substantially better statistical model for the data than the lognormal approximation.

To examine the fit of the data to the lognormal or log-Pearson type III distributions, the user may construct a graph of the data with the frequency factors on the X-axis and the streamflow values on the Y-axis (plotted with a logarithmic scale). Also, the user may calculate the expected streamflow value of each plotting position by using the population statistics calculated using the QSTATS program (appendix 4) with the associated frequency factor calculated by the MkPP program and then comparing the actual and expected values. The expected streamflow (Y) for each frequency factor in the output file would be calculated as follows:

$$Y = 10^{(\overline{X}+S\square Z)} , \qquad (7)$$

where

Y is the expected streamflow value,
$\overline{X}$ is the average of the common logarithms of the streamflow data,
S is the standard deviation of the common logarithms of the streamflow data, and
Z is the standard normal frequency factor or the Pearson Type III frequency factor.

The user also may examine a three-parameter lognormal distribution to improve the fit of data to a lognormal distribution (Haan, 1977; Stedinger and others, 1993). The third parameter in the three-parameter lognormal distribution is a displacement factor, which is subtracted from each streamflow value before the value is transformed into log space for calculation of the average and standard deviation of the modified lognormal data set. The transformation of each streamflow value (Y) is calculated by the equation

$$X = Log_{10}(Y - df) , \qquad (8)$$

where

X is the common logarithm of the adjusted streamflow value,
df is the displacement factor, and
Y is the measured streamflow value.

The MkPP program calculates the quantile-lower-bound estimator of the displacement factor (Stedinger and others, 1993) and prints this value in the heading lines of the output file. The quantile-lower-bound estimator is calculated as

$$df = \frac{(Y_{\min} \times Y_{\max} - Y_{median}^2)}{(Y_{\min} + Y_{\max} - 2 \times Y_{median})} , \qquad (9)$$

where

Y_{min} is the minimum of measured streamflow values,
Y_{max} is the maximum of measured streamflow values,
Y_{median} is the median of measured streamflow values, and
df is the quantile-lower-bound estimator for the displacement factor.

Statistical Methods

Graphical examination of environmental data is commonly considered the first step in the statistical analysis of this data (Haan, 1977; U.S. Interagency Advisory Committee on Water Data, 1982; Vogel, 1986; Vogel and Kroll, 1989; Helsel and Hirsch, 2002). Graphs as visual summaries of data describe essential information more quickly and completely than tables of numbers or summary statistics, and thus provide an understanding of the data that is difficult to obtain in any other way (Helsel and Hirsch, 2002). If the choice of a distribution is critical, however, quantitative statistical tests are available to determine if a data set fits a statistical distribution (Haan, 1977; Vogel, 1986; Helsel and Hirsch, 2002).

The choice of plotting-position factor depends on the expected distribution of the data and the intended use(s) for the probability plots. The Weibull (1939) plotting-position factor (a=0) is commonly used in hydrology and was used by the U.S. Interagency Advisory Committee on Water Data (1982) to examine flood-frequency data that were expected to have a log-Pearson Type III distribution. The Weibull (1939) plotting position formula, however, is only unbiased for a uniform distribution (Vogel and Kroll, 1989). The Blom (1958) plotting-position factor (a=0.375) is commonly considered as unbiased for normally distributed data (Stedinger and others, 1993; Helsel and Hirsch, 2002). The Cunnane (1978) plotting-position factor (a=0.4) is commonly recommended for general use because it is approximately unbiased for many statistical distributions (Vogel and Kroll, 1989; Stedinger and others, 1993; Helsel and Hirsch, 2002). The Gringorten (1963) plotting-position factor (a=0.44) is optimized for a Gumbel distribution (Vogel and Kroll, 1989; Stedinger and others, 1993). The Hazen (1914) plotting-position factor (a=0.5) is commonly used because it was developed first and has traditional use (Stedinger and others, 1993; Helsel and Hirsch, 2002).

The standard normal frequency factors calculated by the MkPP program from the plotting positions may be used to construct a probability plot for the data. Plotting the standard normal frequency factor (Z) on the X-axis and the associated streamflow on a log-scale Y-axis indicates if the streamflow data are best approximated by a lognormal distribution, a log Pearson Type III distribution, or if these data are unsuitable for either distribution. If the data are lognormal, the majority of the data points fall along a straight line. Data sets that may be approximated by a log Pearson Type III distribution exhibit an upward curvature if the skew is substantially greater than zero or a downward curvature if the skew is substantially less than zero. If linear regression is used to fit a straight line to the lognormal frequency factor (as the independent variable) and the common logarithms of the streamflow data (as the dependent variable), then the intercept approximates the average, and the slope of the line approximates the standard deviation of the common logarithm of the streamflow data (Helsel and Hirsch, 2002). Similarly, a regression line fit to the log-Pearson Type III frequency factor (as the independent variable) and the common logarithm of the streamflow data (as the dependent variable) provide the same statistics for a log-Pearson Type III distribution. The regression line with the best fit residual statistics indicates which distribution is a better approximation for the available streamflow data. Use of this regression method with visual inspection of the data is a good screening tool. Other methods, however, such as the probability-plot correlation-coefficient test (Vogel, 1986) or use of L-moment ratios (Stedinger and others, 1993) may be better for identifying an appropriate distribution.

Use of the MkPP Program

MkPP is a Visual Basic program that creates plotting-position files from NWISWeb daily mean streamflow-data files retrieved with output options specified by the GNWISQ program (table 2-1). The user interface consists of one interactive form identifying the program, providing input- and output-file information, and providing the graphical user interface controls necessary to specify user input and provide feedback on the streamflow-data-analysis. The MkPP program uses two input-file formats and creates one output-file format. MkPP can be run to calculate plotting positions for data from an individual station or can be run in batch mode for data from multiple stations. The program outputs tab-delimited text files that can be used with many computer applications including text editors, spreadsheets, graphing software, and statistical software. MkPP, if installed properly, should be compatible with commonly used Microsoft Windows operating systems.

Installation and Removal

The MkPP program depends on a number of software drivers and dynamic-link libraries that may not be installed and available on the user's computer. This program must be installed by someone who has administrative rights on the program user's computer. The administrator must test the installation with the user's profile to ensure that all permissions are set properly. A readme.txt file with installation instructions and three installation files are in the folder MkPPInstall on the CD-ROM containing this manual. Three files—setup.exe, setup.LST, and MkPPv1.CAB—are needed to install the program. The file setup.exe is the installation program. The setup.LST file is a text file that provides the necessary installation specifications. The file MkPPv1.CAB is the file containing the program and support files. The setup program is a standard Microsoft installation wizard that should be familiar to the user or system administrator. These three files must be located in the same directory on the CD-ROM or in a directory on the user's computer. The person installing the program should follow all the standard choices for installation. The installation program creates the directory C:\Program Files\FHWA\HEP\MkPP\ and includes the MkPP program in the computer's registry. If desired, a shortcut to the program can be added to the desktop manually after installation. Sample files are saved

Table 2-1. Summary of input and output files for the Make Plotting Position (MkPP Version 1.0) program.

[GNWISQ, Get NWIS streamflow (Q) software version 1; NWIS, National Water Information System; RDB, relational database]

File type	Example files on the CD-ROM directory: CDROM:\Applications\MkPPEx\	Comments
	Input files	
Station-list input file	IndexStations.txt	Lists long-term index-station numbers
NWISWeb daily value RDB streamflow data file	01118000Q.txt	The 2005 daily value format NWISWeb RDB file, formatted by GNWISQ
	Output files	
MkPP plotting-position file	CPP01118000Q.txt	Tab-delimited ouput file with selected percentiles and frequency factors The plotting-position method is denoted by the prefix of the filename
Interpolation output file	CPP01118000-INTQ.txt	Tab-delimited ouput file with selected percentiles and frequency factors User-selected option for data sets with fewer than 1,000 values Streamflow values are interpolated to produce even percentiles

on the CD-ROM containing this manual and should be copied to a directory in which the user has read, write, and execute rights. The user may uninstall the MkPP program and its support files by use of the standard Microsoft Windows Add or Remove Programs wizard on the control panel of the user's computer.

Input-File Formats

Data from the NWISWeb database is transmitted in a tab-delimited RDB format (Manis and others, 1988). RDB files are tab-delimited ASCII text files that can be imported into many software packages. The USGS RDB file format begins with comment lines that are intended to describe the source of the data and the content of the file. These comment lines are denoted by a pound symbol (#) in the first character location of each line. The comments are followed by a line of NWIS variable names for each type of information in the file. The variable name line is followed by a data-format line which indicates the maximum character length for each variable and the type of variable ("d" for date, "n" for number, and "s" for string). All remaining lines in the RDB files contain the daily mean streamflow data. If data elements are qualified, such as being estimated, being less than a listed value, or being greater than a listed value, a numerical value will appear within a data field with an additional character code; alternatively, if the measurement cannot be quantified, the data field may contain just the character code (U.S. Geological Survey, 2008). The documentation and format lines in these files and the qualification codes associated with the data are essential for transmitting meaningful data in an efficient format. Example daily mean streamflow files are also provided in the example data directory for the MkPP program on the CD-ROM containing this manual.

Station-List Input-File Format

Streamflow data for multiple stations can be processed in batch mode to facilitate the analysis of local, regional, and national hydrologic data sets. In batch mode, the program reads a station-list file. This list file does not contain any header lines. The list file consists of 8- to 12-character streamflow-gaging-station numbers on separate lines without leading or following spaces. The station-list file should not include any text except the station numbers of interest and should not end in a blank line. The station numbers in this file are used with a "Q.txt" suffix to identify NWISWeb data files for each streamflow-gaging station.

Streamflow Data Input-File Format

The MkPP program is designed to read the 2005 NWISWeb daily-value format. The GNWISQ program (appendix 1) may be used to download and reformat daily discharge files from NWISWeb (U.S. Geological Survey, 2008) to the 2005 NWISWeb dv format. The streamflow-data file output from GNWISQ provides the daily mean streamflow values for the requested period of record. The streamflow data-file names include the station number followed by the letter "Q" (a common abbreviation for streamflow in the hydrologic literature) and the suffix ".txt" (for example, 01022500Q.txt). The streamflow data file includes the agency code, USGS

station number, the date of each value, the daily mean streamflow value in cubic feet per second, and any daily mean streamflow-value qualification codes.

Output-File Format

The output-file format is a tab-delimited text file that identifies the streamflow-gaging station, the plotting-position formula, and the calculated values. Each output file is named by a plotting-position abbreviation prefix, the station number, and a ".txt" suffix (for example CPP01117500.txt). The abbreviations are "BPP" for the Blom plotting-position formula, "CPP" for the Cunnane plotting-position formula, "GPP" for the Gringorten plotting-position formula, "HPP" for the Hazen plotting-position formula, and "WPP" for the Weibull plotting-position formula. If the input data set has fewer than 1,000 individual streamflow data points, the user can create an interpolation output file with the same format and naming convention as the regular output file but with an additional suffix "-INT" added to the filename.

The output files have seven header lines to define the MkPP output. The streamflow-gaging station name and number are identified from the identification line from the input file and are printed as the first lines of both output file formats. The interpolation file format has an additional (second) line indicating that it is a "Table of interpolated percentiles." The next line in the output files indicates whether all data, data by calendar year, data by water year (October–September), or data by climatic year (April–March) is used. The following line in the output file lists the beginning date, ending date, and the number of missing days in the record. The fourth line in the regular file and the fifth line in the interpolation file, list the number of points in the selected record set. The next line lists the plotting-position formula that was used for analysis of the data. The next format line indicates the value of the percentile-skew of the original data and the base-ten logarithms of the data. The last format line is the tab-delimited heading line, which identifies the contents of each tab-delimited column below. All remaining lines include the tab-delimited plotting positions (as fractions and percentiles) and the standard normal frequency factors. If the percentile skew of the common (base-ten) logarithms of the data is greater than 0.05, the Log-Pearson type III frequency factors also are printed in the next column. The streamflow values are printed in the last column in order of increasing streamflow.

The MkPP program is designed to print plotting positions, percentiles, and frequency factors for the streamflow data values ranked as smallest and largest in the data set and for a selection of values between the extremes. The selection is determined by the number of sampled values in the data set. If there are fewer than or equal to 1,000 data values, MkPP program prints each value with the associated plotting positions, percentiles, and frequency factors. The user may also select an interpolated-value output format that provides data with integer percentiles. If the number of data values is more than 1,000 and less than or equal to 10,000, MkPP prints the maximum and minimum values as well as 999 selected values that represent each sequential plotting position in increments of 0.001 (0.1 percentile) from 0.001 to 0.999. Finally, if a data set includes more than 10,000 values, MkPP prints the minimum data value, 9 selected values with plotting positions from 0.0001 to 0.0009, 999 selected values with plotting position in increments of 0.001 from 0.001 to 0.999, 9 selected values with plotting positions from 0.9991 to 0.9999, and the maximum value in the data set. Streamflow values associated with these incremental percentile values are calculated by use of linear interpolation between adjacent streamflow and plotting-position values in the original ranked data set. The minimum and maximum plotting positions are printed to six decimal places; this degree of precision should be adequate to characterize more than 200 years of daily mean streamflow values. The other percentiles are printed to four decimal places. MkPP program prints the interpolated streamflow values to three decimal places. Daily mean streamflow values are commonly reported with only two or three significant figures. These values, however, may be reported as a fraction of a cubic foot per second for some streams.

Graphical User Interface

The graphical user interface consists of one interactive form and two file-specification dialog forms. A screen resolution of at least 1024 by 768 pixels is needed to properly display the MkPP interface. The program is initiated by double-clicking on the program file or a Microsoft Windows shortcut to the executable program-file location. Initially, the user is presented with two choices, an individual NWIS RDB file or a batch file (fig. 2-3). If the user selects the individual-file option, the program then requires that the user select a USGS streamflow-gaging-station data file (fig. 2-4). If the user selects the batch-file option, a file-specification dialog box appears so that the user may select the name of a file containing one valid station number per line. In batch-file mode, the files must follow the streamflow-file naming convention (for example, 01117500Q.txt). In either case, the form indicates the choice and waits for the command to process data (fig. 2-5). The user can select an option within the "Data Set" frame to use all data, data by calendar year (from the first January through the last December), data by water year (from the first October through the last September), or data by climatic year (from the first April through the last March). The program begins and ends the record that is used for analysis on the basis of the user's selection. However, the program does not use this to exclude available data for parts of a year if there are missing days (or missing years) within the period of record. Once the user selects the Process Data button (fig. 2-5), the user must enter a letter to specify the desired plotting-position formula (fig. 2-6). Finally, the program displays station information and streamflow statistics (fig. 2-7). If the program is run in batch mode, information and statistics for each station are displayed successively on the form as the data are processed.

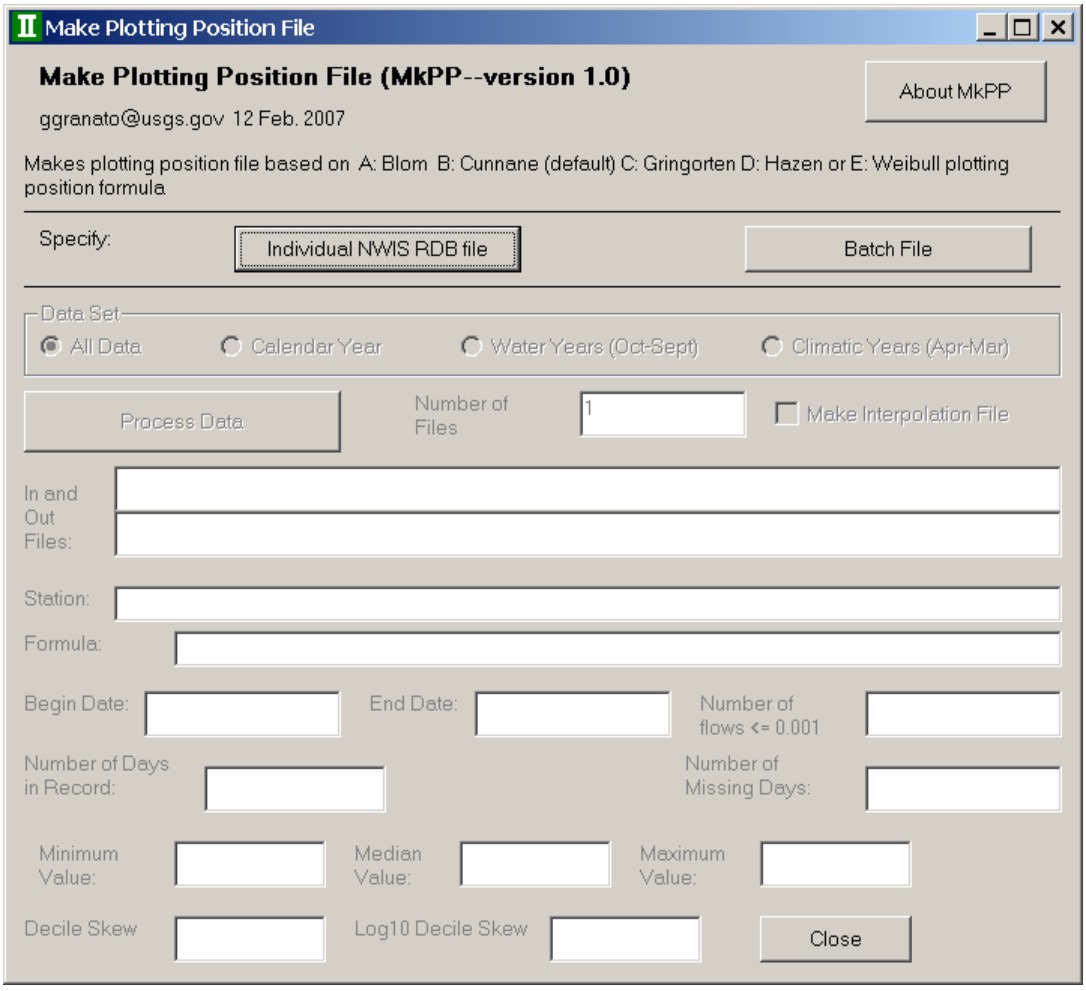

Figure 2-3. Initial appearance of the MkPP program input form when the program opens.

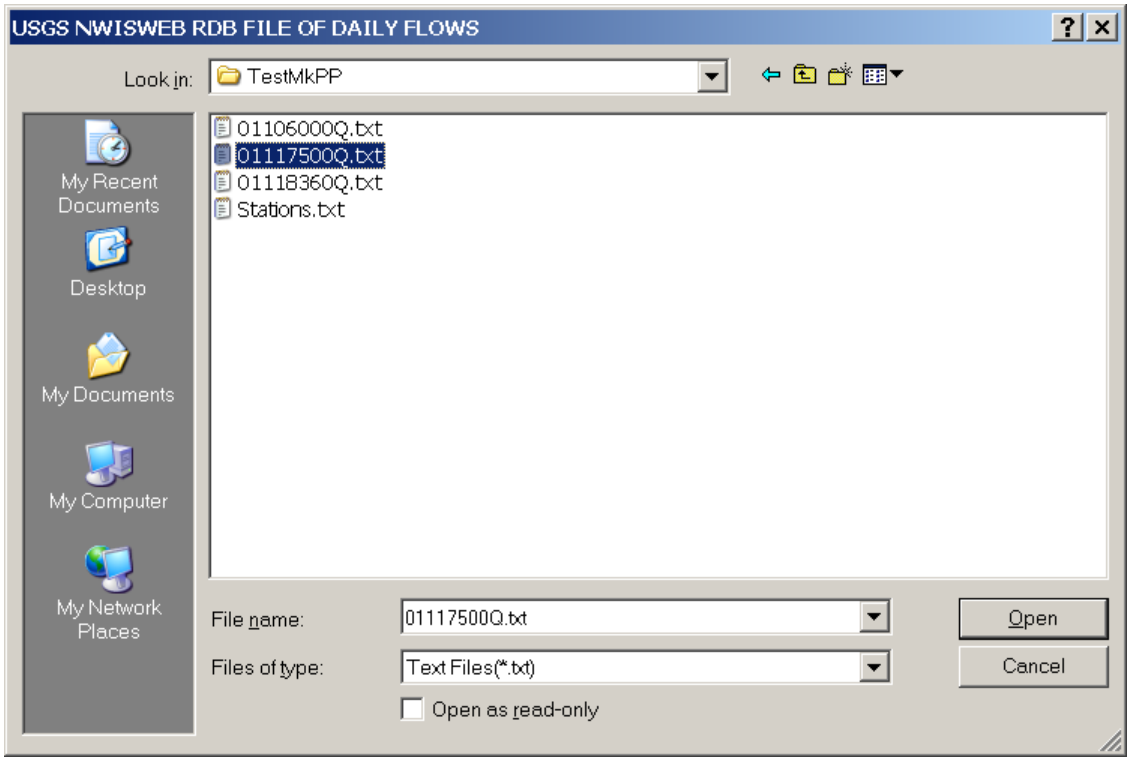

Figure 2-4. Example of the MkPP program file-specification dialog box.

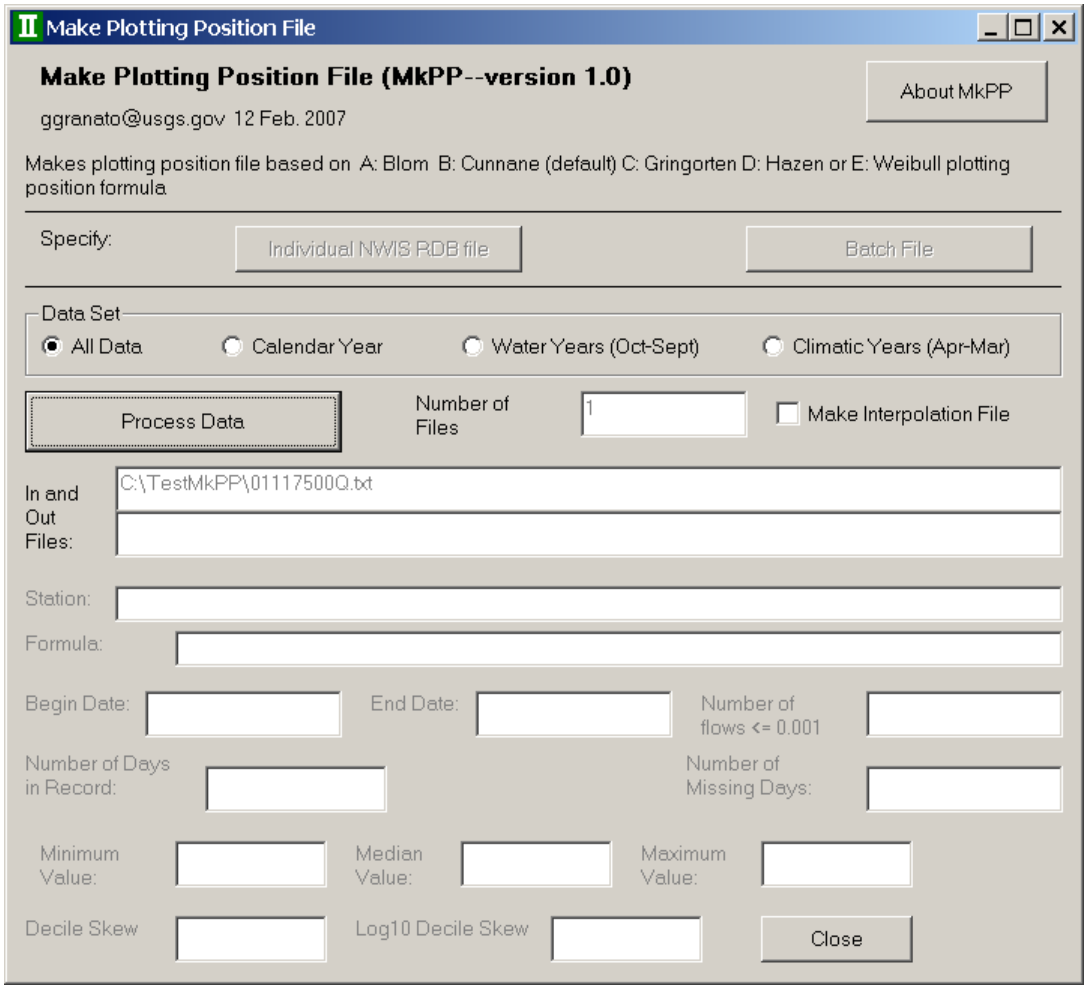

Figure 2-5. Example of the MkPP program input form when ready for data processing.

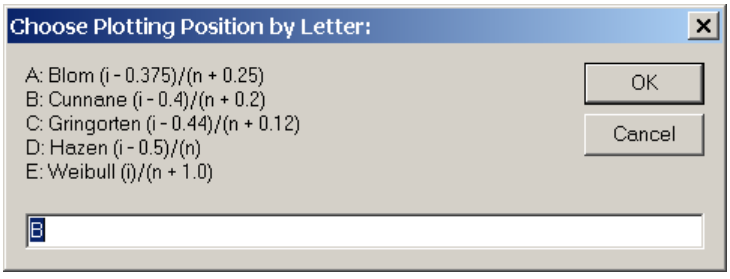

Figure 2-6. Example of the MkPP program plotting-position equation-input form.

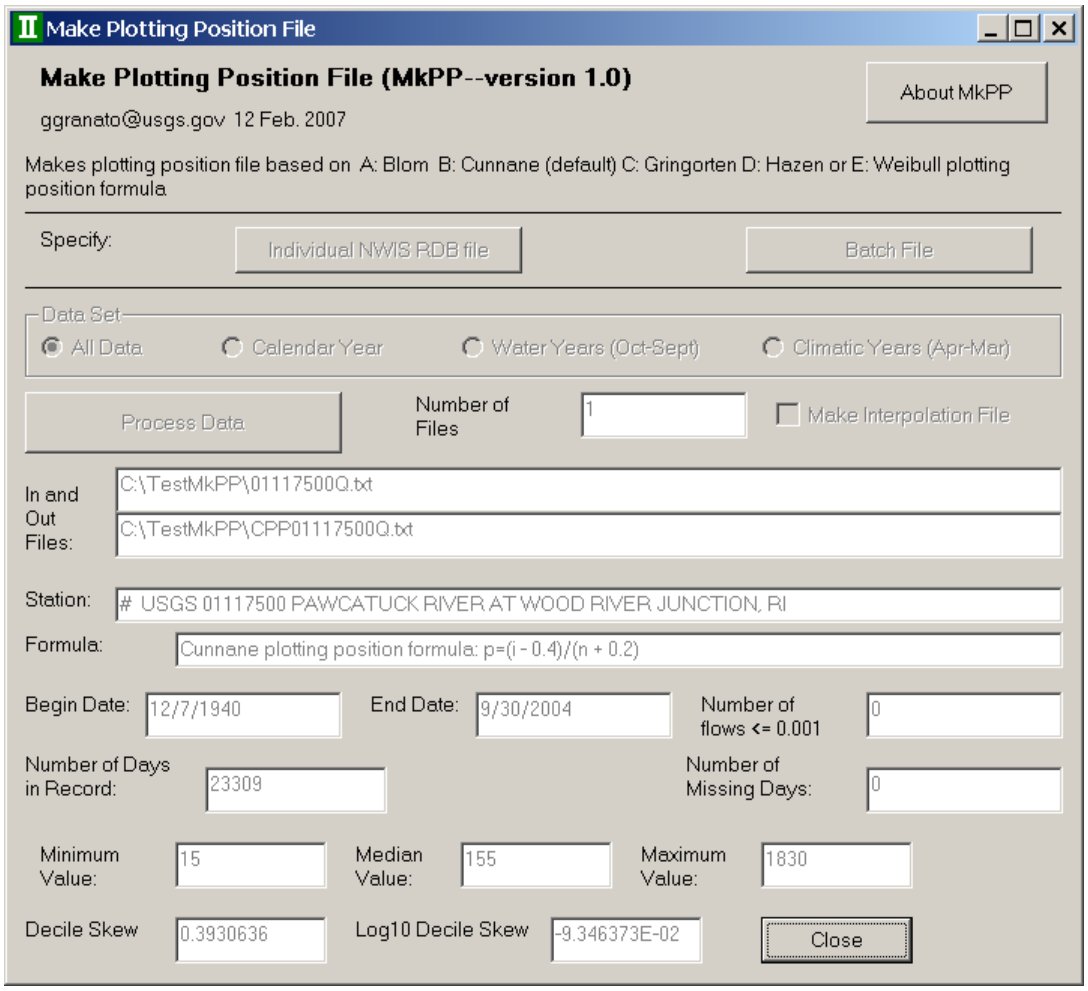

Figure 2-7. Example of the MkPP program input form after completion of the data-processing activities for streamflow-gaging-station data.

If the input data set has fewer than 1,000 individual streamflow data values the user can create an interpolation output file. Once the "Make Interpolation File" check box is checked, the program will interpolate the input values and associated percentiles to produce a file with integer percentile values. The user must check the "Make Interpolation File" check box (fig. 2-5) before the Process Data button is clicked, because the checkbox is locked once the process data button is clicked (fig. 2-7).

Summary

The Make Plotting Position (MkPP Version 1.0) program was developed by the U.S. Geological Survey (USGS) in cooperation with the Federal Highway Administration to facilitate visual analysis of daily mean streamflow data. Streamflow statistics are important for many planning and design activities including water supply, habitat protection, bridge and culvert design, surface- and ground-water models, and water-quality assessments. Graphical examination of environmental data is commonly considered the first step for statistical analysis of this data. Graphs, as visual summaries of data, describe essential information more quickly and completely than do tables of numbers or summary statistics, and thus provide an understanding of the data that is difficult to obtain in any other way. MkPP was written to calculate the plotting positions, normal frequency factors, and Log-Pearson Type III frequency factors for daily mean streamflow-value files retrieved from the USGS National Water Information System (NWIS) Web site. MkPP condenses information in large daily mean streamflow files (one 40-year record may contain about 14,610 daily mean values and require about 400,000 bytes of computer disk space) into a short plotting-position file of about 1,000 selected streamflow values over the complete range of flow. MkPP uses standard equations and statistical methods to facilitate graphical-data analysis. The user may choose from among several of the most commonly used plotting-position formulas, including the Blom (1958), Cunnane (1978), Gringorten (1963), Hazen (1914), and Weibull (1939) formulas.

This manual describes the implementation, use, and interpretation of results from the MkPP program. MkPP is a Visual Basic program with a graphical user interface that should be familiar to the user because it follows standard Microsoft Windows interface conventions. The program uses daily mean streamflow files that can be obtained from USGS NWISWeb site. The program outputs a tab-delimited text file that can be used with many computer applications, including text editors, spreadsheets, graphing software, and statistical software. An executable version of the program, example files, and the Visual Basic source code are documented in the MkPP directory on the computer disk containing this manual.

Acknowledgments

The author thanks Elizabeth Ahearn and Julie Kiang of the USGS for providing thoughtful and thorough technical reviews that helped the author improve the MkPP program and the associated documentation. Gardner Bent and Andrew Waite of the USGS did initial tests of the program and recommended several improvements included in the current version of MkPP.

References Cited

Benson, M.A., and Carter, R.W., 1973, A national study of the streamflow data-collection program: U.S. Geological Survey Water-Supply Paper 2028, 44 p.

Blom, G., 1958, Statistical estimates and transformed beta variables: New York, John Wiley, p. 68–75, 143–146.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988, Applied hydrology: New York, McGraw-Hill Publishing Company, 572 p.

Cunnane, C., 1978, Unbiased plotting positions—a review: Journal of Hydrology, v. 37, p. 205–222.

DiToro, D.M., 1984, Probability model of stream quality due to runoff: Journal of Environmental Engineering, v. 110, no. 3, p. 607–628.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990, Pollutant loadings and impacts from highway stormwater runoff volume I: design procedure: Federal Highway Administration Final Report FHWA-RD-88-006, 67 p.

Glysson, G.D., 1987, Sediment-transport curves: U.S. Geological Survey Open-File Report 87-218, 47 p.

Gringorten, I. I., 1963, A plotting rule for extreme probability paper: Journal of Geophysical Research, v. 68, p. 813–814.

Haan, C.T., 1977, Statistical Methods in Hydrology: Ames, IA, The Iowa State University Press, 378 p.

Harter, H.L. 1969, A new table of percentage points of the Pearson type III distribution: Technometrics, v. 11, no. 1, p. 177–187.

Harter, H.L. 1971, More percentage points of the Pearson distribution: Technometrics, v. 13, no. 1, p. 203–204.

Hazen, A., 1914, Storage to be provided in the impounding reservoirs for municipal water supply: Transactions of the American Society of Civil Engineers, v. 77, p. 1547–1550.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: Techniques of Water-Resources Investigations of the U.S. Geological Survey, chap. A3, book 4, 510 p.

Hirsch, R.M., Helsel, D.R., Gilroy, E.J., and Cohn, T.A., 1993, Chapter 17—Statistical analysis of hydrologic data, *in* Maidment, D.R., ed., Handbook of Hydrology: New York, McGraw-Hill Book Company, p. 17.1–17.55.

Kirby, W.H., 1972, Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson Type 3 distribution: Water Resources Research, v. 8, no. 5, p. 1251–1254.

Limbrunner, J.F., Vogel, R.M., and Brown, L.C., 2000, Estimation of harmonic mean of a lognormal variable: Journal of Hydrologic Engineering, v. 5, no. 1, p. 59–66.

Manis, Rod, Schaffer, Evan, and Jorgensen, Robert, 1988, Unix relational database management—application and development in the Unix environment: Englewood Cliffs, NJ, Prentice Hall, 476 p.

Markovic, R.D., 1965, Probability functions of best fit to distributions of annual precipitation and runoff: Fort Collins, CO, Colorado State University Hydrology Papers, n. 8, 33 p.

Mathey, S.B., ed., 1998, National Water Information System: U.S. Geological Survey Fact Sheet FS-027-98, 2 p.

McCuen, R.H., Johnson, P.A., Ragan, R.M., 2002, Highway hydrology—Hydraulic design series no. 2, 2d ed.: U.S. Department of Transportation, Federal Highway Administration, FHWA-NHI-02-001, 424 p.

Nash, D.B., 1994, Effective sediment-transporting discharge from magnitude-frequency analysis: The Journal of Geology, v. 102, p. 79–95.

National Research Council, 2004, Assessing the National Streamflow Information Program: Washington, DC, The National Academies Press, 164 p.

O'Connor, D.J., 1976, The concentration of dissolved solids and river flow: Water Resources Research, v. 12, no. 2, p. 279–294.

Riggs, H.C., 1968, Frequency curves: Techniques of Water-Resources Investigations of the United States Geological Survey, chap. A2, book 4, 15 p.

Stedinger, J.R., Vogel, R.M., Georgiou, E.F., 1993, Chapter 18—Frequency analysis of extreme events, *in* Maidment, D.R., ed., Handbook of Hydrology: New York, McGraw-Hill Book Company, p. 18.1–18.66.

Thomas, M.P., 1966, Effect of glacial geology upon the time distribution of stream flow in eastern and southern Connecticut: U.S. Geological Survey Professional Paper 550-B, p. 209–212.

U.S. Army Corps of Engineers, 1993, Engineering and Design—Hydrologic frequency analysis: Washington, DC, Department of the Army, CECW-EH-Y, U.S. Army Corps of Engineers, Engineer Manual No. 1110-2-1415, 147 p.

U.S. Army Corps of Engineers, 1994, Engineering and Design—Flood-Runoff Analysis: Washington, DC, Department of the Army, CECW-EH-Y, U.S. Army Corps of Engineers Engineer Manual No. 1110-2-1417, 214 p.

U.S. Environmental Protection Agency, 1986, Technical guidance manual for performing wasteload allocation—Book 6, Design conditions, Chapter I Stream design flow for steady-state modeling: Washington, DC, U.S. Environmental Protection Agency: Report EPA 440/4-86-014, 64 p.

U.S. Geological Survey, 2002, NWIS Web: New Site for the Nation's water data: U.S. Geological Survey Fact Sheet FS-128-02, 2 p.

U.S. Geological Survey, 2008, NWISWeb data for the Nation, accessed February 12, 2008, at *http://waterdata.usgs.gov/nwis/*

U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency, Bulletin 17B of the Hydrology Subcommittee: Reston, VA, U.S. Geological Survey, Office of Water Data Coordination, 185 p.

U.S. Natural Resources Conservation Service, 1998, Tables of percentage points of the Pearson Type III Distribution: U.S. Department of Agriculture, Natural Resources Conservation Service, Conservation Engineering Division, Technical Release 38, 17 p.

Van Buren, M.A., Watt, W.E., and Marsalek, J., 1997, Application of the lognormal and normal distributions to stormwater quality parameters: Water Research, v. 31, no. 1, p. 95–104.

Vogel, R.M., 1986, The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses: Water Resources Research, v. 22, no. 4, p. 587–590.

Vogel, R.M., and Kroll, C.N., 1989, Low-flow frequency analysis using probability-plot correlation coefficients: Journal of Water Resources Planning and Management, v. 115, no. 3, p. 338–357.

Vogel, R.M., and McMartin, D.E., 1991, Probability plot goodness of fit and skewness estimation for the Pearson type 3 distribution: Water Resources Research, v. 27, no. 12, p. 3149–3158.

Vogel, R.M., Rudolph, B.E., and Hooper, R.P., 2005, Probabilistic behavior of water-quality loads: Journal of Environmental Engineering, v. 131, no. 7, p. 1081–1089.

Vogel, R.M., Stedinger, J.R., and Hooper, R.P., 2003, Discharge indicies for water-quality loads: Water Resources Research, v. 39, no. 10, p. 1273–1281.

Wahl, K.L., Thomas, W.O., Jr., and Hirsch, R.M., 1995, Stream-gaging program of the U.S. Geological Survey: U.S. Geological Survey Circular 1123, 22 p.

Weibull, W., 1939, The phenomenon of rupture in solids: Stockholm, Ingeniors Vetenskaps Akademien Handlinga 153, p. 17.