U.S. Department of Transportation
Federal Highway Administration
Office of Project Development and Environmental Review

Statistical Characterization of Streamflow Data

By Gregory E. Granato

Appendix 2 *in*

Granato, G.E., 2010, Methods for development of planning-level estimates of stormflow at unmonitored sites in the conterminous United States: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-HEP-09-005, 90 p. with CD-ROM

Prepared by the
U.S. Department of the Interior
U.S. Geological Survey
Northborough Massachusetts, 2010

USGS
science for a changing world

Foreword

The mission of the Federal Highway Administration (FHWA) is to continually improve the quality of our Nation's highway system and intermodal connections in a manner that protects and enhances the natural environment and communities affected by transportation. The FHWA must develop and disseminate research products that help FHWA and its partners implement surface transportation programs in a manner that protects and enhances the natural and human environment. More specifically, the Office of Office of Project Development and Environmental Review of the FHWA strives to develop and disseminate skills, tools, and information to redesign Federal environmental and transportation decisionmaking, and to ensure an integrated process at the Federal, State, tribal, and local levels. These tools, techniques and methods are designed to reduce direct and indirect adverse impacts of highways on water quality, habitat, and ecosystems to preserve and enhance human health, biological productivity, and ecological diversity.

This report, the associated computer applications, and data provide tools and techniques for developing planning-level estimates of stormflows at sites receiving highway runoff. This information is vital for assessing the potential for adverse effects of runoff on receiving waters throughout the Nation and it should provide transportation agencies with the tools and information to improve project delivery without compromising environmental protection.

Patricia Cazenas Highway Engineer
Office of Project Development and Environmental Review
Federal Highway Administration

Susan Jones Highway Engineer
Office of Project Development and Environmental Review
Federal Highway Administration

For more information on FHWA-USGS Runoff-Quality Studies:

World Wide Web: http://ma.water.usgs.gov/fhwa/

World Wide Web: *http://www.fhwa.gov/environment/* *http://environment.fhwa.dot.gov/*

Suggested citation

Granato, G.E., 2010, Statistical characterization of streamflow data--Appendix 2 in Granato, G.E., 2010, Methods for development of planning-level estimates of stormflow at unmonitored sites in the conterminous United States: Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, FHWA-HEP-09-005, 90 p. with CD-ROM

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Contents

Statistical Characterization of Streamflow Data

By Gregory E. Granato

Introduction

The stochastic empirical loading and dilution model (SELDM) uses streamflow statistics that define prestorm streamflow from the basin upstream of a highway-runoff discharge point. The statistics used by SELDM are derived from National streamflow monitoring networks. Statistical characterization of streamflow data is necessary for modeling planning-level estimates of prestorm streamflows for water-quality analysis. Statistical characterization includes calculating statistics and using information from these statistics for selecting an appropriate probability distribution. Once a distribution is selected, frequency-factor methods can be used to stochastically generate a population of mean-daily streamflows. This appendix is a brief review of the literature that describes methods for estimating streamflow statistics and related probability distributions that were considered for use in the SELDM development process. Methods for doing statistical characterization of streamflow data, estimating streamflow statistics at ungaged sites, and estimating streamflow statistics at sites with limited data using streamflow correlation are discussed in detail and referenced to supplement material in the main body of this report.

Statistical Characterization of Streamflow Data

Statistical characterization of streamflow data is necessary for modeling planning-level estimates of prestorm streamflows for water-quality analysis. Statistical characterization includes calculating statistics and using information from these statistics for selecting an appropriate probability distribution. Once a distribution is selected, frequency-factor methods can be used to stochastically generate a population of mean-daily streamflows. The common logarithms of mean-daily flow values are used to model streamflows because these values typically have a lower bound of zero and vary by two or more orders of magnitude (Chow, 1954; Haan, 1977; Chow and others, 1988; Stedinger and others, 1993). Frequency-factor methods are used to calculate individual data values for different probability distributions (Haan, 1977; Chow and others, 1988; Stedinger and others, 1993). Some streamflow records, however, have a substantial proportion of zero flows that must be modeled with conditional probability methods if logarithms are used to model non-zero mean-daily flow values. Trends in are not calculated for the statistics used with SELDM to characterize mean-daily flow values, but trends may be important for some analyses.

Probability Distributions

For water-quality studies, the population of mean-daily streamflows commonly is approximated by use of a lognormal distribution (Di Toro, 1984; Driscoll and others, 1990b; Nash, 1994; Van Buren and others, 1997; Schwartz and Naiman, 1999; Limbrunner and others, 2000; Novotny, 2004; Vogel and others, 2005). The common logarithms of mean-daily flow values are used to model streamflows because these values typically have a lower bound of zero and vary by two or more orders of magnitude (Chow, 1954; Haan, 1977; Chow and others, 1988; Stedinger and others, 1993). Chow (1954) provides a detailed explanation of the lognormal distribution and its use for hydrological applications, and Blackwood (1992) provides a summary of the important properties of the lognormal distribution for analysis of environmental data. Di Toro (1984) used a lognormal distribution for streamflows in a theoretical urban-runoff mixing model. Driscoll and others (1990b) approximated receiving-water flows by use of annual statistics and lognormally distributed streamflows. Nash (1994) found that a lognormal-streamflow approximation was useful for examining sediment transport. Van Buren and others (1997) used the lognormal distribution for discharge and concentration for use in probabilistic urban-stormwater studies. Schwartz and Naiman (1999) examined methods for determining planning-level load estimates and determined that use of a single characteristic streamflow and concentration (the mean or geometric mean) was inadequate for characterizing potential effects of runoff. Schwartz and Naiman (1999), therefore, used lognormal assumptions for streamflow, concentration, and loads to evaluate planning-level runoff-quality estimates. Limbrunner and others (2000) examined L-moment statistics for mean-daily streamflows from 1,571 streamgages across the United States and concluded that a two-parameter lognormal distribution will provide a reasonable approximation for these mean-daily streamflows. Novotny (2004) used a Monte Carlo approach with lognormal distributions for concentrations and flows to examine the margin of safety and to estimate the probabilities of water-quality standard excursions in total maximum daily load (TMDL) studies. Vogel and others (2005) used a lognormal distribution of streamflows as a first-order approximation to examine the probabilistic structure of observations of streamflow, concentrations, and constituent loads.

Other statistical distributions also may be applicable for modeling mean-daily streamflow. L-moment diagrams provide a robust method for identifying the underlying probability distribution of a data set (Vogel and Fennessey, 1993). The QSTATS program (Granato, 2009) calculates traditional statistics, L-moment statistics, and the L-moment ratios to facilitate the selection of an appropriate distribution. For example, Quader and Guo (2005) indicate that a one or two-parameter exponential distribution may be appropriate for modeling mean-daily streamflows in small (less than about 19 square miles) urban basins because a lack of base flow may cause a high proportion of daily flows to be at or near zero. Vogel and Fennessey (1993) and Castellarin and others (2004) used the generalized Paraeto distribution to model daily streamflow populations. Ashkar and Ouarda (1996) provide information for fitting data to a generalized Pareto distribution. A three parameter lognormal distribution also is commonly used to model environmental data (Sangal and Biswas, 1970; Stedinger, 1980; Stedinger and others, 1993). The three-parameter lognormal distribution will produce streamflow estimates that are similar to estimates from a properly specified log-Pearson type 3 distribution. For small samples, however, uncertainties in calculated skew coefficients may adversely affect estimates made using the log-Pearson type 3 distribution (Sangal and Biswas, 1970; Stedinger, 1980; Stedinger and others, 1993). The three-parameter lognormal distribution is not used in this study to estimate prestorm flows because the records were large (7,000-14,000 daily values) and the log-Pearson type 3 is more flexible than the three-parameter lognormal distribution. The log-Pearson type 3 can be used to model data sets with logarithmic skew values within the range of minus nine to nine (Haan, 1977; Rao, 1980; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988). Other researchers have used the four-parameter Kappa distribution to model daily streamflows, but the statistics for this distribution do not have a direct physical interpretation and are difficult to derive (Stedinger and others, 1993).

The log-Pearson type 3 distribution is commonly used to characterize hydrologic data (Bobee, 1975; Haan, 1977; Rao, 1980; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993). The lognormal, the exponential, and the generalized Pareto distributions each are a special case of the log-Pearson type 3 distribution so that, with the proper statistics, each may be characterized within this family of distributions (Bobee and Ashkar, 1991; Ashkar and Ouarda, 1996). The exponential and the generalized Pareto distributions are sometimes preferred for stochastic data generation because these distributions have an explicit analytical equation. The lognormal distribution also is preferred for stochastic data generation because random-normal variates are available from many statistical-software packages (Haan, 1977; Cheng and others, 2007). Cigizoglu (2000) examined use of the log-Pearson type 3 and other distributions to describe flow-duration curves of daily streamflow data. A number of investigations (for example, Vogel and McMartin, 1991; Vogel and Wilson, 1996) indicate that the log-Pearson type 3 is a very flexible distribution that will provide a good fit to many different types of hydrologic data even if the underlying population is not a pure log-Pearson type 3 distribution.

Frequency-Factor Methods

Frequency-factor methods can be used to model a sample of individual streamflow values that follow a lognormal or log-Pearson type 3 distribution (Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993; U.S. Army Corps of Engineers, 1993; 1994; McCuen and others, 2002; Cheng and others, 2007). Frequency-factor methods also can be used to calculate individual data values for other probability distributions (Haan, 1977; Chow and others, 1988; Stedinger and others, 1993). The frequency factor method uses the mean and the standard deviation from a data sample to predict values from the underlying population by use of a distribution-specific frequency factor. For a lognormal or log-Pearson type 3 distribution, the equation for the frequency factor method is

$$X_i = X_m + S \times K_i \quad ,(1)$$

Where

X_i is the value of the logarithm of the i^{th} streamflow value;

X_m is the geometric mean streamflow value (in logarithmic space);

S is the standard deviation of the logarithms streamflow values; and

K_t is the distribution-specific frequency factor.

The distribution-specific frequency factor relates the probability of occurrence of a value to a multiple of the standard deviation above or below the mean value. The frequency factor equals the standard normal variate if the skew value of the population is zero (Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993). In this case, standard normal probability tables, which are listed in most textbooks on statistics, can be used to estimate the frequency factor for one or more individual values (Haan, 1977; Chow and others, 1988; McCuen and others, 2002). Algebraic approximations that relate exceedance probabilities to standard-normal variates are available for numerical implementation (Abramowitz and Stegun, 1965; Odeh and Evans, 1974; Beasley and Springer, 1977; Chow and others, 1988; Wichura, 1988). Once a value is calculated in logarithmic space, it is inverted to obtain the streamflow of interest.

Data from a Pearson Type 3 (gamma) and log-Pearson type 3 distribution also can be estimated with frequency factors (Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993; U.S. Army Corps of Engineers, 1993; 1994; Cheng and others, 2007). These frequency factors account for the skew in the data. If the skew of a population equals zero, the frequency factor is the standard normal variate. As skews deviate from zero, the relation between exceedance probability and the associated frequency factor shifts to reflect the distribution of values above and below the median value. Tables of frequency factors for different values of skew are commonly available for selecting a frequency factor for a given exceedance probability (Harter, 1969; 1971; Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; U.S. Army Corps of Engineers, 1993; Natural Resources Conservation Service, 1998; McCuen and others, 2002).

Several algebraic approximations are commonly used to approximate the Pearson type 3 frequency factor as a function of the normal frequency factor for a given exceedance probability and the estimated population skew. The Wilson-Hilferty approximation (Wilson and Hilferty, 1931) is commonly used to estimate frequency factor values from a log-Pearson type 3 distribution (Kirby, 1972; Interagency Advisory Committee on Water Data, 1982; U.S. Environmental Protection Agency, 1986; Rossman, L.A., 1990b; Bobee and Ashkar, 1991; Stedinger and others, 1993; U.S. Army Corps of Engineers, 1993; U.S. Environmental Protection Agency, 2004a). The Wilson-Hilferty approximation is

$$K_{LP3} = (2/G) \times ((1 - (G/6)^2 + (G/6)K_N)^3 - 1) \qquad .(2)$$

Where

K_{LP3}	is the log-Pearson type 3 frequency factor
G	is the skew coefficient; and
K_N	is the standard normal variate for a given exceedance factor.

The Wilson-Hilferty approximation is very accurate for a coefficient of skew within plus or minus 1.0, is considered sufficiently accurate for a coefficient of skew within plus or minus 2.0, and the accuracy deteriorates as the coefficient of skew approaches a value of plus or minus 3 (Kirby, 1972; Kite, 1977; Bobee and Ashkar, 1991). Another approximation is a polynomial approximation suggested by Kite (1977), which also is in common use (Chow, 1998; U.S. Army Corps of Engineers, 1994; Cheng and others, 2007). The Kite (1977) approximation is:

$$K_{LP3} = K_N + (K_N^2 - 1)(G/6) + (1/3)\,(K_N^3 - 6\,K_N)(G/6)^2 - (K_N^2 - 1)(G/6)^3 + K_N(G/6)^4 + (1/3)(G/6)^5 \quad . \text{(3)}$$

The Kite (1977) approximation is of comparable accuracy to the Wilson-Hilferty algorithm if the skew coefficient is within the range of about plus or minus 1.4. Bobee and Ashkar (1991) provide several other polynomial approximations for the frequency factor.

Kirby (1972) developed a modified Wilson-Hilferty algorithm that provides acceptable estimates of log-Pearson type 3 frequency factors for samples with a coefficient of skew that is within the range of about plus or minus 9. This modified Wilson-Hilferty algorithm is essentially equivalent to the Wilson-Hilferty for small skew values and is substantially better than the Wilson-Hilferty (1931) and Kite (1977) approximations for skews beyond the range of plus or minus 2. Almost 99 percent of the streamgages used in this study have (common logarithm) skew coefficients within the range of plus or minus two. Only a few stations have skew values beyond this range, but these stations, or other stations with relatively large skew values may be selected by an analyst for local use. Therefore, the modified Wilson-Hilferty algorithm (Kirby, 1972) was selected for use in developing planning-level estimates of prestorm streamflows for water-quality analysis. The paper describing the modified Wilson-Hilferty algorithm is Kirby72.pdf is included on the CD-ROM accompanying this report.

Conditional Probability

Conditional probability methods are used to account for data below a measurement threshold (Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993). For example, Delong and Wells (1988) used conditional probability methods to determine water-quality loads from ephemeral and intermittent streams in Wyoming to account for the proportion of dry days in these drainage basins. Croker and others (2003) used conditional-probability methods to develop flow-duration curves for gaged ephemeral streams in Portugal and used these statistics to estimate streamflows (or the lack thereof) at ungaged sites. The measurement threshold that is used to determine a nonzero mean-daily flow is commonly about 0.01 ft^3 (Rantz, 1982). Dry days are defined as days with no detectable streamflow measurements and are assigned a mean-daily streamflow of zero. Although this condition is analogous to the occurrence of water-quality concentrations below a laboratory detection limit, natural surface-water flows have a physical lower limit of zero rather than a continuum of infinite dilution. Conditional probability methods are necessary for this national synthesis of planning-level prestorm flow statistics because about thirty percent of the 2,783 streamgages included in this synthesis have one or more dry days and more than 50 percent of the streamgages have one or more dry days in 28 of the 84 ecoregions in the conterminous United States.

Conditional probability methods (Haan, 1977; Interagency Advisory Committee on Water Data, 1982; Chow and others, 1988; Stedinger and others, 1993; McCuen and others, 2002) may be used to adjust streamflow statistics to account for the proportion of zero streamflows at a site of interest in the stochastic data-generation process. If conditional probability methods are used, statistics are calculated for nonzero streamflows and the probabilities are adjusted for the proportion of days with zero flows (William Kirby, U.S. Geological Survey Office of Surface Water, Written commun., 2003). Methods for stochastic data generation with conditional probabilities depend on the objective of the analysis. If the analyst needs to recreate a historical record that matches the population statistics, the serial correlation, the seasonal cycle, and inter-annual variability, then methods described by Salas (1993) must be used. If the analyst just needs random samples that do not account for such temporal variations in flow, then probability of selecting a zero flow value from the period of record is equal to the proportion of dry days in the record, and the probability of selecting a nonzero flow value is equal to one minus the proportion of dry days in the record. The probability of exceeding any nonzero flow value is equal to the probability of exceeding the selected streamflow value from among the nonzero streamflow values times the probability of selecting a nonzero flow. For example, if the proportion of zero flows is 0, 10, 25, or 50 percent of the record, the probability of exceeding the median of nonzero flows is 50, 45, 37.5, or 25 percent, respectively.

Streamflow Trends

Several reports indicate that there may be trends in long-term streamflow records. Lins and Michaels (1994) examined mean monthly streamflow from 559 streamgages from the Hydro-Climatic Data Network (HCDN) within the conterminous United States for the 1941-1988 period and found upward trends in autumn and winter streamflows. Lins and Slack (1999) used nonparametric trend tests on 395 selected streamgages from the HCDN for different time periods in the interval from 1914-1993. McCabe and Wolock (2002) examined the annual minimum, median, and maximum of mean-daily streamflows for the period 1941–1999 at 400 streamgages in the conterminous United States. They detected a noticeable increase in annual minimum and median daily streamflow around 1970, but did not detect increases in maximum streamflows. They determined that this step change, which was predominant in the eastern United States, coincided with changes in precipitation during the same period. They determined that, in general, flow quantiles that are less than or equal to the 70^{th} percentile flow duration are increasing, and the highest flows show little to no trends. Therefore, Lins and Slack

(1999) conclude that streamflows are increasing, but high flows are not becoming more extreme. Hodgkins and others (2003) examined data from 27 rural, unregulated river gaging stations in New England, USA with an average of 68 years of record and concluded that changes in the timing of streamflow statistics are correlated with increasing temperatures over the last 30 years. Asquith and others (2007), however, did not detect trends in annual maximum, minimum or average streamflows in 80 to 90 percent of 712 selected streamgages in Texas. Cohn and Lins (2005) indicate that recent trends in hydrologic data may be substantial but the effect of anthropogenic climate change may be difficult to separate from natural long-term cycles.

The streamflow data records used to represent prestorm flows in this study are from stations with at least 24 years of record that was collected during the period 1960-2004. McCabe and Wolock (2002) indicate a step change in streamflow statistics that occurs around 1970. The maximum proportion of pre-1970 data would be about 58 percent for a given streamgage, assuming that a 24-year record started in 1960. About 23 percent of a 44-year record during 1960-2004would be pre-1970 data. The statistics were not adjusted for the potential effects of trends because the magnitude and timing of trends differ from region to region and site to site (Lins and Slack, 1999; McCabe and Wolock, 2002; Hodgkins and others, 2003). Furthermore, site specific trends may be affected by changes in land use and water use during the period because the current study is not restricted to the relatively natural HCDN sites. Finally, uncertainties in estimating streamflows by region or from nearby sites are, potentially, much larger than the effects of subtle trends in long-term streamflow data. However, if long-term temporal trends or shifts in streamflow are of concern for a given analysis, then the user may update the data and statistics used for future analyses by using the streamflow-data processing software included on the CD-ROM accompanying this report.

Alternative Methods for Estimating Streamflow Statistics at Ungaged Sites

One of the primary objectives of the USGS streamflow gaging program is to provide the information and data necessary to estimate streamflows at any site in the United States (Benson and Carter, 1973; National Research Council, 2004). Although many streamgages are located at or near the point where a road crosses a river, the number of current and historical streamgages (about 24,000) is small in comparison to the estimated number of small stream segments in the conterminous United States. For example, Leopold and others (1992) estimate that there are about 2 two-million stream segments nested within basins with mean drainage areas that are less than 20 square miles. It is therefore necessary to use available flow statistics from selected streamgages to estimate streamflows at an ungaged site of interest. There are three primary methods to estimate streamflows at sites without any streamflow data from the site of interest. These methods are runoff maps, regression on basin characteristics, and statistical transfer with drainage-area ratios. The primary method to estimate streamflows at sites with limited streamflow data is correlation of streamflows.

Runoff Maps

Runoff maps provide a good overview of the spatial variability of annual average streamflow, but do not provide the statistics necessary for characterizing random prestorm streamflows. Historically, the USGS has produced a series of annual runoff maps for the United States (Langbein, 1949; Busby, 1966; Gebert and others, 1987) and for selected regions of the United States (for example, Randall, 1996). The USGS also has produced maps defining estimates of runoff components (Wolock 2003a; b; c). As previously indicated, Driscoll and others (1990a; b) used streamflow estimates that were generalized from a national annual runoff map. The USEPA also has developed runoff-contour maps to assist planning-level analysis of water-quality data (Bishop and Church, 1995). The USGS maps provide contours of equal average annual runoff (total streamflow) in inches. Average-annual streamflow values are equivalent to the long-term average of mean-daily streamflows. Although the authors of such maps clearly indicate that these maps are not for the purpose of estimating streamflows at any particular location, the maps are the result of an expert interpretation of information on climate, geology, topography, and vegetation with streamflow data from thousands of streamgages (Langbein, 1949; Busby, 1966; Gebert and others, 1987; Krug and others, 1989; 1990). Rochelle and others (1989) indicate that such maps could be used to estimate long-term average annual streamflow with an uncertainty of about 15 percent. Busby (1966), however, indicates that the accuracy in any location is proportional to the spatial density of the gaging stations and inversely proportional to the spatial variability in runoff of a given area. Such maps can be used to help select index-site statistics based on hydrological similarity as well as geographic proximity. In areas with little data, such maps may be used to adjust statistics from nearby sites to account for differences in hydrologic features. For example, Bishop and Church (1995) used topographically interpolated precipitation maps (Daly and others, 1994) to adjust streamflow contours to reflect orographic effects. Therefore, detailed runoff maps may be useful for refining initial, regional, planning-level streamflow estimates or for selecting the most representative station(s) for a more focused planning-level streamflow estimate.

Regression on Basin Characteristics

Regression on basin characteristics is a method for using physical, hydrological, and meteorological characteristics of basins with streamflow data to develop multivariate regression models that are used to predict selected flow statistics at any location along a stream in the area for which equations are developed. Regression on basin characteristics is generally considered the best method for estimating selected streamflow statistics at sites for which there is no streamflow data (Thomas and Benson, 1970; Benson and Carter, 1973; Hirsch, 1979; Tasker and Stedinger, 1989; Martin and Ruhl, 1993; Jennings and others, 1994; Vogel and others, 1999; Ries and Friesz, 2000; Eng and others, 2005). Many local studies (for example, Tasker, 1972; Williams, and Pearson, 1985; Parrett and Cartier, 1990; Vogel and Kroll, 1992), state studies (for example, Koltun and Schwartz,1987; Sumioka and others, 1998; Ries and Friesz, 2000; Hortness and Berenbrock, 2001; Berenbrock, 2002; Koltun and

Whitehead, 2002; Ahearn, 2004; Stuckey, 2006) , regional studies (for example, Jennings and others, 1994; Lichty and Karlinger, 1995; Eng and others, 2005), and national studies (for example, Thomas and Benson 1970; Vogel and others, 1998; Kroll and others, 2004; Ries and Crouse, 2007) have used the regression on basin characteristics method to estimate various streamflow statistics.

Physical, hydrological, and geological information about the basins upstream of the streamgage are used to develop the multivariate regression equations to predict values of selected streamflow statistics at any site of interest within the study area. These estimates are made using the characteristics of the basin upstream of the site of interest. Drainage area (or contributing drainage area) is commonly the most significant variable for estimating streamflow statistics within a hydrologic region (Thomas and Benson, 1970; Benson and Carter, 1973; Hirsch, 1979; Tasker and Stedinger, 1989; Vogel and Kroll, 1992; Martin and Ruhl, 1993; Jennings and others, 1994; Vogel and others, 1999; Ries and Friesz, 2000; Eng and others, 2005). Other physical features that are commonly tested or used in regional regression equations include basin or channel slope, channel length, basin width, channel sinuosity (commonly quantified as channel length divided by basin length), surface storage (lakes, ponds, and wetlands), basin elevation, drainage density, soil type, land use, and geohydrological metrics (for example aquifer properties). Hydrological and meteorological characteristics that are commonly tested or used in regional regression equations are precipitation characteristics, temperature metrics, evapotranspiration, and seasonal snow-storage metrics. Drainage area (or contributing drainage area) commonly is used for estimating low flows, measures of central tendency, and high flows. Other metrics, however, commonly are different among the regression equations developed for estimating different flow regimes. For example, basin or channel slope drainage density, soil type, and land use commonly are used for flood flow equations, whereas factors like soil type, geohydrological metrics, temperature metrics, and evapotranspiration commonly are used to estimate low-flows. The probability of having zero flows also may be estimated by regression on basin characteristics. For example, Bent and Steeves (2006) developed and implemented a logistic regression equation to estimate the probability of a site flowing perennially in Massachusetts based on drainage area, soil, the percentage of forest land, and a binary variable that is used to identify different regions within the state.

U.S. Geological Survey has conducted many studies since the 1970s to define regression on basin characteristics and is currently (2009) in the process to develop a Streamstats application (U.S. Geological Survey, 2007b) in cooperation with various state agencies. This internet application is designed to facilitate application of the local regression on basin characteristics equations. These studies are commonly done on a state-by-state basis rather than regions of hydrologic similarity, but commonly include streamgages in neighboring states that are near the state boundaries. Although equations are developed on a state-by-state basis there may be multiple equations for different regions in highly heterogeneous states or there may be state-wide equations that include a regional factor. For example, Ries and Friesz (2000) include a binary factor to adjust equations for regions defined as the eastern or western part of Massachusetts (which roughly correspond to ecoregions 59 and 58, respectively). The Streamstats application is a Web-based geographic information system that allows the user to zoom in to any point along a defined stream reach in a given state and click on the reach to obtain upstream basin characteristics and estimates of streamflow statistics from application of predefined regression equations (Ries and others, 2004; U.S. Geological Survey, 2007b). The Streamstats application also provides uncertainty estimates for regression results.

Regression on basin characteristics is a powerful tool, but implementation of the method has limitations. If a Streamstats application is not available for the area of interest with estimates of the statistic of interest, the effort required to delineate the drainage basin and determine the values of the predictor variables in a way that is consistent with methods used to formulate the regression models can be substantial. Although the Streamstats application compiles the necessary information and provides easy access to existing regression equations, substantial effort is required to develop and implement new equations. Many States independently generate the GIS coverages that are used to develop these regression equations, leading to differences in methods for determining variables and differences in the map-resolution of explanatory variables from state to state. Many national scale coverages are too generalized for quantitative local characterization (Peter Steeves, GIS coordinator, USGS Streamstats program, written commun., 2010). Kroll and others (2004) describe the need for a national, standard, database of basin characteristics at a reasonable map scale for developing low-flow regression equations.

Streamflow estimates from regression equations may include substantial uncertainty (Thomas and Benson, 1970; Benson and Carter, 1973; Tasker and Stedinger, 1989; Vogel and Kroll, 1992; Martin and Ruhl, 1993; Jennings and others, 1994; Vogel and others, 1999; Ries and Friesz, 2000). Errors in predictions of low-flow statistics at different sites may range from 0 to 700 percent. Errors in predictions of central tendency statistics, such as annual average flows, commonly are in the range of 10 to 40 percent. Errors in predictions of high-flow statistics at different sites may range from 20 to 40 percent. It should be noted, however, that although percent errors are large for low-flow statistics, the magnitude of volumetric errors (in ft^3) may be orders of magnitude higher for flood-flow statistics than for low-flow statistics. Errors on the order of 40 percent are well within the data-quality objectives for planning-level streamflow estimates, but implementation of regression on basin characteristics methods are not well suited for stochastic estimation of random prestorm streamflow. This is because these regression studies are commonly focused on low flows, such as the 7-day 10-year low flow (7Q10), or high flows, such as the 2-year, 10-year, 50-year or 100-year flood. Even when these studies define multiple percentiles in the flow-duration curve, multiple regression equations for different flow percentiles would be difficult to use for a continuum of mean-daily streamflow values needed for stochastic data generation. It should be noted, however, that equations could be developed for the mean, standard deviation and skew of the logarithms of nonzero mean-daily streamflow data (for example, Castellarin and others, 2004).

Drainage-Area Ratios

The drainage-area ratio method is a technique for using available streamflow data to estimate streamflows at any location along a stream (Hirsch, 1979; Koltun and Schwartz, 1987; Stedinger and others, 1993; Thomas and others 1994; Sumioka and others, 1998; Ries and Friesz, 2000; U.S. Geological Survey, 2002a; Berenbrock, 2002; Parrett and Johnson 2004; Emerson and others, 2005; Hortness, 2006). Hirsch (1979) indicates that the advantage of the drainage-area ratio method is that it requires only a record of mean-daily streamflow data at an index station and a drainage area for the site of interest. However, the basin containing the index station must be hydrologically similar to the basin upstream of the site of interest for the successful application of the drainage-area ratio method. The assumption of hydrologic similarity is implicit in the application of the drainage-area ratio method because basin characteristics are not explicitly included in the predictive equation. The general equation for the drainage-area ration method is

$$Q_y = (A_y/A_x)^Z \times Q_x \qquad ,(4)$$

Where

Q_y is the streamflow at the site of interest;

Q_x is the streamflow at the index site;

A_y is the drainage area of the site of interest;

A_x is the drainage area of the index site; and

Z is an exponent to adjust for systematic differences in the drainage-area to flow ratio.

Stedinger and others (1993) indicate that the exponent is usually assumed to be one or may be determined by regression analysis with data from nearby index stations. The drainage-area ratio method may be a simple, robust method to derive planning-level prestorm streamflow estimates for sites of interest if **equation 4** is properly formulated with streamflow data from one or more hydrologically similar index sites.

The drainage-area ratio method has received mixed reviews in the scientific literature. In a national synthesis of annual runoff in the United States, Langbein (1949) noted that the size of drainage area in of itself does not affect the annual streamflow per unit area. Generally, within similar climatic areas, streamflows from small basins more strongly reflect detailed effects of geology and topography than large areas. In many areas of the country headwater streams tend to lose water as they enter valley-fill aquifer systems and gain water further down the valley. Small headwater areas, even in humid parts of the country, may be perched and therefore drained by ephemeral or intermittent streams which may flow only during storms during some seasons. Variation in annual runoff is primarily associated with climate, but also is affected by geology, topography, vegetation and other natural and anthropogenic factors. The influence of geology alone may account for halving or doubling of runoff (Langbein, 1949).

The degree of success in the application of the drainage-area ratio method in different studies is dependent upon the choice of an index station or stations that are hydrologically similar to the site of interest. For example, Hirsch (1979) compared results from the regression on basin characteristics, drainage area, and flow correlation methods. In that study, the results from the drainage-area ratio method (with a bias of about 15 percent of the historic average value) were comparable to results from regression on basin characteristics and streamflow correlation when the index site had a high degree of hydrologic similarity. However, Hirsch (1979) also demonstrated potential effects of the misapplication of the method (with a bias of about 30 percent of the historic average value) if the index site has different hydrologic characteristics than the site of interest.

Many studies use drainage area ratio methods to estimate streamflows and streamflow statistics at an ungaged location from a streamflow record collected from a hydrologically similar basin with a similar drainage area (Hirsch, 1979; Koltun and Schwartz, 1987; Stedinger and others, 1993; Ries and Friesz, 2000; U.S. Geological Survey, 2002a; Berenbrock, 2002; Parrett and Johnson 2004; Emerson and others, 2005; Hortness, 2006). Ries and Friesz (2000) compared regression on basin characteristics and the drainage area ratio method with MOVE.1 based flow correlation estimates for sites in Massachusetts. They found that drainage-area ratio estimates were generally more accurate than regression on basin characteristics estimates for hydrologically similar basins with drainage areas within the range of about 0.3 to 1.5 times the index-site drainage area. Similarly, the U.S. Geological Survey (2002a) documents the results of an analysis of sites in Pennsylvania that includes 92 comparisons to examine the relative accuracy of the drainage area ratio and regression methods. This analysis indicates that, for hydrologically similar basins, 82 percent of the drainage-area ratio estimates had absolute percent differences less than or equal to available regression estimates if drainage areas of the site of interest were within the range of about 0.3 to 3 times the times the index-site drainage area. Thomas and others (1994) used the drainage-area ratio method to estimate peak flood-flows at sites along streams with drainage area ratios between 50 and 150 percent of the gaged area. They found that the exponent Z (**eq. 4**) ranged from 0.4 in arid areas to 0.8 in the highlands in the southwestern United States. Similarly, Berenbrock (2002) examined the drainage-area ratio method for estimating peak-flow statistics in Idaho for sites with drainage-area ratios between 0.5 and 1.5 times the drainage area of the index site and found that this method provided acceptable results if the exponent Z (**eq. 4**) was adjusted for site characteristics. He found that the exponent Z in **equation 4** varied from 0.65 to 0.94 (with a median of 0.84) for 9 regions of hydrologic similarity. These exponents indicate higher flood flows per unit area in smaller basins and may reflect orographic effects in headwater basins. Thompson and Cavallo (2005) indicated that the drainage-area ratio method provided results that were consistent with the accuracy of regression equations for a variety of flow statistics at sites in Pennsylvania; except for some streams with reaches that were losing water to the underlying aquifer. Emerson and others (2005) successfully used regression equations to estimate drainage-area ratio parameters for two regions in North Dakota and Minnesota.

Asquith and others (2006) did an extensive evaluation of the drainage-area ratio method for estimating selected streamflow statistics at ungaged sites in Texas. They evaluated the drainage-area ratio method by pairing statistics from each of the 712 USGS streamgages located throughout the state of Texas with every other station within the selected data set. They found substantial variation in the drainage-area ratio exponents. Asquith and others (2006), however, did not apply hydrologic similarity to constrain the station-pairs used to evaluate the drainage-area ratio method in this study. For example, the study area for this study (Asquith and others, 2006) includes 12 of 84 ecoregions (U.S. Environmental Protection Agency, 2003; 2004b) and 5 of 9 rain zones (Palecki and others, 2005) defined within conterminous United States. Similarly, Omernik (2003) describes the hydrologic dissimilarities of the different ecoregions within the different water-resource regions in Texas. Ecoregions are defined by physiography, geology, soils, climate, land use and hydrology; all factors that determine hydrologic similarity. Furthermore, implicit regionalization indicated by the distance between sites may have a more substantial effect on streamflow characteristics than some basin characteristics that may be used to assess hydrologic similarity between basins (Merz and Bloschl, 2004). Asquith and others (2006) did note improvements in drainage-area ratio predictions for streamgages that were within 100 miles of each other. Much of the remaining variability may be caused by a lack of hydrologic similarity for nearby streamgages that are in different ecoregions or in different rain zones, or the effect of large distances on precipitation patterns.

Results of regression on basin characteristics studies also indicate the potential utility of the drainage-area ratio method for planning-level estimates of streamflow statistics. Many studies that focus on the regression on basin characteristics method indicate that drainage area (or contributing drainage area) is commonly the most significant variable for estimating streamflow statistics within a hydrologic region (Thomas and Benson, 1970; Benson and Carter, 1973; Tasker and Stedinger, 1989; Martin and Ruhl, 1993; Vogel and others, 1999; Ries and Friesz, 2000). For example, Vogel and others (1999) formulated regression models for average annual streamflows and the variability in average annual streamflows and found that drainage area is the most significant predictor (with R-square values from 27 to more than 99 percent, with a median R-square value of about 70 percent) for these flow statistics for all 18 water-resource regions in the conterminous United States. The 8 water-resource regions with the highest degree of internal hydrologic similarity had r-square values that exceed 88 percent. The region with the lowest R-square value for the drainage-area regression-model includes hydrologically dissimilar areas from the Rocky Mountains, through the Great Plains, to the Arkansas Valley. Lins (1997) indicates that the water-resource regions, which are based on drainage divides of major river basins (or coastal areas) are not areas of hydrologic similarity. Koltun and Whitehead (2002) formulated regression equations for various flow statistics in Ohio and found that exponents for drainage area were approximately 1, and that the addition of other basin characteristics improved prediction errors, but only by (a median) of about 8 percent. Simon and others (2004) examined flood-flows to estimate sediment-transport by ecoregions and determined that drainage-area is the most significant variable for estimating these streamflows. Emerson and others (2005) examined regression equations for streamflow prediction in North Dakota and Minnesota. They found that drainage area is the most significant explanatory variable in this area.

The occurrence of zero streamflows may be a concern for streamflow statistics in some stream basins. Information about the occurrence and statistical treatment of zero streamflows is not prevalent in the published literature. Bent and Steeves (2006) indicate that, in Massachusetts, the estimated flow status of a stream site (intermittent or perennial) depends on drainage area, surficial geology, percent of forest cover, and streamflow region. The regions described by Bent and Archfield (2002) and Bent and Steeves (2006) for estimating the flow status of a stream site were selected by characteristics of streamflow statistics, but generally correspond to the delineation of ecoregions 58 (Northeastern Highlands), 59 (Northeastern Coastal zone), and 84 (Atlantic Coastal Pine Barrens) in Massachusetts. Bent and Steeves (2006) indicate that there is a high probability that streams throughout Massachusetts with drainage areas greater than about 2 square miles will be perennial. However, a study from Wyoming to estimate yields of dissolved constituents from intermittent and ephemeral streams included sites with drainage areas of up to 836 square miles that were dry about 30 percent of the time (Delong and Wells, 1988). Similarly, in arid areas ephemeral stream sites may have drainage areas as large as 150 to more than 200 square miles (Hejl, 1980; Coes and Pool, 2005). Hydrologic studies in Kansas (Studley, 2000; 2001), indicate that relations between the percentage duration of appreciable flow (defined as 0.1 ft^3) is a linear function of basin drainage area in log-log space in the range from 10 to 3,000 square miles.

Estimating Streamflow Statistics at Sites with Limited Data Using Streamflow Correlation

Streamflow correlation is generally considered to be a robust method for estimating selected streamflow statistics at sites of interest for which there is limited streamflow data (Matalas and Jacobs, 1964; Hirsch, 1979, 1982; Vogel and Stedinger, 1985; Ries and Friesz, 2000). Streamflow correlation is done by developing regression equations with data from a long-term continuous streamgage and the site of interest. Streamflow correlation is used for record augmentation and record extension. Streamflow record augmentation is commonly defined as a method for generating long-term estimates of statistics such as the mean and variance of streamflows at short record sites (Matalas and Jacobs, 1964; Vogel and Stedinger, 1985). Record extension is defined as a method for estimating individual discharge values at the short-record site from concurrent mean-daily streamflow measurements at one or more longer-record index site(s) (Hirsch, 1979; 1982; Hirsch and Gilroy, 1984; Vogel and Stedinger, 1985). Record extension also may be done to estimate mean-daily streamflow values to fill a gap in the record at the site of interest. The theory supporting record augmentation and record extension is that the record of streamflow from an index (or predictor) station (the long-term continuous streamflow gaging station) may be used as surrogate for streamflow information at the site of interest and that the drainage basins upstream of the short and long-term record sites are hydrologically similar.

There are several methods that are accepted for record extension and augmentation. Hirsch (1982) demonstrated that the regression method commonly referred to as the line of organic correlation (Helsel and Hirsch 2002), is a method suitable for record extension with maintenance of variance. Hirsch and Gilroy (1984), and Helsel and Hirsch (2002) discuss the relative merits of regression equations developed with the ordinary least squares, the line of organic correlation, and least normal squares methods, indicating that the line of organic correlation is the best method of the three for record augmentation and record extension. Vogel and Stedinger (1985) provide methods to adjust the record extension equations for differences in the mean and variance of streamflows in the short-term concurrent period and the long-term streamflow record. Vogel and Stedinger (1985) designated the adjusted method as MOVE.3. Similar methods designated as MOVE.2 (Hirsch, 1982) and MOVE.4 (Vogel and Stedinger, 1985) have been developed but are more complex than MOVE.1 and MOVE.3, respectively and do not offer significantly greater predictive capability. As with other regression-based methods for estimating streamflows of interest, statistics from streamflow correlation methods can be used to estimate the uncertainties in the predictions. Ries and Friesz (2000) provide a detailed description of this type of uncertainty analysis for results from the MOVE.1 method.

Record augmentation or extension at a site of interest must be done with an appreciation for the potential effects of outliers. If partial record stations are used, potential differences between instantaneous measurements and the mean-daily flow available for historical records must be examined. Record augmentation or extension of partial-record data is commonly used to estimate low-flow statistics, so partial-record data is commonly collected only during times of stable-streamflow recession (Riggs, 1968, 1972; Ries and Friesz, 2000). Measurement of streamflows over a wide range of flows is necessary for extending a complete record. Collection of such a record, however, may be difficult without multiple instantaneous discharge measurements on days with high, rapidly changing streamflows. Record extension for short-term continuous streamgage records is more practicable because the daily flows are more comparable between stations than are instantaneous measurements. Also, proper development of a stage-discharge relation for such a station reduces the potential for poor measurements in the data set. This is because each measurement can be compared to the stage-discharge relation to determine if an individual measurement is a potential outlier. Such a comparison is used to evaluate a measurement in the field and make repeated measurements to verify a potential change in the stage-discharge relation (Rantz, 1982).

Requirements for selecting one or more index stations to be used for streamflow correlation depend on the objectives of the analysis. Hydrogeologic similarity, relative basin size, and the proportion of consumptive water use between the index station and the site of interest are important considerations for both record extension and augmentation because some statistical characteristics (such as higher moments of the index-streamflow population) are transferred by record extension/augmentation techniques (Timothy Cohn, U.S. Geological Survey, Office of Surface Water, written commun., 2006). In general, index-station drainage-basins should be in close proximity to the site of interest, have similar physiographic and hydrogeologic characteristics, and have similar drainage areas. For both augmentation and extension, the analyst must consider regression statistics (from the line of organic correlation) and visual examination of the relation between concurrent measurements to ensure that the relation is linear and residuals are approximately normal and homoscedastic. If the objective is record augmentation, to adjust short-term statistics at one site so they represent long-term statistics, then the selection of one or more index stations may be more heavily weighted on regression statistics than proximity to the site of interest to get the best possible estimates of streamflow statistics. In this case, the proximity and relative size of drainage areas of the site of interest and the index station or stations is not the most critical consideration. If, however, the objective is record extension, to generate a sequence of daily, weekly, monthly, or annual flow series that represent a time-series of streamflow data with the correct flow statistics and serial correlation of streamflows, then consideration of proximity and basin size may be critical. Proximity is important for extending daily flow records to reflect the effects of individual precipitation events (or lack thereof) from one basin to the next.

For record extension, drainage basin size is important to represent the watershed response to a given precipitation event. The tendency for high flows to follow high flows and low flows to follow low flows in a streamflow data set, commonly called serial correlation, is considered an important feature to preserve in streamflow extension techniques (Hirsch, 1982). For example, a commonly used rule-of-thumb for hydrograph separation is that that the number of days required for a basin to return to base flow after a storm peak is the drainage area (in square miles) raised to the power 0.2 (Linsley and others, 1975; Sloto and Crouse, 1996). Therefore, one may expect stormflows to persist, once the stormflow has peaked, for about 1 day for a 1 mi^2 basin, 2 days for a 32 mi^2 basin, and 3 days for a 243 mi^2 basin. Similarly, the time to the hydrograph peak, commonly expressed as a function of stream length and distance to the upstream divide, is related to basin size (Chow and others, 1988). Basin size may be less critical for record augmentation than for extension, because the regression equation coefficients should account for differences in responses of the mean and variance of streamflows between basins and the MOVE equation is not being used to estimate a streamflow for a particular day.

Hydrologists may want to use two or more index stations for record extension or record augmentation because different streamflow sites rarely are similar in all physiographic, hydrogeologic, land-cover, land-use, and water-use and features. Use of multiple stations is done in an attempt to improve individual estimates and compensate for perceived limitations in hydrologic similarity or the relative proximity of the one index station with the best regression relation. For example, Ries and Friesz (2000) use MOVE.1 regression statistics to derive a weighted-average estimated streamflow value at the site of interest from individual MOVE.1 prediction values at different stations. Use of multiple index stations, with similarly high regression statistics, is commonly done for record augmentation in an attempt to generate estimates that are representative of streamflow characteristics in adjacent areas. Estimates with multiple stations are commonly weighted by the strength of regression statistics. Although the effect of random differences in spatial and temporal patterns in precipitation on the relative timing and magnitude of streamflow variation from day to day will tend to equal out over long periods of time (for record augmentation), it is thought that use of multiple stations for estimating individual streamflow values may better reflect the occurrence of individual storm events (for record extension). The magnitude of local storm events, however, may be muted by including estimates from neighboring index stations that did not receive precipitation during an individual event.

References Cited

Abramowitz, M., and Stegun, I.A. , 1965, Handbook of mathematical functions, Applied Mathematical Series Number 55, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., 1046 p.

Ahearn, E.A., 2004, Regression equations for estimating flood flows for the 2-, 10-, 25-, 50-, 100-, and 500-year recurrence intervals in Connecticut: U.S. Geological Survey Scientific Investigations Report 2004-5160, 62 p.

Ashkar, F., Ouarda, T.B.M.J., 1996, On some methods of fitting the generalized Pareto distribution: Journal of Hydrology: v. 177, no. 1-2, p. 117-141.

Asquith, W.H., Roussel, M.C., Cleveland,T.G., Fang, Xing, and Thompson, D.B., 2006, Statistical characteristics of storm interevent time, depth, and duration for eastern New Mexico, Oklahoma, and Texas: U.S. Geological Survey Professional Paper 1725, 299 p.

Asquith,W.H., Vrabel, Joseph, and Roussel, M.C., 2007, Summary of annual mean, maximum, minimum, and L-scale statistics of daily mean streamflow for 712 U.S. Geological Survey streamflow-gaging stations in Texas through 2003: U.S. Geological Survey Data Series 248, 721 p.

Beasley J.D., and Springer S.G., 1977, Algorithm AS 111--Percentage points of the normal distribution: Applied Statistics, v. 26, no. 1, p. 118-121.

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

Bent, G.C., and Archfield, S.A., 2002, A logistic regression equation for estimating the probability of a stream flowing perennially in Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 2002-4043, 45 p.

Bent, G.C., and Steeves, P.A., 2006, A revised logistic regression equation and an automated procedure for mapping the probability of a stream flowing perennially in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2006-5031, 107 p.

Berenbrock, Charles, 2002, Estimating the Magnitude of Peak Flows at Selected Recurrence Intervals for Streams in Idaho: U.S. Geological Survey Water Resources-Investigations Report 02-4170, 59 p.

Bishop, G.D., and Church M.R., 1995, Mapping long-term regional runoff in the eastern United States using automated approaches: Journal of Hydrology, v. 169, p. 189-207.

Blackwood, L.G., 1992, The lognormal distribution, environmental data, and radiological monitoring: Environmental Monitoring and Assessment, no. 21, p. 193-210.

Bobee, Bernard, 1975, Log Pearson type 3 distribution and its application in hydrology: Water Resources Research, v. 11, no. 5, p. 681-689.

Bobee, Bernard, Ashkar, Fahim, 1991, The gamma family and derived distributions applied in hydrology: Littleton, CO, Water Resources Publications, 203 p.

Busby, M.W., 1966, Annual runoff in the conterminous United States, 1951-80: U.S. Geological Survey Hydrologic Investigations Atlas HA-212, 1 pl.

Castellarin, A.R., Vogel R.M., and Brath, Armando, 2004, A stochastic index flow model of flow duration curves, Water Resources Research, v. 40, no. 3, p. w0310401-w0310410.

Cheng, K.S, Ciang, J.L., Hsu, C.W., 2007, Simulation of probability distributions commonly used in hydrological frequency analysis: Hydrological Processes, v. 51, p 51-60.

Chow, V.T., 1954, The log-probability law and its engineering applications: Proceedings of the American Society of Civil Engineers: v. 80 Separate No. 536, 25 p.

Chow, V.T., Maidment, D.R., and Mays, L.W., 1988: Applied Hydrology: New York, McGraw-Hill Inc., 572 p.

Cigizoglu, H.K., 2000, A method based on taking the average of probabilities to compute the flow duration curve: Nordic Hydrology v. 31, no. 3, p. 187-206.

Coes, A.L., and Pool, D.R., 2005, Ephemeral-stream channel and basin-floor infiltration and recharge in the Sierra Vista subwatershed of the Upper San Pedro basin, southeastern Arizona: U.S. Geological Survey Open-File Report 2005-1023, 84 p.

Cohn, T.A., and Lins, H.F., 2005, Nature’s style--Naturally trendy: Geophysical Research Letters, v. 32, L23402, doi:10.1029/2005GL024476.

Croker, K.M., Young, A.R., Zaidman, M.D., and Rees, H.G., 2003, Flow duration curve estimation in ephemeral catchments in Portugal: Hydrological Sciences Journal, v. 48, no. 3, p. 427-439.

Daly, Christopher, Neilson, R.P., Phillips, D.L., 1994, A Statistical-topographic model for mapping climatological precipitation over mountainous terrain: Journal of Applied Meteorology, v.33, no. 2, p. 140-158.

DeLong L.L., Wells, D.K., 1988, Estimating average dissolved-solids yield from basins drained by ephemeral and intermittent streams—Green River Basin, Wyoming: U.S. Geological Survey Water Resources-Investigations Report 87-4222, 14 p.

Di Toro, 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., 1990a, Pollutant loadings and impacts from highway stormwater runoff, Volume I, Design procedure: U.S. Federal Highway Administration Final Report, FHWA-RD-88-006, 67 p.

Driscoll, E.D., Shelley, P.E., and Strecker, E.W., 1990b, Pollutant loadings and impacts from highway stormwater runoff, Volume III--Analytical investigation and research report: Federal Highway Administration Final Report FHWA-RD-88-008, 160 p.

Emerson, D.G., Vecchia, A.D., and Dahl, A.L., 2005, Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota: U.S. Geological Survey Scientific Investigations Report 2005-5017, 13 p.

Eng, Ken, Tasker, G.D., and Milly, P.C.D., 2005, An analysis of region-of-influence methods: Journal of the American Water Resources Association, v. 41, no. 1, p. 135-143.

Gebert, W.A., Graczyk, D.J., and Krug, W.R., 1987, Average annual runoff map of the United States, 1951-80: U.S. Geological Survey Hydrologic Investigations Atlas HA-710, 1 pl.

Granato, 2009, Computer programs for obtaining and analyzing daily-mean streamflow data from the U.S. Geological Survey National Water Information System Web: U.S. Geological Survey Open-File report 2008-1362, CD-ROM.

Haan, C.T., 1977, Statistical methods in hydrology: Ames, Iowa, 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.

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.

Hejl, H.R., Jr., 1980, Preliminary appraisal of ephemeral-streamflow characteristics as related to drainage area, active-channel width, and soils in northwestern New Mexico: U.S. Geological Survey Open-File Report Number 81-64, 19 p.

Hirsch, R.M., 1979, An evaluation of some record extension techniques: Water Resources Research, v. 15, no. 6, p. 1781-1790.

Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: Water Resources Research, v. 18, no. 4., p. 1081-1088.

Hirsch, R.M., and Gilroy, E.J., 1984, Methods of fitting a straight line to data—examples in water resources: Water Resources Bulletin, v. 20, no. 5, p. 705-711.

Hodgkins, G.A., Dudley, R.W., and Huntington, T.G., 2003, Changes in the timing of high river flows in New England over the 20th Century: Journal of Hydrology, v. 278, p. 244–252.

Hortness, J.E., 2006, Estimating low-flow frequency statistics for unregulated streams in Idaho: U.S. Geological Survey Scientific Investigations Report 2006-5035, 40 p.

Hortness, J. E., and Berenbrock, Charles, 2001, Estimating monthly and annual streamflow statistics at ungaged sites in Idaho: U.S. Geological Survey Water Resources-Investigations Report 01-4093, 36 p.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood-flow frequency: Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, Va., 183 p., available at: http://water.usgs.gov/osw/bulletin17b/bulletin_17B.html.

Jennings, M.E., Thomas, W.O., Jr., Riggs, H.C., 1994, Nationwide summary of US Geological Survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites, 1993: U.S. Geological Survey Water-Resources Investigations Report 94-4002, 196 p. with 3 1/2 in. computer disk

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.

Kite, G.W., 1977, Frequency and risk analysis in hydrology: Fort Collins, CO, Water Resources Publications 257 p.

Koltun, G.F., and Schwartz, R.R., 1987, Multiple-regression equations for estimating low flows at ungaged stream sites in Ohio: Water-Resources Investigations Report 86-4354, 39 p.

Koltun, G.F., and Whitehead, M.T., 2002, Techniques for estimating selected streamflow characteristics of rural unregulated streams in Ohio: Water-Resources Investigations Report: 2002-4068, 50 p.

Kroll, Chuck, Luz, Joana, Allen, Brad, and Vogel, R.M., 2004, Developing a watershed characteristics database to improve low streamflow prediction: Journal of Hydrologic Engineering, V. 9, no. 2, p. 116-125.

Krug, W.R., Gebert, W.A., and Graczyk, D.J., 1989, Preparation of average annual runoff map of the United States, 1951-80: U.S. Geological Survey Open-File Report 87-535, 414 p.

Krug, W. R., Gebert, W. A., Graczyk, D. J., Stevens, D. L., Jr., Rochelle, B. P., Church, M. R., 1990, Map of mean annual runoff for the Northeastern, Southeastern, and Mid-Atlantic United States, water years 1951-80: U.S. Geological Survey Water-Resources Investigations Report 88-4094, 11 p.

Langbein, W.B. (ed.), 1949, Annual runoff in the United States: U.S. Geological Survey Circular 52, 14 p., 1 pl.

Leopold, L.B., Wolman, M.G., and Miller, J.P., 1992, Fluvial processes in geomorphology: New York, Dover Publications Inc., 522 p.

Lichty, R.W., and Karlinger, M.R., 1995, Spatial trends in Pearson type III statistical parameters: Journal of Hydraulic Engineering v. 121, no. 9, p. 672-678.

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.

Lins, H.F., 1997, Regional streamflow regimes and hydroclimatology of the United States: Water Resources Research, v. 33, no. 7 p. 1655-1667.

Lins, H.F., and Michaels, P.J., 1994, Increasing U.S. streamflow linked to greenhouse forcing: EOS, Transactions, American Geophysical Union, v. 75, no. 25, p. 281, 284-285.

Lins, H.F., and Slack, J.R., 1999, Streamflow trends in the United States: Geophysical Research Letters, v. 26, no.2, p. 227-230.

Linsley, R.K., Jr., Kohler, M.A., and Paulhus, J.L.H., 1975, Hydrology for engineers 2nd Ed.: New York, McGraw-Hill Book Company, 482 p.

Martin, G.R., and Ruhl, K.J., 1993, Regionalization of harmonic-mean streamflows in Kentucky: U.S. Geological Survey Water-Resources Investigations Report 92-4173, 47 p.

Matalas, N.C., and Jacobs, Barbara, 1964, A correlation procedure for augmenting hydrologic data: U.S. Geological Survey Professional Paper 434-E, 7 p.

McCabe, G.J., and Wolock, D.M., 2002, A step increase in streamflow in the conterminous United States: Geophysical Research Letters, v. 29, no. 24, 2185, doi:10.1029/2002GL015999

McCuen, R.H., Johnson, P.A., and Ragan, R.M., 2002, Highway hydrology—Hydraulic design series number 2, Second Edition: Washington D.C., Federal Highway Administration Report FHWA-NHI-02-001 HDS-2, 424 p.

Merz Ralf, and Bloschl, Gunter, 2004, Regionalisation of catchment model parameters: journal of hydrology, v. 287, p. 95-123.

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, National Academies Press, 164 p.

Natural Resources Conservation Service, 1998, Tables of percentage points of the Pearson type III distribution: United States Department of Agriculture, Natural Resources Conservation Service Technical Release 38 (TR–38), 17 p.

Novotny, Vladimir, 2004, Simplified data based total maximum daily loads, or the world is log-normal: Journal of Environmental Engineering, v. 130, no. 6, p. 674-683.

Odeh, R. E., and Evans, J. O., 1974, Algorithm AS 70--Percentage points of the normal distribution: Applied Statistics v. 23, no. 1, p. 96-97.

Omernik, J.M., 2003, The misuse of hydrologic unit maps for extrapolation, reporting, and ecosystem management: Journal of the American Water Resources Association, v. 39, no. 3, p. 563-573.

Palecki, M.A., Angel, J.R., and Hollinger, S.E., 2005, Storm precipitation in the United States--Part I--Meteorological characteristics: Journal of Applied Meteorology, v. 44, no. 6, p. 933-946.

Parrett, Charles, and Cartier, K.D., 1990, Methods for estimating monthly streamflow characteristics at ungaged sites in western Montana: U.S. Geological Survey Water-Supply Paper 2365, 30 p.

Parrett, Charles; Johnson, D.R., 2004, Methods for estimating flood frequency in Montana based on data through water year 1998: U.S. Geological Survey Water-Resources Investigations Report 2003-4308, 101 p.

Quader, A., and Guo, Y., 2005, Drainage area and flow frequency distribution pattern of small Ontario streams, *in* Proceedings of the 33rd Annual General Conference of the Canadian Society for Civil Engineering Toronto, Ontario, Canada June 2-4, 2005, p. GC.337.1-GC.337.10.

Randall, A.D., 1996, Mean annual runoff, precipitation, and evapotranspiration in the glaciated Northeastern United States, 1951-80: U.S. Geological Survey Open-File Report 96-395, 2 pl. available at: *http://ny.water.usgs.gov/pubs/of/of96395/OF96-395.html*

Rantz, S.E. (ed.), 1982, Measurement and computation of streamflow: Volume 1 and 2, Computation of Discharge: U.S. Geological Survey Water Supply Paper 2175, 631 p.

Rao, D.V., 1980, Log Pearson type 3 distribution—A generalized evaluation: American Society of Engineers, v. 106, no. HY5, p. 853-872.

Ries, K.G., III, and Crouse, M.Y., 2007, NFF, Version 3.2--National Flood Frequency program: accessed July 16, 2007 at URL http://water.usgs.gov/software/nff.html

Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 2000-4135, 81 p.

Ries, K.G., III, Steeves, P.A., Coles, J.D., Rea, A.H., and Stewart, D.W., 2004, StreamStats--A U.S. Geological Survey Web application for stream information: U.S. Geological Survey Fact Sheet FS 2004-3115, 4 p. available at: http://md.water.usgs.gov/publications/fs-2004-3115/html/index.html

Riggs, H. C., 1968, Some statistical tools in hydrology: U.S. Geological Survey Techniques of Water-Resource Investigation Book 4, Chap. A1, 39 p.

Riggs, H.C., 1972, Low-flow investigations: U.S. Geological Survey Techniques of Water-Resource Investigation Report Book 4, Chap. B1, 18 p.

Rochelle, B.P, Stevens, D.L., Jr., Church, M.R., 1989, Uncertainty analysis of runoff estimates from a contour map: Water Resources Bulletin v. 25, no. 3 p. 491-498.

Rossman, L.A., 1990b, DFLOW User's Manual: Risk Reduction Engineering Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, 26 p.

Salas, J.D., 1993, Chapter 19--Analysis and modeling of hydrologic time series, *in* Maidment, D.R., ed., Handbook of hydrology: New York, McGraw Hill, p. 19.1-19.72.

Sangal B.P., and Biswas, A.K., 1970, The 3-parameter lognormal distribution and its applications in hydrology: Water Resources Research, v. 6, no. 2, p. 505-515.

Schwartz, S.S., and Naiman, D.Q., 1999, Bias and variance of planning level estimates of pollutant loads: Water Resources Research: v. 35, no. 11, p. 3475-3487.

Simon, Andrew, Dickerson Wendy, and Heins, Amanda, 2004, Suspended sediment transport rates at the 1.5-year recurrence interval for ecoregions of the United States--Transport conditions at the bankfull and effective discharge?: Geomorphology, v. 58, p. 243-262.

Sloto R.A., and Crouse, M.Y., 1996, HYSEP: a computer program for streamflow hydrograph separation and analysis: U.S. Geological Survey Water-Resources Investigations Report 96-4040, 46 p.

Stedinger, J.R., 1980, Fitting log normal distributions to hydrologic data: Water Resources Research, v. 16, no. 3, p. 481-490.

Stedinger, J.R., Vogel, R.M., and Foufouls-Georgiou, Efi, 1993, Chapter 18--Frequency analysis of extreme events; in Maidment, D.R., ed., Handbook of Hydrology: New York, McGraw-Hill Inc., p. 18.1-18.66.

Stuckey, M.H., 2006, Low-flow, base-flow, and mean-flow regression equations for Pennsylvania streams: U.S. Geological Survey Scientific Investigations Report 2006-5130, 84 p.

Studley, S.E., 2000, Estimated flow-duration curves for selected ungaged sites in the Cimarron and lower Arkansas River basins in Kansas: U.S. Geological Survey Water-Resources Investigations Report 2000-4113, 43 p.

Studley, S.E., 2001, Estimated flow-duration curves for selected ungaged sites in Kansas: U.S. Geological Survey Water-Resources Investigations Report 2001-4142, 90 p.

Sumioka, S.S., Kresch, D.L., and Kasnick, K.D., 1998, Magnitude and frequency of floods in Washington: U.S. Geological Survey Water-Resources Investigations Report 97-4277, 91 p.

Tasker, G.D., 1972, Estimating low-flow characteristics of streams in southeastern Massachusetts from maps of ground water availability: U.S. Geological Survey Professional Paper 800D p. D217-D220.

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, no. 1-4, p. 361-375.

Thomas, B.E., Hjalmarson, H.W., and Waltemeyer, S.D., 1994, Methods for estimating magnitude and frequency of floods in the southwestern United States: U.S. Geological Survey Water Supply Paper 2433, 195 p.

Thomas, D.M., and Benson, M.A., 1970, Generalization of streamflow characteristics from drainage-basin characteristics: U.S. Geological Survey Water Supply Paper 1975, 55 p.

Thompson, R.E., and Cavallo, G.J., 2005, Streamflow statistics for the Paradise and Pocono Creek Watersheds and selected streamflow-gaging station in Monroe County, Pennsylvania: U.S. Geological Survey Scientific Investigations Report 2005-5162, 21 p.

U.S. Army Corps of Engineers, 1993, Engineering and design manual--Hydrologic frequency analysis: Washington, DC: Department of the Army, Engineer Manual EM 1110-2-1415, 149 p.

U.S. Army Corps of Engineers, 1994, Engineering and design manual-- Flood run-off analysis: Washington, DC: Department of the Army, Engineer Manual EM 1110-2-1417, 213 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, D.C., U.S. Environmental Protection Agency: Report EPA 440/4-86-014, 64 p.

U.S. Environmental Protection Agency, 2003, Level III ecoregions of the continental United States: Washington, D.C., U.S. Environmental Protection Agency National Health and Environmental Effects Research Laboratory, 1 pl.

U.S. Environmental Protection Agency, 2004a, Flow 101-- Basic information provides information on the development and uses of DFLOW and the differences between DFLOW 3.0 and earlier versions: accessed on November 24, 2004 at: http://epa.gov/waterscience/models/dflow/flow101.htm

U.S. Environmental Protection Agency, 2004b, Level III ecoregions: Corvallis, OR, U.S. Environmental Protection Agency, Western Ecology Division, accessed on June 28, 2005, available at: http://www.epa.gov/wed/pages/ecoregions/level_iii.htm

U.S. Geological Survey, 2002a, Computing low-flow statistics for ungaged locations on Pennsylvania Streams by use of drainage-area ratios: U.S. Geological Survey, New Cumberland Pennsylvania, accessed November 24, 2006 at URL http://pa.water.usgs.gov/pc38/flowstats/.

U.S. Geological Survey, 2007b, Welcome to StreamStats: accessed July 16, 2007 at URL http://water.usgs.gov/osw/streamstats/index.html

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. and Fennessey, N.M., 1993, L-Moment diagrams should replace product-moment Diagrams: Water Resources Research, v. 29, no. 6, p 1745-1752.

Vogel, R.M., and Kroll, C.N., 1992, Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics: Water Resources Research, v. 28, no. 9, p. 2451-2458.

Vogel, R.M. and McMartin, D.E., 1991, Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type III 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. and Stedinger, J.R., 1985, Minimum variance streamflow record augmentation procedures: Water Resources Research, v. 21, no. 5, p. 715-723.

Vogel, R.M., Tsai, Y., and Limbrunner, J.F., 1998, The regional persistence and variability of annual streamflow in the United States: Water Resources Research, v. 34, no. 12, p. 3445-3459.

Vogel, R.M., and Wilson, Ian, 1996, The probability distribution of annual maximum, minimum and average streamflow in the United States, Journal of Hydrologic Engineering, ASCE, v. 1, no. 2, p. 69-76.

Vogel, R.M., Wilson, Ian, and Daly, Chris, 1999, Regional regression models of annual streamflow for the United States: Journal of Irrigation and Drainage Engineering, v. 125, no. 3, p. 148-157.

Wichura, M.J., 1988, Algorithm AS 241--Percentage points of the normal distribution: Applied Statistics v. 37, no. 3, p. 477-484.

Williams, J.R., and Pearson, H.E., 1985, Streamflow statistics and drainage-basin characteristics for the Southwestern and Eastern Regions, Washington: U.S. Geological Survey Open-File Report 84-145-A, 424 p.

Wilson, E.B., and Hilferty, M.M., 1931, The distribution of the Chi-square: Proceedings of the National Academy of Sciences, v. 17, no. 2, p. 684-688.

Wolock, D.M., 2003a, Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 2003-263, digital data available at: http://water.usgs.gov/lookup/getspatial?bfi48grd

Wolock, D.M., 2003b, Infiltration-excess overland flow estimated by TOPMODEL for the conterminous United States: U.S. Geological Survey Open-File Report 2003-310, digital data available at: http://water.usgs.gov/lookup/getspatial?ieof48

Wolock, D.M., 2003c, Saturation overland flow estimated from TOPMODEL for the conterminous United States: U.S. Geological Survey Open-File Report 2003-264, digital data available at::://water.usgs.gov/lookup/getspatial?satof48