US Army Corps
of Engineers
Hydrologic Engineering Center

Delaware River Basin Regional Skew Analysis

September 2009

Approved for Public Release. Distribution Unlimited.

PR-70

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
September 2009

2. REPORT TYPE
Project Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Delaware River Basin Regional Skew Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David Goldman/Mike Konieczki
David Ford Consulting Engineers Inc.
2015 J Street, Suite 200
Sacramento, CA 95811

Beth Faber, CEIWR-HEC-WRS

5d. PROJECT NUMBER

5e. TASK NUMBER

5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center (HEC)
609 Second Street
Davis, CA 95616-4687

8. PERFORMING ORGANIZATION REPORT NUMBER
PR-70

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
Philadelphia District
Wanamaker Building, 100 Penn Square East
Philadelphia, PA 19107-3391

10. SPONSOR/ MONITOR'S ACRONYM(S)

11. SPONSOR/ MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The US Army Corps of Engineers (USACE), Philadelphia District (NAP), has engaged the Hydrologic Engineering Center (HEC) in its study of the Delaware River Basin. The study, entitled "*Multi-jurisdictional Use and Management of Water Resources for the Delaware River Basin, NY, PA, NJ, and DE*", included a statistical analysis of annual peak flows along the mainstem Delaware River following procedures outlined in "*Guidelines for Determining Flood Flow Frequency, Bulletin 17B*" (Interagency Advisory Committee on Water Data (IACWD), 1982) done jointly by the US Geological Survey (USGS) and NAP. The purpose of this report is to recommend, and describe the methods for estimating, regional skew values required by *Bulletin 17B* in peak flow frequency analysis.

15. SUBJECT TERMS

Delaware River Basin, skew, analysis, regional, Philadelphia District, study, statistical, peak flows, Bulletin 17B, frequency curves, flow, quantiles, Log Pearson III, distribution, parameters, annual, methodology, procedures, results, streamgage, sites, calculate, multiple, time-base, sampling, error, spatial, weighted, isoline map, predictive equations, GLS, L-moment, low outliers, historical, systematic record, gages, regression, average skew

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
U	U	U	UU	144	**19b. TELEPHONE NUMBER**

Delaware River Basin Regional Skew Analysis

September 2009

Prepared for:
US Army Corps of Engineers
Philadelphia District
Wanamaker Building
100 Penn Square East
Philadelphia, PA 19107-33910

Prepared by:
David Ford Consulting Engineers, Inc.
2015 J Street, Suite 200
Sacramento, CA 95811

US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center
609 Second Street
Davis, CA 95616

(530) 756-1104
(530) 756-8250 FAX
www.hec.usace.army.mil

PR-70

Table of Contents

Chapters

1 **Study Overview**

2 **Methods and Procedures**

3 **Analysis and Results**

Table of Contents
(continued)

Chapters

Appendices

Table of Contents
(continued)

Appendices

List of Tables

Number

List of Tables
(continued)

Number

List of Figures

Number

Abbreviations

ArcGIS® – ESRI GIS and Mapping Software

AVP – average variance of prediction

CV - coefficient of variation

Elev - mean basin elevation

GIS – geographic information systems

GLS – Generalized Least Squares

H(1), H(2), H(3) - heterogeneity statistics

HEC – Hydrologic Engineering Center

HEC-FFA – Flood Frequency Analysis

HUC – hydrologic unit code

IACWD – Interagency Advisory Committee on Water Data

IDW – inverse distance weighting

L - L-moment

MAP – mean annual precipitation

MSE – mean square error

NAP – Philadelphia District

NCRS - Natural Resource Conservation Service

OLS – ordinary least squares

SCS - Soil Conservation Service (Natural Resource Conservation Service)

ScSS - Natural Resource Conservation Service (formerly Soil Conservation Service) soil storage parameter in inches

SPSS – Statistical Package for the Social Sciences

TIN – triangular irregular network

USACE – U.S. Army Corps of Engineers

USGS – U.S. Geological Survey

WATSTORE – National Water Data Storage and Retrieval System

WLS – weighted least square

Acknowledgements

This work has been conducted under the general and technical direction of Christopher N. Dunn, Director, Hydrologic Engineering Center; Michael K. Deering, Chief, Water Resource Systems Division, Hydrologic Engineering Center as project manager; and, Beth Faber of the Hydrologic Engineering Center as technical reviewer. The analysis was completed and the report prepared by David Ford Consulting Engineers Inc, including Mike Konieczki, David Ford and David Goldman.

Executive Summary

Situation

The US Army Corps of Engineers (USACE), Philadelphia District (NAP), has engaged the Hydrologic Engineering Center (HEC) in its study of the Delaware River Basin. The study, entitled "*Multi-jurisdictional Use and Management of Water Resources for the Delaware River Basin, NY, PA, NJ, and DE"*, included a statistical analysis of annual peak flows along the mainstem Delaware River following procedures outlined in "*Guidelines for Determining Flood Flow Frequency, Bulletin 17B"* (Interagency Advisory Committee on Water Data (IACWD), 1982) done jointly by the US Geological Survey (USGS) and NAP. The purpose of this report is to recommend, and describe the methods for estimating, regional skew values required by *Bulletin 17B* in peak flow frequency analysis.

Bulletin 17B states that a regional skew value should be included in developing frequency curves for annual peak flows with the log-Pearson III frequency distribution. As suggested in the *Bulletin*, candidate regional skew estimates are calculated by applying area averaging, isoline mapping, and regression methods to the skew estimates from individual gages within a defined region. Among these methods, the regional skew estimate with the smallest mean square error (MSE) is chosen as the best estimate.

Regional skew analysis methods use data from many streamgage sites in a region to estimate regional skew values. Use of multiple streamgage sites approximates an analysis based on a much longer period of record. The approach exchanges space for time, reducing time-based sampling error in the skew estimate, while introducing a lesser spatial sampling error.

NAP's and USGS's analysis included development of frequency curves and flow quantiles at eight gage locations along the main stem Delaware River. During that analysis, it was determined that the current regional skew values for the Delaware River Basin are out-of-date. HEC originally completed a regional skew study in 1983 entitled "*Generalized Skew Study of the Delaware River Basin"* (USACE 1983). In the twenty-five years since that study's completion, more annual peak flows have been recorded and the methods for determining regional skew values have been updated. The purpose of this study is to update the regional skew values for the Delaware River Basin.

Tasks

This study required gathering streamgage data for the greater Delaware River Basin, and completing a regional skew analysis using three methods:

- **Method 1:** Region average skew. This method was implemented four ways:
    - **Method 1a** - average skew of the entire basin;

  - **Method 1b** - average skews of homogenous regions (defined geographically and verified with L-moment analysis);
  - **Method 1c** – weighted-average skews of homogeneous regions (weighted by record length);
  - **Method 1d** – Generalized Least Squares (GLS) constant-only regression on homogeneous regions.

- **Method 2:** Skew isoline map.

- **Method 3:** Predictive equations using GLS regression.

Results from the three methods were compared and the most appropriate method was selected for calculation of regional skew.

Actions

To complete the regional skew study of the Delaware River Basin, the analysis:

- Updated annual peak records for 215 streamgage records. These streamgages were considered in the previous regional skew study of the Delaware River Basin (USACE 1983). This task included collecting streamgage data from 1983 through the 2006 water year and verifying seven watershed parameters: drainage area, 10 to 85 percent slope, basin length, mean basin elevation, percent lake storage, percent forested area, and mean annual precipitation (MAP).

- Gathered annual peak data - recorded through the 2006 water year - for an additional 477 streamgages in and around the Delaware River Basin. These gages were not included in the original 1983 study because they either did not exist at that time, or failed to meet the criteria specified in the 1983 study.

- Analyzed 692 records to ensure data quality and homogeneous records, and eliminated 444 streamgages because of tidal or anthropogenic effects. This was done by noting USGS codes in the peak flow record, and comparing mean, standard deviation, and skew to drainage area for remaining gages. The slope and R^2 values from a linear regression of annual peak flows to water year were also examined.

- Calculated sample statistics, including station skew values, for the remaining 248 streamgages considered in this study using *Bulletin 17B* procedures. Special attention was given to records with historical information, as peaks that are historically weighted can have a significant impact on station statistics.

- Narrowed the list to 163 streamgages using the following criteria: absence of anthropogenic effects (regulation, urbanization, and so on); minimum of twenty-five years of systematic record length; the streamgage is located within the Delaware River Basin, or has a majority of its watershed within twenty-five miles of the basin; less than ten percent of the watershed is urbanized; and the gage is absent of tidal effects.

- Verified, and in some cased determined, watershed parameters for the 163 streamgages.
- Calculated regional average skew and MSE for these 163 streamgages (Method 1a).
- Determined eleven plausibly homogeneous regions using river subbasins (used in Methods 1b, 1c and 1d). Region heterogeneity and stream gage discordance statistics were calculated to find acceptably homogeneous regions. Computed average and weighted average, and GLS-constant skew for those regions.
- Developed, and calculated MSE for, a regional skew contour map using inverse distance weighting (IDW), modified using engineering judgment based on basin physiography and hydrology (Method 2).
- Calculated regional skew coefficients and their average prediction errors using a GLS procedure (Method 1d and Method 3). The GLS procedure used gages for which watershed parameters were available.

Analysis and Results

A total of 163 streamgages – 115 of the gages used in the 1983 study and an additional 48 gages – were used in completing this study. Each streamgage has at least twenty-five years of unregulated annual peak flows whose records are considered absent of both tidal and anthropogenic effects per *Bulletin 17B* guidelines. Table 1 summarizes the results of the regional skew analysis by noting skew values and mean squared errors for the full Delaware River basin (where possible) for the purpose of comparison.

Table 1. Summary of regional skew analysis results

Method (1)	**Associated Regional Skew (average or GLS constant) (2)**	**Record MSE[a] (3)**	**Simulated MSE[b] (4)**	**AVP[c] (5)**
Method 1a (entire region)	0.184	0.142	0.241	n/a
Method 1a (Delaware River Basin only)	0.217	0.155	0.259	n/a
Method 1b (entire region)	0.184	0.133	0.232	n/a
Method 1b (Delaware River Basin only)	0.217	0.146	0.251	n/a
Method 1c (entire region)	0.176	0.117	0.203	n/a
Method 1c (Delaware River Basin only)	0.191	0.130	0.220	n/a
Method 1d (Delaware River Basin only)	0.151	n/a	n/a	0.044
Method 2	n/a	0.147	n/a	n/a
Method 3 (northern region)	computed from eqn	n/a	n/a	0.027
Method 3 (southern region)	computed from eqn	n/a	n/a	0.019

a. Calculated using procedures outlined in *Bulletin 17B,* if applicable.
b. Calculated using Monte Carlo simulation to account for sampling error, if applicable.
c. Applicable only for GLS regressions; AVP – average variance of prediction

As noted in Table 1, averaging all 163 station skews into a single region results in a regional average skew of 0.184, and MSE of 0.142. This *Bulletin 17B* recommendation for estimating MSE assumes that all gage skew values are perfectly estimated (no sampling error). An estimate of MSE equal to 0.241 was obtained using Monte Carlo simulation to include the sampling error of gage skew estimates (Method 1a).

For the homogeneous regions verified using L-moment analysis (Hosking and Wallis, 1997), region skews were computed by averaging the station skew values in each region. The weighted average of those region skews (weighted by the number of gages in a region) for all regions are 0.181, the MSE is 0.133, and the simulated MSE (including time-sampling error) is 0.232. The weighted-average skew for regions within the Delaware River Basin boundary is 0.221, the MSE is 0.146, and the simulated MSE is 0.251 (Method 1b).

For the same homogeneous regions, region skews were also computed by weighting the station skew average by record length. The weighted average of those region skews (weighted by number of gages in a region) for all regions are 0.176, the MSE is 0.117, and the simulated MSE (including time-sampling error) is 0.203. The weighted-average skew for regions within the Delaware River Basin boundary is 0.191, the MSE is 0.130, and the simulated MSE is 0.220 (Method 1c).

GLS regression of the regions using only a constant effectively obtained regional average skew values. The constant provides a direct comparison with the regional average obtained using standard methods outlined above, while also accounting for inter-gage correlation and differences in gage record length. In this approach, average variance of prediction (AVP) is used as a measure of prediction error in place of MSE and simulated MSE. The GLS-constant region average approach results in a weighted-average constant (based on the number of gages in a region) of 0.151, which would be used as the regional skew value. The method has a weighted-average AVP of 0.044 (Method 1d).

A skew isoline map was developed by calculating skew isolines using an inverse distance squared interpolation. The isolines were then modified using engineering judgment based on consideration of region physiography and hydrology, as shown in Figure 17. The MSE for this skew isoline map is 0.147 (Method 2).

GLS regression of all gages in the Delaware River Basin resulted in no regression model prediction error, with all error attributed to limited record length. This was felt to be an unreasonable result because no model error implies a perfect regression model prediction if the gage skew values were perfectly estimated i.e., no sampling error. This is unlikely to occur in skew prediction. More significant results were achieved, however, by dividing the basin into northern and southern regions. A regression using only mean elevation identified a regression model error and had an AVP equal to 0.027 for the northern region. A regression using mean annual precipitation resulted in an AVP of 0.019 for the southern region, but no regression model error could be defined (Method 3).

Recommendation

For determining a regional skew for the Delaware River basin, HEC recommends the results of Method 1d (GLS constant-only regression on basin parameters). This method (based on approximately homogeneous regions verified by L-moment analysis) yields region skew values shown in Table 2, that average to 0.151, with a corresponding MSE of 0.044. The GLS constant-only method is recommended because:

- The simplicity of using only a constant and the comparably small AVP makes this method preferable to the GLS regression equations or skew contour map.
- The method produces improvements to the recommendations of *Bulletin 17B*, as presented in this report.
- The minimum error of the method, AVP, will promote the greatest consistency in the application of the *Bulletin 17B* guidelines.

Table 2 contains the regional skew value and its average variance of prediction (to be used in place of mean square error (MSE)) for each of the regions within the Delaware River Basin.

Table 2. GLS-constant region average skew and errors for regions within the Delaware River Basin Boundary

Region Number (1)	Constant (GLS-Constant Regional Skew) (2)	MSE (from AVP, Average Prediction Error) (3)
2	0.087	0.026
3	0.203	0.077
4	0.165	0.030
5	0.178	0.033
6	0.001	0.064
7	0.287	0.034
Weighted average	**0.151**	**0.044**

The focus on consistency is an important aspect of the study recommendations. In the original testing to develop *Bulletin 17B* (IACWD 1982, Appendix 15) split sample testing demonstrated that the log-Normal distribution (a zero-skew distribution) performed as well as the log-Pearson III distribution when *substituting* regional skew for the computed skew. This implies that regional information had no impact on prediction accuracy. However, the regional skew approach was selected because it promoted greater consistency in the estimate of infrequent quantiles (p=0.01 quantiles) obtained from either of the gage split record samples. The result is greater consistency both at a gage and within a region as future frequency studies are performed. Consequently, the consistency principle is important in promoting reasonably stable flood plain maps going into the future.

Chapter 1

Study Overview

This document is organized in four main sections. The first is the **Study Overview**, which gives general background information on the regional skew study. The second is the **Methodology** section, which outlines the procedures used in this skew analysis. The third is the **Analysis and Results** section, which presents findings and results. The fourth is the **Conclusions** section, which contains HEC's recommended method for conducting the regional skew analysis.

1.1 Background

The US Army Corps of Engineers (USACE), Philadelphia District (NAP), has engaged HEC in a study of the Delaware River Basin. The study, entitled "*Multi-jurisdictional Use and Management of Water Resources for the Delaware River Basin, NY, PA, NJ, and DE*", included a statistical analysis of peak flows along the main stem Delaware River following procedures outlined in "*Guidelines for Determining Flood Flow Frequency, Bulletin 17B"* (IACWD 1982) done jointly by the USGS and NAP. USGS's and NAP's analysis included development of frequency curves and flow quantiles at eight gage locations along the main stem Delaware River. *Bulletin 17B* includes a requirement that a regional skew value should be used in the log-Pearson III frequency distribution of annual peak flows.

During USGS's and NAP's analysis, it was determined that the current regional skew values for the Delaware River Basin are out-of-date. HEC originally completed a regional skew study in 1983 entitled "*Generalized Skew Study of the Delaware River Basin*" (USACE 1983). In the twenty-five years since that study's completion, more annual peak flows have been recorded and the methods for determining regional skew values have been updated. The purpose of this study is to update the regional skew values for the Delaware River Basin.

1.2 Procedure

A regional skew analysis uses annual peak flow data from multiple stream gage sites in a region to calculate regional skew values. Using multiple sites approximates an analysis based on a longer period of record. The approach exchanges space for time, reducing time-based sampling error in the skew estimate, while introducing a lesser spatial sampling error. Regional skew values, in conjunction with flow records of appropriate length (greater than twenty-five years) and weighted with station skew, is thought to yield a better (more consistent) estimate of flow quantiles for a given frequency curve.

This study followed the guidelines and methods outlined in *Bulletin 17B*. *Bulletin 17B* recommends that the chosen method be the one with the lowest calculated MSE from one of three following methods:

- **Method 1:** Region average skew. Implemented four ways:
    - **Method 1a** - average skew of the entire basin;
    - **Method 1b** - average skews of homogenous regions (defined geographically and verified with L-moment analysis);
    - **Method 1c** – weighted-average skews of homogeneous regions (weighted by record length);
    - **Method 1d** – GLS constant-only regression on homogeneous regions.
- **Method 2:** Skew isoline map.
- **Method 3:** Predictive equations using regression.

1.3 Region Description

The Delaware River Basin is comprised of a 13,430 square mile watershed in the northeast states of New York, Pennsylvania, New Jersey, Delaware, and Maryland. The Delaware River's headwaters begin in the Catskill Mountains as two streams - the East and West branches - which flow south to their confluence at Hancock, New York. The river continues approximately 200 miles until it reaches Delaware Bay. The Delaware River Basin is on the eastern seaboard of the United States, and is therefore subject to hydrometeorologic events resulting from hurricanes, tropical storms, and convective precipitation. Figure 1 is a map of the Delaware River Basin and surrounding region.

1.4 Previous Work

The "*Generalized Skew Study of the Delaware River Basin*" (USACE 1983), completed in 1983 by USACE, considered 215 streamgages in its regional skew analysis. Following *Bulletin 17B* guidelines, the 1983 study used procedures that accounted for low outliers, high outliers, and historical information (historical annual peak flows or non-exceedance periods). There were fifteen stations identified as having low outliers in their annual peak flow records. Twenty-eight stations were identified as having historical flood peaks, 39 stations had historical non-exceedance information, and eight stations had both historical records and information.

Gages with annual peaks that were known to be regulated - based on USGS National Water Data Storage and Retrieval System (WATSTORE) Codes 5 and 6 - were discarded from the analysis. WATSTORE Code 7 aided in identifying historical peaks.

The study analyzed three different methods - skew isoline map, predictive (regression) equations, and average skew - and determined that regions of average skew were most appropriate in defining the adopted skew coefficients. Because the adopted skew used this final method, only 132 stations were used in the study's final stages. Of the 132 gages, 65 gages were within the

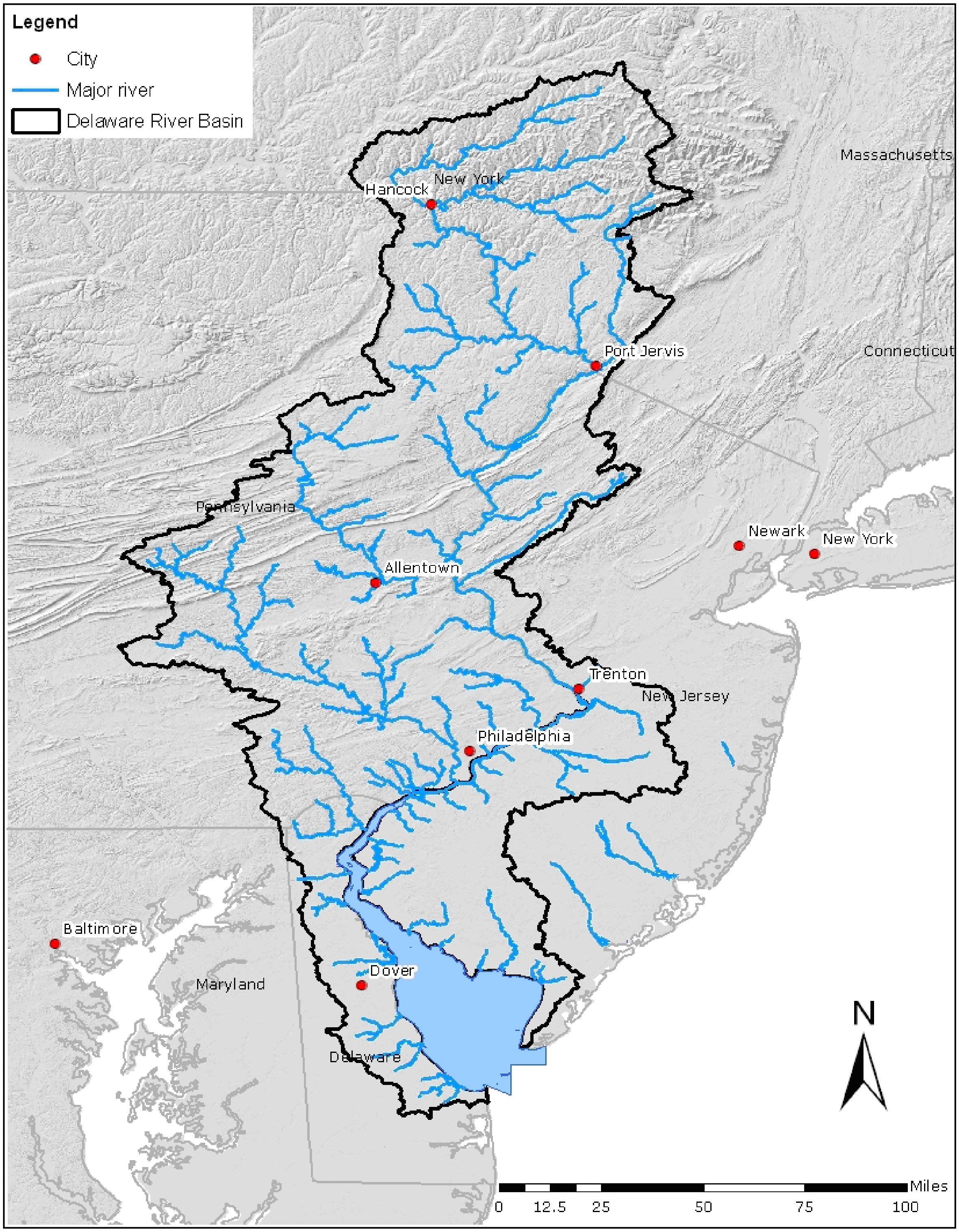

Figure 1. Delaware River Basin

Delaware River Basin, and 67 had drainage areas either adjacent to or mostly within 25 miles of the Delaware River Basin. Appendix A, lists the gages considered, their adopted skew coefficients, and their computed station skews, as calculated in 1983. Table 3 compares the MSE associated with each region, shown in Figure 2, developed in the 1983 study.

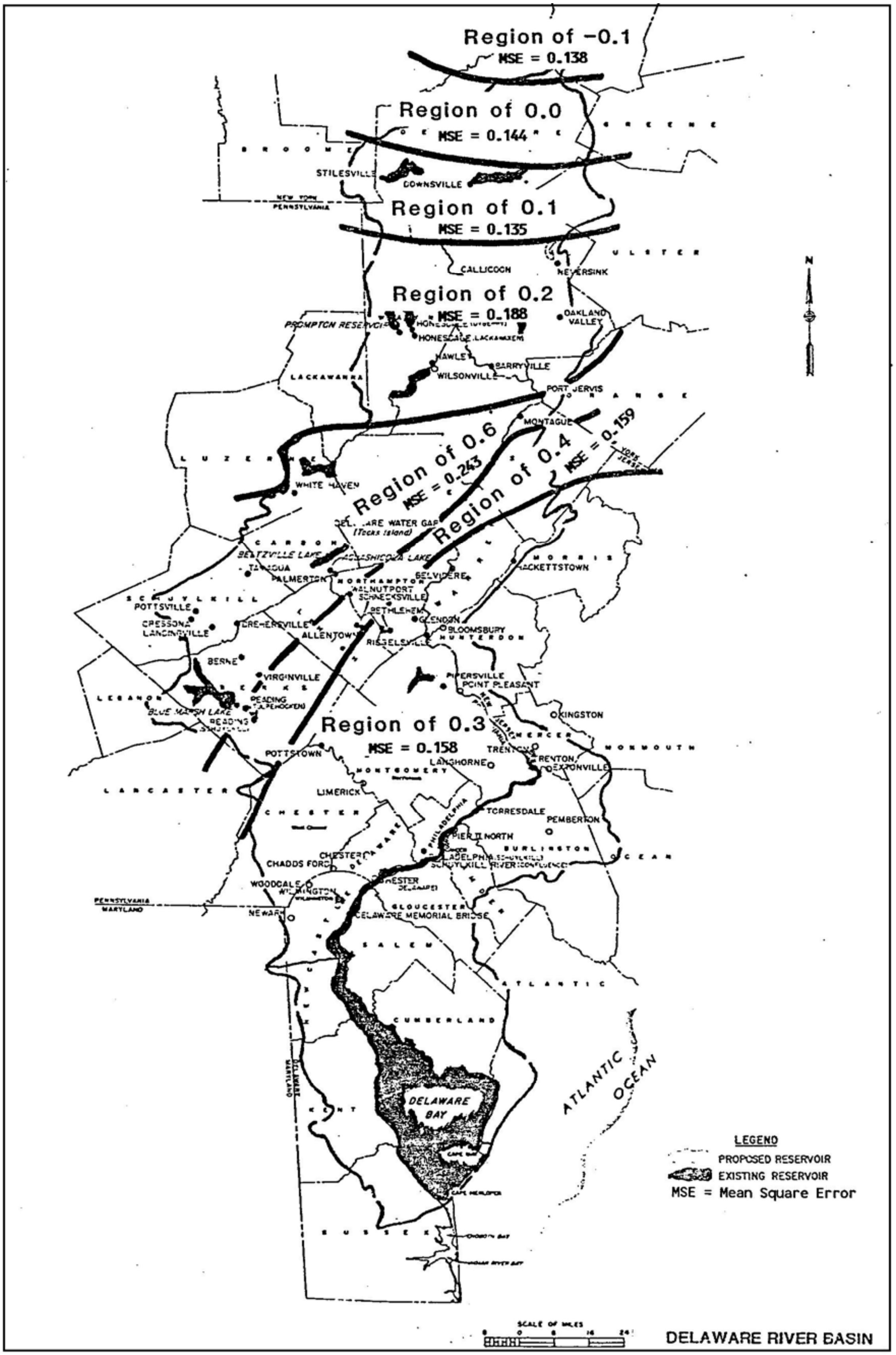

Figure 2. Previous study adopted skew coefficient regions (USACE 1983)

Table 3. Average skew regions defined in 1983 study (USACE 1983)

Region Skew (1)	River Basin Description (2)	Number of Gages (3)	Computed MSE (4)	Adopted MSE (5)
-0.1	Upper Delaware	11	0.820	0.138
0.0	Upper Delaware	15	0.089	0.144
+0.1	Upper Delaware	9	0.050	0.135
+0.2	Upper Middle Delaware	12	0.188	0.188
+0.6	Middle Delaware	20	0.243	0.243
+0.4	Upper and Lower Delaware	11	0.140	0.159
+0.3	Lower Delaware	54	0.158	0.158

The USGS completed a similar study in 2006, entitled "*Magnitude and Frequency of Floods on Nontidal Streams in Delaware*". The area for the study encompasses the entire state of Delaware and portions of Maryland, Pennsylvania, and New Jersey, as shown in Figure 3. The method of choice used in that study is regional average skew based on two physiographic regions: the Coastal Plains and the Piedmont. The mean skews for the Coastal Plains and the Piedmont are 0.204 and 0.107, respectively.

An MSE was not calculated to determine method of choice. This is because plotting of gages, as shown in Figure 3, did not yield an apparent pattern of station skews, thus an effective skew map could not be drawn. A weighted least squares regression of the region did not result in statistically significant relations at the 95 percent confidence level.

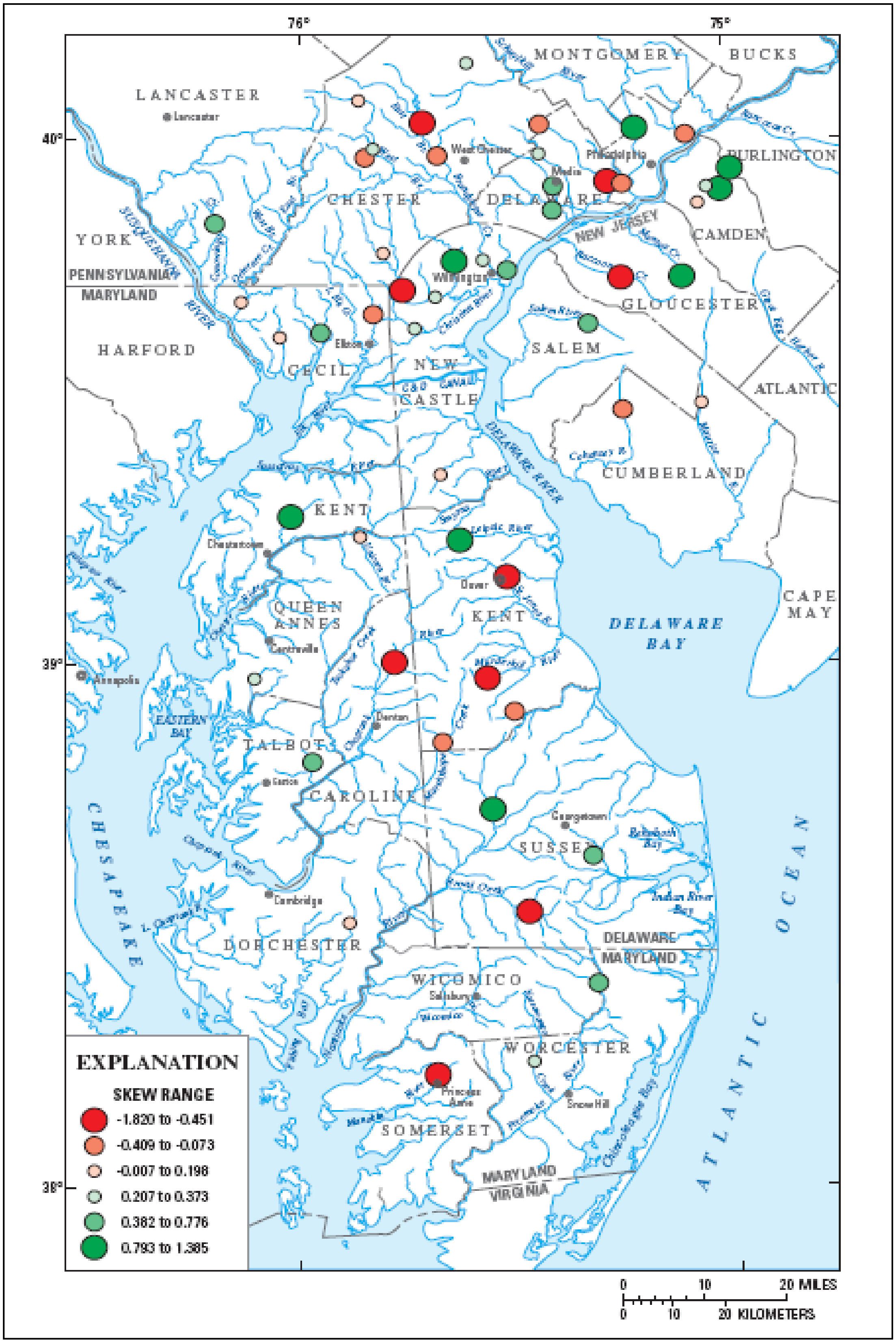

Figure 3. Skew ranges for streamgaging stations in Delaware and surrounding states with twenty-five or more years of record (USGS 2006)

Chapter 2

Methods and Procedures

This section briefly describes the methods used in this study to complete the regional skew analysis for the Delaware River Basin. In general, the methods used follow the procedures outlined in *Bulletin 17B.*

2.1 Gage Data Quality Assurance

Skew analysis procedures demand a high degree of confidence in the data quality of the gage records. Therefore, the records of 692 streamgages (the original 215 and an additional 477) were analyzed to assess the effects of regulation, determine the existence of outliers, and identify historical information. This analysis involved simple statistical tests, looking at the WATSTORE flags, processing information about various gages from water supply papers, and personal conversations with USGS and USACE Philadelphia District (NAP) staff.

2.2 Computing Station Skew

HEC's Flood Frequency Analysis (HEC-FFA) software was used to compute station skew. HEC-FFA implements *Bulletin 17B* procedures for identifying low outliers, high outliers, and historical information when computing station statistics of the logarithmic transforms of the annual peak flows. Also, historical information was used in the calculation of station statistics wherever possible. Procedures for incorporating the historical information are detailed in *Bulletin 17B*.

A skew bias correction was not used in calculating station skew values in order to maintain consistency with Table 1 of *Bulletin 17B*, which includes non-bias-corrected skews in the estimation of MSE.

2.3 Computing Regional Skew and Mean Square Error

Bulletin 17B recommends three methods for estimating regional skew: computing a regional average skew, developing a skew isoline map, and developing a regression prediction using hydrometeorological variables. *Bulletin 17B* recommends using the method that yields the lowest calculated mean square error (MSE). The following sections outline the procedures used for these three methods.

2.3.1 Region Average Skew (Method 1)

The basic assumption in the Method 1 approach is that a homogeneous watershed region has a single value of skew, and the various estimates of skew in the region are combined to estimate that value. Region average skew values were estimated for:

- A single region encompassing the entire study area (Method 1a).
- Approximately homogeneous subregions, defined geographically and checked for statistical homogeneity using L-moment procedures described by Hosking and Wallis (1997) (Methods 1b, 1c and 1d). *(Regions are defined in Appendix C, Table 32, page C-8.)*

In Method 1a, regional skew and MSE were obtained by averaging station skew of all gages in the region. In Method 1b, station skews and MSE were averaged within homogeneous subregions described below. Method 1c used a weighted-average of station skews and MSE within each subregion, weighting by record length to give more weight to stations with less sampling error. Method 1d computed a GLS regression with only a constant value, producing the equivalent of a regional average.

While the skew estimate for a region is obtained as some form of average of the individual station skew values in that region, MSE of skew for a region was computed in two ways: by the method recommended in *Bulletin 17B,* and by adding to that MSE an estimate of sampling error in the gage estimates of skew. The *Bulletin 17B* method estimates MSE as the average sum of squared differences between the region average skew and each of the station skews. However, this method assumes that each station skew is estimated perfectly, i.e., without the sampling error due to limited record length. The second computation method uses Monte Carlo simulation to estimate MSE including this sampling error. (This method is detailed in Appendix D, Section D.4.1.1, page D-7.)

In order to define several (approximately) statistically homogeneous subregions for averaging skew, this study used the procedure for L-moment analysis outlined in *Regional Frequency Analysis* (Hosking and Wallis, 1997). Initially, subregions were defined using river subbasins, and then those subregions were tested for homogeneity using L-moment and discordancy statistics for each gage and heterogeneity statistics for each subregion. Discordancy values indicate whether a specific gage statistically fits in a grouping (subregion), and heterogeneity measures indicate whether a grouping is acceptably homogeneous.

The heterogeneity measure is used to determine if a group of gages can be considered a homogeneous subregion by measuring the similarity in the shape of the probability distributions. Regions that were not acceptably homogeneous were then examined for discordant gages, or in the absence of obviously discordant gages, were redefined. Regions that appeared to be homogeneous were also checked for discordant gages. Regions were considered "acceptably homogeneous" "possibly heterogeneous" and "definitely heterogeneous" depending upon the value of the heterogeneity statistic, as described further in Appendix D, Section D.4.3 (page D-13).

When discordant gages were found, regions were redefined to include a different subset of drainage areas. A stream gage was considered discordant if its discordancy statistic was greater than a critical value. With any change, discordancy and homogeneity measures were recalculated and the region re-examined. L-moment analysis and discordancy procedures are detailed in Appendix D, Section D.4.3 (page D-13).

Gages with skews greater than one were omitted from the calculation of regional average skew, record MSE, and simulated MSE. In the Delaware River Basin, skews greater than one resulted from the impact of Hurricane Agnes on gages with short records. The resulting sampling error caused different exceedance probability estimates for that single event, based solely on the length of available record at each gage, a result thought to be incorrect. The impact of Hurricane Agnes on longer record gages remains in the analysis.

2.3.2 Skew Isoline Map (Method 2)

ArcGIS® tools were used to develop skew isolines (contours) for the Delaware River Basin. First, skew values were plotted at their corresponding gage locations. This choice is consistent with the methodology used by the USGS (USGS, 2006). Then isolines were developed from the plotted station skew values using three methods: linear interpolation, inverse distance weighting (IDW), and engineer's judgment-assisted IDW.

To develop isolines using linear interpolation, the skew values were inspected to see if there was a pattern. Finding such a pattern would allow more reasonable drawing of isolines on a skew map than just using only mathematical algorithms. In the absence of a pattern, mathematical algorithms would still be used to create regional skew contour values. However, additional inspection of the computed skew contours and their comparison to basin physiography and hydrology would be required to assure the computed contours are reasonable and rational and not just "lines connecting the dots".

A triangular irregular network (TIN) was developed using ArcGIS® tools and the station skew values plotted at their respective gaging locations. Skew contours were then linearly interpolated from the TIN.

To develop isolines using IDW, an algorithm was used to compute initially a gridded surface of skew, from which skew contours could be computed. The grid size for the surface was 1.21 square miles, which is smaller than the smallest gaged drainage area. The IDW algorithm is:

$$z_j = \frac{\sum_i \frac{z_i}{(d_{ij})^n}}{\sum_i \frac{1}{(d_{ij})^n}} \qquad (1)$$

where:

z_j is the value at an unknown point,
z_i is the value at a known point,
d_{ij} is the distance between a known and the unknown point, and,
n is a user-defined exponent.

The number of known points used to determine an unknown value can be specified by the user with ArcGIS® tools.

An exponent *n* of one (1) was used and the nearest forty gages were used for z_i. The exponent was chosen so as to maximize the influence of nearby gages in calculating skew values, while forty gages were used to be consistent with *Bulletin 17B* guidelines. From this gridded surface, the skew isolines were calculated. Grids created using an exponent of two and using the nearest ten gages were also examined. These contours exaggerated the effects of local gages, resulting in more localized extremes (peaks and valleys in the contour map) than the number of localized extremes originally developed, therefore their used in the development of a skew map was discontinued.

The mathematical algorithms used to calculate isolines in ArcGIS® using both linear interpolation and IDW result in exact solutions at the sites where skews are plotted, and therefore both methods have an MSE equal to zero (0).

The third mapping method uses the IDW-created isolines as a starting point. These isolines, and the station skew values used to create them, were then compared with an elevation map of the region. The isolines then were modified using engineering judgment based on region physiography and hydrology. In general, skew contours representing local minima or maxima around a single gage were removed. Contours were also redrawn to establish skew contours around regions of similar physiographic and hydrologic characteristics.

MSE for a skew map is calculated by averaging the squared differences between the station skew, at the gage and the skew interpolated from the skew map isolines.

2.3.3 Predictive Equations Using Generalized Least Squares (Method 3)

The assumption underlying Method 3 is that skew can be described as a function of various watershed parameters, and that a predictive relationship can be developed. GLS regression was used in developing a series of predictive equations for skew. A standard regression equation takes the form:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \varepsilon \qquad (2)$$

The GLS regression considers the difference in record lengths (sampling error) and the inter-gage correlation in peak flows (which reduces the effective record length at each gage) when estimating regression equation parameters. The GLS methodologies and technical considerations used are detailed in Appendix D.

For the GLS analysis, eight watershed parameters were used that were identical to those used in the 1983 study of the Delaware River Basin. These parameters include drainage area, 10 - 85 percent slope, basin length, mean basin elevation, percent lake storage, percent forested area, and mean annual precipitation (MAP), and SCS soils index number. Parameter values used in this study are listed in Appendix B.

AVP was used as the analog to the simulated-MSE computed in the previous two methods. The accuracy of GLS regression is quantified by the AVP. The AVP for a regression equation is the average of the square root of the MSE of the individual MSE of prediction for all gages used in developing the regression. Calculation of AVP is detailed in Appendix D.

2.4 Split-Sample Testing

Split-sample testing was completed to compare the *Bulletin 17B* (MSE-based) and the GLS (AVP-based) estimates of regional skew to at-site estimates. Split-sample testing involved splitting the period of record into two data sets. Frequency curve estimates were obtained from one data set, with and without using regional skew estimates. The estimates of the p=0.1 and the p=0.01 exceedance quantiles from these frequency curves were compared to the observed exceedance frequency of these quantiles in the remaining (reserved) data. The split-sample testing methods used are detailed in Appendix D. In this study, the records were split using the following methods:

- **Forecast method:** first half of the record is used to estimate frequency curve, remaining record is reserved.
- **Back cast method:** second half of record is used to estimate frequency curve, remaining record is reserved.
- **Alternating Method 1:** alternate years in the record are used to estimate the frequency curve, remaining record is reserved.
- **Alternating Method 2:** the reserved record in Alternating Method 1 is used to estimate the frequency curve and the remaining data are now the reserved.

The alternating methods for splitting the data were employed to remove the impact of apparent trends or cycles in the stream flow data, as was done by the Water Resources council in the split sample testing performed in the development of *Bulletin 17B* (IACWD 1982, Appendix 14). The forecast and back cast methods for splitting the data were employed to simulate actual application of frequency curves.

Chapter 3

Analysis and Results

3.1 Gage Data Quality Assurance

Many different criteria were used to assess gage data quality to identify gages in need of further examination. These criteria included information found in the WATSTORE data (such as regulation effects), analysis of sample moment statistics, and trend analysis of peak flow versus water year.

Assessing data quality is a parallel process rather than a sequential one. The analysis started with a common set of stream gages when assessing the dataset for various criteria such as anthropogenic effects (regulation and urbanization) or sufficient record length. Information gained from parallel assessment of gage data gives us a better sense of the region's physiography and hydrology, understanding which is important in a regional skew study.

The initial dataset for assessing data quality included the 215 gages used in the USACE 1983 study and 67 additional gages.

The 67 gages are the subset of 477 gages in and around the Delaware River Basin that were not included in the USACE 1983 study. These 477 gages were not originally included because they either did not exist at that time or failed to meet the criteria specified in the 1983 study. Of the 477 additional gages, 380 of these were found to have record lengths of less than twenty-five years or to have tidal influence, and were therefore unsuitable for analysis. It was also found that thirty of the remaining 97 are on the mainstem of the Delaware River and were removed from the data set because of regulation and urbanization effects, in addition to the inappropriateness of mainstem gages for a regional skew analysis. This yielded 67 additional gages for consideration in the study.

3.1.1 Sample Moment Statistics

Initially the 215 gages considered in 1983 were analyzed by comparing the mean, standard deviation, and skew of the annual peak flow to watershed drainage area; plots are shown in Figure 4 through Figure 6. These comparisons indicate a lack of correlation as measured by R^2, the square of the correlation coefficient, also shown in Figure 4 through Figure 6.

Gages used in the 1983 study and identified as being regulated were not yet removed from the dataset in comparing sample moment statistics. Similarly, sample moment statistics were compared for additional streamgages identified as being regulated, and which had at least twenty-five years of record. These regulated gages were included in the statistical comparison to assess qualitatively the difference in the effects of minor (WATSTORE Code 5) and major

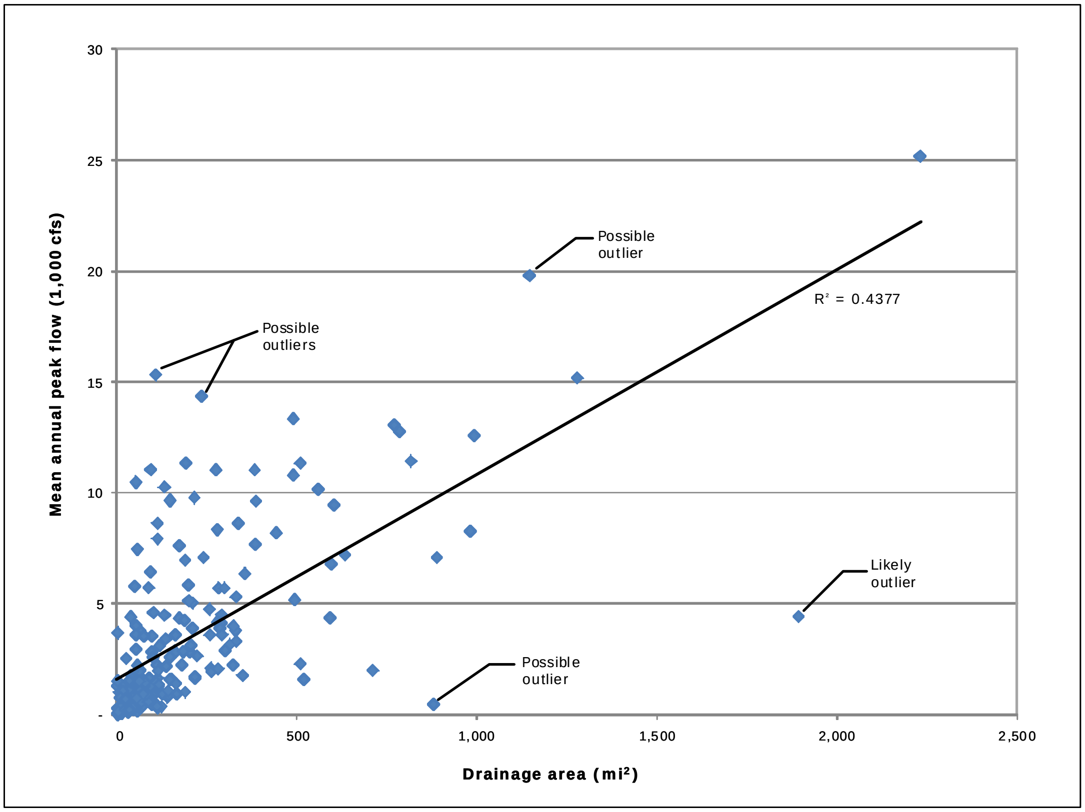

Figure 4. Mean annual peak flow versus drainage area

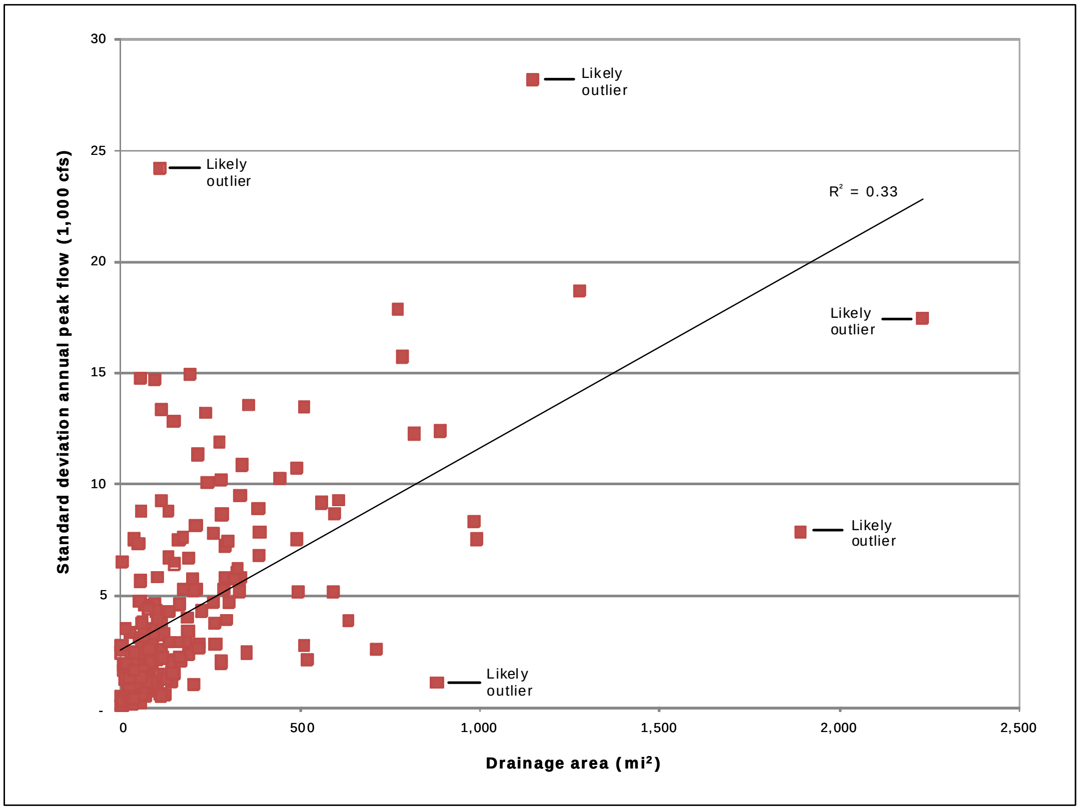

Figure 5. Standard deviation of annual peak flow versus drainage area

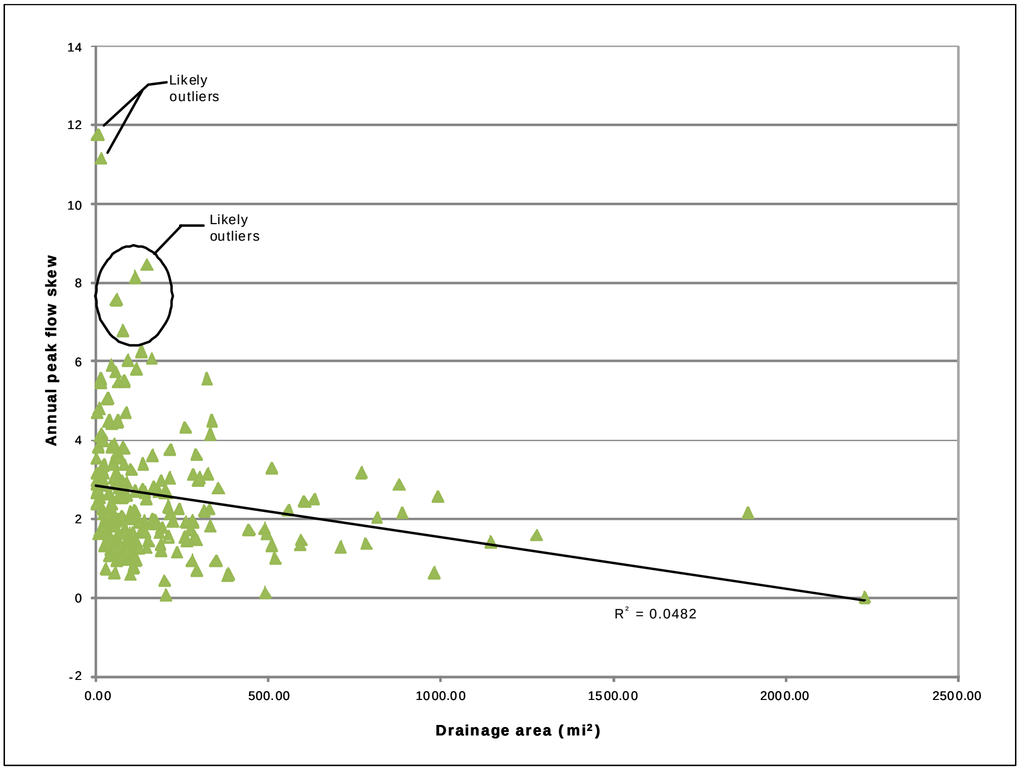

Figure 6. Skew of annual peak flow versus drainage area

(WATSTORE Code 6) regulation. It should be noted that the original study completed in 1983 included gages affected by minor or an unknown degree of regulation, as indicated by WATSTORE Code 5.

This comparison gives a clear visual representation of gages whose moment estimators vary greatly from those of similar drainage area. Comparing the mean annual peak flow to the drainage area, as shown in Figure 4, suggests that values for at least one site, and possibly as many as five sites, are outliers. Comparing the standard deviation of annual peak flow to the drainage area, as shown in Figure 5, suggests that values for at least five sites may be outliers. Comparing the skew of annual peak flow to the drainage area, as shown in Figure 6, statistical outliers indicated by moment statistics suggests that values for at least six sites may be outliers. Outliers are summarized in Table 4 through Table 6.

After analyzing the original 215 gages, the real-space statistics were analyzed for the additional 67 gages in the same manner, as shown in Figure 7 through Figure 9. As expected, these comparisons have low R^2 values. Comparing the mean annual peak flow and drainage areas suggests that three sites may be outliers for these additional gages, as shown in Figure 7. Comparing the standard deviation of annual peak flow to the drainage area of the additional gages, as shown in Figure 8, suggests that values for at least three additional sites may be outliers. Comparing the skew of annual peak flow to the drainage area for the additional gages, as shown in Figure 9, does not suggest the presence of any outliers. Outliers are summarized in Table 4 through Table 6.

The gages indicated as outliers through the comparison of their sample moment statistics all had some degree of regulation, as indicated through WATSTORE Codes 5 and 6 (see Section 3.1.3).

Table 4. Statistical outliers indicated by moment statistics

Comparing Drainage Area to: (1)	Possible Outliers of the Original 215 Gages (2)	Possible Outliers of the Additional 67 Gages (3)
Mean annual peak flow	01350000 01474500 01471500 01499000 01472000 (Figure 4)	01454700 01471510 01472162 (Figure 7)
Standard deviation of annual peak flow	01471500 01499000 01472000 01503000 01474500 (Figure 5)	01454700 01471510 01472162 (Figure 8)
Skew of annual peak flow	01411500 01530500 01475000 01583500 01482500 01657000 (Figure 6)	none indicated (Figure 9)

Table 5. Statistical outliers of original 215 gages indicated by moment statistics

ID (1)	Indicator statistic (2)
01350000	Mean
01411500	Skew
01471500	Mean, standard deviation
01472000	Mean, standard deviation
01474500	Mean, standard deviation
01475000	Skew
01482500	Skew
01499000	Mean, standard deviation
01503000	Standard deviation
01530500	Skew
01583500	Skew
01657000	Skew

Table 6. Statistical outliers of additional 67 gages indicated by moment statistics

ID (1)	Indicator statistic (2)
01454700	Mean, standard deviation
01471510	Mean, standard deviation
01472162	Mean, standard deviation

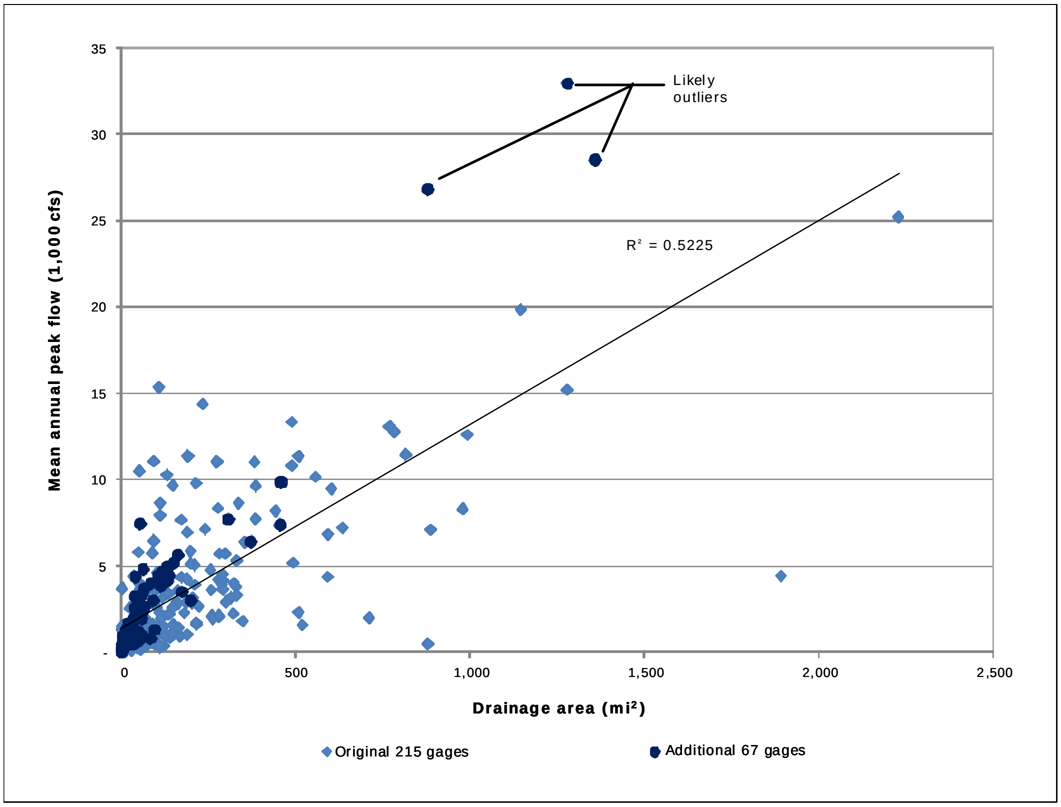

Figure 7. Mean annual peak flow versus drainage area, including additional gages

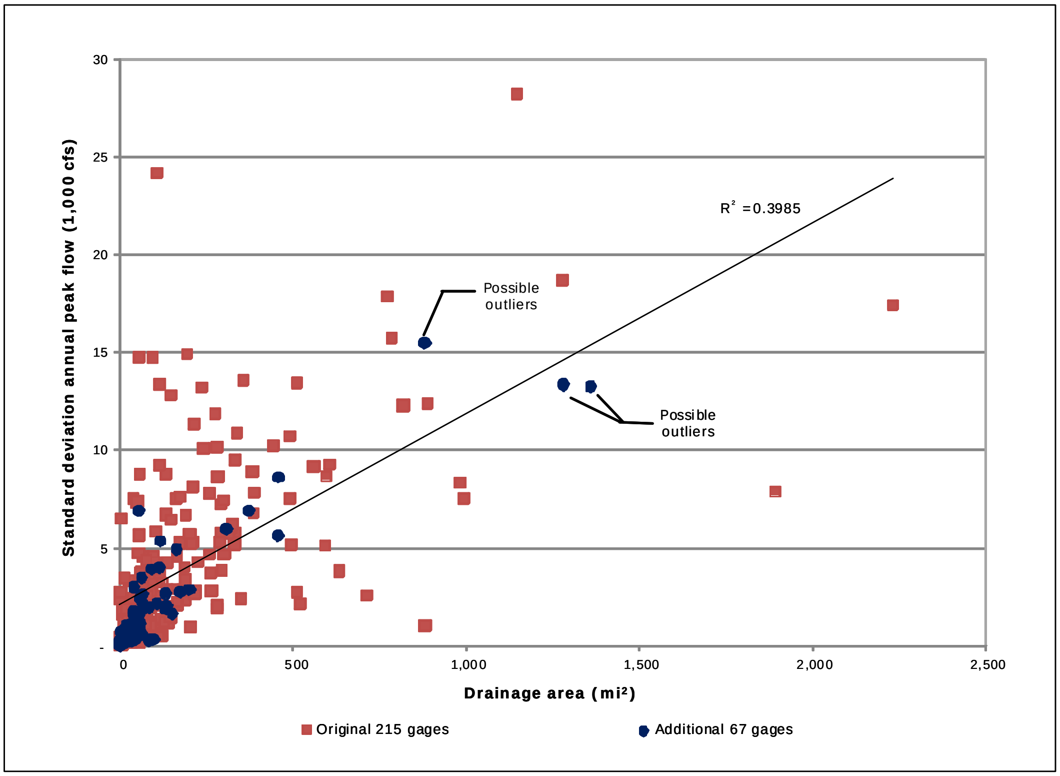

Figure 8. Standard deviation of annual peak flow versus drainage area, including additional gages

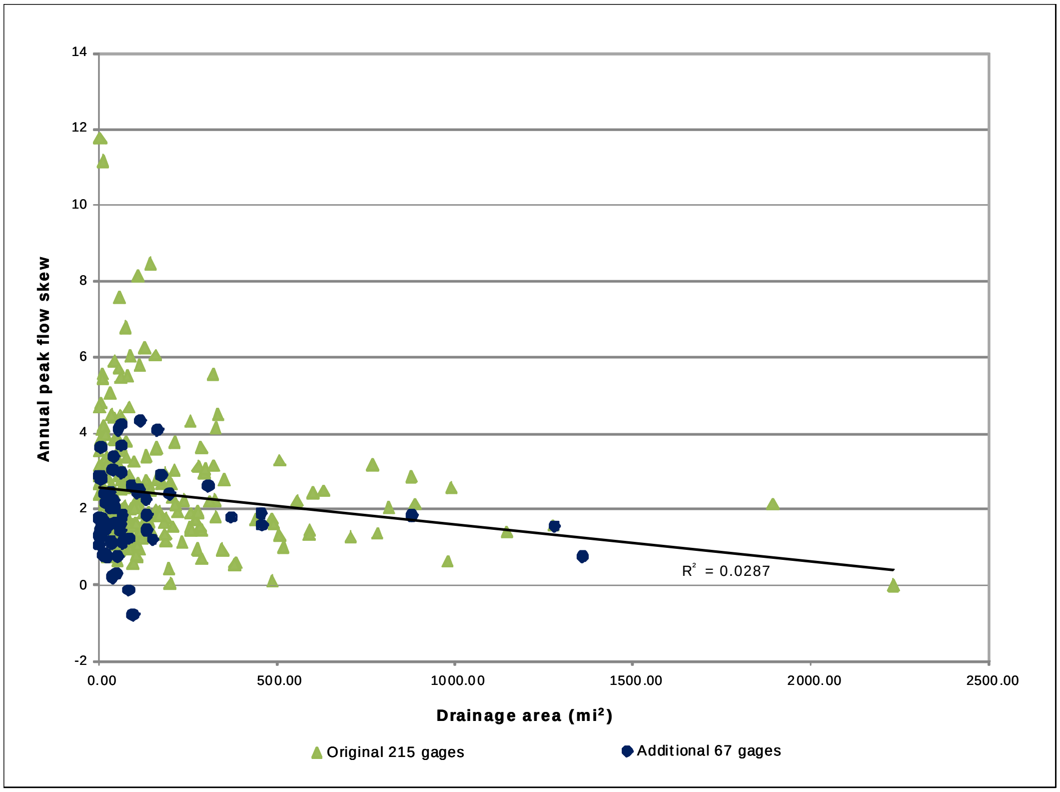

Figure 9. Skew of annual peak flow versus drainage area, including additional gages

3.1.2 Trend Analysis

Linear regression of peak flow versus water year was used to investigate trends in the data. This analysis can identify anomalous gage records, where linear regression produces a large regression coefficient or high R^2 value and no strong trends are apparent, further investigation of the gage record may be warranted. The converse is also true.

The regression coefficients exhibited significant scatter and only three sites (gage IDs 01407830, 01425500, and 01467160) had an R^2 value greater than 0.35, as shown in Figure 10. Because the regression coefficients showed a high degree of scatter, values greater or less than a single standard deviation away from zero were investigated further. Records that had an R^2 value greater than 0.35 were also examined. It was found that some of these outlying points had regulation effects. Those that did not were considered acceptable for inclusion because of the high degree of scatter.

3.1.3 Effects of Regulation

The 215 gages used in the USACE 1983 study are shown in Figure 11. *Bulletin 17B* recommends removal of stream gages that are affected by anthropogenic effects such as regulation or urbanization, or gages that have less than 25 years of record. There were 32 gages flagged as having some degree of regulation (defined by Codes 5 and 6 of the USGS WATSTORE information). However, 15 of those 32 gages had at least 25 years of unregulated record, and therefore could still be included in the dataset. The 32 gages flagged as having

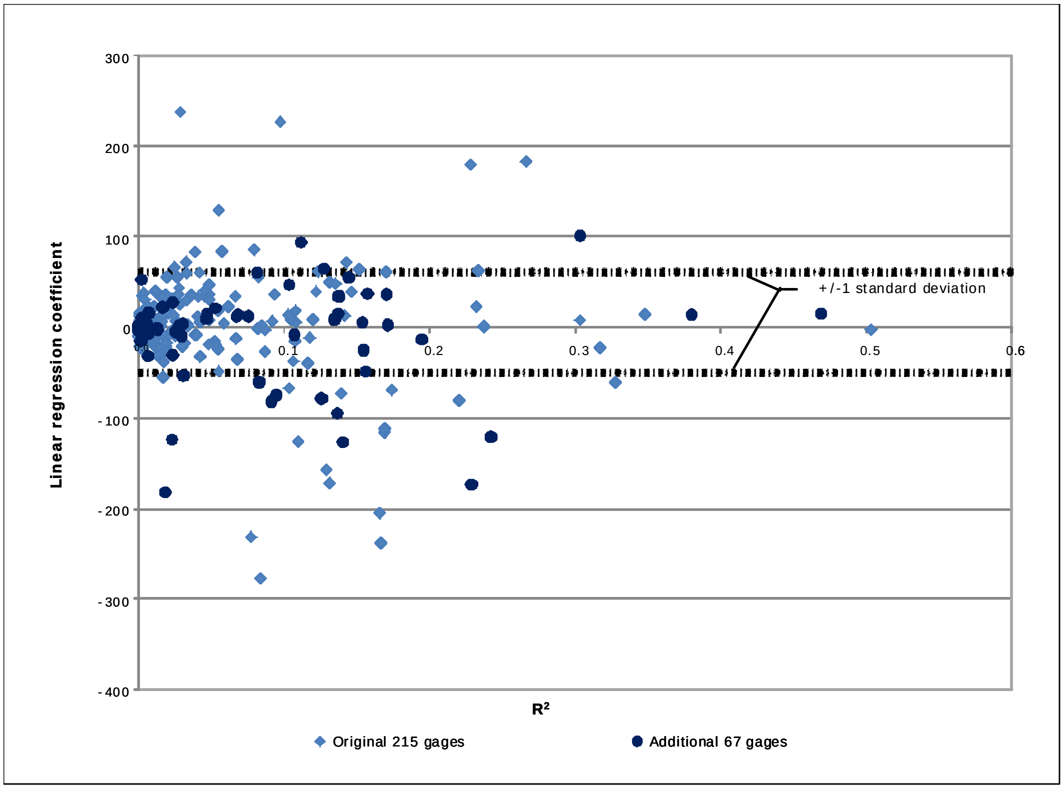

Figure 10. Coefficient of linear regression (annual peak flow and water year) versus R^2

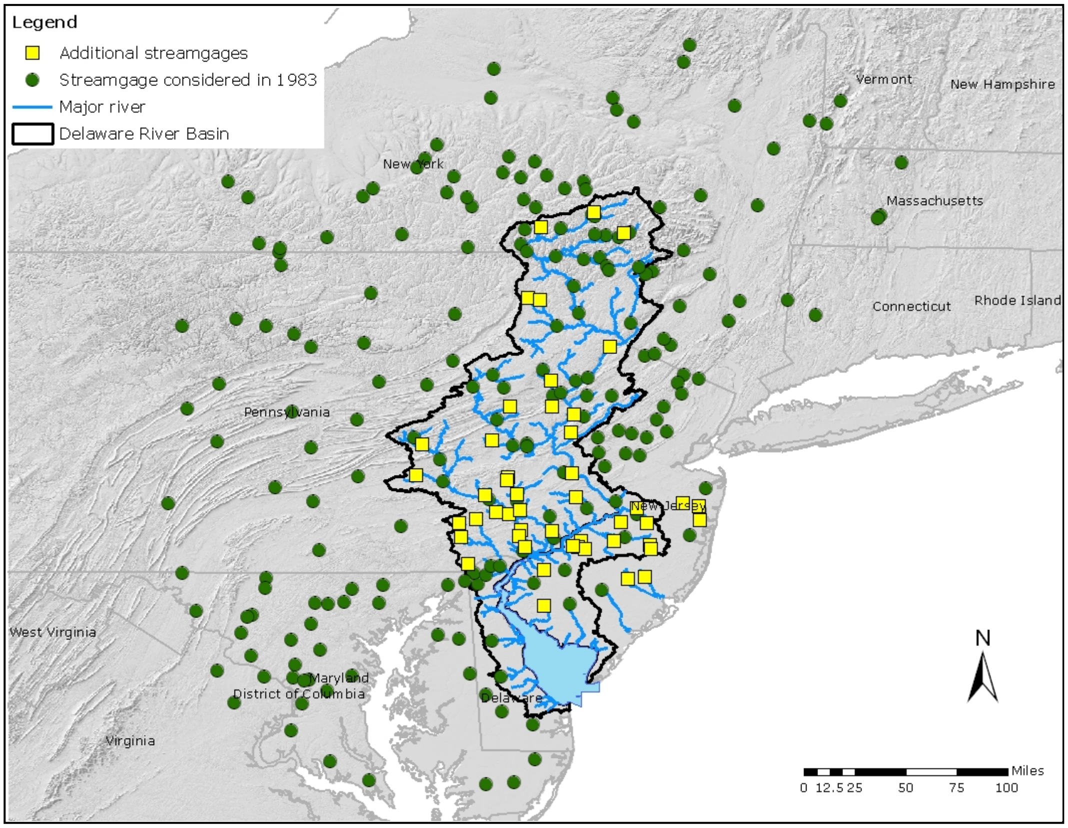

Figure 11. Locations of streamgages considered for analysis

regulation are listed in Table 7. Of the gages identified in the USACE 1983 study, thirteen were also found as being affected by backwater. These thirteen gages, listed in Table 8, were included in the dataset, because no WATSTORE information was found that would merit their removal.

Table 7. Streamgages affected by regulation

17 Gages removed because of regulation (USGS ID, Location) (1)	**15 Regulated gages included (at least twenty-five years of record) (USGS ID, Location) (2)**
01180500 Middle Br Westfield at Goss Heights, MA	01367500 Rondout Creek at Rosendale, NY
01197500 Housatonic near Great Barrington, MA	01397000 South Branch Raritan River at Stanton, NJ
01199000 Housatonic River at Falls Village, CT	01407500 Swimming River near Red Bank, NJ
01200500 Housatonic River at Gaylordsville, CT	01421000 East Br Delaware River at Fishs Eddy, NY
01332500 Hoosic River near Williamstown, MA	01426500 West Br Delaware River at Hale Eddy, NY
01402000 Millstone River at Blackwells Mills, NJ	01431500 Lackawaxen River at Hawley, PA
01451000 Lehigh River at Walnutport, PA	01437000 Neversink River at Oakland Valley, NY
01456000 Musconetcong near Hackettstown, NJ	01450500 Aquashicola Creek at Palmerton, PA
01457000 Musconetcong River near Bloomsbury, NJ	01453000 Lehigh River at Bethlehem, PA
01469500 Little Schuylkill River at Tamaqua, PA	01467000 North Branch Rancocas at Pemberton, NJ
01481000 Brandywine Creek at Chadds Ford, PA	01470500 Schuylkill River at Berne, PA
01518000 Tioga River at Tioga, PA	01471000 Tulpehocken Creek near Reading, PA
01520000 Cowanesque near Lawrenceville, PA	01472000 Schuylkill River at Pottstown, PA
01520500 Tioga River at Lindley, NY	01473000 Perkiomen Creek at Graterford, PA
01548000 Bald Eagle Cr at Beech Creek Station, PA	01474500 Schuylkill River at Philadelphia, PA
01574000 W Conewago Creek near Manchester, PA	
01574500 Codorus Creek at Spring Grove, PA	

Table 8. Streamgages affected by backwater

USGS ID (1)	**Location (2)**
01368000	Wallkill River near Unionville, NY
01369000	Pochuck creek near Pine Island, NY
01369500	Quaker Creek at Florida, NY
01370000	Wallkill River at Pellets Island Mountain, NY
01379000	Passaic River near Millington, NJ
01400500	Raritan River at Manville, NJ
01459500	Tohickon Creek near Pipersville PA
01500500	Susquehanna River at Unadilla NY
01445000	Pequest River at Huntsville NJ
01483500	Leipsic River near Cheswold DE
01497500	Susquehanna Creek at Colliersville NY
01499000	Otego Creek near Oneonta NY
01446000	Beaver Brook near Belvidere NJ

Of the 67 additional gages, fifty had at least twenty-five years of unregulated record. These fifty gages were added to the dataset. Their locations are shown in Figure 11.

3.1.4 Effects of Urbanization

Seven gages were identified as having possible effects of urbanization. Gages with ten percent or greater of the drainage area urbanized were identified as having urbanization effects. Values of urbanization were obtained from a report by the USGS entitled "*Analysis of Flood-Magnitude and Flood-Frequency for Streamflow-Gaging Stations in the Delaware and North Branch Susquehanna River Basins in Pennsylvania*" (USGS 2007). Of these seven gages, five are within the Delaware River Basin, and the other two are within twenty-five miles of the basin. These seven gages, listed in Table 9, were removed from the dataset yielding a total of 241 streamgages for comparison.

Table 9. Streamgages with greater than Ten Percent of their watershed urbanized

USGS ID (1)	Location (2)
01440300	Mill Creek at Mountainhome, PA
01452500	Monocacy Creek at Bethlehem PA
01465500	Neshaminy Creek near Langhorne PA
01473900	Wissahickon Creek at Fort Washington, PA
01477000	Chester Creek near Chester PA
01534000	Tunkhannock Creek near Tunkhannock PA
01538000	Wapwallopen Creek near Wapwallopen PA

3.1.5 Study Boundaries

The study boundaries defined by the USACE 1983 study were that streamgages had to be within 25 miles of the Delaware River Basin. Gages further from the basin boundary are not believed to be as representative of the basin as gages within it. There were 163 gages (of the 241 gages) identified that met this criterion.

3.1.6 Summary of Data Quality

After determining record lengths, comparing gage sample moment statistics, assessing trends, and assessing effects of regulation and urbanization, one can be confident in the quality of data for 115 gages of the original 215 and for 48 additional gages. These 163 gages include 105 gages within the Delaware River Basin and 48 gages that have a majority of their watershed within 25 miles of the Delaware River Basin. Each station has at least t years of unregulated annual peak flows whose records are considered absent of both tidal and anthropogenic effects.

These 163 gages, which are shown in Figure 12 and listed in Appendix B, were used to complete the regional skew analysis.

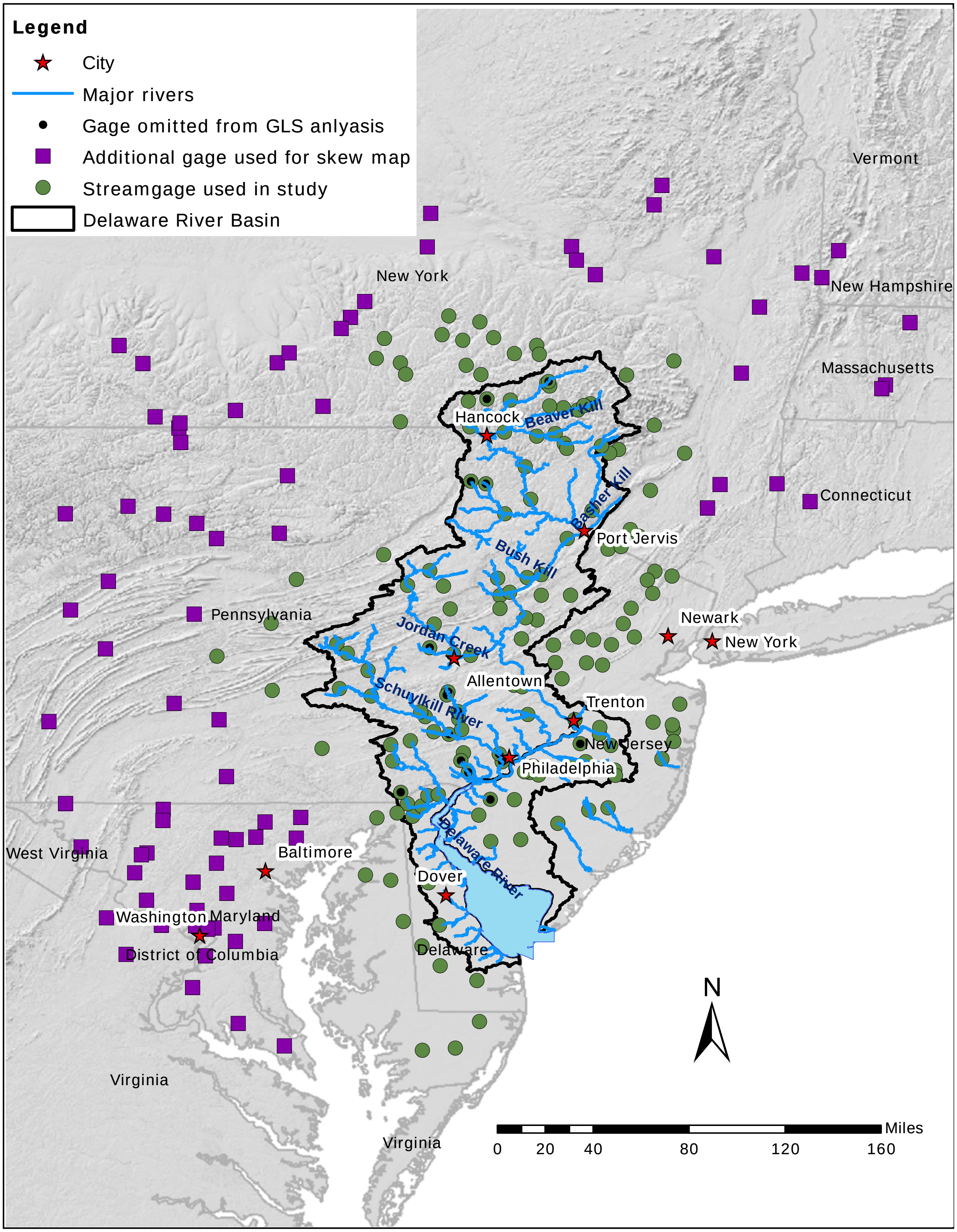

Figure 12. Streamgages used in study

For the remaining 78 gages information was used that also meet the data quality criteria described above, and that lie further outside the Delaware River Basin, in completing the skew isoline map (Method 2). These additional gages are shown in Figure 12. These gages are included because information they provide is useful in drawing contours near the basin's border. However, they were not included in error calculations.

An additional sixteen gages were omitted from the GLS regression analysis because of an incomplete set of parameters. However, these gages were retained in the other analysis methods. These omitted gages are shown in Figure 12.

3.2 Station Skew

The station statistics for the 163 gages used in this analysis are tabulated in Appendix C. Also in Appendix C are the statistics for the 78 additional gages considered.

All instances of historical information, as indicated by WATSTORE Code 7, were initially used in calculating station statistics for the thirty-five streamgages with historical information. All gages were then checked where their statistics resulted in an increase in the occurrence of high outliers when historical information was incorporated. Low historical flow values can cause large flows in the systematic record to be weighted as historical information in calculating station statistics. In cases where a large number of high outliers occurred and the historical information was within five years of the start of the systematic record, the historical information was treated as systematic.

For cases where the historical information occurred less than five years from the systematic record, the effect of removal of the historical information on station statistics was examined, with an emphasis on station skew.

In most cases, removing historical information from the calculation of station statistics did not change the values significantly, and thus the historical information was still incorporated. In the single case where inclusion of historical information resulted in a large increase in station skew, in addition to eight high outliers, the historical information was treated as systematic in the calculation of station statistics.

MSE for stations with historical information was computed using a simulation method (detailed in Appendix D). This method represents an improvement over the *Bulletin 17B* recommended method which uses the full historical record length in computing MSE.

3.3 Regional Average Skew (Method 1)

Averaging all 163 station skews into a single region (Method 1a) results in a regional average skew of 0.184, an MSE of 0.142 (without sampling error), and a simulated MSE (including sampling error) of 0.241. When considering only gages within the Delaware River Basin boundary, regional average skew is 0.217 with MSE of 0.155. The remainders of the Method 1 approaches (1b, 1c, and 1d) address skew in separate approximately homogeneous subregions within the Delaware Basin.

3.3.1 Region Development: L-Moment Analysis

The procedure for L-moment analysis outlined in "*Regional Frequency Analysis*" (Hosking and Wallis 1997) was used to define multiple statistically homogeneous regions for averaging skew (Methods 1b, 1c and 1d). The procedures for averaging station skew outlined in *Bulletin 17B* did not provide a statistical method for defining a region. Hosking et al's L-moment analysis checks a region's homogeneity using its L-moment statistics. Procedures for verifying homogeneous regions using L-moments are found in Appendix D.

Initially regions were formed based on smaller river basins within the study area. Nominally, the river basins were identified in a north-south direction as a measure of distance from the coast. Forming regions based on distance from the coast has the advantage of grouping gages with similar elevation, and with similar influence from hurricanes, events which may significantly affect station skew values. Additionally, the size and number of the river basins used were selected to each have a significant number of gages. Hosking and Wallis (1997) recommend having at least twenty gages in a region for identifying a candidate flood frequency distribution. However, in this application, the goal was to define regions where the gage flow frequency distribution has similar shapes, particularly for more infrequent quantiles, such as 1% chance exceedance (100-year return period). The number-of-gages criterion was consequently relaxed, and the H(3) statistic was used to focus on similar shape.

Table 10 shows the results of discordancy and heterogeneity statistics for the aggregations of gages forming each region. Region 1, which is generally to the east of the Delaware River Basin, has the most gages. The Delaware River Basin was the focus of forming regions, so a great deal of detail was not needed in examining the out-of-basin gages. All the gage regions were found to be acceptably homogenous given the low to moderate heterogeneity values. Only the most northern region - outside of the Delaware River Basin - has two highly discordant values. This is considered acceptable given the relatively large size of this region (Hosking and Wallis 1997). Final regions are shown in Figure 13.

Table 10. Statistical test results for L-moment-defined regions

Region Number (1)	Number of Gages (2)	Heterogeneity (3)	Number of Moderately Discordant Gages (4)	Number of Highly Discordant Gages (5)
1	37	Moderate	1	2
2	18	Low	0	0
3	17	Moderate	0	0
4	12	Moderate	0	0
5	14	Low	0	0
6	16	Low	0	0
7	15	Low	0	0
8	13	Low	1	0
9	9	Low	1	0
10	6	Low	0	0
11	6	Moderate	1	0

The areal extents of the L-moment-verified regions depicted in Figure 13 are based on the watersheds in which the included stream gages are located. For regions outside the Delaware

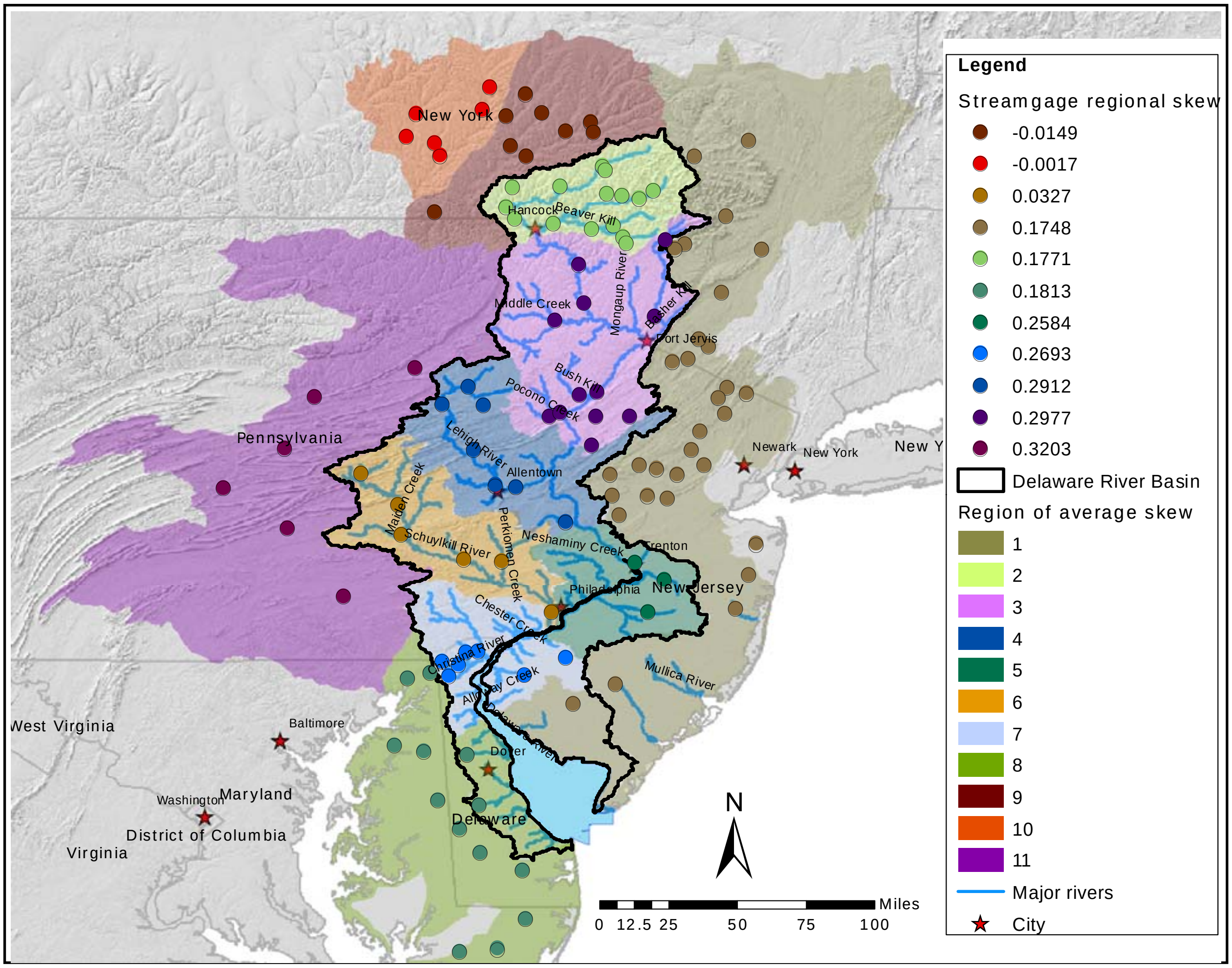

Figure 13. L-moment-defined regions of average skew

River Basin, the areal extents were based on watersheds defined by the USGS using the eight-digit hydrologic unit code (HUC8) in which the region resides.

3.3.2 Region Average Skew (Methods 1b and 1c)

For the regions verified using L-moment analysis described in Section 3.3.1, region skews were computed as an average of the gages in the region (Method 1b), and as a weighted average of the gages, weighted by the gage record length (Method 1c) to provide more weight to longer record stations. For comparison between methods, a single Delaware Basin skew was then computed for each as a weighted average of the region skews (each region weighted by the number of gages in the region).

For Method 1b, the weighted average of all regions is 0.181, the record MSE is 0.133, and the simulated MSE is 0.232. The weighted average skew for gages within the Delaware River Basin boundary is 0.221, the record MSE is 0.146, and the simulated MSE is 0.251. Table 11 lists the region-specific skews and MSE. For Method 1c, the weighted average of all regions is 0.176, the record MSE is 0.117, and the simulated MSE is 0.203. The weighted average skew for gages within the Delaware River Basin boundary is 0.191, the record MSE is 0.130, and the simulated MSE is 0.220. Table 12 lists the region-specific skews and MSE.

Table 11. Average skew coefficients of L-moment-defined regions (Method 1b)

Region Number (1)	Number of Gages (2)	Region Skew (3)	Record MSE (4)	Simulated MSE (5)
1	35	0.175	0.113	0.205
2	18	0.177	0.099	0.202
3	16	0.298	0.157	0.254
4	11	0.291	0.148	0.241
5	12	0.258	0.147	0.260
6	15	0.033	0.242	0.331
7	14	0.269	0.087	0.221
8	12	0.181	0.187	0.273
9	9	-0.015	0.078	0.175
10	6	-0.002	0.027	0.147
11	4	0.320	0.130	0.216
All Regions	152	0.184	0.133	0.232
Delaware River Basin	86	0.217	0.146	0.251

Table 12. Weighted-average skew coefficients of L-moment-defined regions (Method 1c)

Region Number (1)	Number of Gages (2)	Region Skew (3)	Record MSE (4)	Simulated MSE (5)
1	35	0.212	0.097	0.175
2	18	0.101	0.079	0.158
3	16	0.310	0.168	0.262
4	11	0.275	0.120	0.191
5	12	0.199	0.120	0.219
6	15	0.012	0.207	0.278
7	14	0.289	0.086	0.212
8	12	0.157	0.176	0.258
9	9	-0.030	0.061	0.136
10	6	0.016	0.020	0.132
11	4	0.308	0.112	0.189
All Regions	152	0.176	0.117	0.203
Delaware River Basin	86	0.191	0.130	0.220

3.3.3 GLS-Constant Region Average Skew (Method 1d)

In Method 1d, skew for each region is developed with a constant-only GLS regression on basin parameters. To determine an MSE-equivalent measure, the GLS regression approach (detailed in Appendix D) disaggregates the error in estimating the gage skew with a regional estimate as:

gage skew = *regional skew* + (*model error and time sampling error*)

The time sampling error is the error due to having a limited period of record to estimate gage skew. The model error measures the error in predicting regional skew with a regression relationship even if there is no time sampling error in the gage skew values. The GLS regression for regional skew would typically include a regression constant plus independent parameters such as drainage area and mean annual precipitation. However, only a constant was used in this GLS-constant regional average approach, detailed herein.

The constant in the regression equation provides a direct comparison with the regional average, which is obtained using standard methods outlined above. The GLS method considers both the difference in sampling error and the correlation between gaged annual peak flows when weighting gage skew estimates to obtain a regional skew. This weight is different than the equal weighting that gage skew estimates are given when computing the regional skew as an average of gage skew values as recommended in Bulletin 17B. As in the GLS regression with independent parameters, the AVP is used as a measure of the prediction error, instead of the simulated MSE. (Simulated-MSE and AVP can be directly compared in this study, as presented in Appendix D.) The AVP is the average squared prediction error (model error and time sampling error) obtained when using only the regression constant as an estimate of regional skew.

The skew and AVP values found for the Delaware River Basin regions are shown in Table 13. Alternatively, the weighted average constant of 0.151 could be used as the regional skew value. The weighted AVP of 0.044 could be used in place of MSE. While the regions defined in Method 1b and Method 1c are the same, some gages are omitted because of incomplete parameter sets.

Table 13. GLS-constant region average skew and errors for regions within the Delaware River Basin boundary (Regions 2 through 7)

Region Number (1)	**Constant (GLS-Constant Regional Skew) (2)**	**MSE (from AVP, Average Prediction Error) (3)**	**Model Error (4)**
2	0.087	0.026	0.000
3	0.203	0.077	0.052
4	0.165	0.030	0.013
5	0.178	0.033	0.000
6	0.001	0.064	0.032
7	0.287	0.034	0.000
Weighted average	0.151	0.044	0.000

A troublesome aspect of the results is that a model error of zero was estimated for some of the regions. This could be interpreted as meaning that a constant skew value is in fact a perfect model for the region. Alternatively, this could have resulted due to the approximations made in computing the time sampling error. The GLS regression application resulted in zero model error in most of the applications is defined in Appendix D.

The summary of results displayed in Table 14 demonstrates that the GLS-constant procedure would be selected given the *Bulletin 17B* criteria. This is a significant finding in that an AVP of 0.044 is an order of magnitude lower than would result from any application using the *Bulletin 17B* methods. For example, the national regional skew map, *Bulletin 17B (see Plate 1)*, has an MSE of 0.302. The Delaware River Basin MSE value in this study is 0.155. Accepting the GLS-constant AVP would give the regional skew a much greater weight in the *Bulletin 17B* adopted skew calculation than has been used typically.

Table 14. Regional skew results for regions within Delaware River Basin boundary

Method (1)	Regional Skew (2)	Associated Error (MSE or AVP) (3)
Method 1a (Bulletin 17B)	0.217	0.259
Method 1b (average regional skew)	0.217	0.251
Method 1c (weighted regional skew)	0.191	0.220
Method 1d (GLS-constant)	0.151	0.044

3.4 Skew Isoline Map (Method 2)

To develop the skew isoline map, the resulting station skew values were first inspected to see if a pattern could be detected. This was done by binning the data into seven different groups, and plotting the station skew coefficients with symbols based on their bins, as shown in Figure 14. Bins were sized using natural breaks that were then rounded to the nearest tenth of station skew. While skews tend to be positive and have larger magnitudes in the south of the region, no definitive pattern was identified.

Skew contours using both linear and IDW interpolation were developed. Results are shown in Figure 15 and Figure 16, respectively. The methods used to calculate these contours result in an exact solution at the streamgage locations where the station skews were plotted. Therefore MSE is computed as zero for the maps. Despite the fact that MSE equals zero for these two contour methods, neither method is recommended for a regional skew map. The maps are simply a mathematical fit to the data and do not represent any identifiable behavior of streamflow in the basin.

A third version of the skew map was created, considering information from the IDW interpolation, physiography, and hydrology. This improved map, shown in Figure 17, is the result of contour lines modified using judgment based primarily on elevation and a reduction of local extremes from those initially calculated in the IDW method described above. The MSE for this skew map is 0.147. This map has no negative skew contours because within the Delaware River Basin boundary, stations with negative skews were localized extremes, indicated by a series of tightly spaced contours around that gage.

3.5 Generalized Least Squares Regression (Method 3)

3.5.1 Regression Results

Ordinary least squares (OLS) regression (Draper and Smith 1966) is one of the three methods recommended in *Bulletin 17B* for analyzing regional skew. However, this regression approach does not account for differences in the sampling error in gage skew estimates, i.e., the unequal error estimation due to differences in gage record length, nor does it account for the inter-gage correlation of gage annual peak discharges. GLS regression techniques have been developed to account for the sampling error and correlation issues in estimating regional skew. The results of each method are provided for completeness. Appendix B lists the basin parameters considered in the regression, and Appendix D describes the OLS and GLS methods in more detail.

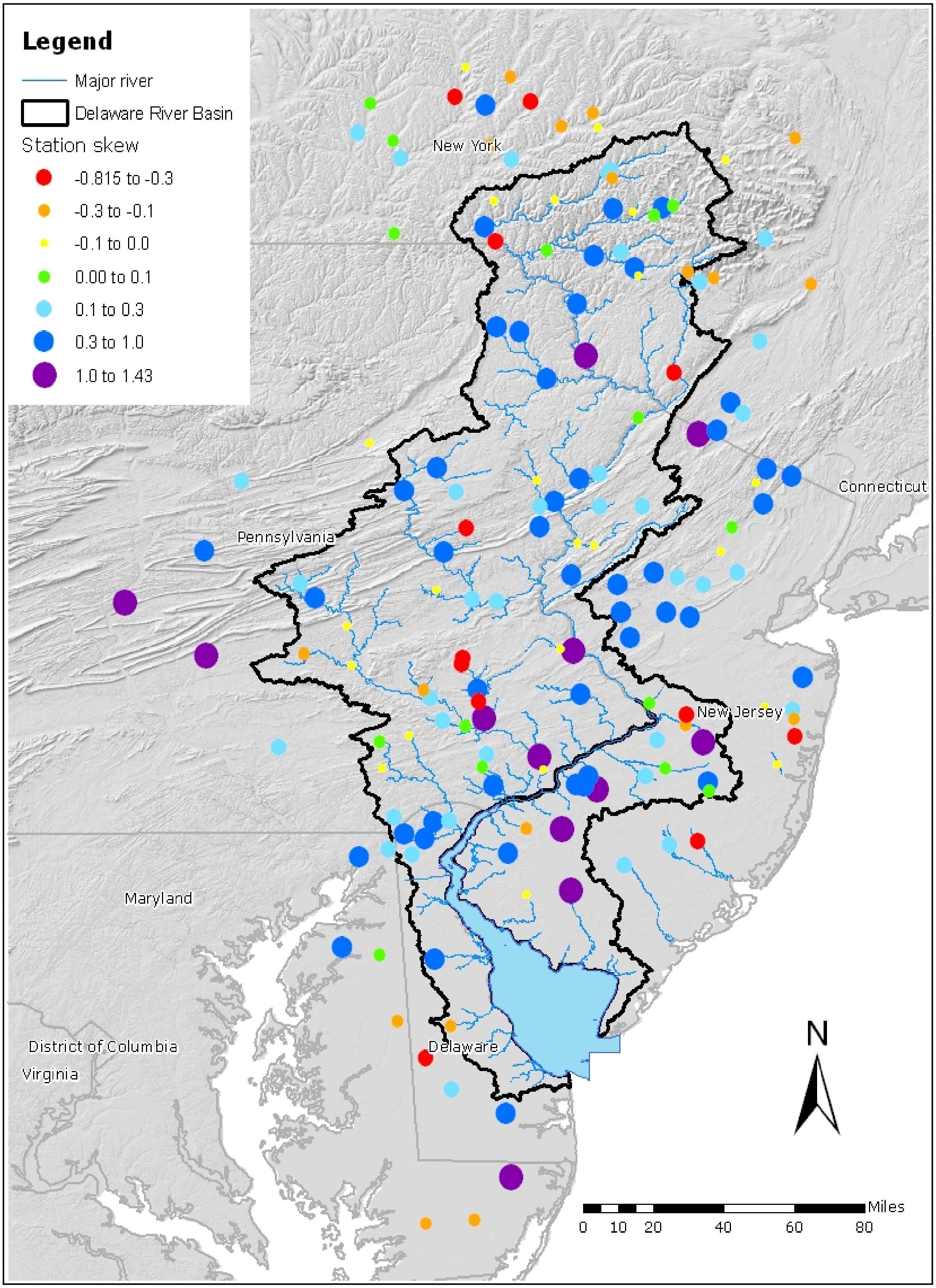

Figure 14. Binned station skew coefficients

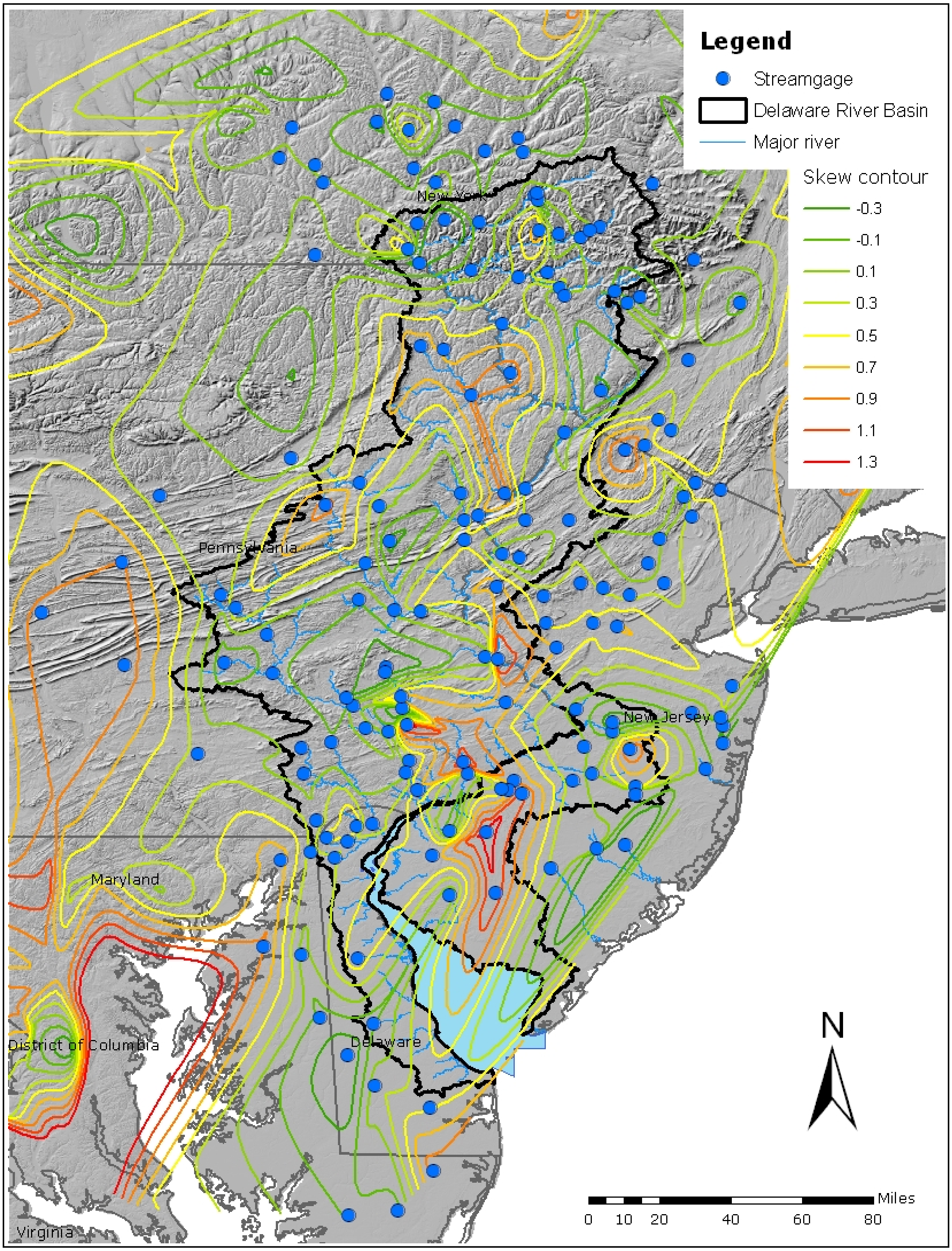

Figure 15. Linearly interpolated skew contours

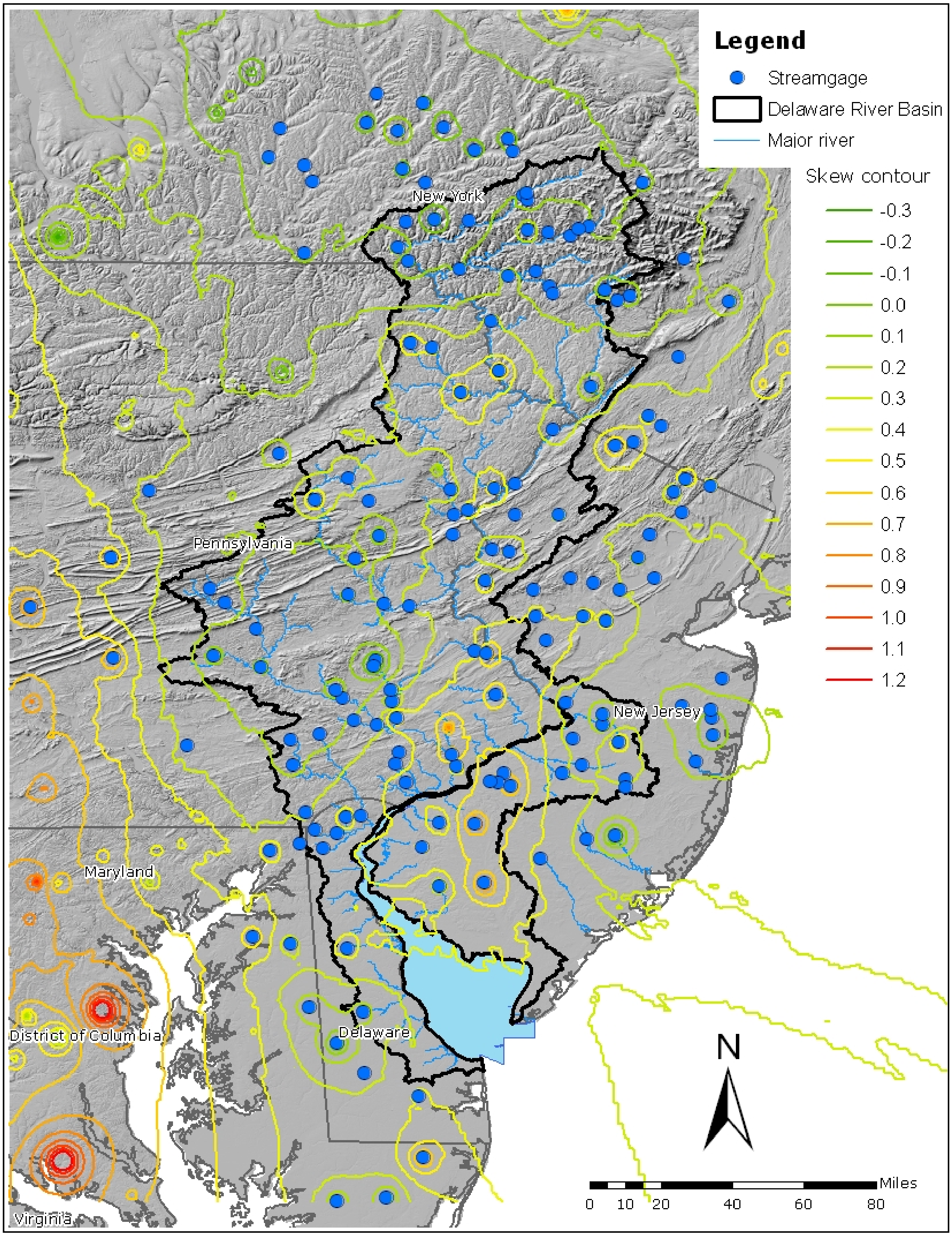

Figure 16. IDW interpolated skew contours

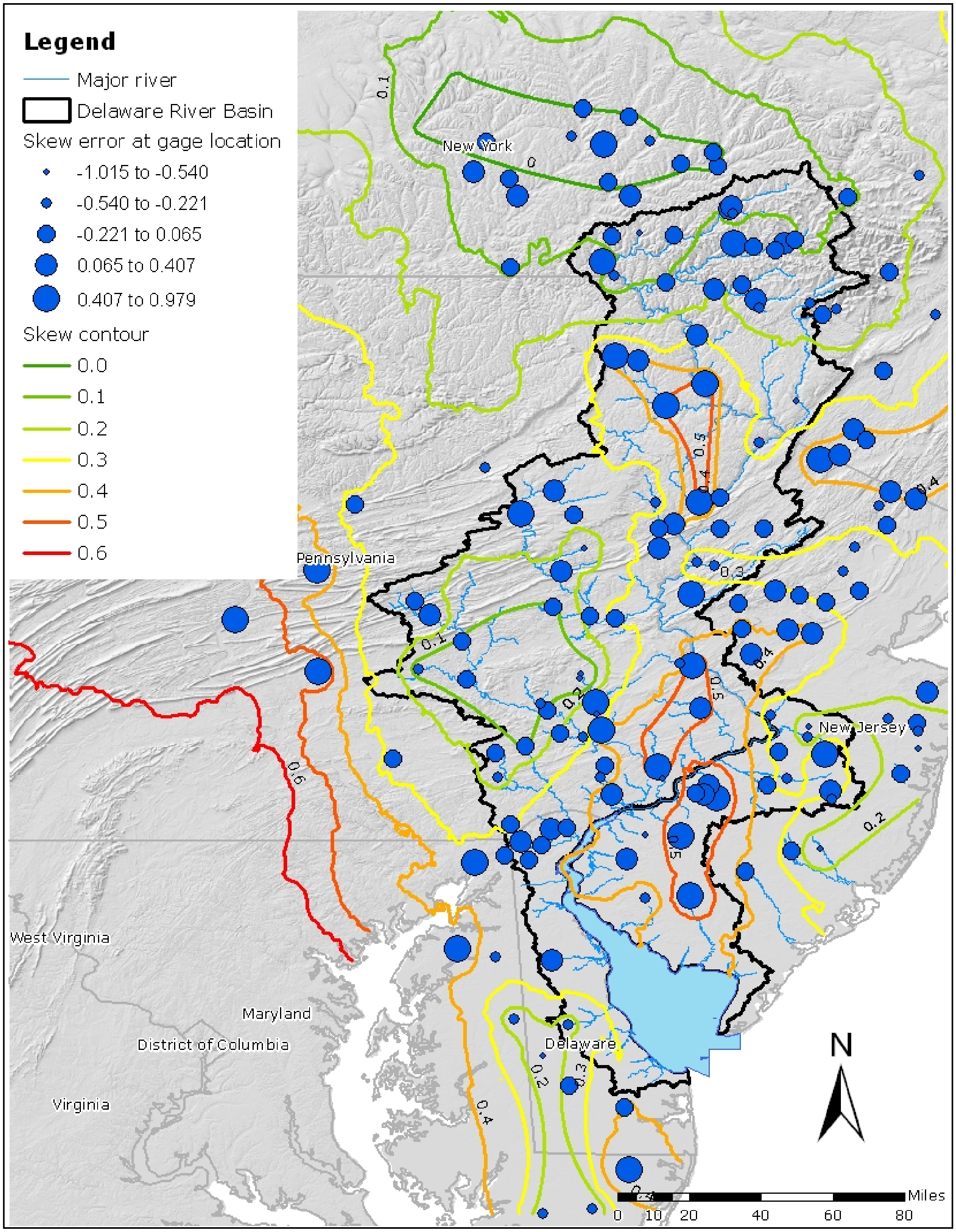

Figure 17. Judgment-edited skew map of the Delaware River Basin

Table 15 and Table 16 provide comparisons between OLS and GLS regression results. Notice that AVP is the sum of the model error and the sampling error (i.e., an error that is a function of the record length used to compute skew). Notice also that traditionally the standard error squared of the OLS would be used as an estimate of MSE of the region predicted by the

Table 15. Regression comparison using mean basin elevation for the Northern Delaware River Basin

Regression Type (1)	Standard Error^2 (2)	Adjusted R^2 (3)	Model Error^2 (4)	Sampling Error (5)	AVP (6)
OLS	0.1379	-0.0485	0.1379	0.0103	0.1482
GLS			0.0079	0.0191	0.0271

Table 16. Regression comparison using mean annual precipitation for the Southern Delaware River Basin

Regression Type (1)	Standard Error^2 (2)	Adjusted R^2 (3)	Model Error^2 (4)	Sampling Error (5)	AVP (6)
OLS	0.1436	-0.0345	0.1436	0.0093	0.1528
GLS			0	0.0189	0.0189

regression, whereas in GLS, the AVP is used. Clearly, the AVP is the smaller of the two values. Application of the GLS result would lead to a much greater weighting of the regional skew in computing the adopted skew in the *Bulletin 17B* methodology.

Table 17 through Table 24; provide the best OLS and GLS regressions, and their associated errors, including a regression using the maximum number of independent variables. Pseudo-R^2 (see Appendix D) values of -999 indicate that model error was not identified (i.e., was set equal to zero).

Regression equations for the Northern and the Southern Delaware River Basin regions were developed. The regression using mean basin elevation in the North was chosen partly because a model error could be estimated and it resulted in a physically reasonable, positive regression coefficient. The negative pseudo-R^2 detracts from the result because it shows that a constant alone provides a better explanation of the skew variance in comparison to using mean elevation. The South equation provides a physically reasonable, positive coefficient for mean annual precipitation; however, no model error could be defined for this region, and, thus, pseudo-R^2 could not be estimated. AVP values for both equations are comparable to other results. Furthermore, the results have physically plausible interpretation. Skew coefficients for Southern basins, closest to the coast, are perhaps more affected by hurricane frequency, leading to a regression with mean annual precipitation. The significance of this regression cannot be qualified given that no model error could be estimated for this equation. The gages in the Northern basin are farther from the coast, but have higher elevation. This might mean that skew is more related to orographic effects influencing thunderstorm rainfall, leading to a regression with mean basin elevation.

3.5.2 Regression Analysis

In the GLS regression analysis, model error could not be identified for any regions tested except one. As Tasker and Stedinger (1989) note, when solving for the GLS regression coefficients,

Table 17. OLS regression results for all gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ $slope^{0.5}$] (11)
0.171	0.357	0.016	0.838	-0.100	-0.141	-0.707	0.158	-0.018	0.195	-0.317
0.165	0.356	0.024	0.612	-0.099		-0.717	0.154	-0.018		-0.312
0.158	0.354	0.031	0.634	-0.084		-0.674		-0.010		-0.318
0.152	0.353	0.037	0.649	-0.083		-0.695	0.156			-0.319
0.147	0.353	0.037	0.690			-0.669	0.075			-0.224
0.141	0.353	0.039	0.754			-0.723				-0.136
0.138	0.356	0.021	0.330							-0.130

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 18. GLS regression results for all gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ $slope^{0.5}$] (10)
0.0104	-999	-0.0093	0.0100	0.2102	-0.4088	0.0630	0.0406	-0.0541	-0.1214
0.0087	-999	-0.0153		0.2090	-0.4086	0.0671		-0.0425	-0.1257
0.0078	-999	0.0034		0.1949	-0.4151	0.0648	0.0386		-0.1230
0.0069	-999	-0.1823		0.3465	-0.4773		0.0409		-0.0496
0.0060	-999	0.0058		0.2220	-0.4643				-0.0375
0.0051	-999	0.3662			-0.4494				-0.0388
0.0044	-999	0.3183							-0.4554

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 19. OLS regression results for Delaware River Basin gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ slope$^{0.5}$] (11)
0.213	0.130	0.021	2.391	-0.338	-1.215	-0.798	0.399	0.008	1.094	-0.671
0.201	0.128	0.036	2.348	-0.344	-1.192	-0.815	0.396		1.157	-0.668
0.190	0.127	0.045	0.375	-0.357		-0.821	0.372		1.172	-0.631
0.180	0.127	0.040	-0.040	-0.377			0.449		0.833	-0.721
0.170	0.128	0.039	0.182	-0.315			0.443			-0.723
0.163	0.132	0.006	0.813			-0.853				-0.113
0.153	0.132	0.003	0.620			-0.724				

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 20. GLS regression results for Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.0104	-999	-0.0093	0.0100	0.2102	-0.4088	0.0630	0.0406	-0.0541	-0.1214
0.0087	-999	-0.0153		0.2090	-0.4086	0.0671		-0.0425	-0.1257
0.0078	-999	0.0034		0.1949	-0.4151	0.0648	0.0386		-0.1230
0.0069	-999	-0.1823		0.3465	-0.4773		0.0409		-0.0496
0.0060	-999	0.0058		0.2220	-0.4643				-0.0375
0.0051	-999	0.3662			-0.4494				-0.0388
0.0044	-999	0.3183							-0.4554

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 21. OLS regression results for northern Delaware River Basin gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ slope$^{0.5}$] (11)
0.263	0.324	0.201	3.334	-0.699	-1.775	-1.935	0.538	-0.016	3.389	-0.829
0.241	0.319	0.224	3.479	-0.669	-1.841	-1.897	0.533		3.190	-0.829
0.222	0.320	0.221	0.436	-0.657			-1.973	0.465	3.318	-0.718
0.205	0.327	0.186	2.228	-0.447	-1.064	-2.321			4.231	
0.184	0.324	0.200	0.474	-0.456		-2.335			4.218	
0.167	0.328	0.182	0.813			-2.330			2.723	
0.192	0.371	-0.049	0.157	-0.060			0.039			

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 22. GLS regression results for northern Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.0315	-999	1.4804	-0.3864	-0.544	-1.5281	0.3874	0.0146	1.7104	-0.6337
0.0264	-999	2.2115		-0.8683	-1.4108	0.3525	0.0364	0.6808	-0.6004
0.0231	-999	2.2291		-0.9036	-1.4765	0.3599		0.9337	-0.6021
0.0201	-999	0.7586			-1.565	0.3261		1.0022	-0.5445
0.0178	-999	0.0897				0.4455		0.0378	-0.7169
0.0156	-999	1.0246			-1.5741	0.0048			
0.0259	-2.043	0.0991	0.1264			0.0144			
0.0132	-999	1.0363			-1.5773				
0.0205	-0.915	0.1253	0.1388						

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 23. OLS regression results for southern Delaware River Basin gages

AVP (1)	Standard error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ $slope^{0.5}$] (11)
0.331	0.388	-0.086	-17.279	-0.561	9.996	1.557	0.247	-0.013	-0.502	-0.371
0.305	0.380	-0.041	-17.628	-0.566	10.183	1.644	0.272	-0.602		-0.398
0.280	0.375	-0.011	-16.506	-0.554	9.525	1.393	0.266			-0.403
0.254	0.369	0.019	-16.703	-0.488	9.778	1.004		-0.053		
0.228	0.365	0.041	-17.364	-0.463	10.209	0.931				
0.202	0.361	0.061	-16.577	-0.542	10.011					
0.184	0.370	0.016	0.098	-0.217						
0.190	0.379	-0.035	0.069		0.087					

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table 24. GLS regression results for southern Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ $slope^{0.5}$] (10)
0.060	-999	-11.343	-0.223	6.415	1.570	0.168	0.082	-0.672	-0.196
0.052	-999	-6.952		3.892	1.399	0.063	0.071	-0.555	-0.078
0.044	-999	-5.841		3.262	1.106	0.009	0.046		-0.023
0.037	-999	-5.162		2.910	0.959	-0.049			0.046
0.030	-999	-5.786		3.228	1.072		0.053		
0.024	-999	-0.383			0.924		0.043		
0.018	-999	0.154							-0.037
0.019	-999	-2.673		1.684					

a. Average Natural Resource Conservation Service soil storage parameter in inches.

"The required estimate of [d^2]…" exists "…if a positive solution for [d] exists. Otherwise, [d]=0". (Note: d is a symbol substituted for the referenced symbol for the square root of the model regression error.)

Estimating d equals zero makes the math work, but it is not the best assumption given the nature of statistical prediction in the hydrologic sciences. It is therefore unlikely that a simple predictive relationship between basin physical parameters and the skew coefficient can be error-free. If the data were divided, or additional observations were added to the problem, it is unlikely that all the regressions would have zero model error. However, the Delaware River Basin skews exhibit the phenomenon of d equals zero. A possible reason for this is that the model used for estimating skew sampling error is not correct. If true, this leads to a poor estimate of the covariance matrix in the GLS formulation, precluding an estimate of d^2 (see Appendix D).

Another problem is that every regression examined had a significant number (five to ten) of gages with large leverage or disproportionate influence. Attempts to create a data set without these data points were not successful even when deleting half the gages in a region. This result points out the problem in assessing the adequacy of gage coverage in a regional analysis. Examination of a map of gages is not sufficient. Measures of leverage and influence are needed to evaluate how the gages cover a parameter space. In GLS regression, the measures of leverage and influence also consider the impact of not observing skew equally due to record length sampling errors. Apparently, the gage coverage in the study area is not as complete as hoped, even though the number of gages is considered adequate by the criteria provided in *Bulletin 17B*.

Another problem is that every regression examined had a significant number (five to ten) of gages with large leverage or disproportionate influence. Attempts to create a data set without these data points were not successful even when deleting half the gages in a region. This result points out the problem in assessing the adequacy of gage coverage in a regional analysis. Examination of a map of gages is not sufficient. Measures of leverage and influence are needed to evaluate how the gages cover a parameter space. In GLS regression, the measures of leverage and influence also consider the impact of not observing skew equally due to record length sampling errors. Apparently, the gage coverage in the study area is not as complete as hoped, even though the number of gages is considered adequate by the criteria provided in *Bulletin 17B*.

A model error was sought for all regressions included in the study; all Delaware River Basin gages, and a north/south division of Delaware River Basin gages. All combinations of regression parameters were investigated for each region.

3.6 Split-Sample Testing

Split-sample tests were completed for the p=0.1 (10-year), p=0.02 (50-year), and p=0.01 (100-year) exceedance quantiles. Testing was completed for 4,600 years of record, thus the number of records, *n*, in a split-sample was 2,300. In this study, the split-samples divided the samples using the following methods:

- **Forecast method:** first half of the record is used to estimate frequency curve, remaining record is reserved.

- **Back cast method:** second half of record is used to estimate frequency curve, remaining record is reserved.

- **Alternating Method 1:** alternate years in the record are used to estimate the frequency curve, remaining record is reserved.

- **Alternating Method 2:** the reserved record in Alternating Method 1 is used to estimate the frequency curve and the remaining data is now the reserved.

The log-Pearson III frequency curve estimation methods explored are:

- Expected probability - no regional skew.

- Computed probability - no regional skew.

- Computed probability using regional skew and mean square error estimates from Method 1b (region average skew).

- Computed probability using regional skew and average prediction error from Method 1d (GLS-constant, Table 13).

Expected probability gives the mean estimate of exceedance probability for normal or log-normal frequency curves estimated from the sample mean and standard deviation. The estimates of probability are approximate for the log-Pearson III frequency curve (*Bulletin 17B*).

The expected probability estimate was used because it is theoretically unbiased with regard to predicting future exceedance values. Expected probability's use has no impact on the comparison to regional skew methods in the testing. This is because each estimate would change proportionally the same if expected probability instead of computed probability was used. However, expected probability is not a universally accepted estimator, and therefore computed probability was also used. The computed probability approximates the median estimate of exceedance probability. This probability is obtained directly from the log-Pearson III sample mean, standard deviation, and skew (see *Bulletin 17B*).

The regional skew methods were chosen to represent both the area average (Method 1b) and GLS-constant (Method 1d) in the split-sample testing. Note that the computation with regional skew differs in that an adopted skew (a weighted average of the station and regional skew values) is used in computing the regional skew gage frequency.

Table 25 through Table 27 show the ratio of the observed flows exceeding the flow quantile to the total number of observations for the given log-Pearson III frequency curve estimation method. The split-sample testing procedure is detailed in Appendix D.

The results show that the expected probability estimate provides the greatest correspondence between observed and predicted future exceedance. However, the sampling error in the observed estimates of future exceedance (the proportion of values observed to be greater than the predicted exceedance level, for example, p=0.1) is larger than the difference in predictions between the methods. The standard error in the proportion-of-exceedance value can be computed as the standard error of the proportion obtained from the binomial distribution $(p(1-p)/n)^{0.5}$, where p is the exceedance probability and n is the number of years of record in the split-sample of gage records. For example, the differences between predictions for the p=0.01 exceedance probability in Table 27 are not different when considering two standard errors in the proportion of 2(0.002) = 0.004.

The expected probability comes closest to producing the expected proportion in the majority of tests (ratios closest to one). However, the observed proportion exceeds the expected proportion in most cases, as is true for the computed probability scenarios (ratios greater than one).

The computed probability scenarios generally agree within the standard errors shown for the expected proportion (the expected proportion being equal to the stated test exceedance probability). This indicates that regional skew provides no advantage in predicting future exceedance values, or equivalently, future flood risk.

Table 25. Split-Sample Testing, p=0.1 (10-year)

Split-Sample Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.131	1.310	0.142	1.420	0.141	1.410	0.142	1.420
Back cast	0.087	0.870	0.092	0.920	0.092	0.920	0.093	0.930
Alternate 1	0.111	1.110	0.125	1.250	0.124	1.240	0.126	1.260
Alternate 2	0.076	0.760	0.084	0.840	0.083	0.830	0.085	0.840

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.1 (10-year) gage flows values.
d. Proportion x 10.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5} = 0.006$, where $p=0.1$, $n=2300$.

Table 26. Split-Sample Testing, p=0.02 (50-year)

Split-Sample Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.035	1.740	0.042	2.080	0.042	2.080	0.048	2.390
Back cast	0.022	1.110	0.027	1.370	0.023	1.150	0.025	1.240
Alternate 1	0.028	1.410	0.040	2.000	0.037	1.870	0.038	1.890
Alternate 2	0.015	0.770	0.022	1.090	0.018	0.900	0.020	1.000

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.2 (50-year) gage flows values.
d. Proportion x 50.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5} = 0.003$, where $p=0.2$, $n=2300$.

Table 27. Split-Sample Testing, p=0.01 (100-year)

Split-Sample Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.024	2.430	0.028	2.820	0.028	2.780	0.032	3.170
Back cast	0.017	1.650	0.020	2.040	0.017	1.650	0.018	1.780
Alternate 1	0.019	1.870	0.025	2.470	0.024	2.390	0.027	2.730
Alternate 2	0.012	1.150	0.015	1.450	0.012	1.190	0.013	1.280

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.01 (100-year) gage flows values.
d. Proportion x 100.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5} = 0.002$, where $p=0.01$, $n=2300$.

Chapter 4

Conclusions

A regional skew analysis of the Delaware River Basin was completed using the following three methods recommended by *Bulletin 17B*:

- **Method 1:** Region average skew. This method was implemented four ways:
  - **Method 1a** - average skew of the entire basin;
  - **Method 1b** - average skews of homogenous regions;
  - **Method 1c** – weighted average skews of homogeneous regions;
  - **Method 1d** – GLS constant-only regression of homogeneous regions.

- **Method 2:** Skew isoline map.

- **Method 3:** Predictive equations using GLS regression.

Bulletin 17B also recommends using a regional skew value that result in the minimum MSE. Results are summarized in Table 28.

Table 28. Summary of Regional Skew Analysis Results

Method (1)	Associated Regional Skew (average or GLS constant) (2)	Record MSE[a] (3)	Simulated MSE[b] (4)	AVP[c] (5)
Method 1a (entire region)	0.184	0.142	0.241	n/a
Method 1a (Delaware River Basin only)	0.217	0.155	0.259	n/a
Method 1b (entire region)	0.184	0.133	0.232	n/a
Method 1b (Delaware River Basin only)	0.217	0.146	0.251	n/a
Method 1c (entire region)	0.176	0.117	0.203	n/a
Method 1c (Delaware River Basin only)	0.191	0.130	0.220	n/a
Method 1d (Delaware River Basin only)	0.151	n/a	n/a	0.044
Method 2	n/a	0.147	n/a	n/a
Method 3 (northern region)	n/a	n/a	n/a	0.027
Method 3 (southern region)	n/a	n/a	n/a	0.019

a. Calculated using procedures outlined in *Bulletin 17B,* if applicable.
b. Calculated using Monte Carlo simulation to account for sampling error, if applicable.
c. Applicable only for GLS regressions.

Analysis of the Delaware River Basin found that Methods 1a, 1b, 1c and 1d yield simulated MSE (or AVP) smaller than that estimated for the skew map in *Bulletin 17B*. MSE decreases from Method 1a to 1b to 1c, and Method 1d, using AVP in place of MSE, yields a very low value.

Method 2 also yields a smaller MSE than that estimated for *Bulletin 17B*'s map. The value is similar to the various MSE of Method 1. However, the estimate of MSE does not realistically consider the impact of sampling error on gage skew estimates (i.e., computation of a simulated MSE is not possible).

Method 3 yields small AVPs for GLS regression equations for the northern and southern regions of the Delaware River Basin. A small AVP will give the regional skew greater weight in an adopted skew calculation, which in turn promotes less variation in adopted skew coefficients and promotes consistency in flood frequency estimates for a region. However, model error for a majority of the regressions could not be identified, which is not reasonable given the nature of statistical predication in the hydrologic sciences.

Therefore, HEC recommends the results of Method 1d (GLS constant-only regression). This method (based on homogeneous regions verified by L-moment analysis) yields region skew values that average to 0.151, with a corresponding MSE of 0.044. Values of skew and MSE for each region are found in Table 13.

The region weighted-average method is recommended because:

- The simplicity of using only a constant and the comparably small AVP makes this method preferable to the GLS regression equations or skew contour map.
- The method produces improvements to the recommendations of *Bulletin 17B*, as presented in this report.
- The minimum error of the method, AVP, will promote the greatest consistency in the application of the *Bulletin 17B* guidelines.

The focus on consistency is an important aspect of the study recommendations. In the original testing to develop *Bulletin 17B* (IACWD 1982, Appendix 15) split sample testing demonstrated that the log-normal distribution (a zero skew distribution) performed as well as the log-Pearson III distribution when substituting regional skew for the computed skew. This implies that regional information had no impact on prediction accuracy. However, the regional skew approach was selected because it promoted greater consistency in the estimate of infrequent quantiles (p=0.01 quantiles) obtained from either of the gage split record samples. The result is greater consistency both at a gage and within a region as future frequency studies are performed. Consequently, the consistency principle is important in promoting reasonably stable flood plain maps going into the future.

References

Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York, NY, pages 230.

Draper, N., and Smith, H. (1966). *Applied Regression Analysis*, John Wiley & Sons, Inc., New York, NY, pages 407.

Griffis, V. W., and Stedinger, J. R. (2007). "*Log-Pearson Type 3 Distribution and its Application in Flood Frequency Analysis. I: Distribution Characteristics*". Journal of Hydrologic Engineering, 12(5), pages 482-491.

Griffis, V. W., and Stedinger, J. R. (2007). "*The Use of GLS Regression in Regional Hydrologic Analyses*", Journal of Hydrology, pages 344, 82-95.

Gruber, Andrea M., Reis, Jr., Dirceu S., and Stedinger, J. R. (2007). "*Models of regional skew based on Bayesian GLS regression*", Paper 40927-3285, World Environmental & Water Resources Conference - Restoring our Natural Habitat, K.C. Kabbes, editor, Tampa, FL, May 15-18.

Haan, C. T. (1977). *Statistical methods in hydrology*, Iowa State University Press, Ames, IA, 378.

Hosking, J. R. M., and Wallis, J. R. (1997). *Regional frequency analysis*, Cambridge University Press, New York, NY.

Interagency Advisory Committee on Water Data (IACWD). (1982). "*Guidelines for determining flood flow frequency*", Bulletin 17B, US Department of the Interior, Geological Survey, http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.

Johnston, J. (1972). *Econometric methods*, McGraw-Hill, New York, NY.

Statistical package for the social sciences [computer program]. (1996). Release 7.5.1, SPSS, Inc., Chicago, IL.

Salas, J. D. (1992). "*Analysis and modeling of hydrologic time series*", Chapter 19, Handbook of Hydrology, ed. David R. Maidment, McGraw Hill, New York, NY.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, Efi. (1992). "*Frequency analysis of extreme events*", Handbook of hydrology, ed. David R. Maidment, McGraw Hill, New York, NY.

Tasker, G. D. (1983). "*Effective record length for the T-year event*", Journal of Hydrology, 1964, pages 39-47.

Tasker, G. D., and Stedinger, J. R. (1986). "*Regional skew with weighted LS regression*", Journal of Water Resources Planning and Management, 112(2).

Tasker, Gary D., and Stedinger, Jery R. (1989). "*An operational GLS model for hydrologic regression*", Journal of Hydrology, 111 361-375.

USACE (1983). *Generalized skew study of the Delaware River Basin*, Hydrologic Engineering Center, Davis, CA.

USACE. (1993). Engineering and design-hydrologic frequency analysis, EM 1110-2-1415, http://www.usace.army.mil/publications/eng-manuals/em1110-2-1415/entire.pdf.

USACE (2000). "*Investigations of methods for obtaining regionally consistent flood distributions*", Upper Mississippi flood frequency study, Rock Island District, US Army, Rock Island, Illinois, 107.

USGS (2006). Magnitude and frequency of floods on nontidal streams in Delaware, US Department of the Interior, Geological Survey, http://pubs.usgs.gov/sir/2006/5146/pdf/sir2006-5146.pdf.

USGS (2007). *Analysis of flood-magnitude and flood-frequency for streamflow-gaging stations in the Delaware and North Branch Susquehanna River basins in Pennsylvania*, US Department of the Interior, Geological Survey, http://pubs.usgs.gov/of/2007/1235/pdf/ofr2007-1235.pdf.

Wallis, J. B., Matalas, N. C., and Slack, J. R. (1974). "*Just a moment!*" Water Resources Research, 10(2) 211-218.

Appendix A

1983 Skew Coefficients

Table 29 lists the gages and watershed parameters of gages used in the USACE 1983 regional skew study of the Delaware River Basin. These statistics are:

- **Systematic record:** observed record length (through 1983) not including historical events.
- **Historical record length:** equivalent record length, accounts for the systematic record length and historical events calculated using *Bulletin 17B* recommendations.
- **Computed skew:** calculated station skew coefficient.
- **Generalized skew:** regional skew coefficient based on regions of average skew.
- Streamgages in Table 29 are listed by average skew region from north to south of the study region.

Table 29. 1983 Study Skew Coefficients

USGS Gage Number (1)	Years of Record		Computed Skew (4)	Generalized Skew (5)
	Systematic (2)	Historical (3)		
01361500	67	77	-0.137	-0.1
01497500	48	48	-0.170	-0.1
01498500	38	40	-0.080	-0.1
01499000	35	38	-0.171	-0.1
01500500	44	47	0.011	-0.1
01501000	49	49	-0.242	-0.1
01501500	36	36	0.671	-0.1
01502000	44	46	0.262	-0.1
01502500	50	50	-0.276	-0.1
01505000	44	46	-0.289	-0.1
01505500	32	42	-0.357	-0.1
01350000	73	78	0.027	0.0
01413500	45	45	-0.292	0.0
01415000	45	45	-0.180	0.0
01415500	26	26	0.836	0.0
01421000	42	51	-0.044	0.0
01422000	38	38	-0.152	0.0
01422500	37	37	-0.183	0.0
01423000	31	31	-0.544	0.0
01425500	34	34	-0.003	0.0
01426500	51	60	0.134	0.0
01503000	69	117	-0.092	0.0
01507000	45	47	0.330	0.0
01507500	34	37	0.031	0.0
01510000	39	47	0.072	0.0
01510500	33	35	0.148	0.0
01362500	50	108	0.217	+0.1
01365000	45	54	-0.126	+0.1
01414500	45	45	0.106	+0.1
01418500	38	38	0.070	+0.1
01419500	37	37	0.518	+0.1
01420000	57	57	-0.062	+0.1
01420500	68	68	0.057	+0.1
01426000	33	33	0.495	+0.1
01435000	31	31	-0.065	+0.1
01365500	43	43	0.197	+0.2
01367500	58	81	-0.251	+0.2
01371000	27	27	-0.050	+0.2
01427500	42	42	0.484	+0.2
01428000	28	28	1.090	+0.2
01431000	36	36	0.290	+0.2
01431500	31	31	0.956	+0.2
01437000	36	36	0.329	+0.2
01534000	66	66	-0.454	+0.2
01537500	40	40	-0.043	+0.2
01538000	60	60	0.034	+0.2
01539000	41	44	0.104	+0.2
01370000	49	49	0.698	+0.6
01371500	57	62	0.620	+0.6
01439500	72	72	1.272	+0.6
01441000	28	45	0.163	+0.6

USGS Gage Number (1)	**Years of Record**		**Computed Skew (4)**	**Generalized Skew (5)**
	Systematic (2)	**Historical (3)**		
01442500	30	30	1.245	+0.6
01447500	37	39	0.928	+0.6
01448000	44	44	0.988	+0.6
01448500	32	32	0.276	+0.6
01450500	41	41	0.205	+0.6
01451000	34	34	0.611	+0.6
01467500	31	31	0.226	+0.6
01469500	61	61	-0.315	+0.6
01470500	33	33	0.098	+0.6
01471000	25	25	0.936	+0.6
01471500	35	174	1.357	+0.6
01472000	53	79	0.161	+0.6
01554500	40	40	0.546	+0.6
01555500	50	50	1.219	+0.6
01573000	61	91	1.148	+0.6
01576500	51	51	1.295	+0.6
01368000	44	46	1.171	+0.4
01369000	40	40	0.488	+0.4
01369500	42	42	0.242	+0.4
01384500	43	43	-0.467	+0.4
01387500	70	70	0.461	+0.4
01440000	58	58	0.345	+0.4
01443500	58	58	0.557	+0.4
01451500	35	35	0.534	+0.4
01452000	36	36	0.463	+0.4
01452500	32	32	0.730	+0.4
01453000	75	75	0.392	+0.4
01379000	63	63	0.243	+0.3
01379500	52	52	0.291	+0.3
01380500	44	44	0.151	+0.3
01381500	60	60	-0.045	+0.3
01386000	43	43	-0.018	+0.3
01388000	60	60	0.269	+0.3
01396500	64	88	0.451	+0.3
01397000	66	66	0.377	+0.3
01397500	45	45	0.494	+0.3
01398000	51	51	0.620	+0.3
01398500	60	60	0.298	+0.3
01399500	60	86	0.217	+0.3
01400000	58	58	0.278	+0.3
01400500	70	70	0.315	+0.3
01402000	61	61	0.378	+0.3
01407500	59	59	0.795	+0.3
01408000	50	50	0.162	+0.3
01408500	53	53	0.272	+0.3
01411000	56	56	0.528	+0.3
01411500	49	49	0.859	+0.3
01445000	41	41	-0.091	+0.3
01445500	60	60	0.317	+0.3
01456000	59	59	-0.048	+0.3
01457000	60	60	0.034	+0.3
01459500	63	63	0.191	+0.3

USGS Gage Number (1)	**Years of Record**		**Computed Skew (4)**	**Generalized Skew (5)**
	Systematic (2)	**Historical (3)**		
01464000	38	38	0.188	+0.3
01464500	58	58	0.065	+0.3
01465500	41	41	0.127	+0.3
01467000	45	45	0.290	+0.3
01481000	40	40	-0.011	+0.3
01473000	42	42	0.416	+0.3
01474500	50	50	0.001	+0.3
01475000	40	40	1.990	+0.3
01477000	49	49	0.782	+0.3
01478000	39	39	-0.026	+0.3
01478500	25	25	0.201	+0.3
01479000	41	41	0.107	+0.3
01480000	39	39	0.498	+0.3
01480000	61	61	0.315	+0.3
01481500	35	35	0.697	+0.3
01482500	42	42	1.194	+0.3
01483500	33	33	0.824	+0.3
01484000	24	24	-0.131	+0.3
01484500	39	39	0.311	+0.3
01485000	32	32	0.967	+0.3
01485500	32	32	0.109	+0.3
01486000	28	28	-0.261	+0.3
01487000	38	38	0.499	+0.3
01488500	36	36	-0.128	+0.3
01491000	34	34	0.054	+0.3
01493000	34	34	-0.300	+0.3
01493500	31	31	0.473	+0.3
01495000	50	50	0.359	+0.3
01496000	33	33	0.322	+0.3

Appendix B

Study Gages and Watershed Parameters

Table 30 lists the gages and watershed parameters of gages used in the regional skew study of the Delaware River Basin. These parameters are:

- **Area:** drainage area of the watershed.
- **10-85 slope:** basin slope parameter defined as $\frac{(e_{0.85L}) - (e_{0.10L})}{0.75L}$,

  where:
  $e_{0.85L}$ is the elevation at 85 percent of the basin length, L,
  $e_{0.10L}$ is the elevation at ten percent of L.
- **Length:** basin length of the drainage area.
- **Lake storage:** percent of total lake surface area in a basin to drainage area.
- **Mean elevation:** average basin elevation.
- **Forested area:** percent of total forest area in a basin to drainage area.
- **MAP:** mean annual precipitation, rounded to the nearest whole inch.
- **SCS soils index:** average Natural Resource Conservation Service soil storage parameter.

Table 30. Study Gages and Watershed Parameters

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01169000	North River at Shattuckville, MA	88.4	65.6	21.0	1.2	1.440	82	49	4.3
01180500	Middle Branch Westfield River at Gross Heights, MA	52.6	79.0	19.5	1.0	1.420	66	48	4.0
01181000	West Branch Westfield River at Huntington, MA	93.7	54.9	21.0	1.5	1.420	87	48	4.0
01200000	Tenmile River near Gaylordsville, CN	203.0	15.0	34.5	1.9	0.849	65	44	4.1
01203000	Shepaug River near Roxbury, CN	132.0	26.4	38.8	2.8	1.029	63	50	4.2
01319000	East Branch Sandaga River at Griffin, NY	114.0	38.8	22.0	2.0	1.595	99	46	7.1
01321000	Sacandaga River near Hope, NY	491.0	33.3	35.2	3.8	1.376	97	47	7.0
01329000	Batten Kill at Arlington, VT	152.0	62.7	22.0	1.1	1.700	82	49	4.1
01330500	Kayaderosseras Creek near West Milton, NY	90.0	26.0	19.5	1.1	0.578	61	39	6.6
01334000	Walloomsac River near North Bennitgton, VT	111.0	125.0	15.6	1.3	1.700	70	46	4.4
01334500	Housic River near Eagle Bridge, NY	510.0	15.0	47.5	1.3	0.683	52	44	3.7
01347500	East Canada Creek at Dolgeville, NY	261.0	53.7	27.1	1.0	1.348	98	55	5.8
01348000	East Canada Creek at East Creek, NY	291.0	42.7	33.8	1.0	1.120	90	49	5.6
01349000	Otsquago Creek at Fort Plan, NY	59.2	75.4	17.0	1.0	0.832	54	40	4.7
01350000	Schoharie Creek at Prattsville, NY	236.0	30.3	28.8	0.6	2.120	85	46	3.8
01358500	Poesten Kill near Troy, NY	89.4	89.2	18.5	1.0	0.952	75	40	3.5
01361000	Kenderhook Creek at Rossman, NY	329.0	25.7	28.0	1.5	0.388	76	40	2.8
01361500	Catskill Creek at Oak Hill, NY	98.0	45.7	14.3	1.3	1.540	71	39	3.5
01362500	Esopus Creek at Coldbrook, NY	192.0	50.4	24.2	0.3	1.550	96	51	4.6
01365000	Rondout Creek near Lowes, NY	38.5	85.3	10.0	0.2	2.000	95	51	5.1
01365500	Chestnut Creek at Grahamsville, NY	20.9	153.0	5.2	1.1	1.480	90	47	4.9
01367500	Rondout Creek at Rosendale, NY	386.0	24.2	52.7	4.2	1.190	86	46	4.7
01368000	Wallkill River near Unionville, NJ	140.0	14.8	24.4	4.3	0.840	43	42	3.3
01369000	Pochuck Creek near Pine Island, NY	98.0	8.1	19.8	7.7	0.760	55	43	3.3
01369500	Quaker Creek at Florida, NY	9.7	42.0	6.1	6.5	0.600	50	43	3.3
01370000	Wallkill River at Pellets Island, NY	385.0	6.1	39.1	5.5	0.780	50	42	3.0

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01371000	Shawangunk Kill at Pine Bush, NY	102.0	21.8	29.4	3.3	0.800	69	44	3.5
01372500	Wappinger Creek near Wappingers Falls, NY	181.0	12.6	29.1	1.7	0.284	48	41	3.4
01373500	Fishkill Creek at Beacon, NY	190.0	10.7	30.1	1.5	0.344	60	43	3.9
01379000	Passaic River near Millington, NJ	55.4	29.9	11.1	18.1	0.343	46	48	3.4
01379500	Passaic River near Chatham, NJ	100.0	11.3	24.6	11.2	0.323	48	48	3.5
01380500	Rockaway River above reservoir at Boonton, NJ	116.0	12.0	28.1	7.2	0.806	82	48	4.3
01381500	Whippany River at Morristown, NJ	28.4	16.0	12.2	1.4	0.559	64	48	4.0
01384500	Ringwood Creek near Wanaque, NJ	19.1	62.0	10.2	7.1	0.780	90	44	4.6
01386000	West Brook near Wanaque, NJ	11.8	151.0	5.8	3.4	0.829	95	45	4.7
01387500	Ramapo River near Mahwah, NJ	118.0	17.2	22.8	7.5	0.430	77	46	4.1
01388000	Ramapo River at Pompton Lakes, NJ	160.0	12.6	34.3	4.8	0.700	78	46	4.2
01396500	South Branch Raritan River near High Bridge, NJ	65.3	28.4	23.0	1.7	0.852	58	48	5.0
01397000	South Branch Raritan River at Stanton, NJ	147.0	27.0	34.2	2.0	0.665	51	47	4.4
01398000	Neshanic River at Reaville, NJ	25.7	21.4	4.5	0.1	0.274	22	45	3.2
01398500	North Branch Raritan River near Far Hills, NJ	26.2	43.2	8.4	0.8	0.649	58	49	5.2
01399500	Lamington (Black) River near Pottersville, NJ	32.8	23.3	15.8	4.8	0.784	52	48	5.1
01400000	North Branch Raritan River near Raritan, NJ	190.0	18.9	21.7	1.0	0.452	47	47	4.6
01400500	Raritan River at Manville, NJ	490.0	28.4	58.2	0.7	0.428	41	46	4.2
01407500	Swimming River near Red Bank, NJ	48.5	16.7	9.0	2.1	0.130	32	45	3.7
01407830	Manasquan River near Georgia, NJ	10.6	10.6	4.1	40.0	0.123	40	46	4.7
01408000	Manasquan River at Squankum, NJ	43.2	6.5	12.8	1.2	0.118	36	46	5.1
01408015	Mingamahone Brook at Farmingdale, NJ	6.2	20.0	5.2	60.0	0.102	60	46	5.0
01408120	North Branch Metedeconk River near Lakewood, NJ	34.9	7.0	17.1	75.0	0.100	75	46	5.4
01408500	Toms River near Toms River, NJ	124.0	6.3	22.0	13.6	0.109	84	48	9.1

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01409400	Mullica River near Batsto, NJ	46.7	6.7	18.9	28.4	0.084	84	43	4.7
01409810	West Branch Wading River near Jenkins, NJ	84.1	5.4	16.5	85.0	0.050	85	43	4.7
01411000	Great Egg Harbor River at Folsom, NJ	56.3	4.9	18.4	13.9	0.127	70	45	6.2
01411500	Maurice River at Norma, NJ	113.0	5.1	17.6	8.9	0.115	58	43	3.8
01412500	West Branch Cohansey River at Seeley, NJ	2.6	25.3	2.9	0.2	0.110	4	41	3.7
01413500	East BR Delaware River at Margaretville, NY	163.0	15.2	19.4	0.4	2.210	85	44	4.1
01414000	Platte Kill at Dunraven, NY	34.9	91.5	11.7	1.0	2.130	78	43	4.1
01414500	Mill Brook near Dunraven, NY	25.2	120.5	10.5	0.2	2.250	93	46	4.3
01415000	Tremper Kill near Andes, NY	33.2	60.7	9.7	1.0	1.990	71	43	3.7
01415500	Terry Clove Kill near Pepacton, NY	13.6	114.0	5.6	0.1	1.950	99	43	3.7
01418500	Beaver Kill at Craigie Clair, NY	81.9	52.7	22.3	0.7	1.932	83	49	4.1
01419500	Willowemoc Creek near Livingston Manor, NY	62.6	64.1	14.2	1.7	2.150	99	54	4.1
01420000	Little Beaver Kill near Livingston Manor, NY	20.1	83.8	8.1	3.8	1.830	82	49	4.0
01420500	Beaver Kill at Cooks Falls, NY	241.0	33.4	33.2	2.0	2.070	94	50	4.0
01421000	East BR Delaware River at Fishs Eddy, NY	784.0	9.3	62.2	0.9	2.010	86	46	4.4
01421900	W BR Delaware River Upstream From Delhi, NY	134.0	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.4
01422000	West Branch Delaware River at Delhi, NY	142.0	21.5	25.5	0.4	2.000	63	42	3.4
01422500	Little Delaware River near Delhi, NY	49.7	57.6	14.6	0.4	2.180	74	43	3.4
01423000	West Branch Delaware River at Walton, NY	332.0	13.1	42.6	0.4	1.930	67	42	3.3
0142400103	Trout Creek near Trout Creek, NY	20.2	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.9
01425500	Cold Spring Brook at China, NY	1.5	326.0	2.0	0.0	1.670	60	42	4.2
01426000	Oquaga Creek at Deposit, NY	67.6	40.8	17.0	0.6	1.610	81	42	4.3
01426500	West Branch Delaware River at Hale Eddy, NY	595.0	9.3	73.4	0.5	1.790	73	42	4.2

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01427500	Callicoon Creek at Callicoon, NY	110.0	37.9	19.0	2.6	1.470	62	43	3.5
01428000	Tenmile River at Tusten, NY	45.6	29.0	13.3	5.9	1.120	86	42	3.7
01428750	West Branch Lackawaxen River near Aldenville, PA	40.6	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.8
01429300	Dyberry Creek above Reservoir near Honesdale, PA	45.8	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.8
01431500	Lakawaxen River at Hawley, PA	290.0	31.8	36.6	4.6	1.440	86	44	2.9
01435000	Neversink River near Claryville, NY	66.6	70.6	15.1	0.4	2.450	98	58	5.1
01437000	Neversink River at Oakland Valley, NY	223.0	27.2	49.9	4.7	1.760	84	49	4.7
01438300	Vandermark Creek at Milford, PA	5.4	202.0	4.9	3.0	1.040	94	44	3.6
01439500	Bush Kill at Shoemakers, PA	117.0	34.9	30.2	9.4	1.260	100	46	3.0
01440000	Flat Brook near Flatbrookville, NJ	65.1	37.6	23.8	1.5	0.895	86	45	4.6
01440400	Brodhead Creek near Analomink, PA	65.9	80.9	18.7	4.3	1.400	89	47	3.3
01441000	McMichael Creek near Stroudsburg, PA	65.3	34.6	24.6	2.3	0.950	53	49	3.0
01442500	Brodhead Creek at Minisink Hills, PA	259.0	48.4	27.6	3.1	1.160	77	48	3.0
01443500	Paulins Kill at Blairstown, NJ	126.0	8.0	28.6	5.9	0.629	50	45	3.4
01445000	Pequest River at Huntsville, NJ	31.4	8.2	5.1	4.1	0.697	66	44	5.0
01445500	Pequest River at Pequest, NJ	108.0	6.4	24.7	5.0	0.668	64	44	4.9
01446000	Beaver Brook near Belvidere, NJ	36.2	26.0	13.3	3.0	0.568	49	45	4.1
01446600	Martins Creek near East Bangor, PA	10.4	54.5	5.2	4.9	0.940	91	46	3.4
01447500	Lehigh River at Stoddartsville, PA	91.7	28.2	22.2	10.7	1.850	84	45	2.8
01448000	Lehigh River at Tannery, PA	322.0	27.8	38.1	8.3	1.800	83	46	3.5
01448500	Dilldown Creek near Long Pond, PA	2.4	110.4	2.5	2.7	1.890	98	50	3.2
01449360	Pohopoco Creek at Kresgeville, PA	49.9	53.8	14.2	0.7	1.060	51	50	3.2
01450500	Aquashicola Creek at Palmerton, PA	76.7	29.6	23.4	1.5	0.880	58	47	3.4
01451800	Jordan Creek near Schnecksville, PA	53.0	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.1
01452000	Jordon Creek at Allentown, PA	75.8	13.5	29.1	0.1	0.580	15	48	3.2
01453000	Lehigh River at Bethlehem, PA	1279.0	17.7	96.0	2.2	1.120	56	46	2.9
01455200	Pohatcong Creek at New Village, NJ	33.3	36.1	16.3	0.1	0.686	56	45	3.8

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01459500	Tohickon Creek near Pipersville, PA	97.4	11.6	25.7	0.1	0.500	32	45	3.1
01460000	Tohickon Creek at Point Pleasant, PA	107.0	11.8	30.4	0.0	0.460	26	45	3.5
01464000	Assupink Creek at Trenton, NJ	89.4	4.8	20.5	2.6	0.105	34	44	3.1
01464400	Crosswicks Creek at New Egypt, NJ	41.2	3.9	7.5	13.2	0.110	49	42	4.5
01464500	Crosswicks Creek at Extonville, NJ	83.6	3.2	13.8	7.8	0.118	61	44	4.3
01464515	Doctors Creek at Allentown, NJ	17.4	11.3	11.5	5.6	0.142	33	42	4.2
01464538	Crafts Creek at Columbus, NJ	5.4	6.7	5.0	1.2	Unknown	Unknown	Unknown	4.1
01465000	Neshaminy Creek at Rushland, PA	134.0	9.5	27.3	0.0	0.340	19	44	3.9
01465850	South Branch Rancocas Creek at Vincentown, NJ	64.5	6.9	64.5	14.3	0.096	75	42	4.0
01466000	Middle Br Mount Misery Bk in Byrne State Forest, NJ	2.8	12.3	3.8	16.9	0.137	99	45	4.7
01466500	McDonalds Branch in Byrne State Forest, NJ	2.4	18.4	3.2	6.1	0.146	100	43	4.6
01467000	North Branch Rancocas Creek at Pemberton, NJ	111.0	6.6	17.0	14.0	0.117	87	43	3.1
01467081	South Branch Pennsauken Creek at Cherry Hill, NJ	9.0	9.6	7.0	2.0	0.064	52	42	3.9
01467150	Cooper River at Haddonfield, NJ	17.0	11.2	9.9	7.6	0.085	30	43	3.9
01467160	North Branch Cooper River near Marlton, NJ	5.3	14.0	5.4	0.2	0.096	47	43	3.9
01467305	Newton Creek at Collingswood, NJ	1.3	30.3	2.3	3.5	0.070	21	43	3.9
01467500	Schuylkill River at Pottsville, PA	53.4	25.7	13.7	1.0	1.240	72	48	3.2
01468500	Schuylkill River at Landingville, PA	133.0	17.0	22.4	1.0	1.100	62	48	3.2
01470500	Schuylkill River at Berne, PA	355.0	11.7	40.9	1.6	1.020	60	48	3.1
01470779	Tulpehocken Creek near Bernville, PA	66.5	11.3	19.2	0.3	0.540	10	45	3.3
01471000	Tulpehocken Creek near Reading, PA	211.0	7.3	37.1	2.0	0.530	24	45	2.9
01471980	Manatawny Creek near Pottstown, PA	85.5	22.8	21.2	0.0	0.600	35	46	3.1
01472000	Schuylkill River at Pottstown, PA	1147.0	8.4	78.4	1.1	0.660	58	45	3.2
01472157	French Creek near Phoenixville, PA	59.1	25.1	19.0	2.4	0.529	53	45	3.1

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01472162	Schuylkill River at Phoenixville, PA	1280.0	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.0
01472198	Perkiomen Creek at East Greenville, PA	38.0	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.1
01472199	West Branch Perkiomen Creek at Hillegass, PA	23.0	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	3.1
01472500	Perkiomen Creek near Frederick, PA	152.0	21.0	24.6	1.1	0.540	41	45	3.0
01473000	Perkiomen Creek at Graterford, PA	279.0	17.0	28.3	0.1	0.380	29	45	2.7
01473120	Skippack Creek near Collegeville, PA	53.7	13.4	14.3	60.0	0.330	60	44	3.1
01474000	Wissahickon Creek at Mouth, Philadelphia, PA	64.0	13.3	25.4	1.2	0.283	24	44	3.6
01474500	Schuylkill River at Philadelphia, PA	1893.0	47.8	126.9	Unknown	0.540	44	46	3.7
01475000	Mantua Creek at Pitman, NJ	6.1	18.4	7.1	3.3	0.131	39	43	4.1
01475300	Darby Creek at Waterloo Mills near Devon, PA	5.2	36.8	4.1	1.0	0.434	31	46	3.4
01475850	Crum Creek near Newtown Square, PA	15.8	32.0	8.7	1.7	0.419	37	48	3.4
01476500	Ridley Creek at Moylan, PA	31.9	22.3	18.2	1.0	0.367	41	43	3.3
01477120	Raccoon Creek near Swedesboro, NJ	26.9	14.8	10.3	5.7	0.102	30	42	3.7
01478000	Christina River at Coochs Bridge, DE	20.5	21.7	13.4	0.1	0.191	24	45	3.6
01478200	Middle Branch White Clay Creek near Landenberg, PA	12.7	27.5	11.0	0.7	0.452	16	45	3.6
01478500	White Clay Creek above Newark, DE	66.7	17.7	19.5	0.1	0.376	26	45	3.7
01479000	White Clay Creek near Newark, DE	89.1	13.9	27.2	0.1	0.324	25	45	3.6
01480000	Red Clay Creek at Wooddale, DE	47.0	15.9	19.3	0.8	0.336	25	46	3.2
01480300	West Branch Brandywine Creek near Honey Brook, PA	18.7	24.4	9.2	3.2	0.726	22	43	3.3
01480610	Sucker Run near Coatesville, PA	2.6	145.0	2.7	4.7	0.543	32	46	3.3
01480675	Marsh Creek near Glenmoore, PA	8.6	26.9	6.0	5.5	0.598	46	45	3.3
01481500	Brandywine Creek at Wilmington, DE	314.0	11.2	54.4	0.2	0.472	32	46	3.4
01482500	Salem River at Woodstown, NJ	14.6	13.1	7.4	1.3	0.114	21	44	3.1
01483500	Leipsic River near Cheswold, DE	9.4	8.2	5.9	0.1	0.059	9	45	3.7
01484000	Murderkill River near Felton, DE	13.6	6.2	6.9	0.6	0.055	14	47	3.7

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01484500	Stockley Branch at Stockley, DE	5.2	4.2	5.0	0.0	0.044	11	47	3.7
01485000	Pocomoke River near Willards, MD	60.5	1.5	14.9	15.8	0.044	30	47	3.7
01485500	Nassawango Creek near Snow Hill, MD	44.9	3.6	13.1	6.2	0.046	85	47	3.7
01486000	Manokin Branch near Princess Anne, MD	4.8	5.5	4.2	0.0	0.032	57	46	3.7
01487000	Nanticoke River near Bridgeville, DE	75.4	2.6	15.1	1.7	0.049	17	46	3.7
01488500	Marshyhope Creek near Adamsville, MD	43.9	2.5	12.3	0.3	0.056	8	46	3.7
01491000	Choptank River near Greensboro, MD	113.0	3.0	19.0	1.9	0.060	35	45	3.7
01493000	Unicorn Branch near Millington, MD	22.3	6.1	9.9	1.5	0.061	43	44	3.7
01493500	Morgan Creek near Kennedyville, MD	12.7	9.1	6.1	0.2	0.060	8	43	3.7
01495000	Big Elk Creek at Elk Mills, MD	52.6	17.9	23.3	0.1	0.398	14	44	3.7
01496000	Northeast Creek at Leslie, MD	24.3	24.0	14.0	0.1	0.380	22	44	3.7
01497500	Susquehanna Creek at Colliersville, NY	349.0	3.0	35.6	7.4	1.580	53	42	4.2
01498500	Charlotte Creek at West Davenport, NY	167.0	28.3	25.0	1.3	1.840	43	42	4.3
01499000	Otego Creek near Oneonta, NY	108.0	18.7	26.1	1.3	1.470	33	42	3.9
01500500	Susquehanna River at Unadilla, NY	982.0	3.4	62.7	3.8	1.660	63	42	3.9
01501000	Unadilla River near New Berlin, NY	199.0	7.5	32.0	1.3	1.410	33	42	3.3
01501500	Sage Brook near South New Berlin, NY	0.7	318.0	1.3	0.0	1.630	100	41	3.3
01502000	Butternut Creek at Morris, NY	59.7	27.8	21.1	1.1	1.560	42	42	3.4
01502500	Unadilla River at Rockdale, NY	520.0	4.8	57.0	1.3	1.500	33	42	3.4
01503000	Susquehanna River at Conklin, NY	2232.0	3.6	121.8	2.6	1.560	53	41	3.9
01505000	Chenango River at Shurborne, NY	263.0	14.6	23.8	5.2	1.500	48	41	3.2
01505500	Canasawacta Creek near South Plymouth, NY	57.9	65.4	9.8	0.8	1.400	50	43	3.2
01507000	Chenango River at Greene, NY	593.0	5.7	64.3	3.6	1.470	60	41	3.3
01507500	Genegantslet Creek at Smithville Flats, NY	82.3	40.6	16.1	2.8	1.540	58	43	3.3
01508000	Shackham Brook near Truxton, NY	3.0	193.0	2.5	1.0	1.492	91	44	4.5

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01508500	Albright Creek at East Homer, NY	6.8	108.0	5.7	1.0	1.440	38	43	3.4
01509000	Tioughnioga River at Cortland, NY	292.0	6.3	31.5	1.5	1.230	35	42	3.5
01510000	Oteselic River at Cincinnatus, NY	147.0	16.2	30.4	0.6	1.630	44	44	3.7
01510500	Oteselic River near Upper Lisle, NY	217.0	15.0	40.1	1.7	1.560	61	43	3.7
01514000	Owego Creek near Owego, NY	185.0	14.3	30.3	1.0	1.024	52	40	3.3
01518000	Tioga River at Tioga, PA	282.0	44.0	29.8	1.0	1.740	64	37	3.3
01520000	Cowanesque River near Lawrenceville, PA	298.0	20.1	37.3	1.0	1.600	30	35	3.3
01520500	Tioga River at Lindley, NY	771.0	24.4	44.8	1.0	1.447	33	36	3.2
01526000	Tuscarora Creek near South Addison, NY	114.0	35.8	17.5	1.0	1.383	30	33	3.0
01527000	Cohocton River at Cohocton, NY	52.2	29.9	17.4	1.0	1.489	28	34	3.0
01528000	Fivemile Creek near Kanona, NY	66.8	12.6	20.1	1.0	1.305	41	34	3.0
01530500	Newton Creek near Elmira, NY	77.5	45.2	18.9	1.0	1.123	27	35	3.0
01532000	Towanda Creek near Monroeton, PA	215.0	17.7	27.5	1.1	0.966	57	38	3.0
01537500	Solomon Creek at Wilkes-Barre, PA	15.7	291.0	5.6	1.0	1.350	80	41	3.5
01539000	Fishing Creek near Bloomsburg, PA	274.0	39.6	32.3	1.6	1.190	69	44	2.9
01544500	Kettle Creek at Cross Fork, PA	136.0	37.4	22.7	1.0	1.800	94	43	4.8
01546500	Spring Creek near Axemann, PA	87.2	32.1	18.3	1.0	1.300	34	45	4.8
01548000	Bald Eagle Creek at Beech Creek Station, PA	559.0	9.5	44.6	1.0	1.300	81	41	5.5
01548500	Pine Creek at Cedar Run, PA	604.0	22.8	44.5	1.0	1.900	83	40	3.2
01549500	Blockhouse Creek near English Center, PA	37.7	54.8	15.2	1.0	1.350	67	41	3.3
01550000	Lycoming Creek near Trout Run, PA	173.0	45.7	24.2	1.0	1.700	65	43	3.9
01552000	Loyalsock Creek at Loyalsock, PA	443.0	24.4	56.0	1.3	1.600	89	46	3.3
01552500	Muncy Creek near Sonestown, PA	23.8	115.7	10.0	1.0	1.800	80	48	4.2
01554500	Shamokin Creek near Shamokin PA	54.2	42.4	16.7	1.0	1.220	52	44	3.5
01555000	Penns Creek at Penns Creek, PA	301.0	18.9	45.2	1.0	1.300	76	42	4.0
01555500	East Mahantoango Creek near Dalmatia, PA	162.0	11.0	43.7	1.0	0.860	33	43	3.4

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01564500	Aughwick Creek near Three Springs, PA	205.0	14.0	38.8	1.0	1.200	76	41	4.9
01565000	Kishacoquillas Creek at Reedsville, PA	164.0	47.1	20.6	1.0	1.100	60	41	4.0
01568000	Sherman Creek at Shermans Dale, PA	200.0	7.6	43.0	1.0	1.200	58	42	3.9
01571500	Yellow Breeches Creek near Camp Hill, PA	216.0	8.9	53.0	1.0	0.800	32	42	4.6
01573000	Swatara Creek at Harper Tavern, PA	337.0	12.5	44.4	1.4	0.690	35	46	3.2
01574500	Codorus Creek at Spring Grove, PA	75.5	20.1	17.3	1.2	0.660	22	41	4.1
01576500	Conestoga River at Lancaster, PA	324.0	7.4	41.8	1.0	0.520	28	42	3.5
01580000	Deer Creek at Rocks, MD	94.4	17.7	28.6	1.0	0.657	51	45	3.3
01582000	Little Falls at Blue Mount, MA	52.9	33.8	15.6	1.1	0.658	28	45	3.3
01583000	Slade Run near Glydon, MD	2.1	46.1	2.6	2.0	0.591	37	45	3.3
01583500	Western Run at Western Run, MD	59.8	24.5	17.1	1.1	0.544	45	45	3.3
01584500	Little Gunpowder Falls at Laurel Brook, MD	36.1	21.1	15.0	1.1	0.542	41	45	3.3
01586000	North Branch Patapsco River at Cedarhurst, MD	56.6	223.9	14.1	1.1	0.747	39	45	3.3
01587500	South Branch Patapsco River at Henryton, MD	64.4	26.2	16.4	1.0	0.642	23	43	3.4
01590000	North River near Annapolis, MD	8.5	18.1	4.6	1.4	0.112	71	44	3.4
01591000	Patuxent River near Unity, MD	34.8	28.2	12.5	1.0	0.589	27	43	3.4
01593500	Little Patuxent River at Guilford, MD	38.0	22.1	16.1	1.1	0.409	34	43	3.4
01594500	Western Branch near Largo, MD	30.2	7.4	10.6	1.5	0.147	48	44	3.1
01614500	Conococheague Creek at Fairview, MD	494.0	11.2	57.2	1.1	1.050	39	40	3.4
01619500	Antietam Creek near Sharpsburg, MD	281.0	10.8	47.1	1.1	0.781	24	40	3.4
01639000	Monocacy River at Bridgeport, MD	173.0	18.9	31.5	1.0	0.597	25	44	3.1
01639500	Big Pipe Creek at Bruceville, MD	102.0	12.8	28.0	1.0	0.625	18	44	3.1
01642500	Linganore Creek near Frederick, MD	82.3	19.2	20.1	1.0	0.576	21	43	3.1
01643000	Monocacy River at Jug Bridge near Frederick, MD	817.0	5.6	64.7	1.0	0.621	34	44	3.1
01643500	Bennet Creek at Park Mills, MD	62.8	23.8	15.6	1.0	0.521	27	42	3.0

Gage USGS ID Number (1)	Location (2)	Area (mi2) (3)	10-85 Slope (ft/mi) (4)	Length (mi) (5)	Lake Storage (%) (6)	Mean Elevation (1,000 ft) (7)	Forested Area (%) (8)	MAP (in) (9)	SCS Soils Index (in) (10)
01644000	Goose Creek near Leesburg, VA	332.0	8.3	40.6	1.0	0.660	36	40	3.1
01645000	Senaca Creek at Dawsonville, MD	101.0	15.1	20.9	1.0	0.468	21	42	3.0
01646000	Difficult Run near Great Falls, VA	57.9	16.0	13.3	1.0	0.360	61	39	3.1
01648000	Rock Creek at Sherill Drive Washington, DC	62.2	12.6	24.5	1.1	0.387	35	44	3.1
01649500	Northeast Branch Anacostia River at Riverdale, MD	72.8	27.2	15.7	2.5	0.227	57	44	3.1
01650500	Northwest Branch Anacostia River near Colesville, MD	21.1	19.3	8.1	1.0	0.415	32	44	3.0
01651000	Northwest Branch Anacostia River near Hyattsville, MD	49.4	19.7	19.1	1.1	0.323	29	44	3.1
01653500	Henson Creek at Oxon Hill, MD	16.7	22.9	8.5	1.2	0.226	46	44	3.1
01657000	Bull Run near Manassas, VA	148.0	7.5	26.5	1.0	0.380	48	40	3.1
01658000	Mattawoman Creek near Pomonkey, MD	54.8	10.5	17.6	4.2	0.183	70	46	3.1
01661000	Chaptico Creek at Chaptico, MD	10.4	21.2	6.7	1.0	0.133	67	43	2.8
01661500	St. Mary's River at Great Mills, MD	24.0	12.9	8.1	1.3	0.101	83	41	2.8
04233000	Cayuga Inlet near Ithaca, NY	35.2	77.2	8.4	2.5	1.130	55	35	3.4
04234000	Fall Creek near Ithaca, NY	136.0	15.9	31.9	3.3	1.340	35	39	3.5
04242500	East Branch Fish Creek at Taberg, NY	188.0	37.0	40.8	12.9	1.450	76	52	3.8
04243500	Oneida Creek at Oneida, NY	113.0	46.8	24.2	2.8	1.010	30	40	3.5

Appendix C

Station Statistics and Regions

Table 31 lists the station statistics for the streamgages considered in this study. These statistics are:

- **Systematic record:** observed record length (through the 2006 water year) not including historical events.
- **Historical record:** equivalent record length, accounts for the systematic record length and historical events calculated using *Bulletin 17B* recommendations.
- **Mean:** mean of the log-transforms of the peak annual flows calculated using *Bulletin 17B* recommendations.
- **Standard deviation:** standard deviation of the log-transforms of the peak annual flows calculated using *Bulletin 17B* recommendations.
- **Station skew:** skew coefficient of the log-transforms of the peak annual flows calculated using *Bulletin 17B* recommendations.
- **Historical events:** number of recorded historical events, as identified by USGS WATSTORE Code 7.
- **High outliers:** number of statistical high outliers using *Bulletin 17B* recommendations.
- **Low outliers:** number of statistical low outliers using *Bulletin 17B* recommendations.
- Stream gages that are in both Table 31 and Table 32 have different statistics because of the information in the twenty-five years of additional gage record since 1983.

Table 32 lists the gages in each of the eleven subregions of the Delaware River Basin used in Methods 1b, 1c and 1d.

Table 31. Station Statistics

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
01350000	97	103	4.166	0.285	-0.036	1	18	0	Removal of historical information does not significantly change statistics
01361500	88	88	3.623	0.269	-0.142	0	0	0	
01362500	75	75	4.157	0.323	0.165	0	0	0	
01365000	70	70	3.433	0.257	-0.248	0	0	0	
01365500	57	57	3.094	0.284	0.129	0	0	0	
01367500	91	91	4.079	0.196	-0.181	0	0	0	
01368000	51	51	3.253	0.166	1.077	0	0	0	Removed code 7 from 1984, historical information results in increased station skew and 8 low outliers
01369000	40	47	3.082	0.169	0.54	1	12	0	Removal of historical information does not significantly change statistics
01369500	42	47	2.572	0.192	0.273	1	13	0	Removal of historical information does not significantly change statistics
01370000	55	55	3.624	0.161	0.584	0	0	0	Removal of historical information does not significantly change statistics
01371000	26	68	3.467	0.22	0.132	6	9	0	
01379000	88	88	2.916	0.178	0.212	0	0	0	
01379500	77	77	3.117	0.163	0.221	0	0	0	
01380500	69	0	3.311	0.209	0.068	0	0	0	
01381500	85	85	2.965	0.204	-0.08	0	0	0	
01384500	63	63	2.657	0.245	0.704	0	0	0	
01386000	48	48	2.755	0.23	-0.045	0	0	0	
01387500	95	95	3.487	0.266	0.527	0	0	0	
01388000	85	125	3.533	0.257	0.368	3	11	0	Removal of historical information does not significantly change statistics
01396500	89	111	3.282	0.214	0.355	1	0	0	
01397000	91	91	3.661	0.215	0.42	0	0	1	
01398000	76	76	3.534	0.257	0.71	0	0	0	
01398500	86	86	3.152	0.267	0.243	0	0	0	

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
01399500	86	86	2.879	0.263	0.415	0	0	0	Removal of historical information does not significantly change statistics
01400000	83	111	3.956	0.189	0.551	1	6	0	Removal of historical information does not significantly change statistics
01400500	95	95	4.208	0.167	0.744	0	0	0	
01407500	84	88	3.148	0.299	0.467	1	0	0	
01407830	27	27	2.766	0.133	-0.083	0	0	0	
01408000	75	75	3.004	0.187	-0.296	0	0	0	
01408015	37	37	2.261	0.242	0.259	0	0	0	
01408120	33	33	2.68	0.201	-0.34	0	0	0	
01408500	78	78	2.893	0.195	0	0	0	0	
01409400	49	49	2.679	0.236	0.154	0	0	0	
01409810	24	24	2.827	0.213	-0.815	0	0	0	
01411000	81	81	2.485	0.207	0.232	0	0	0	
01411500	74	74	2.732	0.249	1.277	0	0	0	
01412500	53	53	2.042	0.39	-0.089	0	0	1	
01413500	70	70	3.793	0.262	0.044	0	0	0	
01414000	31	65	3.207	0.219	0.301	1	0	0	
01414500	70	31	3.174	0.247	0.082	0	0	0	
01415000	70	70	3.126	0.251	-0.028	0	0	0	
01415500	26	70	2.879	0.248	0.836	0	0	0	
01418500	38	60	3.616	0.225	0.144	1	0	0	
01419500	37	59	3.514	0.227	0.36	1	3	0	2005 no flow recorded
01420000	57	57	3.115	0.234	-0.062	0	0	0	2005 no flow recorded
01420500	93	93	4.067	0.262	0.353	0	0	0	
01421000	94	103	4.34	0.218	0.024	1	1	0	
01421900	49	49	3.607	0.183	0.159	0	0	0	
01422000	38	60	3.585	0.182	0.182	1	0	0	
01422500	48	48	3.315	0.192	-0.192	0	0	0	
01423000	56	56	3.986	0.214	-0.083	0	0	0	

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
0142400103	27	27	4.952	0.708	-0.699	0	0	1	
01425500	34	34	1.908	0.232	-0.003	0	0	0	
01426000	34	34	3.438	0.186	0.6	0	0	0	
01426500	95	95	4.049	0.264	-0.354	0	0	0	
01427500	56	56	3.638	0.189	0.498	0	0	0	
01428000	28	55	3.063	0.236	1.051	1	1	1	2005 no flow recorded
01428750	32	0	3.335	0.228	0.846	0	0	0	
01429300	29	29	3.317	0.207	0.561	0	0	0	
01431500	78	98	3.892	0.24	0.933	1	3	0	
01435000	67	69	3.79	0.245	-0.121	1	0	0	
01437000	46	46	3.8	0.33	-0.35	0	0	0	2005 no flow recorded
01438300	45	45	2.218	0.238	0.049	0	0	0	
01439500	98	98	3.33	0.236	0.994	0	0	0	
01440000	83	83	3.213	0.237	0.226	0	0	0	
01440400	49	49	3.47	0.294	-0.022	0	0	0	
01441000	29	29	3.204	0.213	0.296	0	0	0	
01442500	56	56	3.979	0.268	0.68	0	0	0	
01443500	83	83	3.269	0.195	0.283	0	0	0	
01445000	62	62	2.437	0.203	0.149	0	0	0	
01445500	85	85	2.96	0.166	-0.097	0	0	0	
01446000	77	77	2.701	0.212	-0.029	0	0	0	
01446600	26	26	2.686	0.277	0.478	0	0	0	
01447500	62	65	3.414	0.336	0.616	1	1	0	
01448000	44	44	3.857	0.285	0.988	0	0	0	Treated 1915 historical record as systematic
01448500	48	58	2.149	0.298	0.139	1	0	0	
01449360	40	40	3.032	0.202	-0.401	0	0	0	
01450500	67	67	3.373	0.268	0.326	0	0	0	
01451800	39	39	3.361	0.265	-0.09	0	0	0	
01452000	62	62	3.479	0.29	0.166	0	0	0	

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
01453000	101	101	4.377	0.239	0.233	0	0	0	
01455200	38	38	2.931	0.216	0.772	0	0	1	
01459500	71	71	3.812	0.222	-0.024	0	0	0	
01460000	29	29	3.616	0.162	1.336	0	0	0	
01464000	83	83	3.225	0.187	0.019	0	0	0	
01464400	26	26	3.016	0.201	1.115	0	0	0	
01464500	66	69	3.196	0.248	-0.138	1	2	0	
01464515	27	27	2.644	0.277	-0.437	0	0	0	1978 no flow recorded
01464538	28	28	2.356	0.315	0.146	0	0	0	
01465000	32	32	3.67	0.146	0.797	0	0	0	
01465850	40	40	2.858	0.229	0.202	0	0	1	
01466000	25	25	1.128	0.271	0.599	0	0	0	
01466500	53	53	0.989	0.287	0.026	0	0	0	
01467000	85	85	2.883	0.187	0.032	0	0	1	
01467081	38	38	2.765	0.166	0.889	0	0	0	
01467150	42	42	2.916	0.216	0.621	0	0	0	
01467160	26	26	2.385	0.218	1.177	0	0	0	
01467305	42	42	2.263	0.1	0.345	0	0	0	
01467500	45	53	3.174	0.3	0.142	1	0	0	
01468500	42	42	3.547	0.237	0.482	0	0	0	1942 no flow recorded
01470500	59	65	4.065	0.244	-0.036	1	3	0	
01470779	33	33	3.292	0.343	-0.239	0	0	0	
01471000	56	56	3.561	0.246	-0.065	0	0	0	
01471980	31	34	3.537	0.207	-0.248	1	0	0	
01472000	79	105	4.332	0.198	0.148	1	1	0	
01472157	38	38	3.411	0.301	0.209	0	0	0	
01472162	30	30	4.486	0.166	0.06	0	0	0	
01472198	24	24	3.43	0.265	-0.514	0	0	0	
01472199	24	24	3.158	0.2	-0.761	0	0	0	

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
01472500	29	29	3.699	0.118	0.772	0	0	1	
01473000	92	92	4.168	0.236	-0.564	0	0	0	
01473120	29	29	3.796	0.235	1.379	0	0	1	Removal of historical information does not change statistics
01474000	41	41	3.61	0.221	1.199	0	0	0	Removal of historical information does not change statistics
01474500	75	137	4.617	0.212	-0.095	2	1	0	
01475000	58	67	2.124	0.337	1.429	1	0	0	
01475300	26	28	2.86	0.279	0.266	1	0	0	
01475850	30	30	3.077	0.255	0.083	0	0	0	
01476500	31	31	3.138	0.252	0.697	0	0	0	Removed code 7 from 1978
01477120	40	40	2.864	0.28	-0.218	0	0	0	
01478000	64	64	3.266	0.204	0.277	0	0	0	
01478200	32	32	3.002	0.268	0.174	0	0	0	
01478500	28	28	3.537	0.25	0.601	0	0	0	
01479000	67	67	3.596	0.229	0.322	0	0	0	
01480000	64	64	3.373	0.241	0.729	0	0	0	
01480300	47	47	3.102	0.29	0.086	0	0	0	
01480610	41	41	2.523	0.289	-0.021	0	0	0	1985 no flow recorded
01480675	40	40	2.396	0.267	-0.093	0	0	0	
01481500	60	60	3.917	0.251	0.178	0	0	0	
01482500	60	60	2.873	0.386	0.689	0	0	0	
01483500	33	33	2.359	0.326	0.633	0	0	0	
01484000	31	31	2.491	0.347	-0.117	0	0	0	
01484500	62	62	1.862	0.242	0.465	0	0	0	
01485000	55	55	2.875	0.178	1.04	0	0	1	
01485500	57	57	2.771	0.261	-0.205	0	0	0	
01486000	53	53	2.158	0.28	-0.12	0	0	1	
01487000	63	63	2.817	0.276	0.22	0	0	0	1935 no flow recorded
01488500	59	59	3.003	0.334	-0.532	0	0	0	1935 no flow recorded

Gage USGS ID Number (1)	Systematic Record (yrs) (2)	Historical Record (yrs) (3)	Mean (log) (4)	Standard Deviation (log) (5)	Station Skew (log) (6)	Historical Events (7)	High Outliers (8)	Low Outliers (9)	Comments (10)
01491000	59	59	3.273	0.297	-0.191	0	0	1	
01493000	58	58	2.536	0.318	0.04	0	0	0	
01493500	55	55	2.627	0.421	0.926	0	0	0	
01495000	74	123	3.476	0.257	0.242	1	0	1	
01496000	37	37	3.215	0.25	0.815	0	0	0	
01497500	48	48	3.638	0.143	-0.17	0	0	0	
01498500	38	38	3.602	0.207	-0.064	0	0	0	
01499000	35	35	3.415	0.175	-0.162	0	0	0	
01500500	69	72	4.131	0.158	0.21	2	1	0	
01501000	49	68	3.565	0.154	-0.131	1	0	0	
01501500	36	36	1.55	0.304	0.671	0	0	0	
01502000	58	58	3.287	0.208	-0.312	0	0	0	
01502500	75	75	3.955	0.163	-0.188	0	0	0	
01503000	94	94	4.496	0.148	0.012	0	0	0	
01505000	69	71	3.661	0.192	-0.048	1	0	0	1914 no flow recorded
01505500	32	32	3.421	0.229	-0.339	0	0	0	
01507000	70	70	3.968	0.166	0.129	0	0	0	
01507500	34	34	3.423	0.163	0.039	0	0	1	
01510000	64	64	3.656	0.16	0.056	0	0	0	
01510500	34	34	3.791	0.169	0.153	0	0	0	
01537500	51	51	2.713	0.325	-0.051	0	0	0	1938 no flow recorded
01539000	68	71	3.946	0.255	0.188	1	6	0	
01554500	54	54	2.978	0.217	0.918	0	0	0	
01555500	77	77	3.689	0.276	1.041	0	0	0	
01573000	88	118	4.001	0.222	1.055	1	0	0	
01576500	78	78	3.888	0.249	0.226	0	0	0	

Table 32. Gages included in each of Eleven Subregions

Gage USGS ID Number (1)	Location (2)
Region 1	
1350000	Schoharie Creek at Prattsville, NY
1361500	Catskill Creek at Oak Hill, NY
1362500	Esopus Creek at Coldbrook, NY
1365000	Rondout Creek near Lowes, NY
1365500	Chestnut Creek at Grahamsville, NY
1367500	Rondout Creek at Rosendale, NY
1369000	Pochuck Creek near Pine Island, NY
1369500	Quaker Creek at Florida, NY
1370000	Wallkill River at Pellets Island, NY
1371000	Shawangunk Kill at Pine Bush, NY
1379000	Passaic River near Millington, NJ
1379500	Passaic River near Chatham, NJ
1380500	Rockaway River above reservoir at Boonton, NJ
1381500	Whippany River at Morristown, NJ
1384500	Ringwood Creek near Wanaque, NJ
1386000	West Brook near Wanaque, NJ
1387500	Ramapo River near Mahwah, NJ
1388000	Ramapo River at Pompton Lakes, NJ
1396500	South Branch Raritan River near High Bridge, NJ
1397000	South Branch Raritan River at Stanton, NJ
1398000	Neshanic River at Reaville, NJ
1398500	North Branch Raritan River near Far Hills, NJ
1399500	Lamington (Black) River near Pottersville, NJ
1400000	North Branch Raritan River near Raritan, NJ
1400500	Raritan River at Manville, NJ
1407500	Swimming River near Red Bank, NJ
1407830	Manasquan River near Georgia, NJ
1408000	Manasquan River at Squankum, NJ
1408015	Mingamahone Brook at Farmingdale, NJ
1408120	North Branch Metedeconk River near Lakewood, NJ
1408500	Toms River near Toms River, NJ
1409400	Mullica River near Batsto, NJ
1409810	West Branch Wading River near Jenkins, NJ
1411000	Great Egg Harbor River at Folsom, NJ
1412500	West Branch Cohansey River at Seeley, NJ

Gage USGS ID Number (1)	Location (2)
Region 2	
1413500	East BR Delaware River at Margaretville, NY
1414000	Platte Kill at Dunraven, NY
1414500	Mill Brook near Dunraven, NY
1415000	Tremper Kill near Andes, NY
1415500	Terry Clove Kill near Pepacton, NY
1418500	Beaver Kill at Craigie Clair, NY
1419500	Willowemoc Creek near Livingston Manor, NY
1420000	Little Beaver Kill near Livingston Manor, NY
1420500	Beaver Kill at Cooks Falls, NY
1421000	East BR Delaware River at Fishs Eddy, NY
1421900	W BR Delaware River Upsteam From Delhi, NY
1422000	West Branch Delaware River at Delhi, NY
1422500	Little Delaware River near Delhi, NY
1423000	West Branch Delaware River at Walton, NY
142400103	Trout Creek near Trout Creek, NY
1425500	Cold Spring Brook at China, NY
1426000	Oquaga Creek at Deposit, NY
1426500	West Branch Delaware River at Hale Eddy, NY
Region 3	
1427500	Callicoon Creek at Callicoon, NY
1428750	West Branch Lackawaxen River near Aldenville, PA
1429300	Dyberry Creek above Reservoir near Honesdale, PA
1431500	Lakawaxen River at Hawley, PA
1435000	Neversink River near Claryville, NY
1437000	Neversink River at Oakland Valley, NY
1438300	Vandermark Creek at Milford, PA
1439500	Bush Kill at Shoemakers, PA
1440000	Flat Brook near Flatbrookville, NJ
1440400	Brodhead Creek near Analomink, PA
1441000	McMichael Creek near Stroudsburg, PA
1442500	Brodhead Creek at Minisink Hills, PA
1443500	Paulins Kill at Blairstown, NJ
1445000	Pequest River at Huntsville, NJ
1445500	Pequest River at Pequest, NJ
1446000	Beaver Brook near Belvidere, NJ

Gage USGS ID Number (1)	Location (2)
Region 4	
1446600	Martins Creek near East Bangor, PA
1447500	Lehigh River at Stoddartsville, PA
1448000	Lehigh River at Tannery, PA
1448500	Dilldown Creek near Long Pond, PA
1449360	Pohopoco Creek at Kresgeville, PA
1450500	Aquashicola Creek at Palmerton, PA
1451800	Jordan Creek near Schnecksville, PA
1452000	Jordon Creek at Allentown, PA
1453000	Lehigh River at Bethlehem, PA
1455200	Pohatcong Creek at New Village, NJ
1459500	Tohickon Creek near Pipersville, PA
Region 5	
1464000	Assupink Creek at Trenton, NJ
1464500	Crosswicks Creek at Extonville, NJ
1464515	Doctors Creek at Allentown, NJ
1464538	Crafts Creek at Columbus, NJ
1465000	Neshaminy Creek at Rushland, PA
1465850	South Branch Rancocas Creek at Vincentown, NJ
1466000	Middle Br Mount Misery Bk in Byrne State Forest, NJ
1466500	McDonalds Branch in Byrne State Forest, NJ
1467000	North Branch Rancocas Creek at Pemberton, NJ
1467081	South Branch Pennsauken Creek at Cherry Hill, NJ
1467150	Cooper River at Haddonfield, NJ
1467305	Newton Creek at Collingswood, NJ
Region 6	
1467500	Schuylkill River at Pottsville, PA
1468500	Schuylkill River at Landingville, PA
1470500	Schuylkill River at Berne, PA
1470779	Tulpehocken Creek near Bernville, PA
1471000	Tulpehocken Creek near Reading, PA
1471980	Manatawny Creek near Pottstown, PA
1472000	Schuylkill River at Pottstown, PA
1472157	French Creek near Phoenixville, PA
1472162	Schuylkill River at Phoenixville, PA
1472198	Perkiomen Creek at East Greenville, PA
1472199	West Branch Perkiomen Creek at Hillegass, PA
1472500	Perkiomen Creek near Frederick, PA

Gage USGS ID Number (1)	Location (2)
1473000	Perkiomen Creek at Graterford, PA
1474000	Wissahickon Creek at Mouth, Philadelphia, PA
1474500	Schuylkill River at Philadelphia, PA
Region 7	
1475300	Darby Creek at Waterloo Mills near Devon, PA
1475850	Crum Creek near Newtown Square, PA
1476500	Ridley Creek at Moylan, PA
1477120	Raccoon Creek near Swedesboro, NJ
1478000	Christina River at Coochs Bridge, DE
1478200	Middle Branch White Clay Creek near Landenberg, PA
1478500	White Clay Creek above Newark, DE
1479000	White Clay Creek near Newark, DE
1480000	Red Clay Creek at Wooddale, DE
1480300	West Branch Brandywine Creek near Honey Brook, PA
1480610	Sucker Run near Coatesville, PA
1480675	Marsh Creek near Glenmoore, PA
1481500	Brandywine Creek at Wilmington, DE
1482500	Salem River at Woodstown, NJ
Region 8	
1483500	Leipsic River near Cheswold, DE
1484000	Murderkill River near Felton, DE
1484500	Stockley Branch at Stockley, DE
1485500	Nassawango Creek near Snow Hill, MD
1486000	Manokin Branch near Princess Anne, MD
1487000	Nanticoke River near Bridgeville, DE
1488500	Marshyhope Creek near Adamsville, MD
1491000	Choptank River near Greensboro, MD
1493000	Unicorn Branch near Millington, MD
1493500	Morgan Creek near Kennedyville, MD
1495000	Big Elk Creek at Elk Mills, MD
1496000	Northeast Creek at Leslie, MD
Region 9	
1497500	Susquehanna Creek at Colliersville, NY
1498500	Charlotte Creek at West Davenport, NY
1499000	Otego Creek near Oneonta, NY
1500500	Susquehanna River at Unadilla, NY
1501000	Unadilla River near New Berlin, NY
1501500	Sage Brook near South New Berlin, NY

Gage USGS ID Number (1)	**Location (2)**
1502000	Butternut Creek at Morris, NY
1502500	Unadilla River at Rockdale, NY
1503000	Susquehanna River at Conklin, NY
Region 10	
1505000	Chenango River at Shurborne, NY
1505500	Canasawacta Creek near South Plymouth, NY
1507000	Chenango River at Greene, NY
1507500	Genegantslet Creek at Smithville Flats, NY
1510000	Oteselic River at Cincinnatus, NY
1510500	Oteselic River near Upper Lisle, NY
Region 11	
1537500	Solomon Creek at Wilkes-Barre, PA
1539000	Fishing Creek near Bloomsburg, PA
1554500	Shamokin Creek near Shamokin PA
1576500	Conestoga River at Lancaster, PA

Appendix D

Technical Considerations

The purpose of this Appendix is to describe the assumptions made, statistical methods used, and results obtained in estimating regional skew for the Delaware River Basin. The estimates of regional skew are needed for estimating the log-Pearson III annual peak flow-frequency curves at stream gage locations when using the federal guidelines, "*Guidelines for Determining Flood Flow Frequency*", *Bulletin 17B* (IACWD 1982).

This Appendix is considered a stand-alone document and has text and tables duplicated from the main body of this report.

D.1 Motivation

Bulletin 17B recommends using three methods to compute alternative estimates of regional skew, G_r, and mean square error (MSE): a region average (Method 1), a skew isoline map (Method 2); and a regression relating gage skew and watershed parameters (Method 3).

Bulletin 17B was published in 1982. Therefore, while the recommended methods were applied for this study, also an attempt to improve the regional skew estimates using more recently developed methods was made. Specifically, these improved methods are inclusion of sampling error in calculations of MSE, L-moment analysis for statically checking regions for homogeneity, and generalized least squares (GLS) regression.

Such improvements were attempted because valuing the different regional skew estimates using quantitative measures, such as associated MSE (the MSE of the particular method employed), must be tempered by the underlying fundamental frequency analysis assumptions, which *Bulletin 17B* does not completely address. In other words, the comparative differences in MSE values obtained in the analysis might not be significant given the analysis assumptions.

This Appendix is organized in seven sections. Section D.2 provides an introduction to regional skew calculation. Section D.3 describes the fundamental assumptions made in estimating flow-frequency curves. Region average methodology and application results are described in Section D.4. Section D.5 describes the application of generalized least squares regression. Section D.6 describes the split sample testing methods and their implication to the regional skew analysis. Finally, additional technical considerations for selecting a regional skew estimate method are given in Section D.7.

D.1.1 Frequency Curve Estimation

The frequency curve at a gage location is computed from the sample mean and standard deviation of the annual gage peak flow logarithms and an adopted skew coefficient. The sample mean and standard deviation are computed using the moment estimators:

$$X_m = \sum_{i=1}^{N} \frac{X_i}{N} \tag{D-1}$$

$$S_x = \sqrt{\sum_{i=1}^{N} \frac{(X_i - X_m)^2}{N-1}} \tag{D-2}$$

where:

X_i is the logarithm of the peak flows,
N is the number of annual gage peaks,
X_m is the sample mean, and,
S_X is the sample standard deviation.

The skew coefficient is estimated from an adopted skew, G, which is calculated as a weighted average of regional and gage flow record skew estimates (IACWD 1982, 12):

$$G = \frac{mse_g G_r + mse_r G_g}{mse_g + mse_r} \tag{D-3}$$

where:

G_g is the gage skew coefficient,
mse_g is the associated gage MSE,
G_r is the regional skew, and,
mse_r is the associated regional MSE.

The regional skew (G_r), referred to as generalized skew in the guidelines, and the associated MSE are determined from a regional skew study (IACWD 1982, 10). The gage mean square error (mse_g) for record length (N) has been estimated by simulation methods (IACWD 1982, Table 1). The gage skew coefficient (G_g) is determined using the following moment estimate:

$$G_g = \frac{N\sum_{i=1}^{N} (X_i - X_m)^3}{(N-1)(N-2)S_x^3} \tag{D-4}$$

where:

X_i is annual peak flow,
X_m is the mean of the annual peak flows, and,
S_X is the skew coefficient calculated from the sample record.

D.1.2 Regional Skew Methods

For this study, guidelines and methods outlined in *Bulletin 17B* to calculate regional skew were followed. *Bulletin 17B* recommends that the method of choice should be selected based on the lowest MSE of one of three methods:

- **Method 1:** Region average skew. This method was implemented four ways:
    - **Method 1a** - average skew of the entire basin;
    - **Method 1b** - average skews of homogenous regions;
    - **Method 1c** – weighted average skews of homogeneous regions;
    - **Method 1d** – GLS regression constants for statistically homogeneous regions.

- **Method 2:** Skew isoline map.

- **Method 3:** Predictive equations using GLS regression.

D.2 Regional Skew Analysis Assumptions

The regional skew analysis used to develop the regional skew (G_r) depends on assumptions made in estimating flow-frequency distributions using gage estimates and some assumptions regarding the statistical dependence between annual floods at different gages. The basic flow-frequency analysis assumptions of gage annual peak flows are as follows:

- The flow period of record of interest is statistically stationary, i.e., the statistical characteristics have not changed over the period of record. Factors that influence these characteristics, such as climatic variability and anthropogenic activities, are assumed to have had no effect on the flow record.

- The annual peaks can be described by a single flow-frequency distribution.

- Flows are measured without error.

- The flow-frequency characteristics of the period of record are indicative of the risk of future flooding over some designated planning period.

- Annual peaks are independent from year to year.

The nature of these assumptions has led many researchers to acknowledge the approximate nature of any estimated flow-frequency distribution. For example, consider the comments of Stedinger, Vogel, and Foufoula-Georgiou (1992, 18.22):

> *Several fundamental issues arise when selecting a distribution. One should distinguish between the following questions:*
>
> 1) *What is the true distribution from which the observations are drawn?*
>
> 2) *What distribution should be used to obtain reasonably accurate and robust estimates of design quantiles and hydrologic risk?*
>
> 3) *Is a proposed distribution consistent with the available data from the site?*

> *Question 1 is often asked. Unfortunately, the true distribution is probably too complex to be of practical use. Still L-moment skewness-kurtosis and CV-skewness diagrams… are good for investigating what simple families of distributions are consistent with available data sets for a region. Standard goodness-of-fit statistics, such as probability plot correlation tests… have also been used to see how well a member of each family of distributions can fit a sample. Unfortunately, such goodness-of-fit statistics are unlikely to identify the actual family from which the samples are drawn – rather, the most flexible families generally fit the data best.*

Many distributions provide an acceptable fit to the data. The Water Resource Council found that the log-Pearson III distribution met its acceptability criteria (*Bulletin 17B*, Appendix 14).

However, these acceptability criteria do not address the errors that could occur by not identifying the true distribution, i.e., the true model. If the true distribution is not known, the errors cannot be quantified. Hosking and Wallis (1997, Chapters 6 and 7) investigated some plausible differences that might occur and concluded that sampling error, not modeling error (defined in Section D.5), was the dominant error associated with estimating flood quantiles as large as p=0.01 (100-year) for typical stream flow record lengths. The same study also concluded model error was likely to be dominant for larger quantiles.

Model error is relevant to the regional skew estimation because the computation of both gage error and regional MSE are dependent on: the assumption of a distribution; the estimates of MSE for relatively large quantiles (for example p=0.01); and the upper distribution moments, such as the skew.

In addition to the problems associated with flow-frequency curve distribution selection, assumptions regarding the inter-gage dependence of annual peak floods can detract from the value of a regional analysis. The improvement in estimated skew from a regional analysis effectively depends on pooling the records from all the gages in a region. For example, if forty gages exist in a region, each with twenty-five years of record, then the total pooled record is 1,000 years. However, lack of independence reduces the effective record length available. An innovation of the GLS regression method (detailed in Section D.5) accounts for this reduction in effective record length.

The inter-gage dependence is measured by the correlation between concurrent periods of annual peak flow. Unfortunately, this simple measure does not completely characterize the nature of inter-gage dependence. In the Delaware River Basin, the diverse types of flood-producing storms - hurricanes and thunderstorms - have very different regional effects. Peak flow occurrences due to thunderstorms are likely to be independent, whereas the opposite is true for hurricanes. Categorizing the dependence due to the different regional effects of these two events by a simple correlation coefficient is approximate at best.

In conclusion, the assumptions made in performing flow-frequency analysis and regional skew analysis render measures of accuracy, such as MSE of prediction, approximate. The approximate nature of these measures must be considered when evaluating the accuracy of skew estimates from different methods.

D.3 Skew Isoline Map (Method 2)

ArcGIS® tools were used to develop skew isolines (contours) for the Delaware River Basin. First, skew values were plotted at their corresponding gage locations. This is consistent with the methodology used by the USGS (USGS 2006). Then isolines were developed from the plotted station skew values using three methods: linear interpolation, inverse distance weighting (IDW), and engineer's judgment-assisted IDW.

To develop isolines using linear interpolation the skew values were first inspected to see if there was a pattern. Finding such a pattern would allow more reasonable drawing of isolines on a skew map than just using only mathematical algorithms alone. In the absence of a pattern, mathematical algorithms would still be used to create regional skew contour values. However, additional inspection of the computed skew contours and their comparison to basin physiography and hydrology would be required to assure the computed contours are reasonable and rational and not just "lines connecting the dots".

A triangular irregular network (TIN) was developed using ArcGIS® tools and the station skew values plotted at their respective gaging locations. Skew contours were then linearly interpolated from the TIN.

To develop isolines using IDW, an algorithm was used to compute initially a gridded surface of skew, from which skew contours could be computed. The grid size for the surface was 1.21 square miles, which is smaller than the smallest gaged drainage area. The IDW algorithm is:

$$z_j = \frac{\sum_i \frac{z_i}{(d_{ij})^n}}{\sum_i \frac{1}{(d_{ij})^n}} \tag{D-5}$$

where:

z_j is the value at an unknown point,
z_i is the value at a known point,
d_{ij} is the distance between a known and the unknown point, and,
n is a user-defined exponent.

The number of known points used to determine an unknown value can be specified by the user with ArcGIS® tools.

An exponent n of one (1) was used and the nearest forty gages were used for z_i. The exponent was chosen so as to maximize the influence of nearby gages in calculating skew values, while forty gages were used to be consistent with *Bulletin 17B* guidelines. From this gridded surface, the skew isolines were calculated.

The mathematical algorithms used to calculate isolines in ArcGIS® using both linear interpolation and IDW result in exact solutions at the sites where skews are plotted, and therefore both methods have an MSE equal to zero (0).

The third mapping method uses the IDW-created isolines as a starting point. These isolines, and the station skew values used to create them, were then compared with an elevation map of the region. The isolines then were modified using engineering judgment based on region physiography and hydrology. In general, skew contours representing local minima or maxima around a single gage were removed. Contours were also redrawn to establish skew contours around regions of similar physiographic and hydrologic characteristics.

D.4 Region Area Analysis (Method 1)

The purpose of this section is to describe the calculation methods and the results for the regional area analysis. The method herein is similar to the region average method recommended for regional skew MSE in *Bulletin 17B*.

Bulletin 17B recommends a single average of all gage skews in a region (Method 1a). The MSE is then computed as the average of the sum of squared differences between the gage and regional skew values:

$$MSE = \sum_{i=1}^{n} \frac{(G_i - G_r)^2}{n} \qquad \text{(D-6)}$$

where:
G_r is regional skew, calculated as the average of the gage skew,
G_i for the *n* gages in the region.

The *Bulletin 17B* recommended computations of regional skew and of MSE give equal weight to each gage without considering the varying estimation error in gage skew because of different record lengths. This analysis attempted to improve regional skew estimates by using the methods outlined in this section to estimate gage weights that consider the inter-gage correlation between annual peak flows and differing gage record lengths.

Section D.4.1 describes the model developed to estimate a skew error covariance matrix used to compute the weights for each gage in the region average method. The covariance matrix was used for the region average analysis (Method 1) and for the GLS regression (Method 3, described in Section D.5) for computing regional skew and MSE.

The region average analysis (Method 1) error described in Section D.4.1 effectively adds sampling error into the *Bulletin 17B* recommended approach for obtaining an area average. Sampling error is a function of both the gage record length and the correlation of annual peak flows between gages. The approach is expanded to identify regional skew for subregions analogous to the *Bulletin 17B* recommended contour approach. Section D.4.2 describes the methodology for estimating skew sampling error for historical information. Section D.4.3 details the methodology for refining regions using L-moment analysis. Section D.4.4 describes the results of computing region average skew coefficients. Finally, Section D.4.5 describes the estimation of a skew for a region using only a constant in a GLS regression.

D.4.1 Skew Covariance Error Matrix Estimation

D.4.1.1 Outline of Estimation Methodology

Traditionally, regional skew does not account for sampling error in each skew value (*Bulletin 17B*). However, inclusion of sampling error in calculations of MSE represents an improvement to the *Bulletin 17B* methodology. Sampling error is summarized in a covariance matrix that has the following form for a two-gage region:

$$\mathbf{V} = \begin{matrix} v_{1,1}^2 & v_{1,2} \\ v_{2,1} & v_{2,2}^2 \end{matrix} \tag{D-7}$$

where the diagonal elements of the matrix are the gage skew average squared estimation error, i.e., MSE, and the off-diagonal elements are the skew covariance estimation error.

The covariance matrix was estimated as follows:

1. Computed the log-Pearson III distribution statistics for the streamgages in the region.

2. Estimated the inter-gage correlation from the concurrent period of record for these gages.

3. Developed a regional relationship relating distance between gage and correlation from the inter-gage correlation estimates, as described in Section D.4.1.2.

4. Created many realizations of each region of gages using Monte Carlo simulation, the estimated log-Pearson III statistics, and the inter-gage correlation relationship. (A realization consists of a simulated period of record of annual peak flows at each gage.)
    a. For each realization, computed the skew error as the difference between the gage skew estimated from the observed period of record and the gage skew from the simulated period of record. (Steps a through c were completed for each realization.)
    b. Calculated the squared error for each gage as the square of the skew error.
    c. Calculated covariance error as the product of this skew error at two different gages.

5. Summed the squared errors and covariance errors for all realizations.

6. Calculated the covariance matrix elements as the average of summed squared and covariance errors.

7. Added realizations (continued Step 4) until the matrix errors were estimated with sufficient accuracy; when the estimated errors did not change within a desired tolerance.

The simulation of realizations depends on a simulated random sample of normal deviates for the region. (Normal deviates are samples from a distribution with mean equal to zero and standard deviation equal to one.) The realizations using normal deviates were simulated as follows:

1. Created a correlation matrix for the gage sites in the region using the regional relationship between gage distance and correlation. This matrix had ones on the diagonal and inter-gage correlations for the off-diagonal elements.

2. Simulated correlated normal deviates for each gage period of record using a well known methodology from time series analysis (Salas 1992, 19.55).

3. Converted the gage period of record normal deviates to log-Pearson flows by first computing the exceedance probability for each deviate for a normal distribution (mean equal to zero and standard deviation equal to one); and then, converting the exceedance probability to log-Pearson III flows knowing the period of record statistics at each gage.

D.4.1.2 Estimation of Inter-Gage Correlation Relationship

An inter-gage correlation relationship is estimated using a regional approach to reduce the sampling error in the correlation estimates. This follows the approach by Tasker and Stedinger (1989) to find a regional model for correlation versus distance. The inter-gage dependence relationship is measured by the correlation coefficient, *R*, between concurrent periods of annual peak flow. The correlation coefficient is a function of the inter-gage distance, because a given event affecting a gage is more likely to have impacts on other nearby gages (as described previously in Section D.2).

This relationship was estimated graphically. Initially, the correlations were examined as a function of concurrent record length, as shown in Figure D-1 through Figure D-5. The scatter in the data is related to record length and, to some extent, the number of gages available. The final estimate of the relationship was obtained from gages with at least seventy years of concurrent period, as shown in Figure D-6. This proved to provide the best trade-off between record length and number of gages.

D.4.1.3 Simulation Verification

The computation of the skew covariance error matrix uses the same procedure presented in Wallis et al. (1974) to compute the MSE of skew estimates shown in *Bulletin 17B* (Table 1). If the inter-gage correlation is zero, then the diagonal elements of this matrix (the gage skew average squared sampling error) should be the values shown in this table (*Bulletin 17B*, Table 1). Figure D-7 shows the excellent agreement between the simulated diagonal elements of the covariance matrix and *Bulletin 17B* (Table 1) values.

D.4.2 Estimating Skew Sampling Error for Historical Information

Bulletin 17B recommends that the MSE for station skew be computed assuming equal weight on systematic and historical record lengths. This ignores the loss of information due to the missing data between systematic records and historical information (Tasker, 1983).

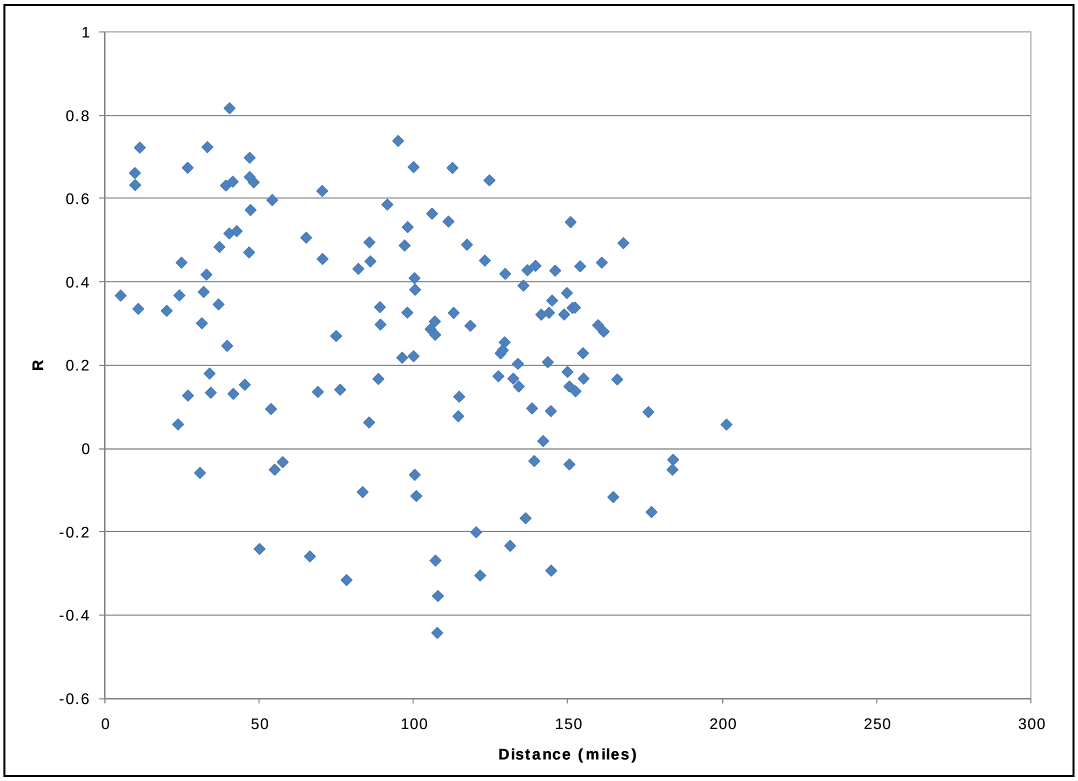

Figure D-1. Inter-gage correlation versus distance between gages for eleven years of concurrent period of record

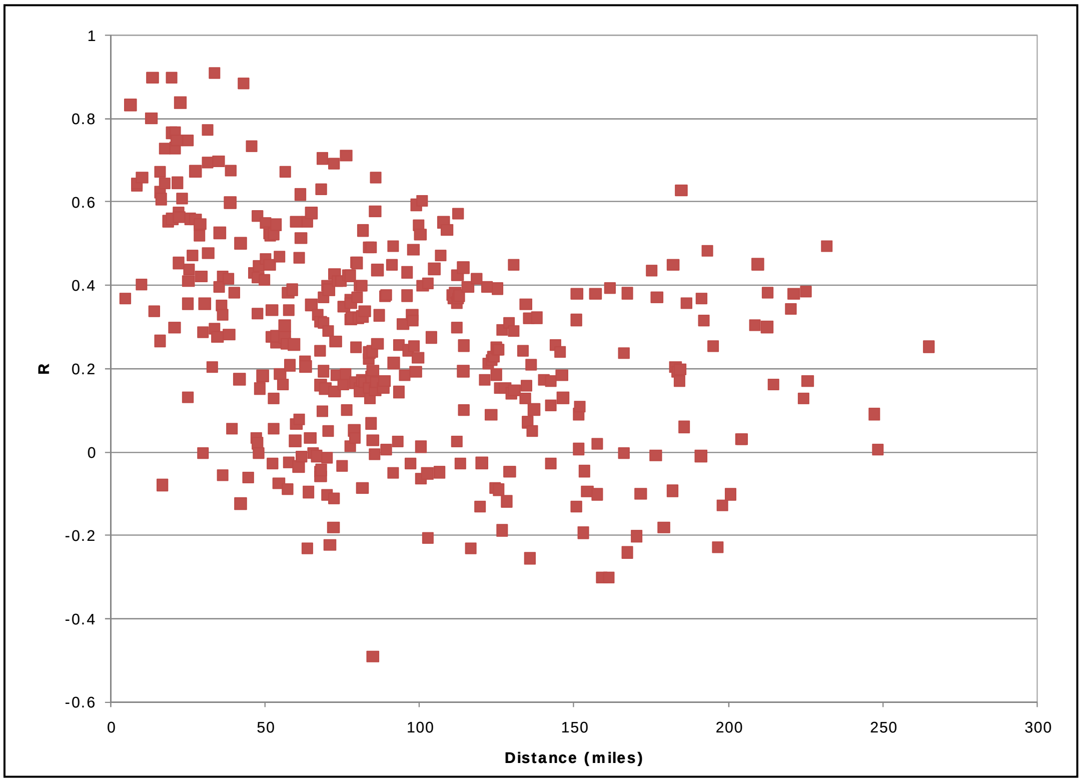

Figure D-2. Inter-gage correlation versus distance between gages for twenty-five years of concurrent period of record

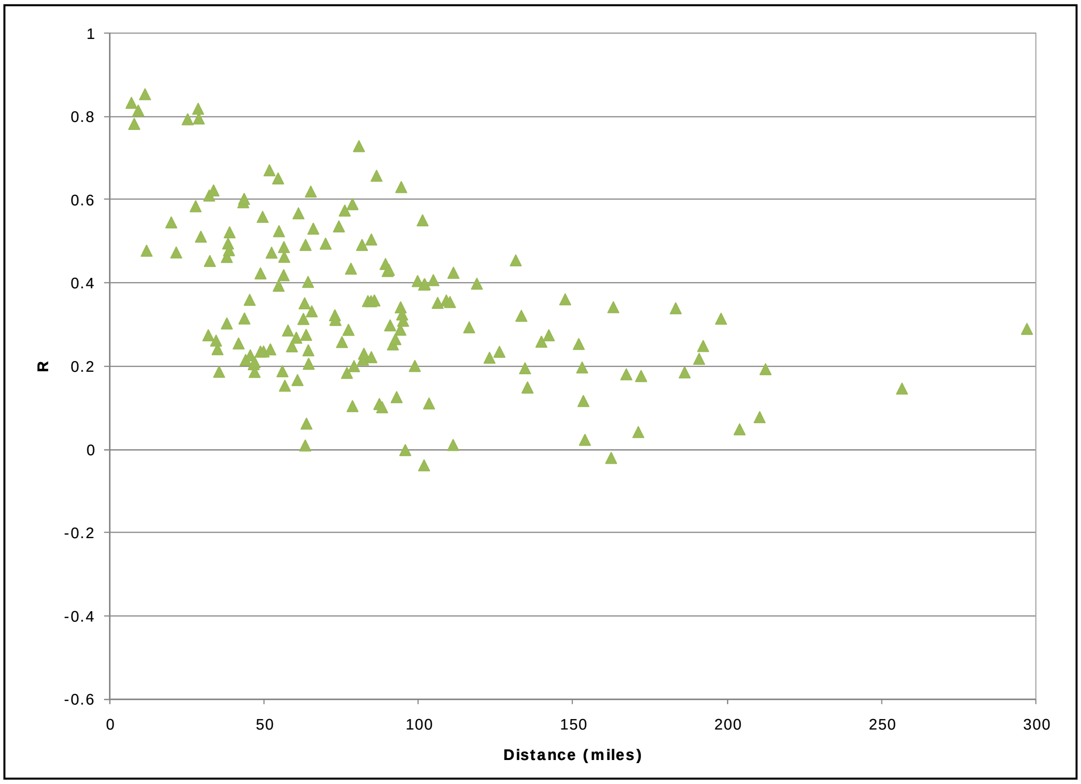

Figure D-3. Inter-gage correlation versus distance between gages for fifty years of concurrent period of record

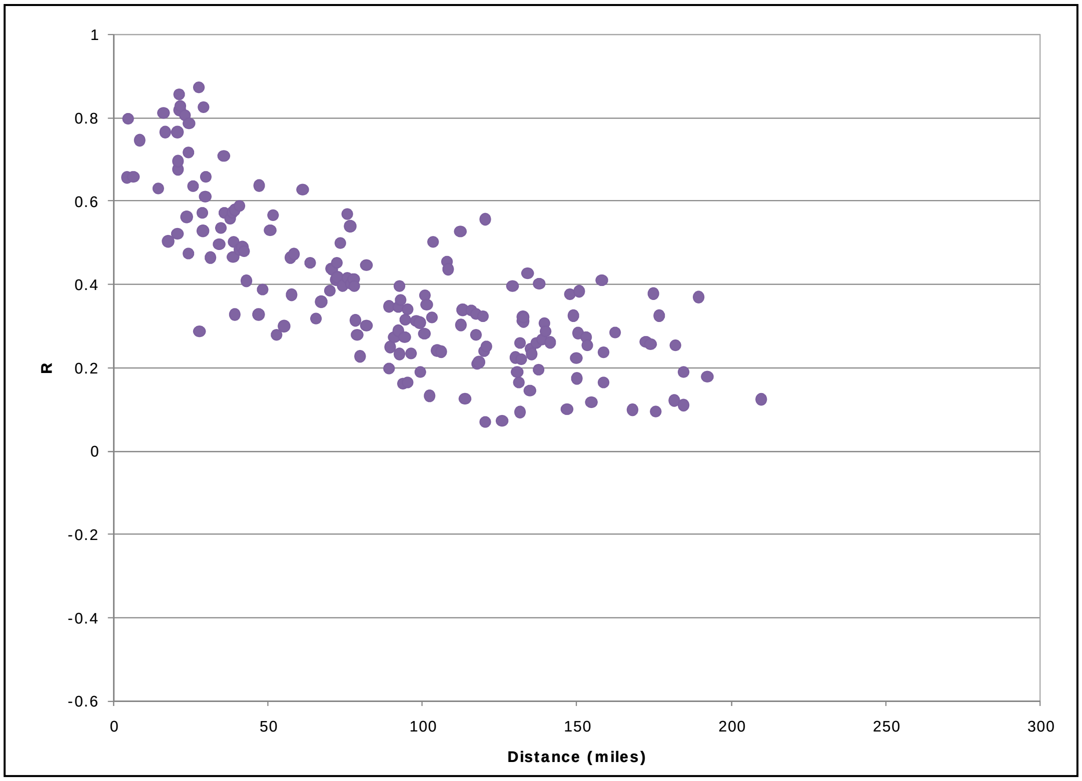

Figure D-4. Inter-gage correlation versus distance between gages for seventy years of concurrent period of record

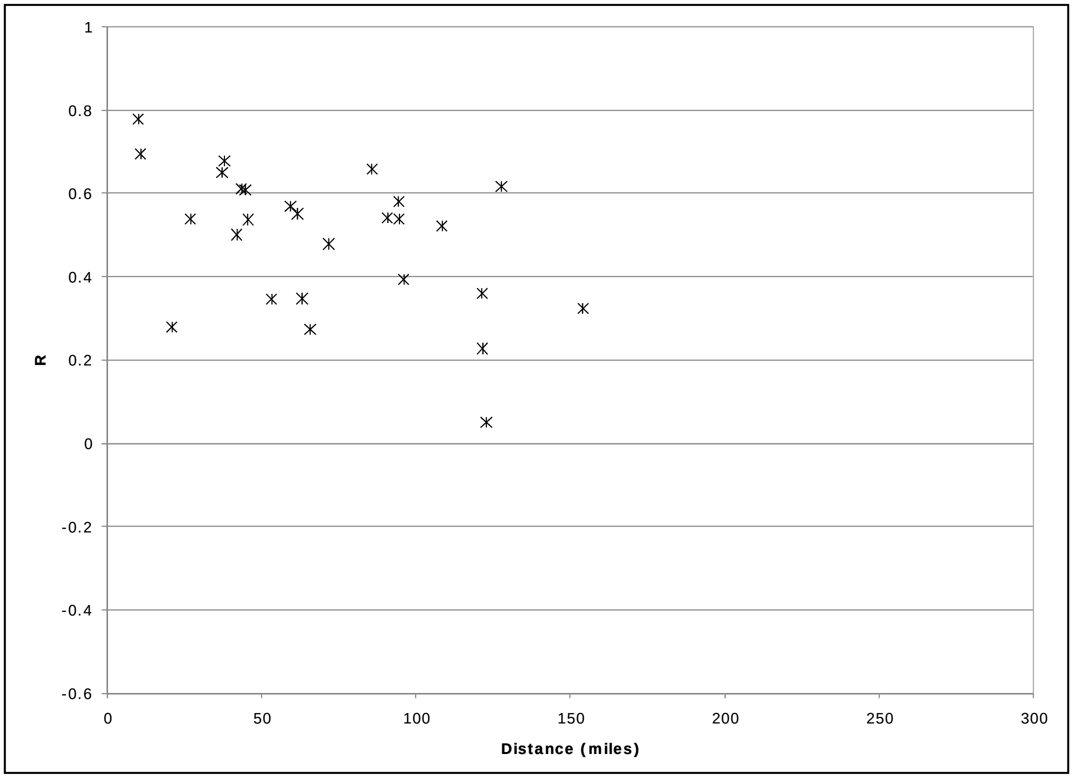

Figure D-5. Inter-gage correlation versus distance between gages for ninety years of concurrent period of record

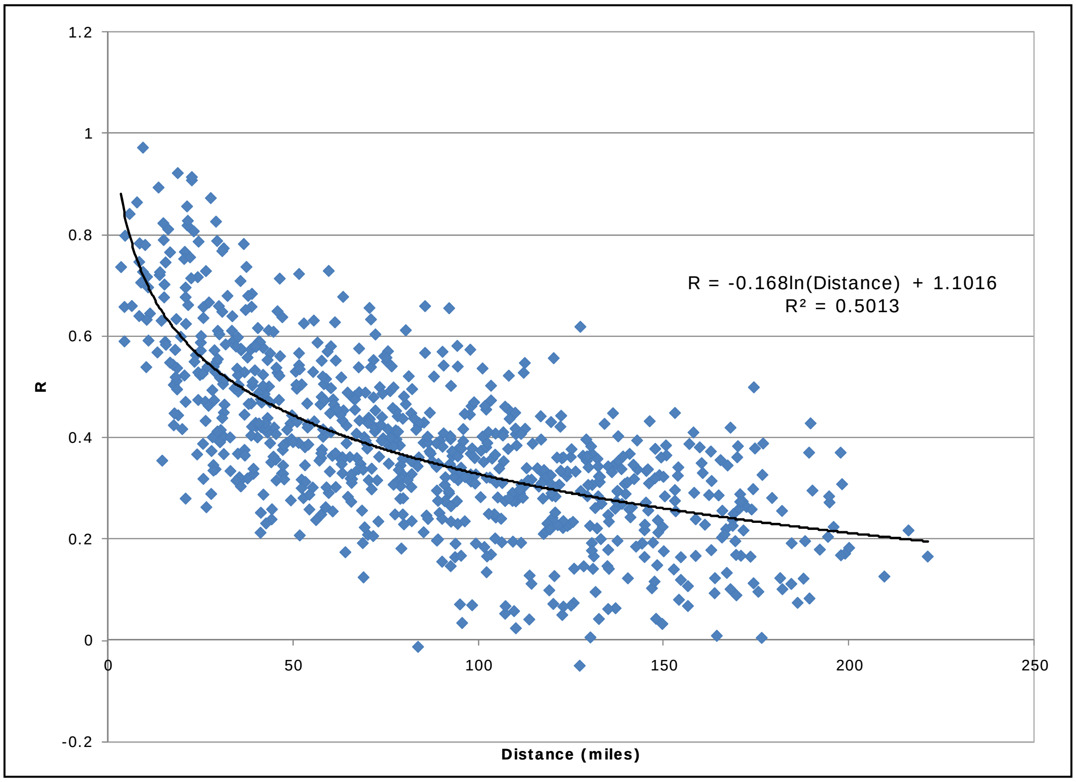

Figure D-6. Selected inter-gage correlation versus distance for a concurrent period greater than seventy years.

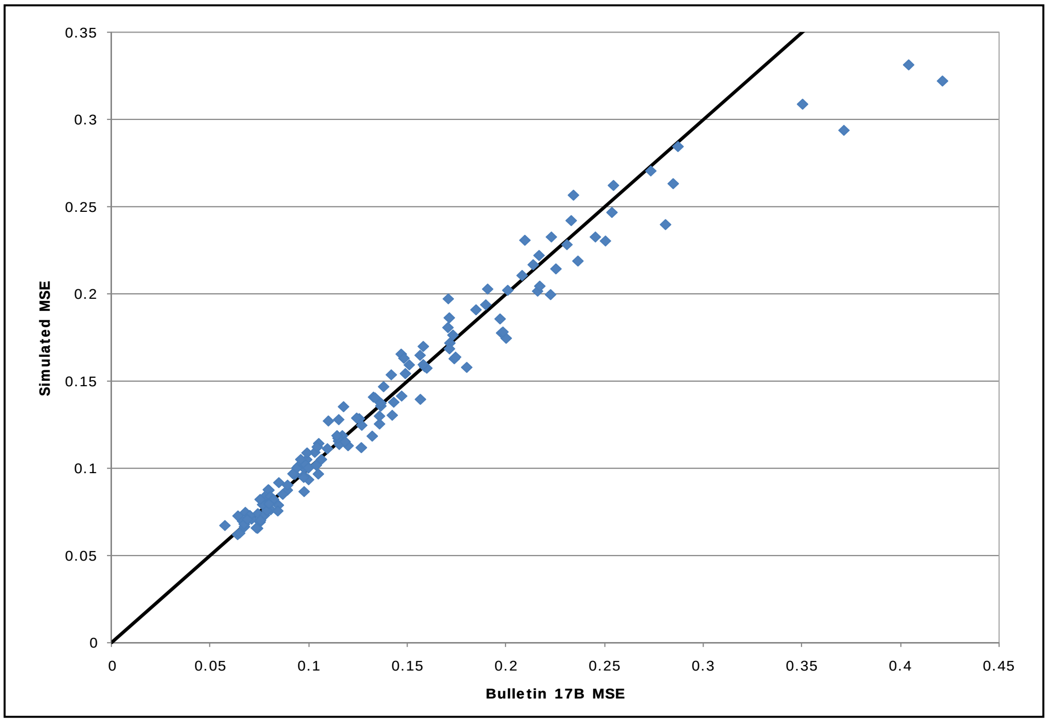

Figure D-7. Comparison of simulated skew MSE and Bulletin 17B Table 1 MSE

Skew sampling error for gages with historical information was computed using part of the simulation procedure described in Section D.4.1.1 as follows:

1. Identified the number, n_h, of largest values that receive historical weighting in the observed gage peak flows (those above the historical threshold).
2. Simulated the number of peak flows equal to the historic period record length, and ranked the flows by magnitude.
3. Chose the n_h largest simulated peak flows from the ranked list of simulated flows.
4. Selected the remaining flows at random from the simulated flows (not ranked) that have had the n_h events removed. The remaining flows below the historical threshold are equal to the systematic period minus n_h.
5. Computed the statistics of this simulated historical sample (the historical weighted mean, standard deviation, and skew) using the historical weighting equations given in *Bulletin 17B* (1982, 6-3, Equation 6-1).
6. Used the simulated historical weighted skew in the same way as a skew simulated for a systematic period in computation of squared error for any realization, as described in Section D.4.1.1.

MSE simulation results for gages with historical information are shown in Table D-1. The value of historical information can be assessed by comparing the skew MSE of estimation for the

Table D-1. Comparison of MSE for systematic and historical periods of record

Gage ID (1)	Skew (2)	Systematic Record Length (yrs) (3)	Historical Record Length (yrs) (4)	Systematic MSE (5)	Historical MSE (6)
1350000	-0.036	97	103	0.057	0.059
1369000	0.540	40	47	0.171	0.150
1369500	0.273	42	47	0.141	0.132
1371000	0.132	26	68	0.202	0.230
1388000	0.368	85	125	0.082	0.061
1396500	0.355	89	111	0.078	0.059
1400000	0.551	83	111	0.096	0.077
1407500	0.467	84	88	0.089	0.087
1414000	0.301	31	65	0.186	0.153
1418500	0.144	38	60	0.144	0.130
1419500	0.360	37	59	0.165	0.138
1421000	0.024	94	103	0.058	0.060
1422000	0.182	38	60	0.147	0.132
1428000	1.051	28	55	0.314	0.225
1431500	0.933	78	98	0.137	0.116
1435000	-0.121	67	69	0.085	0.093
1447500	0.616	62	65	0.126	0.122
1448500	0.139	48	58	0.116	0.105
1452500	0.578	58	62	0.130	0.122
1464500	-0.138	66	69	0.087	0.088
1465500	0.345	71	74	0.094	0.092
1467500	0.142	45	53	0.123	0.119
1470500	-0.036	59	65	0.090	0.096
1471980	-0.248	31	34	0.182	0.193
1472000	0.148	79	105	0.075	0.069
1473900	1.050	25	45	0.339	0.279
1474500	-0.095	75	137	0.075	0.070
1475000	1.429	58	67	0.298	0.275
1475300	0.266	26	28	0.214	0.224
1495000	0.242	74	123	0.085	0.069
1500500	0.210	69	72	0.088	0.080
1501000	-0.131	49	68	0.114	0.110
1505000	-0.048	69	71	0.079	0.085
1539000	0.188	68	71	0.088	0.084
1573000	1.055	88	118	0.147	0.129

systematic period to the skew MSE computed by the simulation of the diagonal elements of the skew covariance error matrix. As the number of unobserved flows (missing data) between historical and systematic observations increases, the historical information becomes less valuable. In fact, the historical information does not improve record length if the amount of missing data is too great. As a result, in those cases the systematic and historical record length MSE are approximately equal (at least within simulation accuracy).

D.4.3 Region Identification

Procedures outlined by Hosking and Wallis (1997) were used to define regions using L-moments because *Bulletin 17B* does not provide statistical criteria for region definition. However, *Bulletin*

17B does provide general recommendations for the number of gages needed to form a region: forty gages, or all gages within 100 miles. The purpose of this section is to describe the application of statistical criteria used by Hosking and Wallis (1997) for forming regions.

In this application, the homogenous regions are tentatively identified based on similar meteorologic and hydrologic characteristics. This homogeneity is examined by computing the L-moment discordancy and heterogeneity statistics for the aggregation of gages within the region. The discordancy statistic provides a means for identifying gages with statistical characteristics which deviate more than would be expected from the average statistical characteristics of gages within a region. A stream gage was considered discordant if its discordancy statistic was greater than the critical value listed in Table D-2. The heterogeneity statistics measure the difference between the average sample statistics of an aggregation of gages (a region) and the sample statistics implied by an index flood distribution (which assumes coefficient of variation and skew are the same for all gages).

Table D-2. Critical values for discordancy (Hosking and Wallis 1997)

Number of Sites in Region (1)	Critical Value (2)
5	1.333
6	1.648
7	1.917
8	2.140
9	2.329
10	2.491
11	2.632
12	2.757
13	2.869
14	2.971
≥15	3.000

More specifically, the heterogeneity statistics are H(1), H(2), and H(3), and are defined as follows:

- H(1) measures the relative difference between the aggregate sample L-CV and the flood distribution L-CV, where L is the L-moment, and CV is the coefficient of variation.
- H(2) measures the difference between the average distance from the centroid of the sample of gages (on a plot of L-skew versus L-CV) and the distance of an individual gage L-CV versus L-skew from the centroid.
- H(3) measures the difference between the average distance from the centroid of the sample of gages (on a plot of L-skew versus L-kurtosis) and the distance of an individual gage L-skew versus L-kurtosis from the centroid.

The heterogeneity statistics are used to determine if an aggregation of gages can be considered a homogeneous region. *A heterogeneity value less than one implies a fairly homogenous region, a value between one and two is marginally acceptable, and a value exceeding three is* *heterogeneous and not likely acceptable.* A technical advisory group reviewing statistical methods for the Upper Mississippi Flow Frequency Study (USACE 2000) recommended focusing on gage discordancy and the H(3) heterogeneity statistic for identifying regions in flow-frequency applications, so this recommendation was followed in this study.

Initially, regions were formed based on smaller river basins within the study area. Nominally, the river basins were identified in a north-south direction as a measure of distance from the coast and increasing basin elevation by using a range of USGS gage ID's. The size and number of the river basins used were selected to have a significant number of gages. Hosking and Wallis (1997) recommend having at least twenty gages in a region for identifying a candidate flood frequency distribution. However, in this application, the goal was to define regions where the gage flow frequency distribution has similar shapes, particularly for more infrequent quantiles, such as, p=0.01 (the 100-year). The number of gages was consequently relaxed, and the H(3) statistic was used to focus on similar shape.

The results of the analysis displayed in Table D-3 shows the resulting discordancy and H(3) statistics for the aggregations of gages forming each region. Region 1, which is generally to the East of the Delaware River Basin, has the most gages. The Delaware River Basin was the focus of forming regions, so a great deal of detail was not needed in examining the out-of-basin gages. All the gage regions were found to be acceptably homogenous given the low to moderate heterogeneity values for the H(3) statistic. Only the most eastern region - outside of the Delaware River Basin - has two highly discordant values. This is considered acceptable given the relatively large size of this region (Hosking and Wallis 1997).

Table D-3. Statistical test results for L-moment-defined regions

Region Number (1)	Number of Gages (2)	Heterogeneity of L-skew versus L-kurtosis (3)	Number of Moderately Discordant Gages (4)	Number of Highly Discordant Gages (5)
1	37	Moderate	1	2
2	18	Low	0	0
3	17	Moderate	0	0
4	12	Moderate	0	0
5	14	Low	0	0
6	16	Low	0	0
7	15	Low	0	0
8	13	Low	1	0
9	9	Low	1	0
10	6	Low	0	0
11	6	Moderate	1	0

D.4.4 Region Average Skew Coefficients (Method 1b)

The skew covariance error matrix (presented in Section D.4.1.1) was computed using the region average skew coefficients and corresponding estimates of MSE (presented in Sections D.4.1 and D.4.2). In this computation a simulated period of record at each gage is obtained as x_k^s for k=1,2,…m, where m is the period of record at the gage. Simulated skew at gage i is:

$$G_i^s = \sum_{k=1}^{m} \frac{m\,(x_k^s - x_m^s)^3}{(m-1)(m-2)(s^s)^3} \tag{D-8}$$

where x_m^s and s^s are mean and standard deviation determined from the simulated period of record, x_k^s (see Equations D-1 and D-2). An average skew for the j^{th} simulation of the region is then computed as:

$$G_r^j = \sum_{i=1}^{n} \frac{G_i^s}{n} \tag{D-9}$$

where n is the number of gages in the region. The associated MSE (mse_r) for the region is then computed using the average skew over the number of simulations N as:

$$mse_r = \sum_{j=1}^{N} \frac{(G_r^j - G_r)^2}{N} \tag{D-10}$$

where G_r is the average sample skew. The number of simulations is determined by the precision selected in determining the mean square error.

The estimates of regional skew and mse_r were compared to those obtained from the *Bulletin 17B* method. Table D-4 shows that the basin's weighted average skew value (weighted by the number of gages in the region) for all the gages in the region does not differ greatly from the weighted average skew for gages within the Delaware River Basin boundary (0.184 vs. 0.217). The difference between the MSE estimated from the gage skews per the *Bulletin 17B* method and the simulated values that account for sampling error is significant. For example, the mean square error for the Delaware Basin is 0.146 versus the simulated value of 0.251.

Table D-4. Regional skew values and associated MSE (Method 1b)

Region (1)	Number of Gages (2)	Region Skew (3)	Record MSE (4)	Simulated MSE (5)	Notes (6)
Study area[a]	163	0.184	0.142	0.241	
1	35	0.175	0.113	0.205	Outside DE Basin
2	18	0.177	0.099	0.202	
3	16	0.298	0.157	0.254	
4	11	0.291	0.148	0.241	
5	12	0.258	0.147	0.260	
6	15	0.033	0.242	0.331	
7	14	0.269	0.087	0.221	
8	12	0.181	0.187	0.273	Outside DE Basin
9	9	-0.015	0.078	0.175	Outside DE Basin
10	6	-0.002	0.027	0.147	Outside DE Basin
11	4	0.320	0.130	0.216	Outside DE Basin
All regions	152	0.184	0.133	0.232	
Delaware River Basin regions	86	0.217	0.146	0.251	

a. All gages initially considered.

b. All gages meeting data quality standard.

c. All gages within the Delaware Basin boundary.

D.4.5 GLS-Constant Region Average Skew (Method 1d)

The GLS regression (detailed in Section D.5) disaggregates the error in estimating the gage skew with a regional estimate as:

gage skew = regional skew + (*model error and time sampling error*)

The time sampling error is the sampling error due to having a limited period of record to estimate gage skew. The model error measures the error in predicting skew with a regression relationship even if no time sampling error is present in the gage skew values. The GLS regression for regional skew would typically include a regression constant plus independent parameters such as drainage area and mean annual precipitation. However, only a constant was used in this GLS-constant region average approach, detailed herein.

The constant in the regression equation provides a direct comparison with the regional average, which is obtained using standard methods outlined above. The GLS method considers the difference in sampling error and the correlation between gage annual peak flows when weighting gage skew estimates to obtain a regional skew. This weight is different than the equal weighting that gage skew estimates are given when computing the regional skew as an average of gage skew values as recommended in *Bulletin 17B*.

As in the GLS regression with independent parameters, the average variance of prediction (AVP) is used as a measure of the prediction error, instead of the standard MSE. (Simulated-MSE and AVP can be directly compared in this study, as described in Section D.5.) The AVP is the average squared prediction error (model error and time sampling error) obtained when using only the regression constant as an estimate of regional skew.

The AVP values found for the Delaware River Basin regions are shown in Table D-5. The weighted average constant of 0.151 would be used as the regional skew value. The weighted AVP of 0.044 would be used in place of MSE.

Table D-5. GLS-constant region average skew and errors for regions within the Delaware River Basin (Regions 2 through 7)

Region Number (1)	Constant (GLS-Constant Regional Skew) (2)	MSE (from AVP, Average Prediction Error) (3)	Model Error (4)
2	0.087	0.026	0.000
3	0.203	0.077	0.052
4	0.165	0.030	0.013
5	0.178	0.033	0.000
6	0.001	0.064	0.032
7	0.287	0.034	0.000
Weighted average	0.151	0.044	0.000

A troublesome aspect of the results is that a model error of zero was estimated for some of the regions. This could be interpreted as meaning that a constant skew value is in fact a perfect model for the region. Alternatively, this could have resulted due to the approximations made in computing the time sampling error. The GLS regression application resulted in zero model error in most of the applications defined in Section D.5. Model error is defined in Section D.5.

D.5 Generalized Least Squares Regression Analysis

The purpose of this section is to describe the generalized least squares regression method used in this study. *Bulletin 17B* recommends regression analysis as an option for estimating regional skew. Section D.5.1 describes the GLS method and how it represents an improvement over the typical application of ordinary least squares (OLS) regression. Statistical measures for evaluating the regression are presented in Section D.5.2. Section D.5.3 describes considerations in calculating regional skew using GLS regression. The application of GLS to the Delaware River Basin study area is detailed in Section D.5.4, which provides an evaluation of the independent variables employed in developing regression predictions. The application of GLS in determining regional skew is unique for each application. Standard statistical software packages do not provide options for using GLS regression. Therefore, specialized software to apply this regression methodology was used. Section D.5.5 presents verification of this software's algorithms. Finally, Section D.5.6 presents the leverage and influence of gage statistics and their significance to the analysis.

D.5.1 GLS Methodology (Method 3)

OLS (Draper and Smith 1966) is one of the three methods recommended in *Bulletin 17B* for analyzing regional skew. However, this regression approach does not account for the sampling error in gage skew estimates, i.e., the unequal error estimation due to differences in gage record length, nor does it account for the inter-gage correlation of gage annual peak discharges. GLS regression techniques have been developed to account for the sampling error and correlation issues in estimating regional skew.

The GLS regression relationship takes the same form as OLS:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \varepsilon \tag{D-11}$$

where:

y is the dependent variable,
x_i are the independent variables,
b_i are regression equation parameters, and,
ε is the regression residual error.

In this study, skew is the dependent variable and x_i are the independent hydrologic and meteorologic variables for the region of interest (stream length, slope, etc.). The residual error represents the inability of the independent variable to explain perfectly the variance of the dependent variable.

In the case of regional skew, the independent parameters are usually taken as the log values of hydrologic and meteorological characteristics. The regression coefficients are typically estimated using ordinary least squares analysis from observed data:

$$y_j = b_0 + b_1 x_{1,j} + b_2 x_{2,j} + \ldots + e_j \tag{D-12}$$

where:

y_j is the j^{th} observation of the dependent variable,

b_i are the i=0 to $p-1$ sample estimates of the coefficients of the $x_{i,j}$ independent variables for each of the j observations (Draper and Smith 1966)

The coefficients of the regressions are estimated by minimizing the sum of squared residuals over all the observations. Equation D-12 can be written in matrix notation:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e} \tag{D-13}$$

where:

Y is a nx1 gage column vector of the dependent variables,

b is a px1 column vector of the regression parameters,

X is a nxp matrix of the observed independent variables, and,

e is a nx1 column vector of regression residual errors.

For example, the matrices would have the following form for a two gage region with two independent variables:

$$\mathbf{Y} = \begin{matrix} y_1 \\ y_2 \end{matrix}, \quad \mathbf{b} = \begin{matrix} b_0 \\ b_1 \\ b_2 \end{matrix}, \quad \mathbf{X} = \begin{matrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \end{matrix}, \quad \mathbf{e} = \begin{matrix} e_1 \\ e_2 \end{matrix} \tag{D-14}$$

In a regional skew analysis, the errors will not be homoscedastic (as assumed in OLS) because of sampling error in gage skew estimates. These sampling errors result because of differing finite record lengths at the gage locations where skew is estimated. Furthermore, the residual errors will be correlated because of the inter-gage correlation in annual peak flow values. This inter-gage correlation is typically modeled as a non-linear function of the distance between gages. Under these circumstances, minimum variance estimates of the regression parameters are obtained using a GLS approach as presented in Draper and Smith (1966) as:

$$\mathbf{b} = (\mathbf{X'}\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X'}\mathbf{V}^{-1}\mathbf{Y}) \tag{D-15}$$

where:

X' is the transpose of **X**,

( )$^{-1}$ is a matrix inverse, and,

V is an nxn covariance matrix of residual errors.

An additional challenge in applying this approach is in estimating **V**. Estimation of the residual error covariance matrix is completed in the study described herein using the method proposed by Stedinger and Tasker (1989). In this approach, the regression residual error in estimating the skew is assumed to be separable into a regression model error and a time sampling error. The regression model error is the error which would result if the flow quantiles were estimated perfectly from the record at each gage. The time sampling error occurs because of the limited record lengths available to estimate the regional skew at each gage. The magnitude of this error is inversely proportional to the square root of the record length.

As an example of how this error model is developed, consider the covariance matrix for the two-gage case where in Equations D-14 and D-15, j=1, 2:

$$\mathbf{V} = \begin{matrix} d^2 + V^2_{1,1} & v_{1,2} \\ v_{2,1} & d^2 + v^2_{2,2} \end{matrix} \qquad \text{(D-16)}$$

where:

d^2 is the estimated regression model error, and,

$v^2_{i,j}$ is the time sampling error covariance squared described previously in Section D.4.1.1.

In the case where $i=k$, the covariance is the error variance for a particular gage; when $i \neq k$, the off-diagonal, matrix error covariance results from the inter-gage correlation of maximum flow values. Note also that the matrix is symmetric with $v_{i,j}=v_{j,i}$.

If the maximum annual flows are not correlated with other gage flows, then the off-diagonal values become zero, and the covariance error matrix becomes:

$$\mathbf{V} = \begin{matrix} d^2 + V^2_{1,1} & 0 \\ 0 & d^2 + v^2_{2,2} \end{matrix} \qquad \text{(D-17)}$$

When residual errors exhibit no inter-gage correlation, then the regression is referred to as weighted least squares (WLS). In the WLS solution for the parameters, $\mathbf{V}^{-1}$ becomes:

$$\mathbf{V} = \begin{matrix} \frac{1}{d^2 + V^2_{1,1}} & 0 \\ 0 & \frac{1}{d^2 + v^2_{2,2}} \end{matrix} \qquad \text{(D-18)}$$

and in Equation D-18 the estimate flow quantiles, **Y**, are weighted inversely proportional to the estimation error when computing the regression parameters **b**. Consequently, the longer the record length is at a particular gage, the smaller $v_{i,i}$, and the larger the weight that is given to a flow quantile at a gage. This weighting does not exist in OLS regression.

The Stedinger and Tasker error model reduces to OLS if the time sampling error is zero, i.e., if the population estimates of flow quantiles are known. In this case, the residual error matrix reduces to:

$$\mathbf{V} = \begin{matrix} d^2 & 0 \\ 0 & d^2 \end{matrix} = (se)^2 \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \qquad \text{(D-19)}$$

where now the regression error $d^2=(se)^2$, and *se* is the usual standard error of the regression. Equation D-15 reduces to Equation D-13.

The GLS regression model error term, d^2, is determined by iteratively solving (Johnston 1972, 210) Equation D-15 and:

$$\mathbf{e}'\mathbf{V}^{-1}\mathbf{e} = n - p \tag{D-20}$$

The residual errors, **e**, are estimated from Equation D-20 after solving Equation D-15 for the regression parameters **b**. The iterative procedure is required because **b** cannot be determined until d^2 is known. The iterative solution proceeds by finding an OLS solution for **b**, assuming d^2 is the standard error of the OLS regression and substituting into Equation D-20. A secant iteration procedure is then followed to adjust estimates of d^2 and **b** to satisfy both Equations D-15 and D-20.

D.5.2 Evaluating the Regression

Measures for evaluating the GLS regression (Method 3) are presented in Sections D.2 and D.3. Each of the measures has an equivalent measure in OLS regression. These measures estimate the statistical significance and accuracy of the regression and identify influential data points, i.e., data points distant from the majority of the independent variables or that have a disproportionate weight in determining the regression coefficients.

D.5.2.1 Measure of Accuracy

A means for comparing alternative OLS regressions is the standard error of the regression. However, a different measure of error is needed for GLS regression because of the non-homoscedastic nature of the residuals. Tasker and Stedinger (1989) recommend the computation of an average prediction error (AVP) to evaluate the accuracy of regression predictions. The AVP for a regression equation is the average of the individual MSE of prediction for all gages used in developing the regression. This is analogous to the OLS standard error squared. The regression MSE of prediction (mse_x) for a particular gage is computed as:

$$mse_x = \mathbf{x}_i'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X} \tag{D-21}$$

where $\mathbf{x}_i$ is a vector of independent variables at a given site *i*. For example, if drainage area (*A*), mean basin elevation (*Elev*), and mean annual precipitation (*MAP*) are the independent variables in the regression, then $\mathbf{x}_i$ would have the form:

$$\mathbf{x}_i = \begin{matrix} 1 \\ A \\ Elev \\ MAP \end{matrix} \tag{D-22}$$

The AVP is the sum of the estimated regression model error and the average mse_x:

$$AVP = d^2 + \sum_{i=1}^{n} \left(\mathbf{x}_i' \left(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{x}_i \right) / n \qquad \text{(D-23)}$$

This measure is substituted for the standard error typically used for OLS regression in evaluating the GLS regression. The OLS standard error is computed as the square root of the average of the sum of squared differences between regression predictions and the observed dependent variables (the gage skew values in this study). This measure of prediction error is useful for OLS application because the prediction error is assumed independent and equal for all combination of the regression independent variables (i.e., the errors are homoscedastic). As described previously in Section D.5.1, this is not true for regressions involving skew.

An additional statistical measure termed *pseudo-*R^2 (Gruber et al. 2007) is analogous to the OLS R^2. In OLS, R^2 is:

$$1 - \frac{se^2}{Sy^2} \qquad \text{(D-24)}$$

where:

se is the standard error, and,
Sy is the standard deviation of the dependent variable.

This measures the improvement of a regression over not having any regression for prediction. *Pseudo-*R^2 provides a similar measure as:

$$pseudo\, R^2 = 1 - d^2 / d_{b0}^2 \qquad \text{(D-25)}$$

where:

d_{b0}^2 is the GLS regression model error when using a constant (no independent variables).

Obtaining the model error, which is a squared error for predicting regional skew, with only a constant is analogous to obtaining the MSE for a region using an average skew value. Consequently, *pseudo-*R^2 measures the relative improvement in prediction using a regression skew over the average skew.

In summary, the AVP replaces the OLS standard error as a measure of the relative value of using an alternative regression equation, and as a measure of the prediction accuracy of an individual regression. *Pseudo-*R^2 measures the improvement obtained by using the regression versus using a region average skew estimate.

D.5.2.2 Leverage and Measures of Influence

The sensitivity of the GLS regression results to individual data points is an important consideration, particularly when data are limited, and a relatively large spread is observed in the independent variable values. To answer this concern, a statistical measure termed "leverage" defines whether or not the residual error associated with any set of gage-independent parameters (such as drainage area, elevation, MAP, and so on) has undue influence on regression parameters.

Mathematically, leverage is defined for GLS regression as the ratio of rate of change of prediction to change in prediction error as (Tasker and Stedinger 1989):

$$H^{*}_{ii} = \frac{\partial(\mathbf{x}_i\mathbf{b})}{\partial(\mathbf{e}_i)} = diag\left[\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]_i \qquad \text{(D-26)}$$

where:

($\mathbf{x}_i\mathbf{b}$)	is an individual regression prediction,
$\mathbf{e}_i$	is the associated residual error,
diag[ ]$_i$	refers to the i^{th} diagonal element of the *n*x*n* matrix inside the brackets, and,
n	the number of gages used in the regression.

In the case of OLS, Equation D-26 reduces to:

$$H^{O}_{ii} = diag\left[\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}^{-1}\right]_i \qquad \text{(D-27)}$$

The average leverage of any data point will equal *p*/*n* where *n* is the number of observations, and *p* (where *p-1* is the number of parameters) is the number of regression constants (for example, **b** in Equation D-13, including the intercept, $\mathbf{b_0}$). Individual sets of observations (x_i) with leverages greater than 2*p*/*n* should cause concern, and leverages greater than 3*p*/*n* should be singled out for closer inspection.

An application to the simple regression (a regression using a single independent parameter) reveals how leverage is the relative contribution of each variable to the regression results. For example, consider the leverage of two observations for the single independent variable case. The diagonal elements of H^O become:

$$H^{O}_{1,1} = \frac{1}{n}\frac{\sum_i (x_i - x_1)^2}{\sum_i (x_i - \overline{X})^2} \qquad \text{(D-28)}$$

$$H^{O}_{2,2} = \frac{1}{n}\frac{\sum_i (x_i - x_2)^2}{\sum_i (x_i - \overline{X})^2} \qquad \text{(D-29)}$$

where:

x_1, x_2	are individual observations of the independent variable (e.g., drainage area for a gage),
$\overline{X}$	is the average of the independent variable estimates, and,
n	is the total number of independent observations.

Leverage of an individual observation is the average of the ratio of the sum of squared differences between all observations and the individual observation to the sum of squared deviations from the mean. Leverage measures the ratio of the variation of all the independent observations from the observation of interest to the variation from the mean of the independent observations.

Cook's statistic, *D*, provides a measure of the influence of an individual observation, which is related to leverage. This statistic is computed for OLS as (Tasker and Stedinger 1989):

$$D_i = \frac{e_i^2 \mathbf{H}_{ii}}{p(1-\mathbf{H}_{ii})^2 se^2} \tag{D-30}$$

where:

$\mathbf{H}_{ii}$	is the leverage for an individual observation,
e_i	is the regression residual for that observation,
p	is the number of parameters for the regression, and,
se^2	is the standard error of the regression.

For GLS, the statistic takes the form:

$$D_i = \frac{1}{p}\left(\frac{\left[\mathbf{H}^*\mathbf{V}\right]_{ii} e_i^2}{\left[\left(\mathbf{I}-\mathbf{H}^*\right)\mathbf{V}\right]_{ii}^2}\right) \tag{D-31}$$

where:

$[\ ]_{ii}$	is the *i*th diagonal element of the contained array,
I	is the identity matrix (i.e., a matrix whose diagonal elements are one's, and all other elements are zero), and,
V and $\mathbf{H}^*$	matrices are defined in Equations D-16 and D-26, respectively.

Observations with Cook's statistic values greater than 4/*n*, where *n* is the number of observations, should be examined to see the sensitivity of regression predictions to these values.

D.5.3 Challenges of GLS Regression (Method 3)

The application of the GLS regression method presents some significant challenges given the characteristics of hydrology and the limited number of gages available. These challenges result from:

- The varying causes of inter-gage correlation between annual peak flows.
- The potential errors in the independent variables.
- The approximate nature of statistical tests of significance.

- The prevalence of gages with large leverage and influence on the regression results because of the limited coverage of gages.

As described previously in Section D.2, linear correlation is not a particularly good indication of inter-gage dependence because of the presence of mixed hydrologic events. (Peak annual floods due to thunderstorms will have less inter-gage dependence than those caused by hurricanes.) The problem with using a simple correlation coefficient in the method is that this correlation is used to formulate the GLS covariance matrix.

Much of the focus of GLS has been on sampling error in the dependent variable, in this case, skew. However, some independent variables, such as drainage area, are likely to be better estimated than others, such as mean annual precipitation. The differences in the accuracy of estimates of independent variables have potential for degrading the regression.

Statistical significance tests for GLS regression are not well developed, although recent research in this area is making great headway (Gruber et al. 2007). This current deficiency of GLS testing methodology precludes the use of well developed OLS statistical procedures for selecting the best regression.

Perhaps the greatest difficulty comes as a consequence of the sparseness of the gage data. Examination of a map of gage locations can be deceiving in this regard. For example, consider the space defined by drainage area and mean annual precipitation. Even though the gage locations may be evenly spread throughout a region, the coverage of the independent parameter space may not be adequate. Leverage and Cook's statistic alert the analyst to this problem. Experience has shown that highly influential points are likely to occur in a region of streamgages. This leads to difficult decisions with regard to the data to minimize leverage and influence. As Cook and Weisberg (1982, 104) note regarding their measure of influence:

> *The techniques developed here are not intended to provide rules for the rejection of data as influential cases are not necessarily undesirable. Often, in fact, they can provide more important information than most other cases.*

Consequently, the fact that a gage is identified as unusually influential may not be a reason to reject that gage.

D.5.4 Data Review

The purpose of this section is to identify gages deleted from the GLS regression analysis of the Delaware River Basin and describe the watershed parameters used as independent variables. Instances of the dependent variable (gage skew coefficient) were deleted if either there was an incomplete set of watershed parameters or the drainage areas were greater than ten percent impervious (USGS 2007). The ten percent censoring level for percent impervious was based on judgment. Table D-6 lists omitted gages.

Table D-6. Stations omitted from regression analysis

Gage ID's for Gages with an Incomplete Parameter Set (1)	Gage ID's for Gages with Impervious Area Greater than 10% (2)
1421900	1440300
142400103	1452500
1428750	1465500
1429300	1477000
1451800	1534000
1464538	1538000
1472162	
1472198	
1472199	
1472500	
1474000	
1474500	
1475850	
1476500	
1477120	
1478200	

The independence of the variables was examined by computing the correlation between parameters, as shown in Table D-7. The correlation in the table shows expected and perhaps unexpected correlation. The most noticeable aspects of these relationships are as follows:

- The high degree of correlation between drainage area and stream length argues against the use of both parameters in the same regression relationship.

- Forested area has sizeable positive correlation with mean elevation, lake storage, soil storage, and mean annual precipitation. The mean area elevation correlation is most interesting in that it indicates that forest dominates away from the low-lying coastal area, where elevations are greater. The positive correlation with both this elevation and mean annual precipitation should cause the associated regression coefficient to be positive. This results because mean elevation is associated with slope and precipitation which causes more peaked flows and thus greater skew. However, the correlation with greater soil storage (as would be expected for forested soils) and lake storage would suggest smaller peak flows and thus smaller skews. These multiple and opposing characteristics of forested area make it difficult to judge how it will relate to skew in the regression.

The difficulty in estimating the various parameters was important both in judging the parameter's measurement accuracy and how easily regression results employing the parameter might be used in the future. Drainage area, basin length, slope, and mean annual precipitation (MAP) were readily available for the Delaware River Basin from geographic information databases. Percent forested area was easily obtainable from land use sources. Percent lake storage is not likely to be as reliable given that the measurement of the lake area was affected by the scale of information available, such as the scale of mapping information. The Natural Resource Conservation Service (NRCS) soil storage parameter probably is least reliable given that it was not available for all gages and was approximated. Furthermore, computations of this storage from soil database information are likely highly inaccurate.

Table D-7. Correlation between independent parameters

Parameter (1)	Slope[a] (2)	Length[b] (3)	% Lake Storage[c] (4)	Mean Elevation[d] (5)	% Forested Area[e] (6)	MAP[f] (7)	ScSS[g] (8)	Area[h] (9)
Slope		-0.32585	-0.13547	0.375842	0.375923	0.054016	-0.03191	-0.20267
Length			-0.08286	0.22228	0.018972	-0.13235	-0.0863	0.884534
% Lake storage				-0.26767	0.180989	-0.01449	0.303465	-0.07885
Mean elevation					0.550619	0.07746	-0.09074	0.232329
% Forested area						0.244964	0.278042	0.030166
MAP							0.209627	-0.15323
ScSS								-0.09116
Area								

a. Stream slope (ft/mi).
b. Basin length (mi).
c. Lake storage as percent of drainage area.
d. Mean basin elevation (ft).
e. Mean annual precipitation (in).
f. Percent of drainage with forest cover.
g. Average Natural Resource Conservation Service soil storage parameter (in).
h. Drainage area (mi^2).

Regression equations with the most easily estimated parameters should be favored in comparisons, assuming measures of prediction error are similar. Regression with drainage area or length, slope, and MAP would be preferable to the other parameters because of their high correlation with skew values.

Finally, it was decided not to use either drainage area or length directly in the regression analysis. Rather, a velocity parameter was created, length/(slope)$^{0.5}$, as an additional parameter. This was done solely to maintain consistency with the original USACE study (1983). However, using a combination of variables has potential to cause analysis problems. Slope should be positively correlated with skew and length, perhaps, more negatively correlated with skew. The ratio of the two variables (noting that 1/slope should be negatively correlated with skew) will enforce a negative correlation with skew. This would be the opposite effect sought from a "velocity" parameter, which should be positively correlated with skew.

D.5.5 Software Verification

The purpose of this section is to provide verification of the software developed for computing GLS regression equations. Tests for weighted least square (WLS) regression were completed. This was done to compare results with Statistical Package for the Social Sciences (SPSS®) and results from Tasker and Stedinger (1986). This comparison is believed to be an adequate test of all the software calculations: matrix multiplication, matrix inversion, and secant search algorithms used in both WLS and GLS operations. The only difference in the computations is that the WLS covariance matrix, unlike GLS, has off-diagonal elements equal to zero.

The matrix multiplication and inversion routines were tested by making comparisons with Microsoft Excel® and SPSS®. For these computations, data taken from a subset of gages was used in the skew study, as shown in Table D-8.

Table D-8. Data for software verification

USGS Gage ID (1)	Station Skew (2)	Simulated MSE for Station Skew (3)	Mean Elevation (4)
1413500	0.044	0.086	0.344
1414000	0.301	0.145	0.328
1414500	0.082	0.086	0.352
1415000	-0.028	0.089	0.299
1415500	0.836	0.277	0.290
1418500	0.144	0.126	0.286
1419500	0.360	0.137	0.332
1420000	-0.062	0.098	0.262
1420500	0.353	0.072	0.316
1421000	0.024	0.061	0.303
1422000	0.182	0.139	0.301
1422500	-0.192	0.115	0.338
1423000	-0.083	0.110	0.286
1425500	-0.003	0.178	0.223
1426000	0.600	0.208	0.207
1426500	-0.354	0.067	0.253

Excel® was used to compute the AVP for this data subset. The first step was to compute the $\mathbf{(X'V^{-1}X)}$ matrix. The covariance matrix, **V**, an nxn matrix, n=16, has a diagonal equal to the mean square error shown in Table D-7. **X** is the nxp parameter matrix, p=2, having a column of ones (for the regression constant) and a second column equal to the mean elevation in Column 4. The inverse, $\mathbf{V^{-1}}$, was calculated by inverting each diagonal element of **V**. (Note: the GLS software does not do this, but actually does an inverse calculation). Table D-9 shows the calculation resulting in the pxp $\mathbf{(X'V^{-1}X)}$ matrix. The only difference between this example and the GLS software calculation was that any model error, d^2, was not identified. If d^2 had been identified, then it would be added to the diagonal elements of the covariance matrix **V**.

Table D-9. Example computation of $\mathbf{(X'V^{-1}X)}$

$\mathbf{X'}$ (1)		$\mathbf{V^{-1}}$ **Diagonal** (2)	**X** (3)		$\mathbf{(X'V^{-1}X)}$ (4)	
1	0.344	11.682	1	0.344	151.1966	45.2583
1	0.328	6.878	1	0.328	45.2583	13.7570
1	0.352	11.682	1	0.352		
1	0.299	11.287	1	0.299		
1	0.290	3.611	1	0.290		
1	0.286	7.911	1	0.286		
1	0.332	7.283	1	0.332		
1	0.262	10.204	1	0.262		
1	0.316	13.966	1	0.316		
1	0.303	16.313	1	0.303		
1	0.301	7.194	1	0.301		
1	0.338	8.673	1	0.338		
1	0.286	9.124	1	0.286		
1	0.223	5.624	1	0.223		
1	0.207	4.815	1	0.207		

Table D-10 shows the next computation step to obtain the invert matrix $\mathbf{(X'V^{-1}X)^{-1}}$.

Table D-10. Computation of the invert matrix $\mathbf{(X'V^{-1}X)^{-1}}$

$\mathbf{(X'V^{-1}X)}$ (1)		$\mathbf{(X'V^{-1}X)^{-1}}$ (2)	
151.1966	45.2583	0.4339	-1.4275
45.2583	13.7570	-1.4275	4.7688

For the final step, shown in, Table D-11, AVP was obtained by computing the average of $\mathbf{x_i(X'V^{-1}X)^{-1}x_i'}$ where $\mathbf{x_i}$ is a 1xp matrix of the constant and single parameter at gage i. The AVP shown, calculated using Excel®, agrees exactly with results found using the GLS software.

The computation of the regression coefficients shown in Table D-12 was verified using SPSS®. These SPSS® results agree exactly with the GLS software results.

The GLS software was also tested in a comparison with the results obtained by Tasker and Stedinger (1986) for the Illinois River Basin. The data used in that study are shown in Table D-13. Tables D-14 and D-15 compare the OLS and WLS results obtained from Tasker and Stedinger to the GLS results for the Delaware River Basin study. Results from the Tasker and

Table D-11. Example computation of AVP

x_i' (1)		$(X'V^{-1}X)^{-1}$ (2)		x_i[a] (3)		MSE[b] (4)
0.344	1	151.1966	45.2583	1	0.344	0.0161
0.328	1	45.2583	13.7570	1	0.328	0.0105
0.352	1			1	0.352	0.0198
0.299	1			1	0.299	0.0066
0.290	1			1	0.290	0.0070
0.286	1			1	0.286	0.0075
0.332	1			1	0.332	0.0117
0.262	1			1	0.262	0.0133
0.316	1			1	0.316	0.0079
0.303	1			1	0.303	0.0067
0.301	1			1	0.301	0.0066
0.338	1			1	0.338	0.0137
0.286	1			1	0.286	0.0075
0.223	1			1	0.223	0.0344
0.207	1			1	0.207	0.0473
0.344	1			1	0.344	0.0169
					AVP	0.0146

[a] Shown as 1x*p* row matrix for convenience, but is a *p*x1 column matrix.
[b] Skew prediction mean square error for a gage.

TableD-12. Example SPSS® WLS regression

Parameter (1)	Coefficient (2)
Constant	-0.226
Mean elevation	1.001

Stedinger study compares closely to the Delaware GLS results. However, some differences exist in WLS coefficients, model error, and AVP. These differences may be due to numerical precision differences and/or the tolerance for convergence in the secant procedure for estimating model error.

D.5.6 Regression Results

Tables D-16 through D-23 provide the best ordinary least squares (OLS) and GLS regressions for regional skew, and their associated errors, for combinations of regressions including a regression using the maximum number of independent variables. Pseudo-R^2 values of -999 indicate that model error was not identified (i.e., was set equal to zero).

Regional skew regression equations for the Northern and Southern Delaware River Basin regions were developed. The regression using mean basin elevation in the North was chosen partly because a model error could be estimated and it resulted in a physically reasonable, positive regression coefficient. The negative pseudo-R^2 detracts from the result because it shows that a constant alone provides a better explanation of the skew variance in comparison to using mean elevation. The South equation provides a physically reasonable, positive coefficient for mean annual precipitation, however, no model error could be defined for this region, and thus pseudo-R^2 could not be estimated. AVP values for both equations are comparable to other results.

Table D-13. Regression data from Tasker and Stedinger (1986)

Skew^a (1)	**Record^b** (2)	**Z1^c** (3)	**Z2^d** (4)	**ln(area)^e** (5)	**ln(slope)^f** (6)	**ln(length)^g** (7)	**ln(forest)^h** (8)	**ln(soil)^i** (9)
-0.12	29	0	0	5.8916	1.2781	0.10436	1.04028	1.09861
-0.34	68	0	0	6.3099	1.0116	0.08618	1.11841	1.25276
-0.25	21	0	0	-2.3026	3.5296	0.00000	0.00000	1.38629
0.58	26	0	0	6.6516	0.793	0.00000	1.38629	1.09861
1.18	26	0	0	2.3979	2.9339	0.00000	0.03922	0.69315
-1.00	27	0	0	5.6204	0.6981	0.09531	1.41342	1.09861
-0.13	33	0	0	6.3315	0.6981	0.04879	1.63705	1.09861
-0.40	27	0	0	6.7650	0.2776	0.05827	1.70293	1.09861
-1.37	64	0	0	7.8702	0.392	0.35767	1.54969	1.09861
-0.35	29	0	0	4.6728	1.6845	0.00000	1.00063	1.09861
-0.60	20	0	0	0.4055	3.6921	0.32208	0.75612	1.60944
0.12	34	0	0	5.8141	0.9517	0.00000	2.72524	1.25276
-1.16	29	0	0	5.3660	1.5369	0.00000	0.83725	1.09861
0.81	18	0	0	1.1184	3.0796	0.00000	2.8679	1.38629
0.30	29	0	0	5.4250	1.8294	0.00000	1.06471	1.09861
0.12	32	0	0	5.7236	1.6332	0.00000	1.34025	1.09861
-0.51	16	0	0	-0.1053	3.8916	0.00000	2.30259	1.09861
0.62	32	0	0	5.8081	1.7509	0.00000	1.82455	0.69315
-0.09	35	0	0	7.4978	0.7975	0.00000	1.89912	1.09861
0.01	21	0	0	-0.0619	3.168	0.00000	0.00000	1.38629
-0.49	27	0	0	3.2771	0.7701	0.00000	0.00000	1.38629
-0.57	59	0	0	8.5356	0.239	0.19885	1.8916	1.09861
-0.58	25	0	1	0.2700	3.7104	0.00000	0.00000	1.38629
-0.06	63	0	1	7.1899	0.6981	0.70310	2.33988	1.38629
-1.92	16	0	1	0.6678	3.3796	0.00000	0.00000	1.38629
0.39	16	0	1	-0.6539	4.5758	0.00000	2.70805	1.60944
-0.41	32	0	1	7.8438	0.6981	1.02245	2.18717	1.38629
-0.15	37	0	1	8.7583	-0.1744	2.21047	2.15987	1.38629
-1.18	16	0	1	0.7930	3.6954	0.00000	1.09861	1.38629
-1.88	15	0	1	4.4438	1.744	1.09861	1.60944	1.38629
-0.20	16	0	1	-0.1744	4.4698	1.38629	2.83321	1.38629
-0.35	37	0	1	6.2879	1.5239	2.59301	1.24415	1.38629
-0.98	21	0	1	0.5128	3.3576	0.00000	0.00000	1.38629
-0.90	37	0	1	5.9584	0.8198	0.00000	1.22671	1.38629
-0.37	18	0	1	0.5365	3.9843	0.00000	1.94591	1.50408

Skew[a] (1)	Record[b] (2)	Z1[c] (3)	Z2[d] (4)	ln(area)[e] (5)	ln(slope)[f] (6)	ln(length)[g] (7)	ln(forest)[h] (8)	ln(soil)[i] (9)
-0.52	37	0	1	7.0022	1.4036	2.32435	1.24127	1.38629
-1.14	37	0	1	4.7622	1.8469	0.00000	0.21511	1.38629
-0.07	18	0	1	0.0000	3.5013	0.00000	1.09861	1.38629
0.70	21	0	1	-1.8971	4.9671	0.00000	0.00000	1.50408
-0.87	37	0	1	4.6347	2.3466	0.23902	0.802	1.38629
-0.87	62	0	1	9.0774	0.0000	2.08318	2.0189	1.38629
-1.29	37	0	1	4.9836	1.454	0.16551	1.38879	1.38629
-0.55	14	0	1	0.3507	4.0975	0.00000	0.00000	1.38629
-0.25	32	0	1	5.0626	1.13635	0.17395	1.8718	1.50408
-0.47	37	0	1	9.1644	0.1044	2.01223	1.99198	1.38629
-1.49	37	0	1	5.3033	1.3481	0.66783	1.51293	1.38629
-1.14	14	0	1	1.5994	3.0402	0.00000	0.00000	1.38629
-1.82	16	0	1	1.6544	2.4105	1.83896	0.00000	1.38629
-2.31	16	0	1	0.1989	3.4818	0.00000	0.00000	1.38629
-1.24	40	0	1	6.9108	0.9282	0.93609	1.58924	1.38629
-0.44	37	0	1	4.1336	2.0069	0.05827	2.53211	1.50408
0.10	21	0	1	-1.5141	4.2056	0.00000	0.00000	1.50408
0.46	17	1	0	0.4824	2.975	0.00000	1.22671	0.00000
-0.95	18	1	0	-0.8440	4.2988	0.69315	3.04452	0.00000
-0.33	62	1	0	7.0309	0.6981	0.32208	3.1086	0.00000
0.72	21	1	0	0.4824	3.5852	0.00000	2.41054	0.00000
-0.30	21	1	0	-2.5257	4.5925	0.00000	3.41411	0.00000
-0.48	21	1	0	-0.844	4.4734	0.00000	2.91831	0.00000
1.36	16	1	0	4.5768	1.4036	0.00000	2.94444	0.00000
-0.33	48	1	0	6.1399	0.6419	0.00000	3.30064	0.00000
-0.06	37	1	0	8.0398	0.1484	0.13103	3.01994	0.00000
0.42	17	1	0	-1.8326	4.4971	1.60944	3.78419	0.00000

d. Unbiased skew (skew(1+6/record length).

e. Record length (years).

f. Dummy variable for location.

g. Dummy variable for location.

h. Natural log area.

i. Natural log slope.

j. Natural log stream length.

k. Natural log (1+%forest area).

l. Natural log (1+soil storage coefficient).

Table D-14. Comparison of GLS software and Tasker and Stedinger (1986) OLS regression estimates

Name (1)		Constant (2)	Z1 (3)	Z2 (4)	ln (area) (5)	ln (slope) (6)	ln (length) (7)	ln(forest) (8)
Case 1	Software[a]	-0.350		-0.558				0.140
	T&S[b]			-0.558				0.140
Case 2	Software	-0.542		-0.579		0.084		0.154
	T&S	-0.542		-0.579		0.085		0.154
Case 3	Software	-1.323	-0.111	-0.638	0.100	0.305	-0.049	0.157
	T&S	1.384	-0.106	-0.608	0.108	0.324	-0.047	0.153

a. GLS software.
b. Tasker and Stedinger (1986).

Table D-15. Comparison of GLS software and Tasker and Stedinger (1986) WLS regression estimates

Name (1)		Constant (2)	Z1 (3)	Z2 (4)	ln (area) (5)	ln (slope) (6)	ln (length) (7)	ln (forest) (8)	Model Error (9)	AVP (10)
Case 1	Software[a]	-0.288		-0.391				0.084	0.073	0.082
	T&S[b]	-0.384		-0.462				0.110	0.089	0.106
Case 2	Software	-0.477		-0.400		0.085		0.106	0.060	0.089
	T&S	-0.620		-0.470		0.106		0.145	0.074	0.100
Case 3	Software	-0.908	-0.190	-0.482	0.050	0.209	0.003	0.136	0.073	0.110
	T&S	-1.273	-0.195	-0.577	0.077	0.294	0.025	0.169	0.084	0.130

a. GLS software.
b. Tasker and Stedinger (1986).

Table D-16. OLS regression results for all gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log drainage area (8)	Log % lake storage (9)	Log % forested (10)	Log [length/ slope$^{0.5}$] (11)
0.171	0.357	0.016	0.838	-0.100	-0.141	-0.707	0.158	-0.018	0.195	-0.317
0.165	0.356	0.024	0.612	-0.099		-0.717	0.154	-0.018		-0.312
0.158	0.354	0.031	0.634	-0.084		-0.674		-0.010		-0.318
0.152	0.353	0.037	0.649	-0.083		-0.695	0.156			-0.319
0.147	0.353	0.037	0.690			-0.669	0.075			-0.224
0.141	0.353	0.039	0.754			-0.723				-0.136
0.138	0.356	0.021	0.330							-0.130

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-17. GLS regression results for all gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.0104	-999	-0.0093	0.0100	0.2102	-0.4088	0.0630	0.0406	-0.0541	-0.1214
0.0087	-999	-0.0153		0.2090	-0.4086	0.0671		-0.0425	-0.1257
0.0078	-999	0.0034		0.1949	-0.4151	0.0648	0.0386		-0.1230
0.0069	-999	-0.1823		0.3465	-0.4773		0.0409		-0.0496
0.0060	-999	0.0058		0.2220	-0.4643				-0.0375
0.0051	-999	0.3662			-0.4494				-0.0388
0.0044	-999	0.3183							-0.4554

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-18. OLS regression results for Delaware River Basin gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ slope$^{0.5}$] (11)
0.213	0.130	0.021	2.391	-0.338	-1.215	-0.798	0.399	0.008	1.094	-0.671
0.201	0.128	0.036	2.348	-0.344	-1.192	-0.815	0.396		1.157	-0.668
0.190	0.127	0.045	0.375	-0.357		-0.821	0.372		1.172	-0.631
0.180	0.127	0.040	-0.040	-0.377			0.449		0.833	-0.721
0.170	0.128	0.039	0.182	-0.315			0.443			-0.723
0.163	0.132	0.006	0.813			-0.853				-0.113
0.153	0.132	0.003	0.620			-0.724				

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-19. GLS regression results for Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.0104	-999	-0.0093	0.0100	0.2102	-0.4088	0.0630	0.0406	-0.0541	-0.1214
0.0087	-999	-0.0153		0.2090	-0.4086	0.0671		-0.0425	-0.1257
0.0078	-999	0.0034		0.1949	-0.4151	0.0648	0.0386		-0.1230
0.0069	-999	-0.1823		0.3465	-0.4773		0.0409		-0.0496
0.0060	-999	0.0058		0.2220	-0.4643				-0.0375
0.0051	-999	0.3662			-0.4494				-0.0388
0.0044	-999	0.3183							-0.4554

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-20. OLS regression results for northern Delaware River Basin gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ slope$^{0.5}$] (11)
0.263	0.324	0.201	3.334	-0.699	-1.775	-1.935	0.538	-0.016	3.389	-0.829
0.241	0.319	0.224	3.479	-0.669	-1.841	-1.897	0.533		3.190	-0.829
0.222	0.320	0.221	0.436	-0.657			-1.973	0.465	3.318	-0.718
0.205	0.327	0.186	2.228	-0.447	-1.064	-2.321			4.231	
0.184	0.324	0.200	0.474	-0.456		-2.335			4.218	
0.167	0.328	0.182	0.813			-2.330			2.723	
0.192	0.371	-0.049	0.157	-0.060			0.039			

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-21. GLS regression results for northern Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.0315	-999	1.4804	-0.3864	-0.544	-1.5281	0.3874	0.0146	1.7104	-0.6337
0.0264	-999	2.2115		-0.8683	-1.4108	0.3525	0.0364	0.6808	-0.6004
0.0231	-999	2.2291		-0.9036	-1.4765	0.3599		0.9337	-0.6021
0.0201	-999	0.7586			-1.565	0.3261		1.0022	-0.5445
0.0178	-999	0.0897				0.4455		0.0378	-0.7169
0.0156	-999	1.0246			-1.5741	0.0048			
0.0259	-2.043	0.0991	0.1264			0.0144			
0.0132	-999	1.0363			-1.5773				
0.0205	-0.915	0.1253	0.1388						

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-22. OLS regression results for southern Delaware River Basin gages

AVP (1)	Standard Error (2)	Adjusted R^2 (3)	Constant (4)	Log Mean Elevation (5)	Log MAP (6)	Log ScSS[a] (7)	Log Drainage Area (8)	Log % Lake Storage (9)	Log % Forested (10)	Log [length/ slope$^{0.5}$] (11)
0.331	0.388	-0.086	-17.279	-0.561	9.996	1.557	0.247	-0.013	-0.502	-0.371
0.305	0.380	-0.041	-17.628	-0.566	10.183	1.644	0.272	-0.602		-0.398
0.280	0.375	-0.011	-16.506	-0.554	9.525	1.393	0.266			-0.403
0.254	0.369	0.019	-16.703	-0.488	9.778	1.004		-0.053		
0.228	0.365	0.041	-17.364	-0.463	10.209	0.931				
0.202	0.361	0.061	-16.577	-0.542	10.011					
0.184	0.370	0.016	0.098	-0.217						
0.190	0.379	-0.035	0.069		0.087					

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Table D-23. GLS regression results for southern Delaware River Basin gages

AVP (1)	Pseudo R^2 (2)	Constant (3)	Log Mean Elevation (4)	Log MAP (5)	Log ScSS[a] (6)	Log Drainage Area (7)	Log % Lake Storage (8)	Log % Forested (9)	Log [length/ slope$^{0.5}$] (10)
0.060	-999	-11.343	-0.223	6.415	1.570	0.168	0.082	-0.672	-0.196
0.052	-999	-6.952		3.892	1.399	0.063	0.071	-0.555	-0.078
0.044	-999	-5.841		3.262	1.106	0.009	0.046		-0.023
0.037	-999	-5.162		2.910	0.959	-0.049			0.046
0.030	-999	-5.786		3.228	1.072		0.053		
0.024	-999	-0.383			0.924		0.043		
0.018	-999	0.154							-0.037
0.019	-999	-2.673		1.684					

a. Average Natural Resource Conservation Service soil storage parameter in inches.

Furthermore, the results have physically plausible interpretation. Skew coefficients for Southern basins, closest to the coast, are perhaps more affected by hurricane frequency, leading to a regression with mean annual precipitation. The significance of this regression cannot be qualified given that no model error could be estimated for this equation. The gages in the Northern basin are farther from the coast, but have higher elevation. This might mean that skew is more related to orographic effects influencing thunderstorm rainfall, leading to a regression with mean basin elevation.

Tables D-24 and D-25 provide comparisons between OLS and GLS regression results. Notice that AVP is the sum of the model error and the sampling error (i.e., an error that is a function of the record length used to compute skew).

Table D-24. Regression comparison using mean basin elevation for the Northern Delaware River Basin

Regression Type (1)	**(Standard Error)^2 (2)**	**Adjusted R^2 (3)**	**(Model Error)^2 (4)**	**Sampling Error (5)**	**AVP (6)**
OLS	0.1379	-0.0485	0.1379	0.0103	0.1482
GLS			0.0079	0.0191	0.0271

Table D-25. Regression comparison using mean annual precipitation for the Southern Delaware River Basin

Regression Type (1)	**(Standard Error)^2 (2)**	**Adjusted R^2 (3)**	**(Model Error)^2 (4)**	**Sampling Error (5)**	**AVP (6)**
OLS	0.1436	-0.0345	0.1436	0.0093	0.1528
GLS			0	0.0189	0.0189

Notice also that traditionally the standard error squared of the OLS would be used as an estimate of MSE of the region predicted by the regression, whereas in GLS, the AVP is used. Clearly, the AVP is the smaller of the two values. Application of the GLS result would lead to a much greater weighting of the regional skew in computing the adopted skew in the *Bulletin 17B* methodology.

D.5.7 Regression Analysis

In the GLS regression analysis, model error could not be identified for any regions tested except one. As Tasker and Stedinger (1989) note, when solving for the GLS regression coefficients, "The required estimate of [d^2]…" exists "…if a positive solution for [d] exists. Otherwise, [d]=0". (Note: d is a symbol substituted for the referenced symbol for the square root of the model regression error.)

Estimating d equal to zero makes the math work, but it is not the best assumption given the nature of statistical prediction in the hydrologic sciences. It is therefore unlikely that a simple predictive relationship between basin physical parameters and the skew coefficient can be error-free. If the data were divided, or additional observations were added to the problem, it is unlikely that all the regressions would have zero model error. However, the Delaware River Basin skews exhibit the phenomenon of d equal to zero. A possible reason for this is that the model used for estimating skew sampling error is not correct. If true, this leads to a poor estimate of the covariance matrix in the GLS formulation, precluding an estimate of d^2.

Another problem is that every regression examined had a significant number (5-10) of gages with large leverage or disproportionate influence. Attempts to create a data set without these data points were not successful even when deleting half the gages in a region. This result points out the problem in assessing the adequacy of gage coverage in a regional analysis. Examination of a map of gages is not sufficient. Measures of leverage and influence are needed to evaluate how the gages cover a parameter space. In GLS regression, the measures of leverage and influence also consider the impact of not observing skew equally due to record length sampling errors. Apparently, the gage coverage in the study area is not as complete as hoped, even though the number of gages is considered adequate by the criteria provided in *Bulletin 17B*.

A model error was sought for all regressions included in the study: all Delaware River Basin gages, and a north/south division of Delaware River Basin gages. All combinations of regression parameters were investigated for each region.

D.5.8 GLS Regression Leverage and Influence

Tables D-26 and D-27 provide predictions obtained by GLS regressions (Method 3) of the Northern and Southern Delaware River Basin. For these two regressions, large leverage and disproportionally large influence was identified by Cook's statistic, *D*. Subsets of the data were created to remove these data points by shifting points with leverage into different regions, but this caused leverage and influence issues with other data points. Therefore the complete dataset was used in the GLS regression (Method 3).

D.6 Split-Sample Testing

The purpose of this section is to describe split-sample testing and to evaluate the log-Pearson III frequency curve using regional skews developed from different methods applied in the Delaware River Basin. Split-sample testing and provides a means for quantifying improvements, if any, of a method of quantile estimation. Split-sample testing was completed by computing the log-Pearson III flow quantile, e.g., the p=0.01 (100-year) exceedance probability flow, for half the record and then comparing this to the observed exceedance for the reserved portion of the record. As an example of how this test works, assume that ten gages are available with 200 years of record each. The p=0.01 flow quantile is estimated from 100 years of record at each gage. The number of flows exceeding this value at each gage is counted for the reserved 100 years of record as the observed number of exceedances. The expectation is that the number of exceedance values should be ten out of a total of 1,000 years of observed record (the product of ten gages and 100 years of reserved record); or equivalently a proportion of 0.01=10/1000.

The issue with split-sample testing is how to split the records. In the original *Bulletin 17B* study (see *Bulletin 17B*, Appendix 14), the records were split by an alternating year method to remove impacts of non-stationarity on the analysis. For example, if the period of record spanned 1950-2000, the first data set would have years 1950, 1952, 1954…, and the reserved portion would

Table D-26. Prediction errors, leverage, and Cook's statistic, northern Delaware River Basin GLS regression with mean basin elevation

Gage ID (1)	**Skew Predicted (2)**	**Skew Observed (3)**	**Error (4)**	**$MSE^{0.5}$ (5)**	**Leverage (6)**	**Cook's Statistic (7)**
1413500	0.17	0.04	-0.13	0.13	0.05	0.02
1414000	0.16	0.30	0.14	0.14	0.04	0.01
1414500	0.16	0.08	-0.08	0.15	0.09	0.01
1415000	0.16	-0.03	-0.19	0.14	0.01	0.04
1415500	0.15	0.84	0.68	0.15	0.02	0.05
1418500	0.16	0.14	-0.02	0.12	-0.02	0.00
1419500	0.17	0.36	0.19	0.14	0.02	0.01
1420000	0.15	-0.06	-0.21	0.14	0.07	0.04
1420500	0.17	0.35	0.18	0.13	0.08	0.05
1421000	0.18	0.02	-0.15	0.14	0.17[a]	0.06
1422000	0.17	0.18	0.01	0.12	0.03	0.00
1422500	0.17	-0.19	-0.36	0.14	0.06	0.08
1423000	0.17	-0.08	-0.25	0.13	0.04	0.03
1425500	0.13	0.00	-0.13	0.20	0.20[a]	0.01
1426000	0.15	0.60	0.45	0.11	-0.01	0.02
1426500	0.17	-0.35	-0.52	0.13	0.13	0.44[b]
1427500	0.15	0.50	0.35	0.11	0.00	0.03
1431500	0.15	0.93	0.78	0.11	0.02	0.14[b]
1435000	0.17	-0.12	-0.30	0.15	0.13	0.13[b]
1437000	0.16	-0.35	-0.51	0.12	0.03	0.06
1438300	0.11	0.05	-0.06	0.15	0.11	0.00
1439500	0.14	0.99	0.85	0.11	0.04	0.25[b]
1440000	0.12	0.23	0.11	0.12	0.06	0.01
1440400	0.14	-0.02	-0.17	0.11	0.02	0.01
1441000	0.12	0.16	0.04	0.12	0.01	0.00
1442500	0.14	0.68	0.54	0.11	0.01	0.07
1443500	0.10	0.28	0.18	0.16	0.11	0.07
1445000	0.10	0.15	0.05	0.15	0.10	0.00
1445500	0.11	-0.10	-0.20	0.15	0.07	0.08
1446000	0.09	-0.03	-0.12	0.17	0.14	0.04
1446600	0.11	0.48	0.37	0.14	0.03	0.02
1447500	0.16	0.62	0.46	0.12	0.05	0.08
1448000	0.17	0.99	0.82	0.12	0.04	0.09

Gage ID (1)	Skew Predicted (2)	Skew Observed (3)	Error (4)	MSE$^{0.5}$ (5)	Leverage (6)	Cook's Statistic (7)
1448500	0.14	0.14	0.00	0.19	0.30[a]	0.00
1449360	0.13	-0.40	-0.53	0.11	0.02	0.05
1450500	0.12	0.33	0.21	0.12	0.04	0.03
1452000	0.10	0.17	0.07	0.16	0.15	0.01
1453000	0.15	0.23	0.08	0.14	0.23[a]	0.02
1455200	0.10	0.77	0.67	0.15	0.02	0.09
1459500	0.09	-0.02	-0.11	0.18	0.29[a]	0.05

a. Leverage test statistic $2p/n$ exceeds 0.15.

b. Cook's test statistic, D, $4/n$ exceeds 0.10.

Table D-27. Prediction errors, leverage, and Cook's statistic, southern Delaware River Basin GLS regression with MAP

Gage ID (1)	Skew Predicted (2)	Skew Observed (3)	Error (4)	$MSE^{0.5}$ (5)	Leverage (6)	Cook's Statistic (7)
1464000	0.2121	0.019	0.1931	0.1903	0.0362	0.0047
1464500	0.2121	-0.138	0.3501	0.1903	0.0362	0.0155
1464515	0.2099	-0.437	0.6469	0.3925	0.1541[a]	0.2300
1465000	0.2120	0.797	-0.585	0.1944	0.0378	0.04512
1465850	0.2103	0.202	0.0083	0.3427	0.1174	0.0000
1466000	0.2129	0.599	-0.3861	0.1868	0.0349	0.0181
1466500	0.2112	0.026	0.1852	0.2525	0.0637	0.0076
1467000	0.2108	0.032	0.1788	0.2955	0.0873	0.0098
1467081	0.2103	0.889	-0.6787	0.3427	0.1174	0.1910
1467150	0.2112	0.621	-0.4098	0.2525	0.0637	0.0374[b]
1467305	0.2112	0.345	-0.1338	0.2525	0.0637	0.0040
1467500	0.2156	0.142	0.0736	0.4438	0.1970[a]	0.0039
1468500	0.2152	0.482	-0.2668	0.3922	0.1538[a]	0.0391
1470500	0.2152	-0.036	0.2512	0.3922	0.1538[a]	0.0346
1470779	0.2127	-0.239	0.4517	0.1804	0.0325	0.0231
1471000	0.2129	-0.065	0.2779	0.1868	0.0349	0.0094
1471980	0.2134	-0.248	0.4614	0.2159	0.0466	0.0347
1472000	0.2128	0.148	0.0648	0.184	0.0338	0.0005
1472157	0.2125	0.209	0.0035	0.1797	0.0323	0.0000
1472500	0.2129	0.772	-0.5591	0.1868	0.0349	0.0380
1473000	0.2127	-0.564	0.7767	0.1818	0.0331	0.0695
1475300	0.2135	0.266	-0.0525	0.2225	0.0495	0.0005
1478000	0.2129	0.277	-0.0641	0.1868	0.0349	0.0005
1478500	0.2129	0.601	-0.3881	0.1868	0.0349	0.0183
1479000	0.2129	0.322	-0.1091	0.1868	0.0349	0.0014
1480000	0.2138	0.729	-0.5152	0.2443	0.0597	0.0554
1480300	0.2112	0.086	0.1252	0.2525	0.0637	0.0035
1480610	0.2134	-0.021	0.2344	0.2098	0.0440	0.0084
1480675	0.2127	-0.093	0.3057	0.1804	0.0325	0.0106
1481500	0.2134	0.178	0.0354	0.2098	0.0440	0.0002
1482500	0.2121	0.689	-0.4769	0.1903	0.0362	0.0287

a. Leverage test statistic $2p/n$ exceeds 0.125.

b. Cook's test statistic, D, $4/n$ exceeds 0.125.

have, 1951, 1953, 1955… In this study, the split-samples were tested using the following methods:

- **Forecast method:** first half of the record is used to estimate frequency curve, remaining record is reserved.

- **Back cast method:** second half of record is used to estimate frequency curve, remaining record is reserved.

- **Alternating Method 1:** alternate years in the record are used to estimate the frequency curve, remaining record is reserved.

- **Alternating Method 2:** the reserved record in Alternating Method 1 is used to estimate the frequency curve and the remaining data is now the reserved.

The value of this approach is that four different sets of gage records are effectively created with different sample estimates of gage frequency curves. This helps eliminate the bias in selecting gages or period of record. The forecast and back cast methods are added to the alternating methods because they are the more realistic tests of frequency analysis assumptions, i.e., the goal of frequency analysis is to predict future exceedance values. Since trends may occur in the future, removal of using the alternating methods would be inconsistent with the goal of frequency analysis. Although a regional skew estimation methodology may ideally improve quantile estimates, the reality may be quite different looking into the future where variation in trends, cycles, or other climatic aspects in variability may occur.

D.6.1 Application to the Delaware River Basin

The log-Pearson III frequency curve estimation methods explored were:

- Expected probability - no regional skew.

- Computed probability - no regional skew.

- Computed probability using regional skew and mean square error estimates from Method 1b (region average skew, Table D-4).

- Computed probability using regional skew and average prediction error from Method 1d (GLS-constant, Table D-5).

Expected probability gives the mean estimate of exceedance probability for normal or log-normal frequency curves estimated from the sample mean and standard deviation. The estimates of probability are approximate for the log-Pearson III frequency curve (*Bulletin 17B*).

The expected probability estimate was used because it is theoretically unbiased with regard to predicting future exceedance values. Expected probability's use has no impact on the comparison to the other regional skew methods in the testing. This is because each estimate would change proportionally the same if expected probability instead of computed probability was used. However, expected probability is not a universally accepted estimator, and therefore computed probability was also used.

The computed probability approximates the median estimate of exceedance probability. This probability is obtained directly from the log-Pearson III sample mean, standard deviation, and skew (see *Bulletin 17B*).

The regional skew methods were chosen to represent both the area average (Method 1b) and GLS approaches (Method 1c) in the split-sample testing. Note that the computed and regional skew methods differ in that an adopted skew (a weighted average skew using the station and regional skew values) is used in computing the regional skew gage frequency curves.

D.6.2 Spilt-Sample Testing Results

Split-sample tests were completed for the p=0.1 (10-year), p=0.02 (50-year), and p=0.01 (100-year) exceedance quantiles. We completed testing for 4,600 years of record, thus the number of records, *n*, in a split-sample was 2,300. Tables D-28 through D-30 show the ratio of the observed flows exceeding the flow quantile to the total number of observations for the given log-Pearson III frequency curve estimation method.

The results show that the expected probability estimate (Column 2 and 3 of Tables D-28 through D-30) provides the greatest correspondence between observed and predicted future exceedance. However, the sampling error in the observed estimates of future exceedance (the proportion of values observed to be greater than the predicted exceedance level, for example p=0.1) is larger than the difference in predictions between the methods. The standard error in the proportion of exceedance value can be computed as the standard error of the proportion obtained from the binomial distribution $(p(1-p)/n)^{0.5}$, where p is the exceedance probability and n is the number of years of record in the split-sample of gage records. For example, the predictions for the p=0.01 exceedance in Table D-30 are not significantly different between methods when considering two standard errors in the proportion of 2(0.002)=0.004.

The expected probability comes closest to producing the expected proportion (ratios closest to one) in the majority of tests, i.e., the flow quantiles adequately predict exceedance. However, the observed proportion exceeds the expected proportion in most cases, as is true for the computed probability scenarios (in which ratios are greater than one), i.e., the flow quantiles generally underestimate the exceedance.

The computed probability scenarios generally agree within the standard errors shown for the expected proportion (the expected proportion being equal to the stated test exceedance probability). This indicates that regional skew provides no advantage in predicting future exceedance values, or equivalently, future flood risk.

Table D-28. Split-sample testing, p=0.1 (10-year)

Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.131	1.310	0.142	1.420	0.141	1.410	0.142	1.420
Back cast	0.087	0.870	0.092	0.920	0.092	0.920	0.093	0.930
Alternate 1	0.111	1.110	0.125	1.250	0.124	1.240	0.126	1.260
Alternate 2	0.076	0.760	0.084	0.840	0.083	0.830	0.085	0.840

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.1 (10-year) gage flows values.
d. Proportion x 10.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5}$ = 0.006, where p=0.1, n=2300.

Table D-29. Split-sample testing, p=0.02 (50-year)

Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.035	1.740	0.042	2.080	0.042	2.080	0.048	2.390
Back cast	0.022	1.110	0.027	1.370	0.023	1.150	0.025	1.240
Alternate 1	0.028	1.410	0.040	2.000	0.037	1.870	0.038	1.890
Alternate 2	0.015	0.770	0.022	1.090	0.018	0.900	0.020	1.000

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.02 (50-year) gage flows values.
d. Proportion x 50.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5}$ = 0.003, where p=0.02, n=2300.

Table D-30. Split-sample testing, p=0.01 (100-year)

Method (1)	Log-Pearson III Frequency Curve Estimation Method							
	Expected Probability		Computed Probability		Regional, Method 1b[a]		Regional, Method 1d[b]	
	Proportion[c] (2)	Ratio[d] (3)	Proportion (4)	Ratio (5)	Proportion (6)	Ratio (7)	Proportion (8)	Ratio (9)
Forecast	0.024	2.430	0.028	2.820	0.028	2.780	0.032	3.170
Back cast	0.017	1.650	0.020	2.040	0.017	1.650	0.018	1.780
Alternate 1	0.019	1.870	0.025	2.470	0.024	2.390	0.027	2.730
Alternate 2	0.012	1.150	0.015	1.450	0.012	1.190	0.013	1.280

a. Adopted skew using regional skew and mean square error (Method 1b).
b. Adopted skew using regional skew and average prediction error (Method 1d).
c. The proportion of 2300 flows observed to exceed the p=0.01 (100-year) gage flows values.
d. Proportion x 100.
Note: The standard error of the proportion = $(p(1-p)/n)^{0.5}$ = 0.002, where p=0.01, n=2300.

D.7 Further Considerations

The original *Bulletin 17B* recommendations to use a regional skew value were because its use improved consistency in predictions, not because its use improved accuracy in predictions. The split-sample testing completed in *Bulletin 17B* (Appendix 14) demonstrates that a log-normal distribution (zero skew) performs as well as the log-Pearson III distribution with a regional skew substituted for the station skew in predicting the future exceedance frequency of annual peak flows.

Improved consistency was found in the original *Bulletin 17B* studies when comparing frequency curve predictions for different halves of gage period of record. Not surprisingly, substituting regional skew for stations skew resulted in superior consistency in applications with the log-Pearson III distribution.

Bulletin 17B recommends the use of adopted skew (average of regional and station skew, weighted by MSE) rather than the regional skew. However, the *Bulletin 17B* split-sample testing used the regional skew rather than the adopted skew in comparison of methods. The application of adopted skew reduces the consistency in predictions that would be expected from the original testing formulated in the guidelines. Furthermore, the split-sample testing completed in this study demonstrated that using adopted skew does not result in improved predictions of future exceedance frequency in the Delaware River Basin.

The choice of the minimum MSE gives the regional skew the greatest weight in the adopted skew computation. This promotes greater consistency in flood frequency estimates when using the *Bulletin 17B* guidelines. However, the split-sample testing does not provide evidence that this results in improved prediction accuracy.

Consequently, if the focus is on obtaining consistent flood frequency estimates, then selecting the regional skew methodology with the minimum MSE makes sense. However, if the regional skew is not viewed as valuable (perhaps because it does not seem to promote greater prediction accuracy), then regional skew should not be used.