USGS
science for a changing world

Methods for Estimating Regional Skewness of Annual Peak Flows in Parts of Eastern New York and Pennsylvania, Based on Data Through Water Year 2013

Scientific Investigations Report 2021–5015

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

Methods for Estimating Regional Skewness of Annual Peak Flows in Parts of Eastern New York and Pennsylvania, Based on Data Through Water Year 2013

By Andrea G. Veilleux and Daniel M. Wagner

Scientific Investigations Report 2021–5015

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

U.S. Geological Survey, Reston, Virginia: 2021

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

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/.

Suggested citation:
Veilleux, A.G., and Wagner, D.M., 2021, Methods for estimating regional skewness of annual peak flows in parts of eastern New York and Pennsylvania, based on data through water year 2013: U.S. Geological Survey Scientific Investigations Report 2021–5015, 38 p., https://doi.org/10.3133/sir20215015.

Associated data for this publication:
Wagner, D.M., and Veilleux, A.G., 2021, Regional flood skew for parts of the mid-Atlantic region (hydrologic unit 02) in eastern New York and Pennsylvania: U.S. Geological Survey data release, https://doi.org/10.5066/P9PGAL0D.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
kilometer (km)	0.6214	mile (mi)
	Volume	
cubic meter (m^3)	35.31	cubic foot (ft^3)
	Flow rate	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

AEP	annual exceedance probability
ASEV	average sampling error variance
AVP_{new}	average variance of prediction at a new streamgage
B17B	Bulletin 17B (see Interagency Advisory Committee on Water Data, 1982)
B17C	Bulletin 17C (see England and others, 2018)
B–GLS	Bayesian generalized least squares
B–WLS	Bayesian weighted least squares
DAR	drainage area ratio
EMA	expected moments algorithm
EVR	error variance ratio
GAGES-II	Geospatial Attributes of Gages for Evaluating Streamflow, Version II
GIS	geographic information system
LP–III	log-Pearson Type III distribution
*MBV**	misrepresentation of the beta variance
MGBT	multiple Grubbs-Beck test
MSE	mean square error
$MSE\left(\hat{G}\right)$	mean squared error of skew
NWIS	National Water Information System
OLS	ordinary least squares
PILF	potentially influential low flood
P_{RL}	pseudo record length
Pseudo ANOVA	pseudo analysis of variance
SD	standardized distance
USGS	U.S. Geological Survey

Methods for Estimating Regional Skewness of Annual Peak Flows in Parts of Eastern New York and Pennsylvania, Based on Data Through Water Year 2013

By Andrea G. Veilleux and Daniel M. Wagner

Abstract

Bulletin 17C (B17C) recommends fitting the log-Pearson Type III (LP–III) distribution to a series of annual peak flows at a streamgage by using the method of moments. The third moment, the skewness coefficient (or skew), is important because the magnitudes of annual exceedance probability (AEP) flows estimated by using the LP–III distribution are affected by the skew; interest is focused on the right-hand tail of the distribution, which represents the larger annual peak flows that correspond to small AEPs. For streamgages having modest record lengths, the skew is sensitive to extreme events like large floods, which cause a sample to be highly asymmetrical or "skewed." For this reason, B17C recommends using a weighted-average skew computed from the skew of the annual peak flows for a given streamgage and a regional skew. This report presents an estimate of regional skew for a study area encompassing parts of eastern New York and Pennsylvania. A total of 232 candidate U.S. Geological Survey streamgages that were unaffected by extensive regulation, diversion, urbanization, or channelization were considered for use in the skew analysis; after screening for redundancy and pseudo record length (P_{RL}) of at least 36 years, 183 streamgages were selected for use in the study.

Flood frequencies for candidate streamgages were analyzed by employing the expected moments algorithm, which extends the method of moments so that it can accommodate interval, censored, and historical/paleo flow data, as well as the multiple Grubbs-Beck test to identify potentially influential low floods in the data series. Bayesian weighted least squares/Bayesian generalized least squares regression was used to develop a regional skew model for the study area that would incorporate possible variables (basin characteristics) to explain the variation in skew in the study area. Ten basin characteristics were considered as possible explanatory variables; however, none produced a pseudo coefficient of determination (pseudo R_{δ}^{2}) greater than 1 percent; as a result, these characteristics did not help to explain the variation in skew in the study area. Therefore, a constant model that had a regional skew coefficient of 0.32 and an average variance of prediction at a new streamgage (AVP_{new}, which corresponds to the mean square error [MSE] of 0.11) was selected. The AVP_{new} corresponds to an effective record length of 68 years, a marked improvement over the Bulletin 17B national skew map, whose reported MSE of 0.302 indicated a corresponding effective record length of only 17 years.

Introduction

Flood-frequency analysis of annual peak flows at streamflow-gaging stations (hereafter referred to as "streamgages") provides engineers, hydrologists, and many others estimates of the magnitudes and frequencies of floods for planning, design, and management of infrastructure along rivers and streams. The Subcommittee on Hydrology of the Federal Advisory Committee on Water Information recently published Bulletin 17C (herein referred to as "B17C"; England and others, 2018), which comprises updated guidelines for flood-frequency analysis. The bulletin recommends the use of the log-Pearson Type III (LP–III) distribution to fit a time series of annual peak flows measured by a streamgage to obtain estimates of flows corresponding to various annual exceedance probabilities (AEPs). In the case of flood-frequency analysis, the LP–III distribution is described by three moments: the mean, the standard deviation, and the skewness coefficient of the logarithms of the flows. The third moment, the skewness coefficient (hereafter referred to as the "skew"), is a measure of the asymmetry of the distribution as shown by the thicknesses of the tails of the distribution. In flood-frequency analysis, the skew is important because the magnitudes of AEP flows estimated by using the LP–III distribution are affected by the skews of the annual peak flows at specific streamgages (hereafter referred to as "station skew"); interest is focused on the right-hand tail of the distribution, which represents annual peak flows corresponding to small AEPs of the larger flood flows.

For streamgages having modest record lengths, approximately in the range of 25 to 100 years, the skew is sensitive to unusually large or small annual peak flows because they cause a sample of such flows to be asymmetrical or skewed (Griffis and Stedinger, 2007). Thus, B17C guidelines recommend using a weighted-average skew that is computed from the skew of the station's annual peak flows and the regional skew. Using the weighted-average skew reduces the sensitivity of the station skew to extreme events, particularly for streamgages with short record lengths of less than approximately 25 years.

The B17C guidelines recommend using the Bayesian weighted least squares/Bayesian generalized least squares (B–WLS/B–GLS) method to estimate regional skew (England and others, 2018, p. 30). Using this procedure, the regional skew is estimated based on the station skew of the logarithms of annual peak-flow data. The B–WLS/B–GLS procedure first uses an ordinary least squares (OLS) regression analysis to generate an initial regional-skew model that is used to compute the variance of the station skew for each streamgage. Next, B–WLS is used to generate estimators of the regional skew model parameters. Finally, B–GLS is used to estimate the precision of the B–WLS parameter values, to estimate the model error variance and its precision, and to compute some diagnostic statistics. The expected moments algorithm (EMA; Cohn and others, 1997) is a generalization of the method-of-moments approach for flood-frequency analysis of annual peak flows from streamgages recommended by B17C. The B–WLS/B–GLS method can account for the complexities introduced by EMA and the cross correlation between annual peak flows at pairs of streamgages (Veilleux, 2011; Veilleux and others, 2011).

To date, the B–WLS/B–GLS method has been used to generate estimates of regional skew for several regions around the Nation (Parrett and others, 2011; Eash and others, 2013; Olson, 2014; Paretti and others, 2014; Southard and Veilleux, 2014; Curran and others, 2016; Mastin and others, 2016; Wagner and others, 2016; Veilleux and Wagner, 2019). In this study, the B–WLS/B–GLS procedure was used to estimate skew for a region encompassing parts of eastern New York and Pennsylvania (fig. 1) to improve estimates of regional skew and flows corresponding to various AEPs across the region.

Purpose and Scope

The purpose of this report is to present the results of a B–WLS/B–GLS analysis of regional skew for parts of eastern New York and Pennsylvania (fig. 1*A*). The scope of the project includes 183 streamgages in the Mid-Atlantic region (hydrologic units 0202, 0204, 0205, 0206, and 0207) and in the Connecticut Coastal region (hydrologic unit 0110) located in the States of Connecticut, Maryland, Massachusetts, New Jersey, New York, Pennsylvania, Vermont, Virginia, and West Virginia (fig. 1*B*). The number of available streamgages that were unaffected by regulation or urbanization in hydrologic unit 0203 in southern New York and northern New Jersey (the Lower Hudson-Long Island region) was insufficient; therefore, the regional skew does not apply there. Flood-frequency analyses for the 183 streamgages were based on annual peak-flow data through water year 2013 (a water year is described as the period of October 1–September 30 and is named for the year in which it ends) and were performed using the U.S. Geological Survey (USGS) peak-flow analysis software (PeakFQ version 7.3; Veilleux and others, 2014; https://water.usgs.gov/software/PeakFQ/). The results were used to analyze the regional skew.

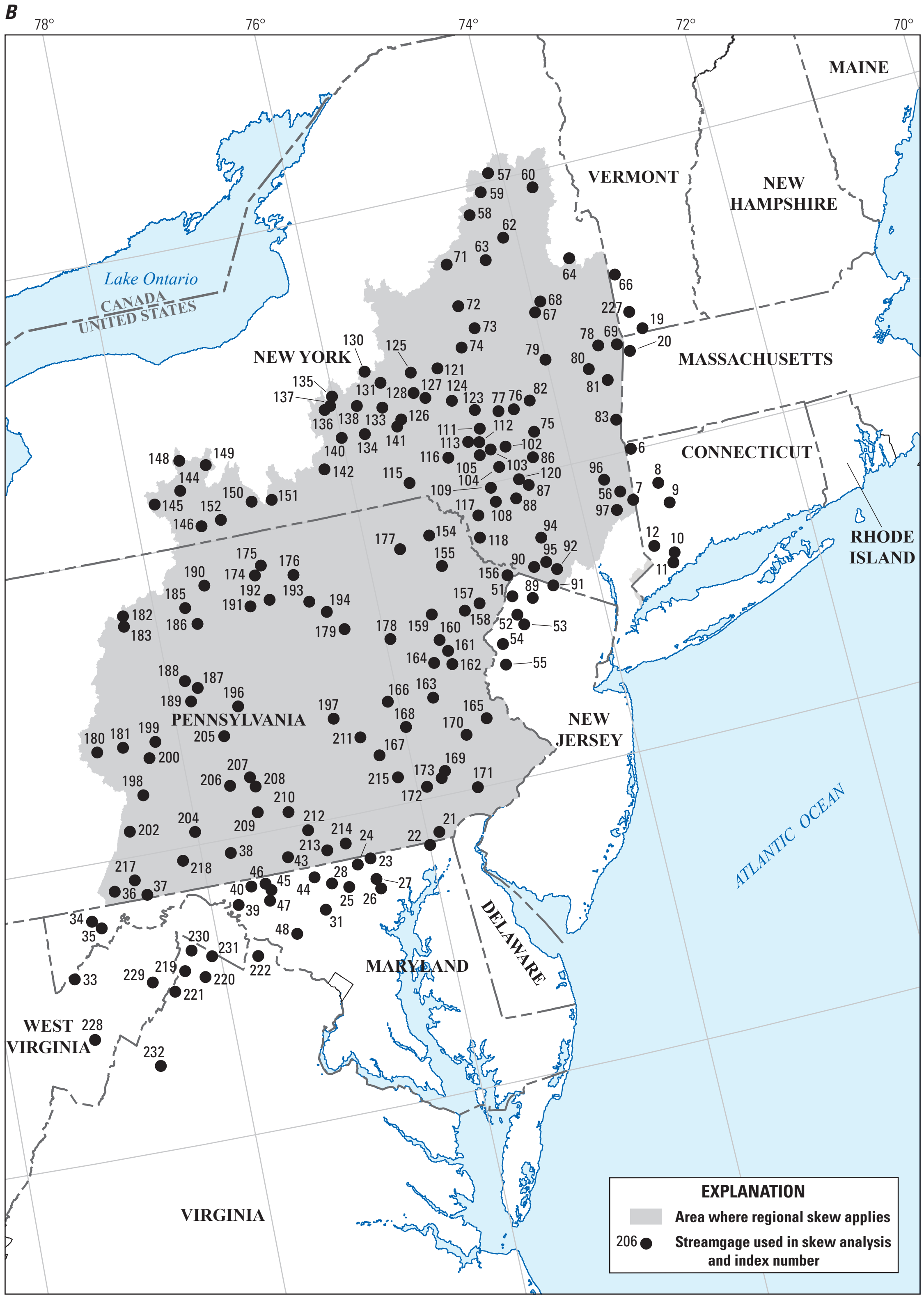
A summary of information for each streamgage used in the regional skew analysis and a description of the basin characteristics considered as potential explanatory variables in the study are provided in the tables in this report. PeakFQ input files (.txt), PeakFQ setup files (.psf), PeakFQ output files (.PRT), a geographic information system (GIS) shapefile (containing streamgage information, basin characteristics, results of flood-frequency analysis, and B–WLS/B–GLS results), a comma-separated values file containing the attributes of the GIS shapefile, and the corresponding metadata are provided in a data release associated with this report (Wagner and Veilleux, 2021).

Description of Study Area

The study area encompasses part of the Mid-Atlantic region (hydrologic unit 02) and the Connecticut Coastal region (hydrologic unit 0110) and includes parts of the States of New York and Pennsylvania (fig. 1*A*). The study area encompasses 49,817 square miles and spans approximately 300 miles from north to south (from northern New York to the southern border of Pennsylvania) and approximately 200 miles from east to west (from west-central Pennsylvania to the eastern border of New York).

The study area is located entirely within the Appalachian Highlands physiographic division, which is characterized by hilly to mountainous terrain of the Adirondack, Valley and Ridge, Appalachian Plateaus, Piedmont, and New England physiographic provinces (Fenneman, 1938). Basin-averaged mean annual precipitation for streamgages in the study area ranged from 35 to 63 inches (88 to 160 centimeters; Falcone, 2011). Based on the 2011 National Land Cover Database, the study area is approximately 60 percent forested, 22 percent agricultural (crops and pasture), 10 percent developed, and 3 percent wetlands, with the remaining 5 percent including open water, barren land, shrub/scrub, and grassland/herbaceous categories (Homer and others, 2015).

Figure 1. Maps of the study area in eastern New York and Pennsylvania showing (*A*) 4-digit hydrologic units (U.S. Geological Survey, 2019), and (*B*) locations of the 183 streamgages used in the skew analysis.



Base from U.S. Geological Survey digital data, various dates and scales
Albers Equal-Area Conic projection

0 50 100 MILES
0 50 100 KILOMETERS

Figure 1.—Continued

Methods

Streamgage Selection

A suite of 232 candidate USGS streamgages were considered for use in the regional skew analysis (table 1, which follows the "References Cited"). Annual peak flows for these streamgages were obtained from the USGS National Water Information System (NWIS; U.S. Geological Survey, 2018). Only streamgage records that were unaffected by extensive regulation, diversion, urbanization, or channelization (based on coding of annual peaks in the peak-flow files) and that had approximately 25 or more gaged peaks were considered for use in the regional skew analysis. Using these criteria, USGS employees who had local knowledge and experience in New York and Pennsylvania selected the candidate streamgages. A review of the study area boundary found that 17 streamgages were located outside the study area and were removed, after which 215 were left for use in the regional skew study (table 1). After that, streamgages that were deemed redundant were then screened and removed from the larger dataset (see the "Redundancy Screening" section for more information).

Redundancy Screening

Two streamgages may be redundant if their drainage basins are nested and similar in size; the drainage basins are considered nested if one entire drainage area is inside the other. If streamgages are redundant, a statistical analysis incorporating data from both streamgages incorrectly represents the information content in the regional dataset (Gruber and Stedinger, 2008). Instead of providing two spatially independent observations that depict how the characteristics of each basin are related to skew, the basins will be assumed to exhibit similar hydrologic responses to a given storm and thus represent only one spatial observation. To determine whether two streamgages are redundant and thus represent the same watershed for the purposes of developing a regional hydrologic model, two types of information are considered: (1) the standardized distance (*SD*) between the centroids of the basins and (2) the ratio of the drainage areas of the basins.

The *SD* between two basin centroids is used to determine the likelihood that the basins are redundant. The *SD* is defined as

$$SD_{ij} = \frac{D_{ij}}{\sqrt{0.5\left(DRNAREA_i + DRNAREA_j\right)}}, \qquad (1)$$

where

D_{ij} is the distance between centroids of basin *i* and basin *j*, in miles;

$DRNAREA_i$ is the drainage area at streamgage *i*, in square miles; and

$DRNAREA_j$ is the drainage area at streamgage *j*, in square miles.

The drainage area ratio (*DAR*) is used to determine if two nested basins are sufficiently similar in size that they represent the same watershed for the purposes of developing a regional hydrologic model (Veilleux, 2009). The *DAR* is defined as

$$DAR = Max\left[\frac{DRNAREA_i}{DRNAREA_j}, \frac{DRNAREA_j}{DRNAREA_i}\right], \qquad (2)$$

where

DAR is the *Max* (maximum) of the two values in brackets,

$DRNAREA_i$ is the drainage area at streamgage *i*, and

$DRNAREA_j$ is the drainage area at streamgage *j*.

Streamgage pairs having a *SD* less than or equal to 0.50 and a *DAR* less than or equal to 5.0 are likely to be redundant for purposes of determining regional skew (Veilleux, 2009). If the *DAR* is large enough, even nested streamgages will reflect different hydrologic responses because storms of different sizes and durations typically affect sites differently.

All possible combinations of streamgage pairs from the 215 streamgages were considered in the redundancy analysis. All streamgage pairs with a $SD \leq 0.5$ and a $DAR \leq 5.0$ were identified as possibly redundant. The drainage area of each streamgage was then investigated to determine if one of the two drainage areas was nested inside the other; if this was true, the preference was generally for the streamgage that had the smaller drainage area and the longer record length. The procedure identified 24 possibly redundant streamgage pairs; of these, 17 streamgages were found to be redundant and were removed from the analysis.

Basin Characteristics

Basin characteristics for the streamgages used in the skew analysis were either obtained from the USGS Geospatial Attributes of Gages for Evaluating Streamflow, Version II (GAGES-II) database or were generated. The GAGES-II database consists of a subset of USGS streamgages that have at least 20 years of streamflow record since 1950 or that were active as of water year 2009 and whose watersheds lie within the United States (Falcone, 2011). For streamgages that were used in the skew analysis but were not in the GAGES-II database, the suite of basin characteristics was generated by using the ArcHydro package in Esri ArcGIS software version 10.3.1 (Environmental Systems Research Institute, 2009; Eash and others, 2013; Wagner and others, 2016). This procedure ensured that a consistent suite of basin characteristics was available for all of the streamgages used in the skew analysis.

Basin characteristics were selected to potentially explain the variation in skew in the study area. These included morphometric (drainage area, latitude and longitude of basin centroid, mean basin slope, mean basin elevation, and basin compactness ratio), climatological (basin-average mean annual precipitation), and pedologic or geologic (areal percentages of open water and forest, and average soil permeability) characteristics (table 2).

Annual Exceedance Probability Analyses

To estimate regional skew for parts of eastern New York and Pennsylvania, a flood-frequency analysis must first be conducted for each streamgage to determine the station skew and its associated mean square error (MSE). The B17C guidelines recommend fitting the LP–III distribution to a series of annual peak flows at a streamgage by using the method of moments (England and others, 2018). In doing so, it is recommended that EMA is employed to extend the method of moments to accommodate interval, censored, and historical or paleo flood data, as well as the use of the multiple Grubbs-Beck test (MGBT) to identify potentially influential low floods (PILFs) in the data series. In this study, the USGS software PeakFQ version 7.3 was used to analyze the flood frequencies (Veilleux and others, 2014).

Hydrologists in the USGS Water Science Centers in New York and Pennsylvania used EMA with PeakFQ version 7.3 for candidate streamgages in their States and for candidate streamgages in the surrounding states of Connecticut, Maryland, Massachusetts, New Jersey, Vermont, Virginia, and West Virginia. Flood frequencies were analyzed by using the station-skew option in PeakFQ software and, with few exceptions (such as a fixed threshold for PILFs that yielded a superior fit of the flood-frequency model to the dataset), the MGBT for PILFs. Historical peaks were included in the analysis; annual peak flows coded as urban or regulated were not. Hydrologists in New York and Pennsylvania assigned perception thresholds to each period of record (including historical and missing periods as well as periods of crest-stage gage operation) and assigned flow intervals to uncertain annual peak flows as appropriate.

Bayesian Weighted Least Squares/Bayesian Generalized Least Squares Analysis

Prior to analyzing regional skew by the B–WLS/B–GLS method, three preliminary steps were completed: (1) calculation of the pseudo record length for each streamgage, given the number of censored observations and concurrent record lengths; (2) removal of structural bias in the estimate of station skew and its MSE; and (3) development of a cross-correlation model of concurrent annual peak flows between streamgages.

Calculating Pseudo Record Length

The pseudo record length of the annual peak-flow series at each streamgage is used in the regional skew study in several steps, including unbiasing the station skew and its mean square error, determining the concurrent record length between two streamgages, and computing the cross correlation of the station skews. Because the dataset includes censored data and historical information, the effective record length used to compute the precision of the skewness estimators is no longer simply the number of annual peak flows at a streamgage. Instead, a more complex calculation based on the availability of historical information and censored values is used. Whereas historical information and records of censored peaks provide valuable information, they often provide less information than records of an equal number of years of gaged peaks (Stedinger and Cohn, 1986). The calculations described in the following paragraphs yield a pseudo record length (P_{RL}) associated with skew, which appropriately accounts for all types of peak-flow data available from a streamgage. If no interval, censored, and (or) historical data are present in the annual peak-flow record of a streamgage, P_{RL} is equal to the gaged record length.

The P_{RL} is defined as the number of years of gaged record that would be required to yield the same mean squared error of the skew ($MSE\left(\hat{G}\right)$) as would the combination of the historical and gaged records that are actually available at a streamgage. Thus, the P_{RL} of the skew is a ratio of the MSE of the station skew when only the gaged record is analyzed ($MSE\left(\hat{G}_S\right)$) to the MSE of the station skew when the entire record, including historical and censored data, is analyzed ($MSE\left(\hat{G}_C\right)$):

$$P_{RL} = \frac{P_s \times MSE\left(\hat{G}_S\right)}{MSE\left(\hat{G}_C\right)}, \tag{3}$$

where

- P_{RL} is the pseudo length of the entire record at the streamgage, in years;
- P_s is the number of years with gaged peaks in the record;
- $MSE\left(\hat{G}_S\right)$ is the estimated MSE of the skew when only the gaged record is analyzed; and
- $MSE\left(\hat{G}_C\right)$ is the estimated MSE of the skew when the entire record, including historical and censored data, is analyzed.

Because the P_{RL} is an estimate, the following conditions must also be met to ensure a valid approximation. The P_{RL} must be nonnegative. If P_{RL} is greater than P_H (the length of the historical period), then P_{RL} should be set to equal P_H. Also, if P_{RL} is less than P_S, then P_{RL} is set to P_S. This ensures that the P_{RL} will not be larger than the complete P_H or less than P_S.

As stated in B17C, the station skew is sensitive to extreme events; therefore, accurate estimates of skew require longer periods of record, typically 50 years or greater. However, 50 years of record are not available for most streamgages, and therefore a minimum of 30 to 40 years has been used in recent studies (Eash and others, 2013; Paretti and

Table 2. Basin characteristics considered for use as explanatory variables in the regional skew analysis.

[GIS, geographic information system; NHDPlus, National Hydrography Dataset Plus; NAD83, North American Datum of 1983; PRISM, Parameter Regression on Independent Slopes Model; NLCD, National Land Cover Database]

Basin characteristic	Units	Source
Drainage area of streamgage basin, delineated by using GIS	Square kilometers	Derived from 30-meter NHDPlus data, http://www.horizon-systems.com/nhdplus/.
Latitude of basin centroid	Decimal degrees, NAD83	Determined from zonal statistics of grids derived from basin polygons in Esri ArcGIS, version 10.3.1.
Longitude of basin centroid	Decimal degrees, NAD83	Determined from zonal statistics of grids derived from basin polygons in Esri ArcGIS, version 10.3.1.
Mean basin elevation	Meters	Determined from 10-meter resolution National Elevation Dataset, https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map.
Basin compactness ratio (area/perimeter^2×100); higher number indicates more compact shape	Unitless	Calculated in Esri ArcGIS, version 10.3.1, by using drainage area and perimeter of GIS-delineated basin polygons.
Basin-averaged mean annual precipitation for the 30-year period of 1971 to 2000	Centimeters	800-meter PRISM data, Oregon State University, http://www.prism.oregonstate.edu/.
Mean basin slope	Percent	Determined from 100-meter resolution National Elevation Dataset, https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map.
Basin-averaged soil permeability	Inches per hour	Wolock (1997, https://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml) and U.S. Department of Agriculture (2008, https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx).
Percentage of streamgage basin in forested land-use categories	Percentage of streamgage basin surface area	2006 NLCD, sum of classes 41, 42, and 43, https://www.mrlc.gov/data?f%5B0%5D=year%3A2006.
Percentage of streamgage basin in open water	Percentage of streamgage basin surface area	2006 NLCD, class 11, https://www.mrlc.gov/data?f%5B0%5D=year%3A2006.

others, 2014; Southard and Veilleux, 2014; Wagner and others, 2016). Thus, after adequate geographic and hydrologic coverage was ensured, streamgages in the dataset that had a P_{RL} less than 36 years were removed from the study. Of the 198 streamgages that remained after removing the 17 streamgages located outside the study area boundary and the 17 redundant streamgages, an additional 15 were removed for having a P_{RL} less than 36 years, leaving 183 streamgages from which a regional skew model was developed (table 1; fig. 2).

Removing the Bias of the At-Site Estimators

The station skew estimates were unbiased by using the correction factor developed by Tasker and Stedinger (1986) and employed by Reis and others (2005). The unbiased station skew estimated by using the P_{RL} is

$$\hat{\gamma}_i = \left[1 + \frac{6}{P_{RL,i}}\right] G_i, \tag{4}$$

where

$\hat{\gamma}_i$ is the unbiased station skew estimate for site i,

$P_{RL,i}$ is the pseudo record length in years for site i as calculated in equations 1 and 2, and

G_i is the traditional biased station skew estimator based on the flood-frequency analysis for site i.

The variance of the unbiased station skew estimate includes the correction factor developed by Tasker and Stedinger (1986):

$$Var\left[\hat{\gamma}_i\right] = \left[1 + \frac{6}{P_{RL,i}}\right]^2 Var\left[G_i\right], \tag{5}$$

where

$Var\left[G_i\right]$ is calculated by using the formula (Griffis and Stedinger, 2009):

$$Var\left(\hat{G}\right) = \left[\frac{6}{P_{RL}} + a\left(P_{RL}\right)\right] \times \left[1 + \left(\frac{9}{6} + b\left(P_{RL}\right)\right)\hat{G}^2 + \left(\frac{15}{48} + c\left(P_{RL}\right)\right)\hat{G}^4\right], \tag{6}$$

where

$$a\left(P_{RL}\right) = -\frac{17.75}{P_{RL}{}^2} + \frac{50.06}{P_{RL}{}^3},$$

$$b\left(P_{RL}\right) = \frac{3.92}{P_{RL}{}^{0.3}} - \frac{31.10}{P_{RL}{}^{0.6}} + \frac{34.86}{P_{RL}{}^{0.9}}, \text{ and}$$

$$c\left(P_{RL}\right) = -\frac{7.31}{P_{RL}{}^{0.59}} + \frac{45.90}{P_{RL}{}^{1.18}} - \frac{86.50}{P_{RL}{}^{1.77}}.$$

For the 183 streamgages in the study area used in the skew analysis, the unbiased station skew ranged from −1.65 to 1.66 in $\log_{10}$ units (table 1; fig. 3).

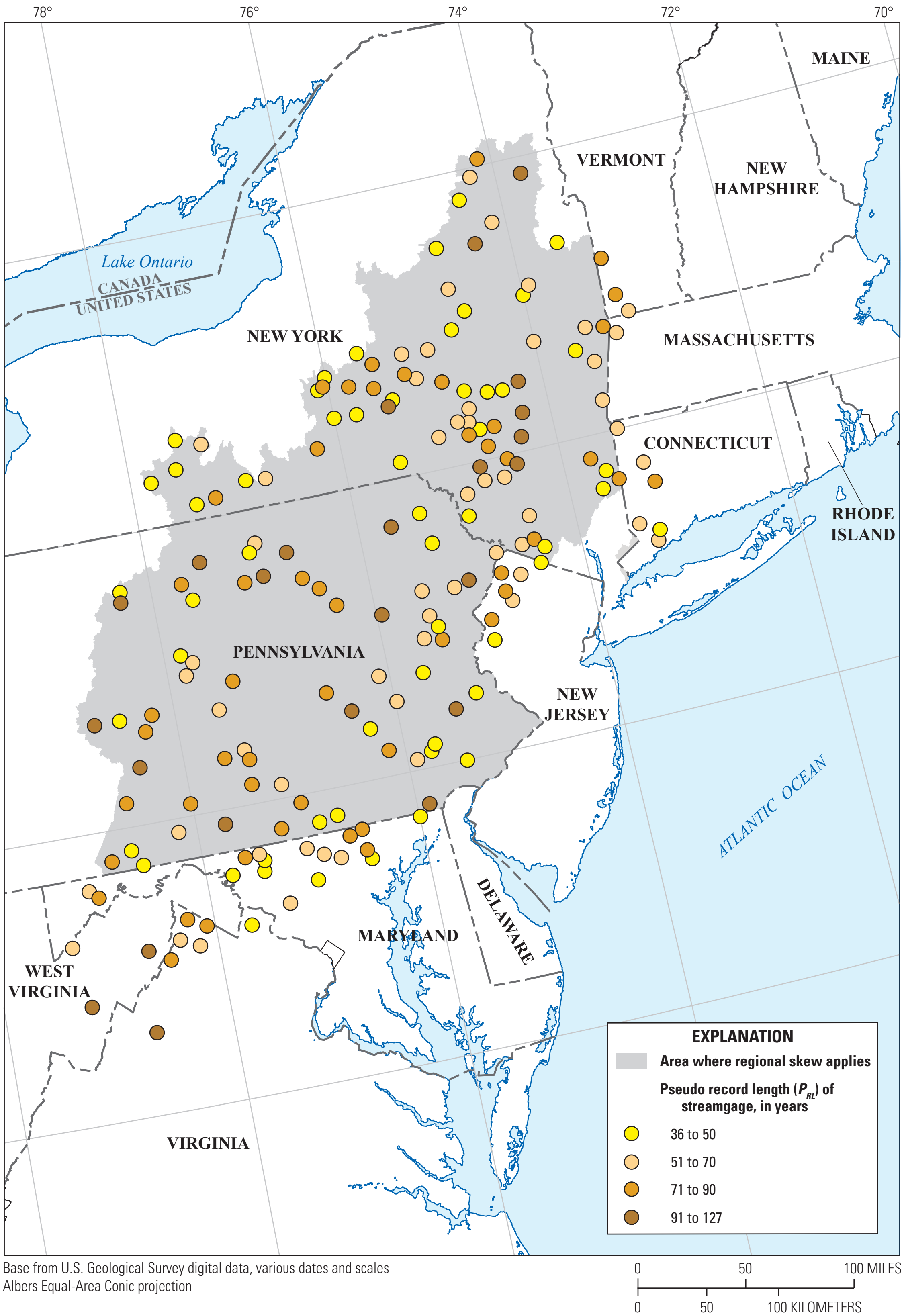


Figure 2. Map showing the pseudo record lengths (P_{RL}) of streamgages that were used in the regional skew analysis for parts of eastern New York and Pennsylvania.

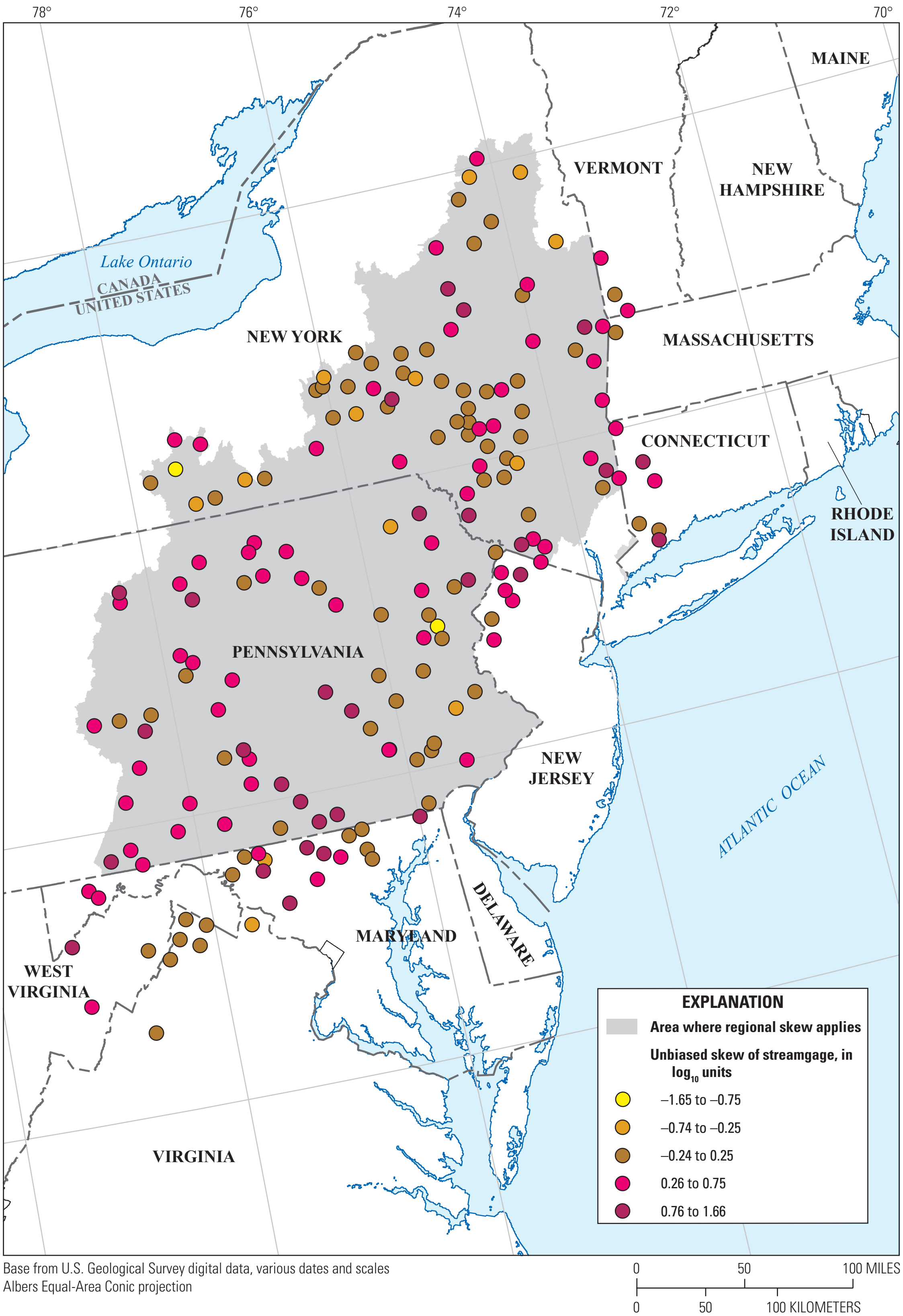


Figure 3. Map showing unbiased station skew of streamgages used in the regional skew analysis for parts of eastern New York and Pennsylvania.

Estimating the Mean Square Error of the Skew

There are several possible ways to estimate $MSE\left(\hat{G}\right)$. The approach used by EMA (taken from eq. 55 in Cohn and others, 2001) generates a first-order estimate of the $MSE\left(\hat{G}\right)$, which should perform well when interval data are available. Another option is to use the Griffis and Stedinger (2009) formula in equations 1–7 (the variance is equated to the MSE) by employing either the gaged-record length or the length of the entire historical period (from the beginning year to the ending year of the record); however, this method does not account for censored data and can lead to an inaccurate and underestimated $MSE\left(\hat{G}\right)$. This issue was addressed by using the P_{RL} instead of the length of the historical period; the P_{RL} accounts for the effects of the censored data and the number of recorded gaged peaks. Thus, the unbiased $MSE\left(\hat{G}\right)$ was used in the regional skewness model because it is more stable and relatively independent of the station skewness estimator (Griffis and Stedinger, 2009). This method also was used in previous regional skew studies (Parrett and others, 2011; Eash and others, 2013; Paretti and others, 2014; Southard and Veilleux, 2014; Wagner and others, 2016).

Developing a Cross-Correlation Model

A critical step for the B–GLS analysis is the estimation of the cross correlation of the station skew coefficient estimators. Martins and Stedinger (2002) used Monte Carlo experiments to derive a relation between the cross correlation of the skew estimators for two streamgages (i and j) as a function of the cross correlation of concurrent annual peak flows ($\hat{\rho}_{ij}$):

$$\hat{\rho}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) = Sign\left(\hat{\rho}_{ij}\right) cf_{ij} \left|\hat{\rho}_{ij}\right|^{\kappa}, \tag{7}$$

where

$\hat{\rho}_{ij}$ is the cross correlation of concurrent annual peak flow for two streamgages,

k is a constant between 2.8 and 3.3, and

cf_{ij} is a factor that accounts for the sample size difference between the concurrent record lengths of the two streamgages and is defined as follows:

$$cf_{ij} = CY_{ij} / \sqrt{\left(P_{RL,i}\right)\left(P_{RL,j}\right)}, \tag{8}$$

where

CY_{ij} is the pseudo concurrent record length, and

$P_{RL,i}$, $P_{RL,j}$ are the pseudo record lengths corresponding to streamgages i and j, respectively.

As shown in equation 8, the pseudo concurrent record length (CY_{ij}) is used to compute the cross correlation of station skews. The pseudo concurrent record length depends on the years of common historical records between the two streamgages as well as on the ratio of the pseudo record length to the historical record length (H_i) for each streamgage. Because censored and historical data are used, calculation of the effective concurrent record length is more complex than simply determining the years during which the two streamgages both recorded peaks.

To compute CY_{ij}, the years of historical record in common between the two streamgages are first determined. For the years in common, the following equation that includes the beginning year (YB_{ij}) and ending year (YE_{ij}) is then used to calculate the concurrent years of record between two streamgages (i and j):

$$CY_{ij} = \left(YE_{ij} - YB_{ij} + 1\right)\left(\frac{P_{RL,i}}{H_i}\right)\left(\frac{P_{RL,j}}{H_j}\right). \tag{9}$$

A cross-correlation model for the annual peak flows in the study area was developed by using the base-10 logarithms of annual peak flows from 43 streamgages that generated 762 streamgage pairs with at least 80 years of concurrent gaged peaks. As shown in figure 4*A*, a logit model, termed the Fisher Z Transformation ($Z = \log[(1+r)/(1-r)]$), provides a convenient transformation of the sample correlations (r_{ij}) from the (−1, +1) range to the (−∞, +∞) range (Fisher, 1915, 1921). Models relating the cross correlations of the concurrent annual peak flows at two streamgages ($\hat{\rho}_{ij}$) to various basin characteristics were considered. The adopted model, which uses only one explanatory variable for estimating the cross correlations of concurrent annual peak flows between two streamgages, is based on the distance, in miles, between basin centroids (D_{ij}):

$$\rho_{ij} = \frac{exp\left(2Z_{ij}\right) - 1}{exp\left(2Z_{ij}\right) + 1}, \tag{10}$$

where

$$Z_{ij} = exp\left(0.64 - 0.14\left(\frac{D_{ij}^{0.34} - 1}{0.34}\right)\right). \tag{11}$$

An OLS regression analysis based on 762 streamgage pairs with at least 80 years of concurrent record indicated that this cross-correlation model is as accurate as having 392 years of concurrent annual peak flows from which to calculate cross correlation. As is standard in an OLS analysis, each station pair in the model was given equal weight. By setting the concurrent-years threshold to 80, the model allowed the complete range of data in the study to be represented, while

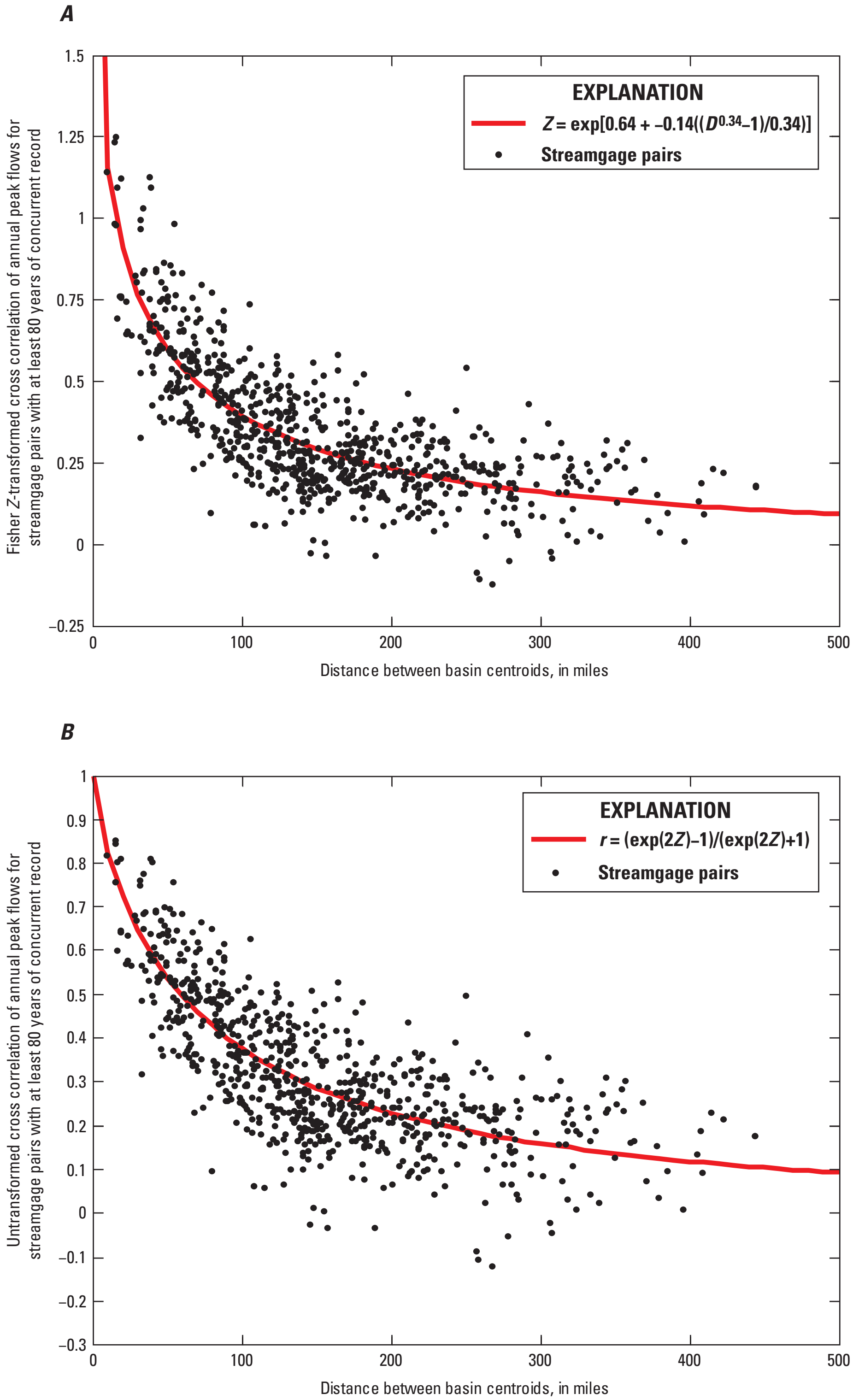


Figure 4. Graphs showing cross correlation of annual peak flows in the study area. *A*, Relation between Fisher *Z*-transformed cross correlation of logarithms of annual peak flows and the distances between basin centroids for 762 streamgage pairs with at least 80 years of concurrent record. *B*, Relation between untransformed cross correlation of logarithms of annual peak flows and the distances between basin centroids for 762 streamgage pairs with at least 80 years of concurrent record. Abbreviations: *r*, cross correlation of concurrent annual peak flows; *D*, distance between basin centroids, in miles; and *Z*, Fisher *Z*-transformed cross correlation of annual peak flows.

also minimizing the influence of station pairs with less accuracy and (or) less data. The fitted OLS regression relation between Z and the distance between basin centroids from the 762 streamgage pairs (fig. 4*A*) shows an exponential decline in the cross correlation for streamgages within 100 miles of each other. A similar decline is found in the cross correlation and distance between basin centroids for the untransformed streamgage pairs (fig. 4*B*). This model was used to estimate cross correlation for concurrent annual peak flows between all streamgage pairs used in the regional skew study.

Regression Analyses

The B–WLS/B–GLS method for computing a regional skew begins with an OLS analysis to develop a regional skew model that is used to generate an estimate of regional skew for each streamgage (Veilleux, 2011; Veilleux and others, 2011, 2012). The OLS-based regional skew estimate is the basis for computing the variance of the skew for each streamgage. Next, B–WLS is used to generate estimators of the regional skew model parameters. Finally, B–GLS is used to estimate the precision of the B–WLS parameter values and the model error variance and its precision, and to compute various diagnostic statistics.

Ordinary Least Squares Analysis

The first step in the B–WLS/B–GLS regional skew analysis is to prepare an initial regional skew model by using OLS regression. The OLS regression analysis yields parameters (such as $\hat{\boldsymbol{\beta}}_{OLS}$) and a model that can be used to generate unbiased regional estimates of the skew for all streamgages:

$$\tilde{y}_{OLS} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{OLS}, \tag{12}$$

where

$\tilde{y}_{OLS}$	are the estimated regional skew values,
$\boldsymbol{X}$	is a $(n \times k)$ matrix of basin characteristics,
$\hat{\boldsymbol{\beta}}_{OLS}$	is a $(k \times 1)$ vector of estimated regression parameters,
n	is the number of streamgages, and
k	is the number of basin parameters, including a column of ones, to estimate the regression constant.

The estimated regional skew values ($\tilde{y}_{OLS}$) are then used to calculate unbiased streamgage regional skew variances by using equation 8 in Griffis and Stedinger (2009). These variances are based on the OLS estimator of the regional skew coefficient instead of the station skew estimator, making the weights in the subsequent steps relatively independent of the station skew estimates.

Weighted Least Squares Analysis

A B–WLS analysis is used to develop estimators of the regression coefficients for the regional skew model. The B–WLS analysis explicitly reflects variations in record length but intentionally neglects cross correlations, thereby avoiding problems experienced with B–GLS parameter estimators (Veilleux, 2011; Veilleux and others, 2011).

The first step in the B–WLS analysis is to estimate the model error variance ($\sigma^2_{\delta,B-WLS}$) (Reis and others, 2005). Using a B–WLS approach to estimate the model error variance precludes the pitfall of estimating the model error variance as zero, which can occur when the method-of-moments weighted least-squares regression is used. Although the B–WLS analysis produces an estimate of the distribution of the model error variance, only the mean model error variance estimator is considered. Given the model error variance estimator, a B–WLS analysis is used to generate the weight matrix ($\boldsymbol{W}$) needed to compute estimates of the final regression parameters ($\hat{\boldsymbol{\beta}}_{B-WLS}$). To compute $\boldsymbol{W}$, a diagonal covariance matrix [$\boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)$] is created (eq. 13). The diagonal elements of the covariance matrix are the sum of the estimated model error variance and the variance of the unbiased station skew ($Var\left[\hat{\gamma}_i\right]$), which depends upon the record length and the estimate of the previously calculated OLS regional skew ($\tilde{y}_{OLS}$). The off-diagonal elements of $\boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)$ are zero because cross correlations among sets of streamgage data are not considered in the B–WLS analysis. Thus, the $(n \times n)$ covariance matrix, $\boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)$, is given by

$$\boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right) = \sigma^2_{\delta,B-WLS}\boldsymbol{I} + diag\left(Var\left[\hat{\gamma}\right]\right), \tag{13}$$

where

$\sigma^2_{\delta,B-WLS}$ is the model error variance,

$\boldsymbol{I}$ is an $(n \times n)$ identity matrix,

n is the number of streamgages in the study, and

$diag\left(Var\left[\hat{\gamma}\right]\right)$ is the $(n \times n)$ matrix containing the variance of the unbiased station skew, $Var\left[\hat{\gamma}_i\right]$, on the diagonal and zeros on the off-diagonal.

By using the covariance matrix, the B–WLS weights are calculated as

$$\boldsymbol{W} = \left[\boldsymbol{X}^T \boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)^{-1} \boldsymbol{X}\right]^{-1} \boldsymbol{X}^T \boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)^{-1}, \tag{14}$$

where

$\boldsymbol{W}$ is the $(k \times n)$ matrix of weights,

$\boldsymbol{X}$ is the $(n \times k)$ matrix of explanatory basin parameters,

$\boldsymbol{\Lambda}_{B-WLS}\left(\sigma^2_{\delta,B-WLS}\right)$ is the $(n \times n)$ covariance matrix, and

k is the number of basin parameters, including a column of ones, to estimate the regression constant.

These weights are used to compute the final estimates of the regression parameters ($\hat{\boldsymbol{\beta}}$) as

$$\hat{\boldsymbol{\beta}}_{B-WLS} = \boldsymbol{W}\hat{\gamma}, \tag{15}$$

where

$\hat{\boldsymbol{\beta}}_{B-WLS}$ is the $(k \times 1)$ vector of estimated regression parameters.

Generalized Least Squares Analysis

After the regression model coefficients ($\hat{\boldsymbol{\beta}}_{B-WLS}$) and weights ($\boldsymbol{W}$) have been determined by using a B–WLS analysis, the degrees of precision of the fitted model and the regression coefficients also are estimated by using a B–GLS analysis. Using the B–GLS regression framework for regional skew, Reis and others (2005) developed the posterior probability-density function for model error variance described as

$$f\left(\sigma^2_{\delta,B-GLS} \mid \hat{\gamma}, \hat{\boldsymbol{\beta}}_{B-WLS}\right) \propto \xi\left(\sigma^2_{\delta,B-GLS}\right) \times \left|\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^2_{\delta,B-GLS}\right)\right|^{-0.5} \times \exp\left[-0.5\left(\hat{\gamma} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{B-WLS}\right)^T \left(\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^2_{\delta,B-GLS}\right)\right)^{-1} \left(\hat{\gamma} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{B-WLS}\right)\right], \tag{16}$$

where

$\hat{\gamma}$ represents the skew data, and

$\xi\left(\sigma^2_{\delta,B-GLS}\right)$ is the exponential prior for the model error variance defined as

$$\xi\left(\sigma^2_{\delta,B-GLS}\right) = \lambda e^{-\lambda\left(\sigma^2_{\delta,B-GLS}\right)}, \ \sigma^2_{\delta,B-GLS} > 0. \tag{17}$$

The value 10 was adopted for lambda (λ) on the basis of a mean model error variance of 1/10. That prior assigns a 63-percent probability to the interval (0, 0.1), 86-percent probability to the interval (0, 0.2), and 95-percent probability to the interval (0, 0.3).

The mean B–GLS model error variance ($\sigma^2_{\delta,B-GLS}$) can then be used to compute the precision of the regression parameters ($\hat{\boldsymbol{\beta}}_{B-WLS}$) that were based on the B–WLS weights ($\boldsymbol{W}$). The B–GLS covariance matrix for the B–WLS estimator ($\hat{\boldsymbol{\beta}}_{B-WLS}$) is simply

$$\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}_{B-WLS}\right) = \boldsymbol{W}\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^2_{\delta,B-GLS}\right)\boldsymbol{W}^T, \tag{18}$$

where

$\boldsymbol{W}$ is the $(k \times n)$ matrix of weights determined by B–WLS analysis,

$\boldsymbol{W}^T$ is the transformation of $\boldsymbol{W}$, and

$\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^2_{\delta,B-GLS}\right)$ is the $(n \times n)$ B–GLS covariance matrix calculated as

$$\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^2_{\delta,B-GLS}\right) = \sigma^2_{\delta,B-GLS}\boldsymbol{I} + \boldsymbol{\Sigma}\left(\hat{\gamma}\right), \tag{19}$$

where

$\boldsymbol{I}$ is an $(n \times n)$ identity matrix, and

$\boldsymbol{\Sigma}\left(\hat{\gamma}\right)$ is a full $(n \times n)$ matrix containing the sampling variances of the streamflow record's unbiased skew, $Var\left[\hat{\gamma}_i\right]$, and the covariances of the skew $\hat{\gamma}_i$.

The off-diagonal values of $\boldsymbol{\Sigma}\left(\hat{\gamma}\right)$ are determined by the cross correlation of concurrent gaged annual peak flows and the *cf* factor (see eq. 8), which accounts for the size differences between pairs of samples collected at different streamgages and their concurrent record length (Martins and Stedinger, 2002). In the calculation of the *cf* factor, only the gaged records and historical peaks are considered when determining the ratio of the number of concurrent peak flows at streamgage pairs to the total number of annual peak flows at both streamgages. Thus, any additional information provided by perception thresholds and censored peaks in the EMA analysis is neglected in the calculation of the cross correlation of annual peak flows and the *cf* factor. Precision metrics include (1) the standard error of the regression parameters [$SE\left(\hat{\boldsymbol{\beta}}_{B-WLS}\right)$]; (2) the model error variance ($\sigma^2_{\delta,B-GLS}$); (3) pseudo coefficient of determination (pseudo R^2_δ); and (4) the average variance of prediction at a new streamgage not used in the regional model (AVP_{new}).

Results and Discussion

Final Bayesian Weighted Least Squares/ Bayesian Generalized Least Squares Model

A constant B–WLS/B–GLS model (having a skew of 0.32 and developed by using data from 183 streamgages with at least 36 years of P_{RL} each) produced the only statistically significant model of skew in the study area (table 3). A constant model does not explain any variability in skew; therefore, the pseudo R^2_δ, a diagnostic statistic that describes the percentage of the variability in the skew from streamgage to streamgage that is estimated by the model (Gruber and others, 2007; Parrett and others, 2011), is 0 percent. All available basin characteristics were evaluated as possible explanatory variables in the B–WLS/B–GLS regression analysis; however, the addition of any of the available basin characteristics or combinations thereof did not produce a pseudo R^2_δ greater than 1 percent, indicating that they did not explain the variation in station skews in the study area. Thus, the addition of basin characteristics as explanatory variables was not warranted because the increase in complexity did not result in a gain in precision.

The posterior mean of the constant model error variance (σ^2_δ) is 0.10. The average sampling error variance (*ASEV*) of the constant model is 0.0078, which represents the average error in the regional skew as calculated from the station skew values measured at the streamgages used in the analysis. The average variance of prediction at a new streamgage (AVP_{new}) corresponds to the MSE used in Bulletin 17B (B17B; Interagency Advisory Committee on Water Data, 1982) to describe the precision of the generalized skew map. The constant model has an AVP_{new} of 0.11, which corresponds to an effective record length of 68 years. An AVP_{new} of 0.11 is a marked improvement over the B17B national skew map, whose reported MSE of 0.302 has a corresponding effective record length of only 17 years (Interagency Advisory Committee on Water Data, 1982). Measured by effective record length, the new regional model includes more than three times the information than the B17B map. Appendix 1 provides a graphical assessment of the B–WLS/B–GLS model of regional skew.

Bayesian Weighted Least Squares/Bayesian Generalized Least Squares Regression Diagnostics

To determine whether a regression model is a good representation of the data and which regression parameters, if any, should be included in the model, diagnostic statistics have been developed to evaluate how well a model fits a regional

Table 3. Regional skew model and model fit for parts of eastern New York and Pennsylvania.

[Standard deviations are in parentheses. Terms: σ^2_δ, model error variance; *ASEV*, average sampling error variance; AVP_{new}, average variance of prediction at a new streamgage; pseudo R^2_δ, fraction of the variability of the station skews explained by each model (Gruber and others, 2007)]

Model	Regression constant	σ^2_δ	*ASEV*	AVP_{new}	Pseudo R^2_δ (percent)
Constant	0.32 (0.09)	0.10 (0.020)	0.0078	0.11	0

hydrologic dataset (Griffis, 2006; Gruber and Stedinger, 2008). In a regional skew study, potential explanatory variables are statistically evaluated to ensure an accurate prediction of skew while also keeping the model as simple as possible.

A pseudo analysis of variance (pseudo ANOVA) contains regression diagnostics and goodness-of-fit statistics that describe how much of the variation in the observations can be attributed to the regional model, and how much of the variation in the residuals can be attributed to modeling and sampling error (table 4; fig. 5). Determining these quantities is difficult; the modeling errors cannot be resolved because the values of the sampling errors (η_i) for each streamgage (i) are not known. However, the total sampling error sum of squares (SS) can be described by its mean value ($\sum_{i=1}^{n} Var[\hat{\gamma}_i]$), where n is the number of equations and where the total variation caused by the model error (δ) for a model with k parameters has a mean equal to $n\sigma_\delta^2(k)$. Thus, the residual variation attributed to the sampling error is $\sum_{i=1}^{n} Var[\hat{\gamma}_i]$, and the residual variation attributed to the model error is $n\sigma_\delta^2(k)$.

For a model with no explanatory parameters other than the mean (the constant model), the estimated model error variance ($\sigma_\delta^2(0)$) describes all of the variation in $\gamma_i = \mu + \delta_i$, where μ is the mean of the estimated station skews. Thus, the total variation resulting from model error (δ_i) and sampling error ($\eta_i = \hat{\gamma}_i - \gamma_i$) in the expected sum of squares should equal $\sigma_\delta^2(0) + \sum_{i=1}^{n} Var(\hat{\gamma}_i)$. For a model type other than constant, the expected sum of squares attributed with k parameters equals $n\left[\sigma_\delta^2(0) - \sigma_\delta^2(k)\right]$ because the sum of the model error variance $n\sigma_\delta^2(k)$ and the variance explained by the model must equal $n\sigma_\delta^2(0)$. This division of the variation in the observations is referred to as a pseudo ANOVA because the contributions of the three sources of error are estimated or constructed rather than determined from the computed residual errors and the observed model predictions, while not accounting for the effect of correlation on the sampling errors.

The error variance ratio (EVR) is a diagnostic statistic used to determine whether a simple OLS regression analysis would suffice, or if a more sophisticated B–WLS or B–GLS analysis would be more appropriate. The EVR is the ratio of the average sampling error variance to the model error variance. Generally, an EVR greater than 0.20 indicates that the sampling error variance is not negligible when compared to the model error variance, suggesting that a B–WLS or B–GLS regression analysis is appropriate. The EVR is calculated as

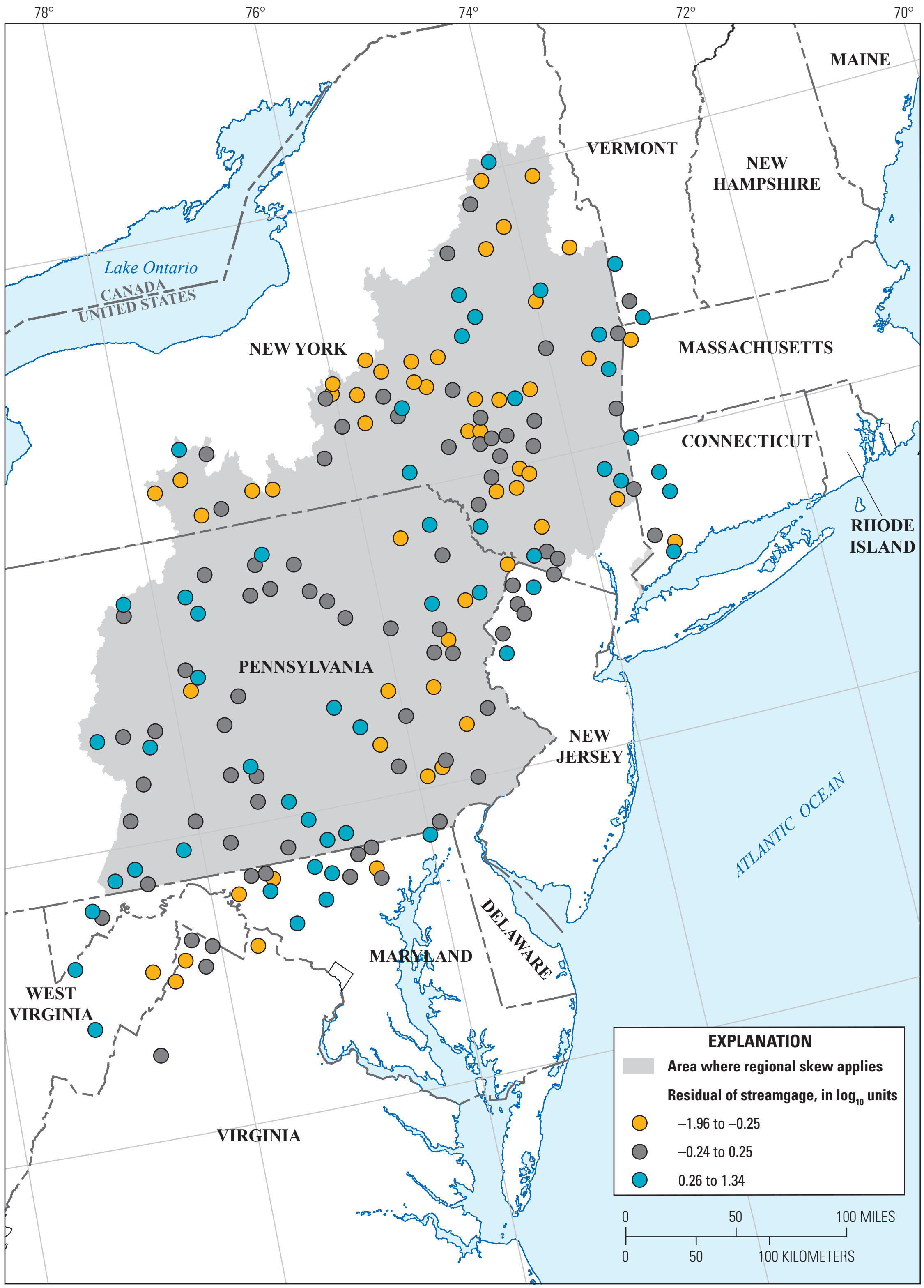
$$EVR = \frac{\mathrm{SS}(\text{sampling error})}{\mathrm{SS}(\text{model error})} = \frac{\sum_{i=1}^{n} Var(\hat{\gamma}_i)}{n\sigma_\delta^2(k)}. \quad (20)$$

The constant model has a sampling error variance of 0.0078 (table 3) and an EVR of 1.3 (table 4), indicating that the sampling error variance is not negligible when compared to the model error variance, and that a B–WLS or B–GLS regression analysis was appropriate. Thus, an OLS model that

Table 4. Pseudo analysis of variance (pseudo ANOVA) table for the constant model of regional skew in parts of eastern New York and Pennsylvania.

[Terms: k, number of estimated regression parameters not including the constant; n, number of streamgages used in regression; $\sigma_\delta^2(0)$, model error variance of a constant model; $\sigma_\delta^2(k)$, model error variance of a model with k regression parameters and a constant; NA, not applicable; $Var(\hat{\gamma}_i)$, variance of the estimated sample skew at site i; EVR, error variance ratio; MBV^*, misrepresentation of the beta variance; b_0^{B-WLS}, regression constant from B–WLS analysis; B–GLS, Bayesian generalized least squares; B–WLS, Bayesian weighted least squares; $\boldsymbol{W}^T$, the transformation of $\boldsymbol{W}$; $\boldsymbol{\Lambda}$, covariance matrix; $\boldsymbol{W}$, the ($k \times n$) matrix of weights determined by B–WLS analysis; $\boldsymbol{W}_i = \frac{1}{\sqrt{\boldsymbol{\Lambda}_{ii}}}$; pseudo R_δ^2, fraction of variability in the true skews explained by each model (Gruber and others, 2007); %, percent]

Source	Degrees of freedom	Equations	Sum of squares	Result
Model	$k = 0$	$n\left[\sigma_\delta^2(0) - \sigma_\delta^2(k)\right]$	0	NA
Model error	$n-k-1 = 182$	$n\left[\sigma_\delta^2(k)\right]$	19	NA
Sampling error	$n = 183$	$\sum_{i=1}^{n} Var(\hat{\gamma}_i)$	24	NA
Total error	$2n-1 = 365$	$n\left[\sigma_\delta^2(k)\right] + \sum_{i=1}^{n} Var(\hat{\gamma}_i)$	43	NA
EVR	NA	$\frac{\sum_{i=1}^{n} Var(\hat{\gamma}_i)}{n\sigma_\delta^2(k)}$	NA	1.3
MBV^*	NA	$\frac{Var\left[b_0^{B-WLS} \mid B-GLS\ analysis\right]}{Var\left[b_0^{B-WLS} \mid B-WLS\ analysis\right]} = \frac{\boldsymbol{W}^T \boldsymbol{\Lambda} \boldsymbol{W}}{\sum_{i=1}^{n} \boldsymbol{W}_i}$	NA	6.3
Pseudo R_δ^2	NA	$1 - \frac{\sigma_\delta^2(k)}{\sigma_\delta^2(0)}$	NA	0%



Base from U.S. Geological Survey digital data, various dates and scales
Albers Equal-Area Conic projection

Figure 5. Map showing residuals from constant model of skew for 183 streamgages that were used in the regional skew analysis for parts of eastern New York and Pennsylvania.

neglects sampling error in the station skew would not provide a statistically reliable analysis of the data. Given the diagnostic statistics and the range of record lengths among streamgages, a B–WLS or B–GLS analysis was warranted to evaluate the final precision of the model.

The misrepresentation of the beta variance (*MBV**) diagnostic statistic is used to determine whether a B–WLS regression is sufficient, or if a B–GLS regression is more appropriate to determine the precision of the estimated regression parameters (Griffis, 2006; Veilleux, 2011). The *MBV** describes the error produced by a B–WLS regression analysis in its evaluation of the precision of b_0^{B-WLS}, which is the estimator of the constant β_0^{B-WLS}, because the covariance among the estimated station skews ($\hat{\gamma}_i$) generally has its greatest effect on the precision of the constant term (Stedinger and Tasker, 1985). If the *MBV** is substantially greater than 1, then a B–GLS error analysis should be employed; conversely, if the *MBV** is not substantially greater than 1, a B–WLS analysis is sufficient. The *MBV** is calculated as

$$MBV^* = \frac{Var\left[b_0^{B-WLS} \mid B-GLS\,analysis\right]}{Var\left[b_0^{B-WLS} \mid B-WLS\,analysis\right]} = \frac{\boldsymbol{W}^T \boldsymbol{\Lambda} \boldsymbol{W}}{\sum_{i=1}^{n} \boldsymbol{W}_i}, \tag{21}$$

where

$$\boldsymbol{W}_i = \frac{1}{\sqrt{\boldsymbol{\Lambda}_{ii}}}.$$

The *MBV** is equal to 6.3 for the constant model (table 4), indicating that the cross correlation among the skew estimators has an effect on the precision with which the regional skew can be estimated. If a B–WLS precision analysis were used for the estimated constant in the model, the variance would be underestimated by a factor of 6.3. Thus, a B–WLS analysis alone would misrepresent the variance of the constant in the skew model. Moreover, a B–WLS model would underestimate the variance of prediction, given that the sampling error in the constant term in both models were sufficiently large to make an appreciable contribution to the average variance of prediction.

Leverage and Influence

Diagnostic statistics for leverage and influence can be used to identify atypical observations and to address lack-of-fit when skew coefficients are estimated. The leverage statistics identify those streamgages in the analysis for which the observed streamflow values have a large impact on the fitted (or predicted) values (Hoaglin and Welsch, 1978). Generally, leverage statistics can determine whether an observation or explanatory variable is unusual and thus likely to have a large effect on the estimated regression coefficients and predictions. Unlike leverage, which highlights points that have the ability or potential to affect the fit of the regression, influence (measured using Cook's distance, or "Cook's D") attempts to describe those points that have an unusual effect on the regression analysis (Belsley and others, 1980; Cook and Weisberg, 1982; Tasker and Stedinger, 1989). An influential observation is one with an unusually large residual that has a disproportionate effect on the fitted regression relations.

Influential observations often have high leverage. If p is the number of estimated regression coefficients (p=1 for a constant model), and n is the sample size (or number of streamgages in the study), then leverage values have a mean of p/n, and values greater than $2p/n$ are generally considered large. Influence values greater than $4/n$ are typically considered large (Veilleux, 2011; Veilleux and others, 2011).

For the constant model of skew in the study area, influence greater than 0.022 (p/n = 4/183) and leverage greater than 0.011 [(2×1)/183] were considered high. No sites in the study area exhibited high leverage; therefore, the differences in the leverage values for the constant model reflect the variation in record lengths among sites. Twelve streamgages in the study area exhibited high influence, and thus had an unusual effect on the fitted regression (table 5). These streamgages also had 12 of the largest residuals (in magnitude) among the 183 streamgages used in the B–WLS/B–GLS analysis.

Table 5. Streamgages with high influence on the constant model of regional skew.

[High influence is defined as Cook's distance (Cook's D) values greater than 4/*n* (or 4/183=0.022). Each of the 183 sites in the regional skew study was assigned a value from 1 to 183 signifying its relative rank, where a rank of 1 corresponds to the largest positive value in each category. The table is sorted from the largest to the smallest Cook's D values. Abbreviations: USGS, U.S. Geological Survey; ERL, effective record length; MSE, mean squared error; PA, Pennsylvania; NY, New York; MD, Maryland]

Index number	USGS streamgage number	State in which streamgage is located	Cook's D	Leverage	Pseudo ERL (years)		Unbiased at-site skew		Unbiased MSE (at-site skew)		Residual	
					Value	Rank	Value	Rank	Value	Rank	Value	Rank
212	01574000	PA	0.079	0.007	85	34	1.7	1	0.09	147	1.3	3
211	01573000	PA	0.050	0.007	95	20	1.3	4	0.08	163	1.0	6
161	01449360	PA	0.050	0.005	47	134	−1.5	182	0.17	50	−1.8	2
170	01473000	PA	0.046	0.007	99	15	−0.6	180	0.07	169	−0.92	11
78	01358500	NY	0.031	0.005	59	98	1.5	2	0.13	83	1.2	4
144	01521596	NY	0.030	0.004	36	175	−1.6	183	0.22	9	−2.0	1
197	01555500	PA	0.029	0.007	84	38	1.1	10	0.09	142	0.82	14
157	01439500	PA	0.029	0.007	105	5	1.0	16	0.07	179	0.70	21
33	01595500	MD	0.029	0.006	63	90	1.4	3	0.12	93	1.1	5
28	01586000	MD	0.028	0.006	68	80	1.3	5	0.11	103	1.0	8
177	01534000	PA	0.023	0.007	100	10	−0.3	174	0.07	174	−0.65	28
48	01643500	MD	0.022	0.006	65	84	1.2	8	0.12	97	0.89	12

Summary

Bulletin 17C (B17C) guidelines recommend fitting the log-Pearson Type III (LP–III) distribution to a series of annual peak flows at a station by using the method of moments. The LP–III distribution is described by three moments: the mean, the standard deviation, and the skewness coefficient. The third moment, the skewness coefficient (hereinafter referred to as “skew”), is a measure of the asymmetry of the distribution or, in other words, the thickness of the tails of the distribution. In flood-frequency analysis, the skew is important because the magnitude of annual exceedance probability (AEP) flows for a streamgage estimated by using the LP–III distribution are affected by the skew of the annual peak flows (hereinafter referred to as “station skew”); interest is focused on the right-hand tail of the distribution, which represents annual peak flows corresponding to small AEPs and the larger flood flows. For streamgages having modest record lengths, the skew is sensitive to extreme events, such as large floods, as they cause a sample to be highly asymmetrical, or skewed. Thus, B17C recommends using a weighted-average skew computed from the station skew for a given streamgage and a regional skew. These choices reduce the sensitivity of the station skew to extreme events, particularly for streamgages with short record lengths.

An estimate of regional skew is generated for a study area that includes part of the Mid-Atlantic region (hydrologic units 0202, 0204, 0205, 0206, and 0207) and the Connecticut Coastal region (hydrologic unit 0110) located in the States of Connecticut, Maryland, Massachusetts, New Jersey, New York, Pennsylvania, Vermont, Virginia, and West Virginia. The study area encompasses 49,817 square miles and spans approximately 300 miles from north to south (from northern New York to the southern border of Pennsylvania) and approximately 200 miles from east to west (from west-central Pennsylvania to the eastern border of New York).

Candidate streamgages in the study area were selected by the U.S. Geological Survey (USGS) in New York and Pennsylvania. Only streamgage records unaffected by extensive regulation, diversion, urbanization, or channelization, and that had approximately 25 or more gaged peaks were considered.

As recommended in B17C, the flood frequency for each candidate streamgage was determined by employing the expected moments algorithm (EMA), which extends the method of moments so that it can accommodate interval, censored, and historical/paleo data, as well as use the multiple Grubbs-Beck test (MGBT) to identify potentially influential low floods (PILFs) in the data series.

A total of 232 candidate streamgages were initially considered for use in the skew analysis; after screening for redundancy and sufficient pseudo record length (P_{RL}), 183 streamgages were selected. The Bayesian weighted least squares/Bayesian generalized least squares (B–WLS/B–GLS) regression method was used to develop a regional skew model for the study area that would incorporate possible explanatory variables (basin characteristics) to explain the variation in skew in the study area. Basin characteristics for candidate streamgages were obtained from the USGS Geospatial Attributes of Gages for Evaluating Streamflow, Version II (GAGES-II) database or were generated by using the ArcHydro package in Esri ArcGIS version 10.3.1. Ten basin characteristics were considered as possible explanatory variables in the B–WLS/B–GLS regression analysis; however, none produced a pseudo coefficient of determination greater than 1 percent, indicating that they did not explain the variation in station skews in the study area. Therefore, a constant skew model was selected. The constant model has a regional skew coefficient of 0.32 and an average variance of prediction at a new streamgage (AVP_{new}) of 0.11, which corresponds to the mean square error (MSE). An AVP_{new} of 0.11 corresponds to an effective record length of 68 years, which is a marked improvement over the Bulletin 17B (B17B) national skew map, whose reported MSE of 0.302 has a corresponding effective record length of only 17 years. Measured by effective record length, the new regional model provides more than three times the amount of information provided by the B17B map.

Acknowledgments

The authors would like to acknowledge Doug Burns and Gary Wall in the U.S. Geological Survey (USGS) New York Water Science Center and Mark Roland in the USGS Pennsylvania Water Science Center for selecting streamgages from their respective States and parts of adjacent States for use in the regional skew analysis and for conducting flood-frequency analyses for those streamgages. The authors would also like to acknowledge Chris Sanoki and Danny Morel in the USGS Upper Midwest Water Science Center for generating basin characteristics for streamgages that were not in the GAGES-II database.

References Cited

Belsley, D.A., Kuh, E., and Welsch, R.E., 1980, Regression diagnostics—Identifying influential data and sources of collinearity: Hoboken, N.J., John Wiley & Sons, Inc., 300 p. [Also available at https://doi.org/10.1002/0471725153.]

Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, no. 9, p. 2089–2096, accessed October 1, 2019, at https://doi.org/10.1029/97WR01640.

Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates: Water Resources Research, v. 37, no. 6, p. 1695–1706. [Also available at https://doi.org/10.1029/2001WR900016.]

Cook, R.D., and Weisberg, S., 1982, Residuals and influence in regression: New York, N.Y., Chapman and Hall, 230 p.

Curran, J.H., Barth, N.A., Veilleux, A.G., and Ourso, R.T., 2016, Estimating flood magnitude and frequency at gaged and ungaged sites on streams in Alaska and conterminous basins in Canada, based on data through water year 2012: U.S. Geological Survey Scientific Investigations Report 2016–5024, 47 p., accessed October 1, 2017, at https://doi.org/10.3133/sir20165024.

Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013–5086, 63 p. with appendix.

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2018, Guidelines for determining flood flow frequency—Bulletin 17C (ver. 1.1, May 2019): U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p. [Also available at https://doi.org/10.3133/tm4B5.]

Environmental Systems Research Institute [Esri], 2009, ArcGIS desktop help, accessed November 23, 2015, at https://desktop.arcgis.com/en/desktop/.

Falcone, J.A., 2011, GAGES-II—Geospatial attributes of gages for evaluating streamflow: U.S. Geological Survey dataset, accessed October 1, 2015, at https://pubs.er.usgs.gov/publication/70046617.

Fenneman, N.M., 1938, Physiography of the Eastern United States (1st ed.): New York, McGraw-Hill Book Co., 714 p.

Fisher, R.A., 1915, Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population: Biometrika, v. 10, no. 4, p. 507–521. [Also available at https://www.jstor.org/stable/2331838.]

Fisher, R.A., 1921, On the “probable error” of a coefficient of correlation deduced from a small sample: Metron, v. 1, p. 3–32.

Griffis, V.W., 2006, Flood-frequency analysis—Bulletin 17, regional information, and climate change: Ithaca, N.Y., Cornell University, Ph.D. dissertation, 241 p.

Griffis, V.W., and Stedinger, J.R., 2007, Evolution of flood frequency analysis with Bulletin 17: Journal of Hydrologic Engineering, v. 12, no. 3, p. 283–297, accessed October 1, 2019, at https://doi.org/10.1061/(ASCE)1084-0699(2007)12:3(283).

Griffis, V.W., and Stedinger, J.R., 2009, Log-Pearson type 3 distribution and its application in flood frequency analysis, III—Sample skew and weighted skew estimators: Journal of Hydrology (Amsterdam), v. 14, no. 2, p. 121–130.

Gruber, A.M., Reis, D.S., Jr., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, *in* Kabbes, K.C., ed., Proceedings of the 2007 World Environmental and Water Resources Congress—Restoring our natural habitat, Tampa, Fla., May 15–18, 2007: Reston, Va., American Society of Civil Engineers, 10 p., accessed October 3, 2019, at https://doi.org/10.1061/40927(243)400.

Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian GLS regression, *in* Babcock, R.W., and Walton, R., eds., World Environmental and Water Resources Congress 2008—Ahupua'a—Proceedings of the congress, Honolulu, Hawai'i, May 12–16, 2008: Reston, Va., American Society of Civil Engineers, p. 5575–5584. [Also available at https://doi.org/10.1061/40976(316)563.]

Hoaglin, D.C., and Welsch, R.E., 1978, The hat matrix in regression and ANOVA: The American Statistician, v. 32, no. 1, p. 17–22.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354, accessed October 12, 2018, at https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=309950.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency—Bulletin 17B (revised and corrected): Washington, D.C., Hydrologic Subcommittee, 28 p.

Martins, E.S., and Stedinger, J.R., 2002, Cross correlations among estimators of shape: Water Resources Research, v. 38, no. 11, p. 34-1 to 34-7. [Also available at https://doi.org/10.1029/2002WR001589.]

Mastin, M.C., Konrad, C.P., Veilleux, A.G., and Tecca, A.E., 2016, Magnitude, frequency, and trends of floods at gaged and ungaged sites in Washington, based on data through water year 2014 (ver. 1.2, November 2017): U.S. Geological Survey Scientific Investigations Report 2016–5118, 70 p., accessed October 1, 2017, at https://doi.org/10.3133/sir20165118.

Olson, S.A., 2014, Estimation of flood discharges at selected annual exceedance probabilities for unregulated, rural streams in Vermont, *with a section on* Vermont regional skew regression, by Veilleux, A.G.: U.S. Geological Survey Scientific Investigations Report 2014–5078, 27 p. plus appendixes, accessed October 1, 2015, at https://doi.org/10.3133/sir20145078.

Paretti, N.V., Kennedy, J.R., Turney, L.A., and Veilleux, A.G., 2014, Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2014–5211, 61 p., accessed October 1, 2015, at https://doi.org/10.3133/sir20145211.

Parrett, C., Veilleux, A.G., Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through water year 2006: U.S. Geological Survey Scientific Investigations Report 2010–5260, 94 p. [Also available at https://doi.org/10.3133/sir20105260.]

Reis, D.S., Jr., Stedinger, J.R., and Martins, E.S., 2005, Bayesian generalized least squares regression with application to the log Pearson type 3 regional skew estimation: Water Resources Research, v. 41, no. 10, W10419, 14 p. [Also available at https://doi.org/10.1029/2004WR003445.]

Southard, R.E., and Veilleux, A.G., 2014, Methods for estimating annual exceedance-probability discharges and largest recorded floods for unregulated streams in rural Missouri: U.S. Geological Survey Scientific Investigations Report 2014–5165, 39 p. [Also available at https://doi.org/10.3133/sir20145165.]

Stedinger, J.R., and Cohn, T.A., 1986, Flood frequency analysis with historical and paleoflood information: Water Resources Research, v. 22, no. 5, p. 785–793. [Also available at https://doi.org/10.1029/WR022i005p00785.]

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis, 1. Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432. [Also available at https://doi.org/10.1029/WR021i009p01421.]

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: Journal of Water Resources Planning and Management, v. 112, no. 2, p. 225–237. [Also available at https://doi.org/10.1061/(ASCE)0733-9496(1986)112:2(225).]

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology (Amsterdam), v. 111, nos. 1–4, p. 361–375. [Also available at https://doi.org/10.1016/0022-1694(89)90268-0.]

U.S. Department of Agriculture, 2008, U.S. general soil map (STATSGO): U.S. Department of Agriculture, Natural Resources Conservation Service database, accessed October 2015 at https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/.

U.S. Geological Survey, 2018, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed October 1, 2018, at https://doi.org/10.5066/F7P55KJN. [Peak streamflow information directly accessible at https://nwis.waterdata.usgs.gov/nwis/peak.]

U.S. Geological Survey, 2019, The National Map downloader: U.S. Geological Survey web page, accessed January 2019 at https://apps.nationalmap.gov/downloader/#/.

Veilleux, A.G., 2009, Bayesian GLS regression for regionalization of hydrologic statistics, floods and Bulletin 17 skew: Ithaca, N.Y., Cornell University, M.S. thesis, 155 p.

Veilleux, A.G., 2011, Bayesian GLS regression, leverage, and influence for regionalization of hydrologic statistics: Ithaca, N.Y., Cornell University, Ph.D. dissertation, 184 p.

Veilleux, A.G., Cohn, T.A., Flynn, K.M., Mason, R.R., and Hummel, P.R., 2014, Estimating magnitude and frequency of floods using the PeakFQ 7.0 program: U.S. Geological Survey Fact Sheet 2013–3108, 2 p., accessed October 1, 2015, at https://doi.org/10.3133/fs20133108.

Veilleux, A.G., Stedinger, J.R., and Eash, D.A., 2012, Bayesian WLS/GLS regression for regional skewness analysis for regions with large crest stage gage networks, *in* Loucks, E.D., ed., Proceedings of the World Environmental and Water Resources Congress 2012—Crossing boundaries, Albuquerque, N. Mex., May 20–24, 2012: Reston, Va., American Society of Civil Engineers, p. 2253–2263.

Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011, Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows, *in* Beighley, R.E., II, and Killgore, M.W., eds., Proceedings of the World Environmental and Water Resources Congress 2011—Bearing knowledge for sustainability, Palm Springs, Calif., May 22–26, 2011: Reston, Va., American Society of Civil Engineers, p. 3103–3123.

Veilleux, A.G., and Wagner, D.M., 2019, Methods for estimating regional skewness of annual peak flows in parts of the Great Lakes and Ohio River Basins, based on data through water year 2013: U.S. Geological Survey Scientific Investigations Report 2019–5105, 26 p., accessed February 1, 2020, at https://doi.org/10.3133/sir20195105.

Wagner, D.M., Krieger, J.D., and Veilleux, A.G., 2016, Methods for estimating annual exceedance probability discharges for streams in Arkansas, based on data through water year 2013: U.S. Geological Survey Scientific Investigations Report 2016–5081, 136 p., accessed August 5, 2016, at https://doi.org/10.3133/sir20165081.

Wagner, D.M., and Veilleux, A.G., 2021, Regional flood skew for parts of the mid-Atlantic region (hydrologic unit 02) in eastern New York and Pennsylvania: U.S. Geological Survey data release, accessed February 1, 2021, at https://doi.org/10.5066/P9PGAL0D.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, accessed October 7, 2019, at https://doi.org/10.3133/ofr97656.

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis: Skew coefficient (log units)	EMA analysis: MSE of skew (log units)	Used in regional skew model
1	01184100	0108	CT	54	NA	0.167	0.1063	no-O
2	01187300	0108	CT	75	NA	0.099	0.0754	no-O
3	01187800	0108	CT	65	NA	0.894	0.1542	no-O
4	01188000	0108	CT	82	NA	0.001	0.0647	no-O
5	01192883	0108	CT	33	NA	−0.26	0.1732	no-O
6	01199050	0110	CT	54	0.61	0.548	0.1331	Yes
7	01200000	0110	CT	81	0.42	0.389	0.0871	Yes
8	01203000	0110	CT	57	1.26	1.143	0.2274	Yes
9	01204000	0110	CT	81	0.74	0.693	0.1084	Yes
10	01208925	0110	CT	41	−0.01	−0.007	0.1246	Yes
11	01208950	0110	CT	54	0.85	0.765	0.1544	Yes
12	01208990	0110	CT	52	0.09	0.078	0.1041	Yes
13	01169000	0108	MA	74	NA	0.457	0.0983	no-O
14	01169900	0108	MA	47	NA	0.709	0.1655	no-O
15	01170100	0108	MA	46	NA	0.706	0.1679	no-O
16	01171500	0108	MA	75	NA	−0.348	0.0901	no-O
17	01181000	0108	MA	78	NA	0.413	0.0913	no-O
18	01198000	0110	MA	31	NA	0.4	0.1949	no-P
19	01332000	0202	MA	62	0.65	0.592	0.1226	Yes
20	01333000	0202	MA	64	−0.06	−0.056	0.0848	Yes
21	01495000	0206	MD	127	0.08	0.074	0.0447	Yes
22	01496000	0206	MD	37	0.87	0.75	0.2004	Yes
23	01580000	0205	MD	87	0.17	0.16	0.069	Yes
24	01582000	0206	MD	80	0.09	0.082	0.0695	Yes
25	01583500	0206	MD	69	0.52	0.477	0.1056	Yes
26	01584050	0206	MD	38	0.08	0.067	0.1382	Yes
27	01584500	0206	MD	86	−0.22	−0.207	0.0714	Yes
28	01586000	0206	MD	68	1.27	1.17	0.2003	Yes
29	01586610	0206	MD	31	NA	−0.304	0.1867	no-P
30	01587500	0206	MD	32	NA	0.864	0.2403	no-P
31	01588000	0206	MD	43	0.63	0.552	0.1621	Yes
32	01591400	0206	MD	35	NA	−0.664	0.2022	no-P
33	01595500	0207	MD	63	1.38	1.259	0.2338	Yes
34	01596500	0207	MD	65	0.67	0.617	0.1218	Yes
35	01599000	0207	MD	84	0.46	0.433	0.0871	Yes
36	01601500	0207	MD	84	0.92	0.858	0.1196	Yes
37	01609000	0207	MD	39	0.47	0.407	0.1573	Yes
38	01614500	0207	MD	94	0.27	0.257	0.0683	Yes

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis		Used in regional skew model
						Skew coefficient (log units)	MSE of skew (log units)	
39	01617800	0207	MD	50	0.04	0.039	0.1055	Yes
40	01619500	0207	MD	86	0.10	0.096	0.0665	Yes
41	01637000	0207	MD	30	NA	0.405	0.2015	no-P
42	01638500	0207	MD	120	NA	0.316	0.0585	no-R
43	01639000	0207	MD	83	0.22	0.206	0.0736	Yes
44	01639500	0207	MD	66	1.18	1.078	0.1831	Yes
45	01640500	0207	MD	53	0.46	0.409	0.1256	Yes
46	01641000	0207	MD	43	−0.34	−0.295	0.1402	Yes
47	01641500	0207	MD	39	0.76	0.662	0.1891	Yes
48	01643500	0207	MD	65	1.21	1.106	0.1934	Yes
49	01386000	0203	NJ	61	NA	0.203	0.0966	no-O
50	01399525	0203	NJ	32	NA	−0.286	0.1802	no-P
51	01440000	0204	NJ	90	0.44	0.41	0.0808	Yes
52	01443500	0204	NJ	89	0.36	0.336	0.0773	Yes
53	01445000	0204	NJ	67	0.35	0.318	0.0972	Yes
54	01446000	0204	NJ	83	0.19	0.177	0.0729	Yes
55	01455200	0204	NJ	40	0.72	0.63	0.1757	Yes
56	01199477	0110	NY	38	0.90	0.774	0.2456	Yes
57	01312000	0202	NY	89	0.60	0.559	0.0911	Yes
58	01313500	0202	NY	41	0.19	0.166	0.1386	Yes
59	01314000	0202	NY	55	−0.25	−0.225	0.1115	Yes
60	01317000	0202	NY	104	−0.30	−0.286	0.0639	Yes
61	01318500	0202	NY	94	NA	−0.056	0.059	no-R
62	01319000	0202	NY	51	0.06	0.052	0.1029	Yes
63	01321000	0202	NY	102	0.05	0.045	0.0546	Yes
64	01328000	0202	NY	37	−0.30	−0.26	0.1567	Yes
65	01329154	0202	NY	35	NA	−0.986	0.2533	no-R
66	01329490	0202	NY	73	0.67	0.615	0.1158	Yes
67	01330000	0202	NY	44	−0.08	−0.07	0.1204	Yes
68	01330500	0202	NY	69	0.58	0.538	0.1099	Yes
69	01333500	0202	NY	73	0.50	0.463	0.0966	Yes
70	01334500	0202	NY	102	NA	0.824	0.1024	no-R
71	01342800	0202	NY	40	0.32	0.274	0.1418	Yes
72	01348000	0202	NY	61	0.89	0.809	0.1462	Yes
73	01348420	0202	NY	39	1.06	0.917	0.2219	Yes
74	01349000	0202	NY	42	0.58	0.504	0.1592	Yes
75	01350000	0202	NY	106	0.15	0.139	0.0568	Yes

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis: Skew coefficient (log units)	EMA analysis: MSE of skew (log units)	Used in regional skew model
76	01350120	0202	NY	38	0.63	0.541	0.1778	Yes
77	01350140	0202	NY	39	−0.21	−0.186	0.1438	Yes
78	01358500	0202	NY	59	1.47	1.336	0.2378	Yes
79	01359528	0202	NY	55	0.40	0.362	0.113	Yes
80	01359750	0202	NY	38	−0.19	−0.164	0.133	Yes
81	01361000	0202	NY	70	0.63	0.577	0.1117	Yes
82	01361500	0202	NY	107	−0.10	−0.091	0.0571	Yes
83	01362100	0202	NY	56	0.41	0.367	0.1168	Yes
84	01362197	0202	NY	33	NA	−0.071	0.157	no-P
85	01362200	0202	NY	69	NA	0.071	0.0784	no-R
86	01362500	0202	NY	96	0.08	0.078	0.0585	Yes
87	01365000	0202	NY	102	−0.29	−0.275	0.0631	Yes
88	01365500	0202	NY	65	0.04	0.033	0.0828	Yes
89	01368000	0202	NY	52	1.27	1.135	0.237	Yes
90	01368500	0202	NY	51	1.01	0.901	0.1785	Yes
91	01369000	0202	NY	40	0.56	0.488	0.1653	Yes
92	01369500	0202	NY	42	0.28	0.242	0.1378	Yes
93	01370000	0202	NY	54	NA	0.636	0.1453	no-R
94	01371000	0202	NY	64	−0.22	−0.2	0.0841	Yes
95	01371500	0202	NY	90	0.39	0.369	0.0785	Yes
96	01372500	0202	NY	85	0.71	0.66	0.102	Yes
97	01372800	0202	NY	44	0.04	0.031	0.1291	Yes
98	01374250	0203	NY	39	NA	0.323	0.1664	no-O
99	01387400	0203	NY	34	NA	0.756	0.2236	no-P
100	01387410	0203	NY	43	NA	0.095	0.1648	no-O
101	01387450	0203	NY	53	NA	0.442	0.1282	no-O
102	01413500	0204	NY	77	0.30	0.279	0.0838	Yes
103	01414000	0204	NY	39	0.49	0.424	0.1632	Yes
104	01414500	0204	NY	77	0.12	0.115	0.0745	Yes
105	01415000	0204	NY	77	0.13	0.118	0.0747	Yes
106	01418500	0204	NY	44	NA	0.103	0.1204	no-R
107	01419500	0204	NY	39	NA	0.443	0.1633	no-R
108	01420000	0204	NY	57	−0.07	−0.062	0.0948	Yes
109	01420500	0204	NY	100	0.30	0.281	0.0669	Yes
110	01421000	0204	NY	45	NA	−0.039	0.1152	no-R
111	01421900	0204	NY	57	0.08	0.071	0.0952	Yes
112	01422500	0204	NY	55	−0.24	−0.217	0.1072	Yes

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis		Used in regional skew model
						Skew coefficient (log units)	MSE of skew (log units)	
113	01423000	0204	NY	63	−0.08	−0.07	0.0869	Yes
114	01425500	0204	NY	34	NA	−0.004	0.1483	no-P
115	01426000	0204	NY	40	0.62	0.542	0.1726	Yes
116	01426500	0204	NY	59	0.13	0.118	0.0937	Yes
117	01427500	0204	NY	55	0.52	0.469	0.1266	Yes
118	01428000	0204	NY	44	1.12	0.984	0.1857	Yes
119	01434025	0204	NY	30	NA	0.661	0.2272	no-P
120	01435000	0204	NY	75	−0.02	−0.019	0.0716	Yes
121	01496500	0205	NY	62	−0.03	−0.029	0.0854	Yes
122	01497805	0205	NY	35	NA	0.961	0.247	no-P
123	01498500	0205	NY	38	−0.08	−0.066	0.1367	Yes
124	01500500	0205	NY	84	0.19	0.176	0.0712	Yes
125	01501000	0205	NY	55	−0.15	−0.135	0.1006	Yes
126	01501500	0205	NY	36	0.82	0.707	0.2023	Yes
127	01502000	0205	NY	58	−0.35	−0.314	0.1089	Yes
128	01502500	0205	NY	80	−0.20	−0.182	0.0756	Yes
129	01503000	0205	NY	114	NA	0.041	0.0489	no-R
130	01503980	0205	NY	49	−0.03	−0.024	0.1065	Yes
131	01505000	0205	NY	77	−0.09	−0.081	0.073	Yes
132	01505500	0205	NY	35	NA	−0.367	0.1699	no-P
133	01507000	0205	NY	77	0.30	0.276	0.0834	Yes
134	01507500	0205	NY	36	−0.27	−0.235	0.1585	Yes
135	01508000	0205	NY	36	−0.38	−0.325	0.166	Yes
136	01508500	0205	NY	38	0.20	0.171	0.146	Yes
137	01509000	0205	NY	76	−0.24	−0.224	0.0814	Yes
138	01510000	0205	NY	72	−0.13	−0.122	0.0799	Yes
139	01510500	0205	NY	43	NA	0.058	0.1208	no-R
140	01510610	0205	NY	37	0.17	0.148	0.1472	Yes
141	01513500	0205	NY	92	0.15	0.144	0.0638	Yes
142	01514000	0205	NY	87	0.39	0.361	0.0797	Yes
143	01520500	0205	NY	54	NA	0.459	0.1265	no-R
144	01521596	0205	NY	36	−1.65	−1.41	0.173	Yes
145	01522500	0205	NY	36	−0.23	−0.193	0.1535	Yes
146	01526000	0205	NY	36	−0.65	−0.555	0.1847	Yes
147	01526500	0205	NY	62	NA	0.456	0.1127	no-R
148	01527000	0205	NY	49	0.74	0.66	0.1552	Yes
149	01528000	0205	NY	59	0.53	0.481	0.1201	Yes

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis: Skew coefficient (log units)	EMA analysis: MSE of skew (log units)	Used in regional skew model
150	01530301	0205	NY	36	−0.40	−0.343	0.1654	Yes
151	01530500	0205	NY	54	−0.24	−0.212	0.1094	Yes
152	01531000	0205	NY	78	0.16	0.144	0.0753	Yes
153	0142400103	0204	NY	34	NA	0.344	0.1754	no-P
154	01428750	0204	PA	39	1.13	0.983	0.2346	Yes
155	01431000	0204	PA	43	0.30	0.266	0.1383	Yes
156	01438300	0204	PA	52	0.02	0.015	0.1217	Yes
157	01439500	0204	PA	105	1.02	0.964	0.1161	Yes
158	01440400	0204	PA	56	−0.04	−0.037	0.0948	Yes
159	01447500	0204	PA	70	0.70	0.645	0.117	Yes
160	01448500	0204	PA	52	0.14	0.128	0.1074	Yes
161	01449360	0204	PA	47	−1.50	−1.334	0.3029	Yes
162	01450500	0204	PA	74	0.22	0.201	0.0821	Yes
163	01451800	0204	PA	47	−0.12	−0.105	0.1205	Yes
164	01453000	0204	PA	55	0.50	0.455	0.1254	Yes
165	01459500	0204	PA	37	0.21	0.182	0.1504	Yes
166	01470500	0204	PA	68	−0.05	−0.046	0.0798	Yes
167	01470779	0204	PA	40	−0.14	−0.126	0.1353	Yes
168	01472000	0204	PA	57	0.22	0.199	0.1025	Yes
169	01472157	0204	PA	45	0.22	0.192	0.127	Yes
170	01473000	0204	PA	99	−0.61	−0.571	0.0847	Yes
171	01475850	0204	PA	37	0.29	0.25	0.1559	Yes
172	01480300	0204	PA	54	−0.05	−0.041	0.0983	Yes
173	01480675	0204	PA	47	−0.12	−0.103	0.116	Yes
174	01516350	0205	PA	47	0.33	0.296	0.1255	Yes
175	01516500	0205	PA	59	0.63	0.571	0.1275	Yes
176	01532000	0205	PA	100	0.52	0.486	0.0786	Yes
177	01534000	0205	PA	100	−0.33	−0.308	0.0683	Yes
178	01538000	0205	PA	94	0.22	0.21	0.0668	Yes
179	01539000	0205	PA	75	0.43	0.4	0.0928	Yes
180	01541000	0205	PA	101	0.73	0.687	0.0911	Yes
181	01541500	0205	PA	47	0.25	0.219	0.124	Yes
182	01542810	0205	PA	50	0.77	0.69	0.1561	Yes
183	01543000	0205	PA	100	0.38	0.354	0.0709	Yes
184	01543500	0205	PA	75	NA	0.398	0.0927	no-R
185	01544500	0205	PA	73	0.66	0.611	0.1103	Yes
186	01545600	0205	PA	49	0.82	0.729	0.1623	Yes

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis		Used in regional skew model
						Skew coefficient (log units)	MSE of skew (log units)	
187	01547700	0205	PA	58	0.71	0.64	0.1351	Yes
188	01547950	0205	PA	46	0.28	0.25	0.1288	Yes
189	01548005	0205	PA	60	−0.15	−0.138	0.0949	Yes
190	01548500	0205	PA	95	0.53	0.503	0.083	Yes
191	01549500	0205	PA	74	0.20	0.184	0.0806	Yes
192	01550000	0205	PA	100	0.37	0.345	0.0703	Yes
193	01552000	0205	PA	88	0.47	0.438	0.0841	Yes
194	01552500	0205	PA	74	0.09	0.083	0.0755	Yes
195	01553700	0205	PA	34	NA	−0.076	0.1538	no-P
196	01555000	0205	PA	84	0.45	0.418	0.0861	Yes
197	01555500	0205	PA	84	1.14	1.061	0.1529	Yes
198	01556000	0205	PA	98	0.41	0.391	0.0737	Yes
199	01557500	0205	PA	72	0.08	0.074	0.0769	Yes
200	01558000	0205	PA	75	1.10	1.018	0.1568	Yes
201	01559000	0205	PA	112	NA	0.586	0.0782	no-R
202	01560000	0205	PA	75	0.45	0.417	0.0948	Yes
203	01562000	0205	PA	103	NA	0.244	0.0631	no-R
204	01564500	0205	PA	83	0.32	0.298	0.0782	Yes
205	01565000	0205	PA	60	0.53	0.486	0.1187	Yes
206	01566000	0205	PA	79	0.17	0.157	0.074	Yes
207	01567500	0205	PA	60	0.81	0.732	0.1397	Yes
208	01568000	0205	PA	85	0.40	0.369	0.0825	Yes
209	01570000	0205	PA	81	0.51	0.473	0.094	Yes
210	01571500	0205	PA	69	1.03	0.946	0.1617	Yes
211	01573000	0205	PA	95	1.32	1.241	0.1842	Yes
212	01574000	0205	PA	85	1.66	1.553	0.2721	Yes
213	01574500	0205	PA	36	0.90	0.769	0.2089	Yes
214	01575000	0205	PA	43	1.02	0.891	0.1964	Yes
215	01576500	0205	PA	85	0.31	0.29	0.0776	Yes
216	01601000	0207	PA	47	NA	0.14	0.1185	no-R
217	01603500	0207	PA	36	0.67	0.572	0.4171	Yes
218	01613050	0207	PA	51	0.60	0.538	0.1403	Yes
219	01613900	0207	VA	53	−0.05	−0.042	0.1001	Yes
220	01615000	0207	VA	65	0.09	0.086	0.0849	Yes
221	01634500	0207	VA	78	0.03	0.025	0.0691	Yes
222	01638480	0207	VA	43	−0.35	−0.304	0.1409	Yes
223	01142500	0108	VT	152	NA	0.568	0.0517	no-O

Table 1. Streamgages that were considered for use in the regional skew analysis for parts of eastern New York and Pennsylvania.—Continued

[Locality abbreviations: CT, Connecticut; MA, Massachusetts; MD, Maryland; NJ, New Jersey; NY, New York; PA, Pennsylvania; VA, Virginia; VT, Vermont; WV, West Virginia. Other abbreviations: USGS, U.S. Geological Survey; EMA, expected moments algorithm; MSE, mean-square error; NA, not applicable because the streamgage was not used in the regional skew analysis; no-O, streamgage not used in regional skew analysis due to 4-digit hydrologic unit code outside the study area; no-P, streamgage not used in regional skew analysis due to Psuedo record length less than 36 years; no-R, streamgage not used in regional skew analysis due to redundancy]

Regional skew index number	USGS streamgage number	Hydrologic Unit Code (4-digit)	State where streamgage is located	Pseudo record length (P_{RL}, years)	Unbiased station skew (log units)	EMA analysis: Skew coefficient (log units)	EMA analysis: MSE of skew (log units)	Used in regional skew model
224	01150900	0108	VT	29	NA	1.491	0.5984	no-O
225	01155350	0108	VT	30	NA	0.714	0.233	no-O
226	01329000	0202	VT	59	NA	0.458	0.117	no-R
227	01334000	0202	VT	82	0.07	0.066	0.0679	Yes
228	01608500	0207	WV	97	0.61	0.578	0.0877	Yes
229	01611500	0207	WV	99	−0.06	−0.057	0.056	Yes
230	01614000	0207	WV	81	0.20	0.185	0.0718	Yes
231	01616500	0207	WV	71	0.14	0.126	0.0803	Yes
232	01636500	0207	WV	110	0.13	0.125	0.0539	Yes

Appendix 1. Assessment of a Regional Skew Model for Parts of Eastern New York and Pennsylvania by Using Monte Carlo Simulations

This appendix provides a graphical assessment of the Bayesian weighted least squares/Bayesian generalized least squares (B–WLS/B–GLS) model of regional skew for parts of eastern New York and Pennsylvania that is described in this report. Observed, unbiased station skews are depicted in figure 1.1 along with contour lines and shading to provide a sense of geographic patterns in the skews.

Monte Carlo simulations were used to determine whether geographic patterns observed in the station skews are evidence of significant model misspecification or an artifact of random-sampling variability that is possibly confounded by the covariance structure of the errors. The Monte Carlo simulations were generated from a multivariate normal distribution with a mean equal to the constant from the regional skew model and a covariance matrix identical to the covariance matrix used in the regional skew model. The constant model of skew in the study area is:

$$\hat{\gamma}_{B-WLS/B-GLS} = 0.32 + \varepsilon\ , \qquad (1.1)$$

where

ε represents the total error, and

$$\varepsilon \sim N\left(0, Var\left(\varepsilon\right)\right), \qquad (1.2)$$

where

N signifies a normal distribution of the total error in the constant regional skew model determined in the B–GLS analysis.

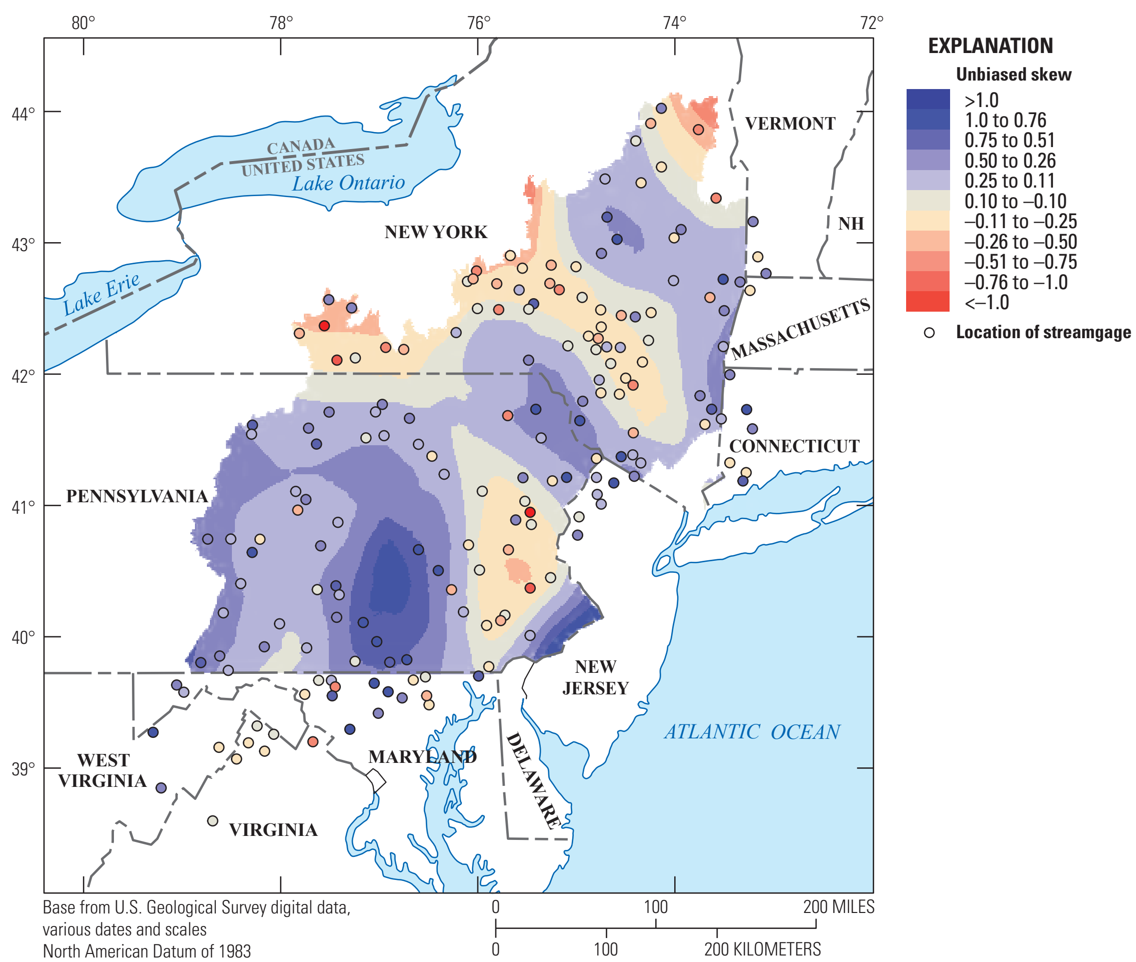


Figure 1.1. Contour map of unbiased station skews for the 183 streamgages that were used in the regional skew analysis for parts of eastern New York and Pennsylvania.

As described in equation 1.2, the $Var(\varepsilon)$ can be described as

$$\left[\varepsilon\varepsilon^{T}\right] = \boldsymbol{\Lambda}_{B-GLS}\left(\sigma^{2}_{\delta,B-GLS}\right) = \sigma^{2}_{\delta,B-GLS}\boldsymbol{I} + \boldsymbol{\Sigma}\left(\hat{\gamma}\right), \quad (1.3)$$

where

ε^{T} is the transformation of ε,

$\boldsymbol{\Lambda}_{B-GLS}\left(\sigma^{2}_{\delta,B-GLS}\right)$ is the $(n \times n)$ B–GLS covariance matrix,

$\sigma^{2}_{\delta,B-GLS}$ is the B–GLS variance of the underlying model error δ,

$\boldsymbol{I}$ is an $(n \times n)$ identity matrix, and

$\boldsymbol{\Sigma}\left(\hat{\gamma}\right)$ is the full $(n \times n)$ covariance matrix of the sampling errors for each streamgage (n).

The covariance matrix of the sampling errors is made up of the sampling variances of the unbiased station skew ($Var\left[\hat{\gamma}_i\right]$) and the covariances of the skewness estimators ($\hat{\gamma}_i$). The off-diagonal values of $\boldsymbol{\Sigma}\left(\hat{\gamma}\right)$ are determined by the cross correlation of concurrent gaged annual peak flows and the *cf* factor (see equation 3 in Martins and Stedinger, 2002). The model error variance σ^{2}_{δ} for the constant model is 0.10 (table 3) and was used in the Monte Carlo simulations. The covariance matrix $\boldsymbol{\Sigma}\left(\hat{\gamma}\right)$ used in the Monte Carlo simulations is the same as that used in the B–WLS/B–GLS regression analysis (see eqs. 18 and 19 in the report).

The results of the Monte Carlo simulations are depicted graphically in 20 realizations (fig. 1.2) of the expected patterns in the station skew if the station skews are normally distributed with a mean equal to 0.32 and a covariance matrix given by equation 1.3. The Monte Carlo simulations reveal no structure in the pattern of the station skews that is consistent with the observed pattern of the station skews in the constant model (fig. 1.1). Therefore, it seems reasonably safe to conclude that, despite the geographic patterns observed in the station skews, there is little evidence of a lack of fit.

Reference Cited

Martins, E.S., and Stedinger, J.R., 2002, Cross correlations among estimators of shape: Water Resources Research, v. 38, no. 11, p. 34-1 to 34-7. [Also available at https://doi.org/10.1029/2002WR001589.]

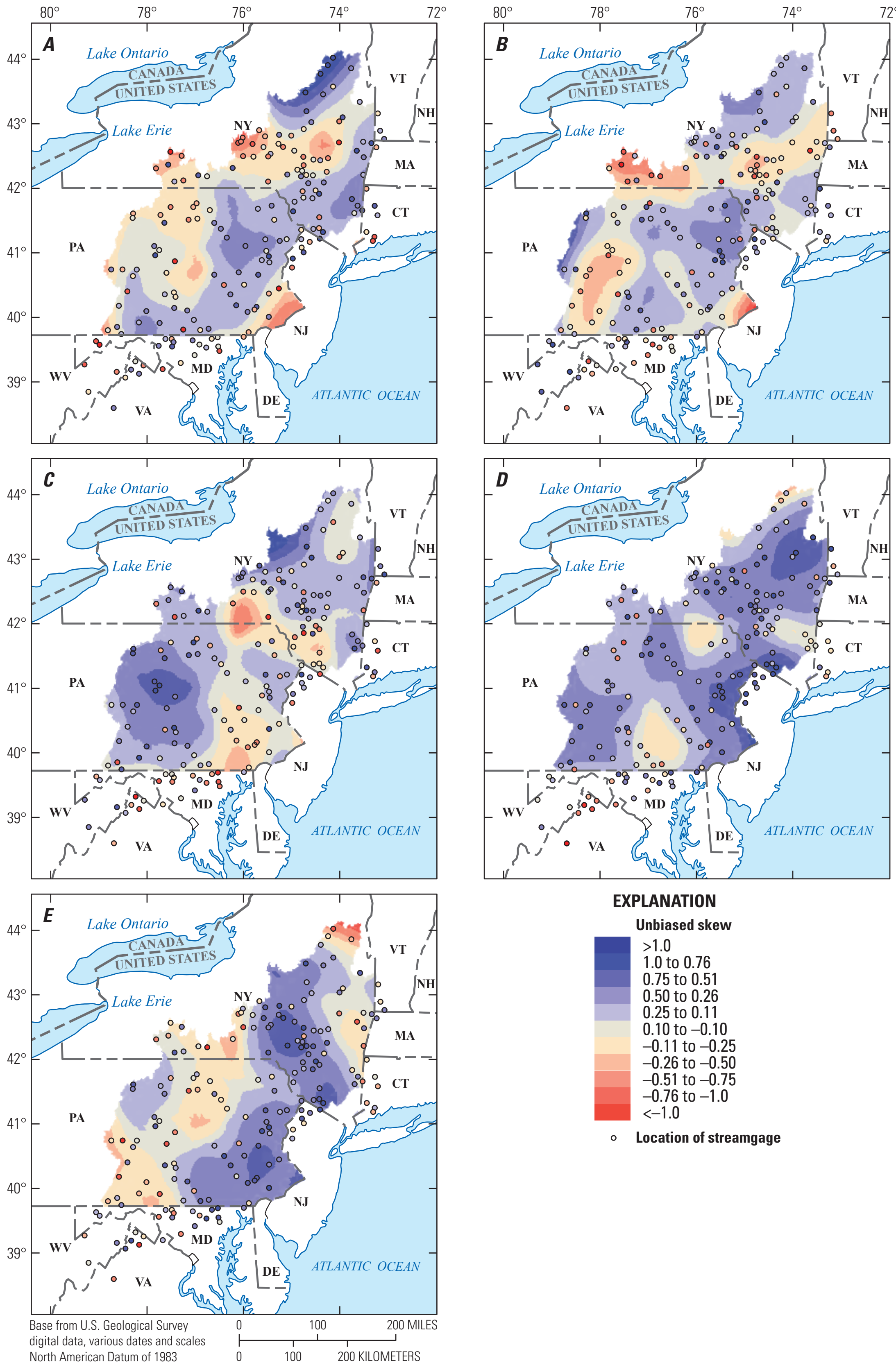


Figure 1.2. Contour maps showing the results of 20 Monte Carlo simulations (*A*–*T*) of unbiased station skew at 183 streamgages that were used in the regional skew analysis for parts of eastern New York and Pennsylvania. Simulations are normally distributed to the constant skew model and covariance matrix.

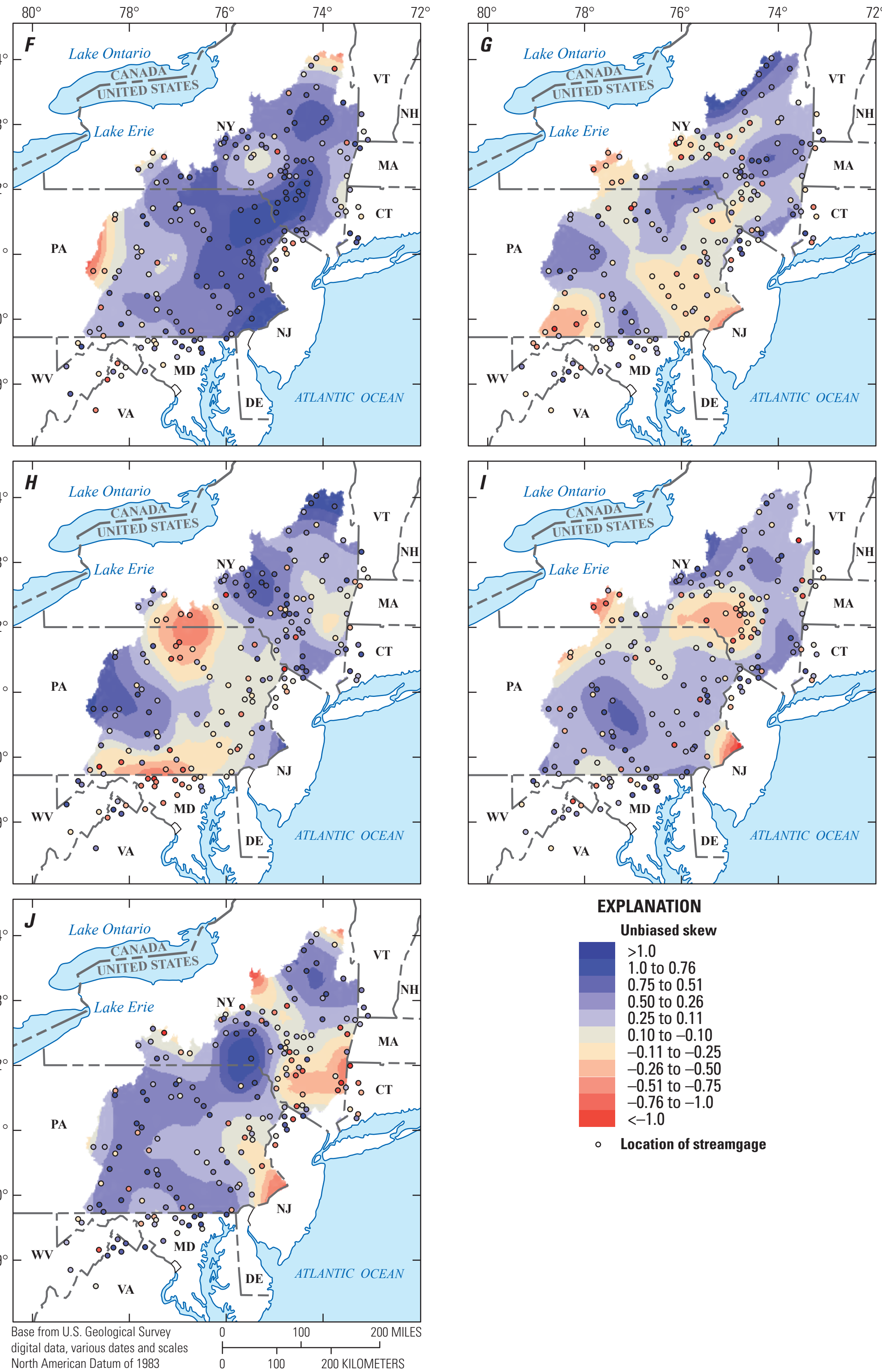


Figure 1.2.—Continued

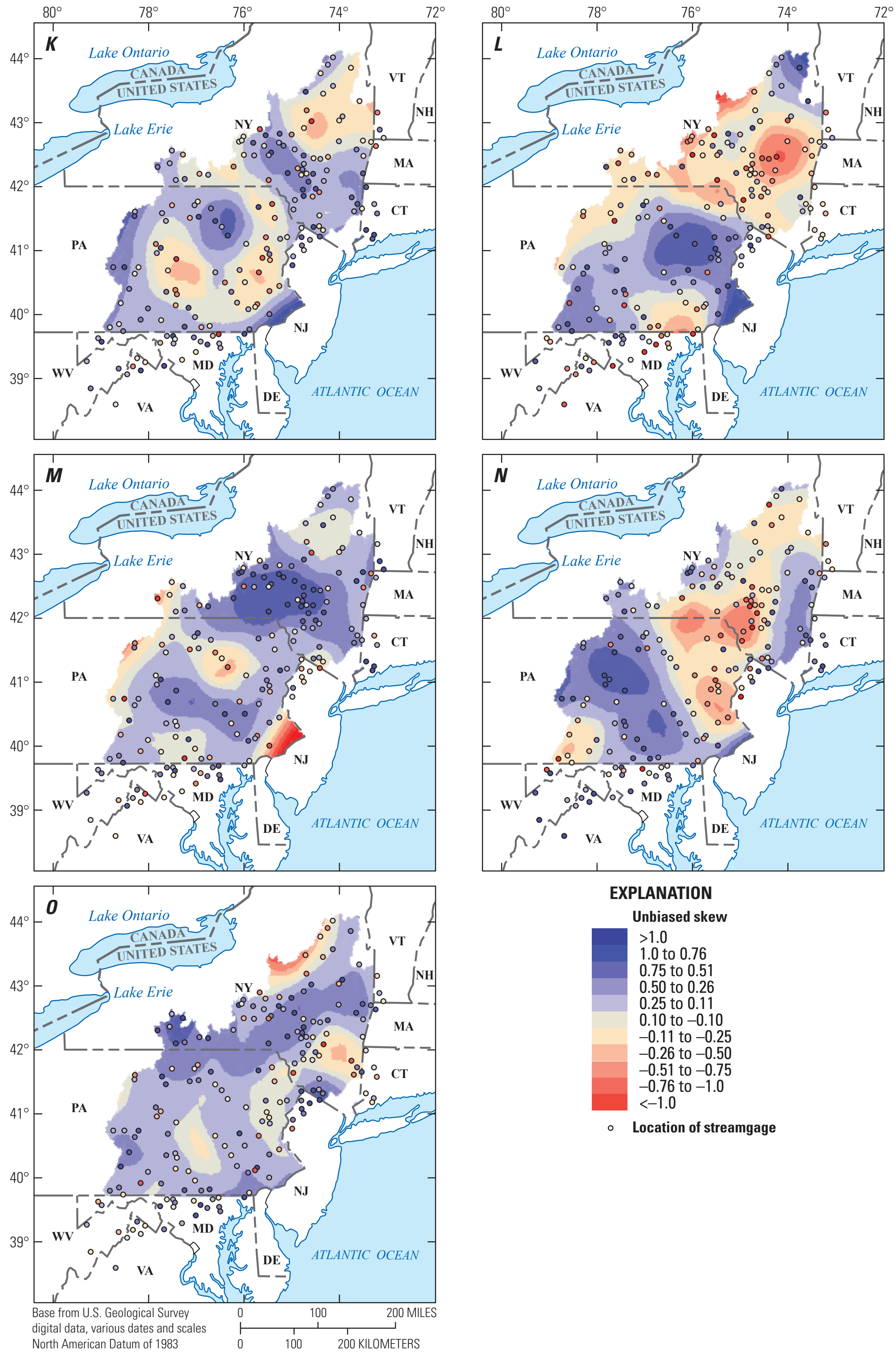


Figure 1.2.—Continued

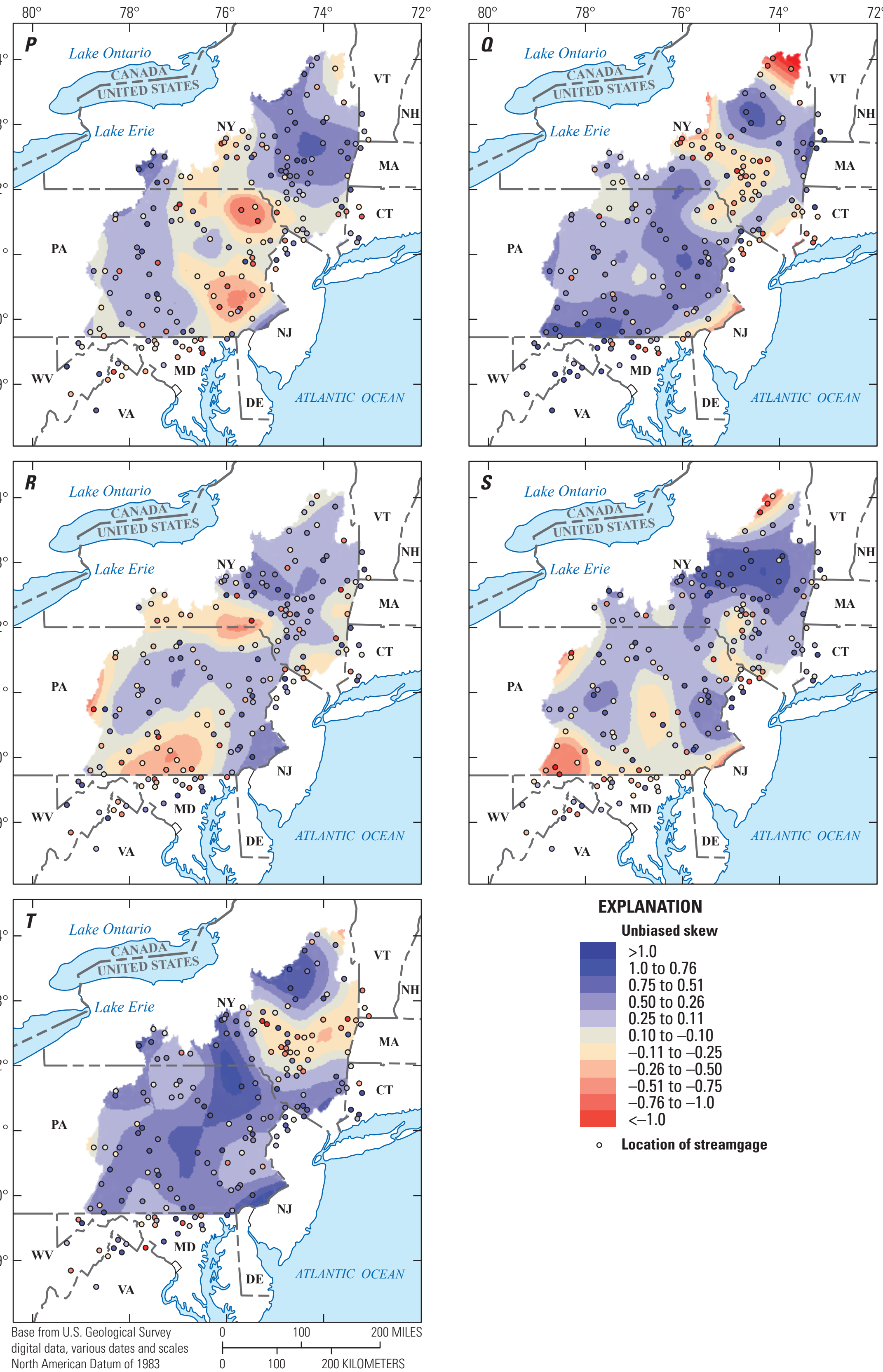


Figure 1.2.—Continued

Veilleux and Wagner—**Methods for Estimating Regional Skewness of Annual Peak Flows in Parts of Eastern New York and Pennsylvania**—SIR 2021–5015

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20215015