USGS
science for a changing world

Development of a Contour Map Showing Generalized Skew Coefficients of Annual Peak Discharges of Rural, Unregulated Streams in New York, Excluding Long Island

U.S. GEOLOGICAL SURVEY
Water-Resources Investigations Report 00-4022

Prepared in cooperation with the
NEW YORK STATE DEPARTMENT OF TRANSPORTATION

Development of a Contour Map Showing Generalized Skew Coefficients of Annual Peak Discharges of Rural, Unregulated Streams in New York, Excluding Long Island

By Richard Lumia and Yvonne H. Baevsky

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 00-4022

Prepared in cooperation with the

NEW YORK STATE DEPARTMENT OF TRANSPORTATION

Troy, New York
2000

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For additional information write to:

District Chief
U.S. Geological Survey
425 Jordan Road
Troy, New York 12180-8349

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of nformation Services
Box 25286
Federal Center
Denver, CO 80225-0286

CONTENTS

FIGURES

TABLES

Development of a Contour Map Showing Generalized Skew Coefficients of Annual Peak Discharges of Rural, Unregulated Streams in New York, Excluding Long Island

By Richard Lumia and Yvonne H. Baevsky

ABSTRACT

Flood-frequency relations that are developed by fitting the logarithms of annual peak discharges to a Pearson Type-III distribution are sensitive to skew coefficients. Estimates of population skew for a site are improved when computed from the weighted average of (1) the sample (station) skew, and (2) an unbiased, generalized skew estimate. A weighting technique based on the number of years of record at each of 226 sites was used to develop a contour map of unbiased, generalized skew coefficients for New York. An attempt was made to group (regionalize) the station skew coefficients into five hydrologically similar areas of New York, but the statewide version proved to be as accurate as the regionalized version and therefore was adopted as the final generalized skew-coefficient map for New York. An error analysis showed the statewide contour map to have lower MSE's (mean square errors) than those computed from (1) the five regional skew-coefficient contour maps, (2) a previously used (1982) nationwide skew coefficient map, and (3) the weighted mean of skew coefficients for sites within each of five hydrologically uniform, but distinct areas of New York.

INTRODUCTION

The effective management of flood-prone areas and the design of structures along rivers and streams requires knowledge of the magnitude and frequency of floods. Discharge-frequency relations for streamflow-gaging stations on rural, unregulated streams in New York were developed by Lumia (1991), who fitted the logarithms of the annual peak discharges to a Pearson Type-III distribution according to guidelines recommended by the Interagency Advisory Committee on Water Data (IACWD) (1982), and updated these frequency curves by including data collected through 1996. The Pearson Type-III distribution requires computation of the mean, the standard deviation, and the skew of the station's annual peak discharges.

The skew coefficient of a station's peak-discharge record is sensitive to extreme values; therefore, accurate values are difficult to obtain for sites with short records. The accuracy of a station's skew coefficient can be improved by weighting that station skew coefficient with a "generalized" skew value that represents pooled skew-coefficient data from nearby stations with long records; generalized skew coefficients can be estimated through regression, mapping, or averaging methods (IACWD, 1982). A nationwide map by the IACWD (1982) provides skew-coefficient contours that represent generalized estimates, but the IAWCD suggests that separate, regional skew analyses be made to obtain more accurate values for local flood-frequency analyses.

In 1998, the U.S. Geological Survey (USGS), in cooperation with New York State Department of Transportation (NYSDOT), began a study to develop a generalized skew-coefficient map of New York, excluding Long Island. Long Island streams were excluded from this study primarily because of the varying degrees of development and urbanization within the gaged basins. An attempt was made to regionalize skew coefficients into five hydrologically uniform, but distinct areas of New York. The resulting five contour maps provided little or no improvement over the statewide contour map, however; therefore the statewide map (excluding Long Island) was adopted as the final source for determining generalized skew coefficients. The statewide map is based on skew coefficients from 226 rural, unregulated streamflow-gaging stations with at least 20 years of annual peak discharges; 194 of these sites have at least 25 years of record.

An error analysis showed the statewide map to have a lower MSE (mean square error, an estimate of sampling variance) than either (1) the IACWD's nationwide map of 1982, (2) the five regional maps, or (3) the weighted mean skew coefficients for each of the five regions. The MSE from the statewide generalized skew-coefficient map can be used in future flood-frequency computations.

Purpose and Scope

This report (1) describes the development of the statewide generalized skew coefficient map, (2) presents methods used, and (3) summarizes error analyses. This report supersedes previous publications that provide techniques or maps for estimation of generalized skew coefficients on rural, unregulated streams in New York.

Acknowledgments

Most of the information presented here is based on data collected through the support of Hudson River-Black River Regulating District, U.S. Army Corps of Engineers, New York State Department of Environmental Conservation, New York State Department of Transportation, New York City Department of Environmental Protection, New York Power Authority, New York State Canal Corporation, Niagara Mohawk Power Corporation, and several other local agencies and private organizations.

STATION SKEW COEFFICIENTS

A discharge-frequency relation for a streamflow-gaging station is developed by fitting the logarithms of the station's annual peak discharges to a Pearson Type-III distribution according to guidelines recommended by IACWD (1982). In this method, an assumption is made that the sample (the recorded values of annual peak discharge at the station) is representative of the population (all recorded and unrecorded annual peak discharges at that station). The Pearson Type-III distribution requires values for the population's mean, standard deviation, and skew coefficient. The population values are estimated from computed station (sample) values. The station skew coefficient is computed as follows:

$$G_s = \frac{N}{(N-1)(N-2)S^3} \sum_{i=1}^{N} (x_i - \bar{x})^3, \qquad (1)$$

where:

G_s = station's skew coefficient;
x_i = station's log-transformed annual peak discharge for year *i*;
$\bar{x}$ = station's log-transformed mean of annual peak discharges;
S = station's log-transformed standard deviation of annual peak discharges; and
N = station's number of years of peak-discharge record.

Several studies have shown that the station's skew coefficient is a biased estimator of the population's skew coefficient. A bias-correction equation based on record length (years) is presented by Tasker and Stedinger (1986) as:

$$C_b = \left(1 + \frac{6}{N}\right), \qquad (2)$$

where:

C_b = station's bias-correction factor; and
N = station's number of years of peak-discharge record.

The station (sample) skew coefficient for each of the 226 streamflow-gaging stations used in this study was multiplied by the bias-correction factor to obtain an unbiased value.

The accuracy of the estimated population skew coefficient at a station with a short record of annual peaks can be improved by weighting the computed station (sample) skew value with a "generalized" skew value that represents pooled skew-coefficient data from nearby stations with long records (IACWD, 1982). The MSE (mean-square error) of the resulting estimate is minimized by weighting the station's skew coefficient and generalized skew in inverse proportion to their individual MSE's. The following equation (Tasker, 1978) is used to compute a station's weighted skew coefficient (estimate of the population skew coefficient):

$$G_w = \frac{MSE_g(G_s) + MSE_s(G_g)}{MSE_g + MSE_s}, \quad (3)$$

where:

G_w = station's weighted skew coefficient;
G_s = station's skew coefficient;
G_g = station's unbiased generalized skew coefficient;
MSE_g = mean-square error of the unbiased generalized skew coefficient; and
MSE_s = mean-square error of the station's skew coefficient.

GENERALIZED SKEW COEFFICIENTS

Generalized skew coefficients are normally estimated from the unbiased skew coefficients for nearby stations through regression, mapping, or averaging methods. The IACWD's nationwide skew-coefficient contour map (1982) provides generalized skew coefficients, but their limited accuracy, and recent improvements in estimating methods, warranted more detailed (statewide or local) skew analyses.

The estimating techniques referred to above assume that the skew coefficients for each gaging station have equal accuracy (uniform sampling variance), but previous studies have shown the variance of station skew coefficients (V_s) to vary with record length (N). The skew-coefficient-mapping procedure presented herein, upon which the statewide generalized skew coefficient contour map is based, uses a weighting method to account for nonuniform variance of the station skew coefficients.

Weighting of Skew Coefficients

Several investigators have developed equations to estimate the variance of station skew coefficients. Landers and Wilson (1991) used the parametric methods of Fisher (1931) and Tasker and Stedinger (1986), corrected for bias and defined as:

$$V_s = \frac{6N(N-1)[1+(6/N)]^2}{(N-2)(N+1)(N+3)}, \quad (4)$$

where N is as defined previously.

A station's skew coefficient is weighted in inverse proportion to the estimated station variance (V_s); therefore, the weight given to each station's skew coefficient is:

$$W = 1/V_s, \quad (5)$$

where W is the weight given to the station's unbiased skew coefficient, and V_s is as defined previously. These weights were applied to the station skew coefficients that were used to develop the statewide generalized skew coefficient contour map presented later in this report.

Mapping of Skew Coefficients

The tendency of skew coefficients to vary spatially, as shown in previous studies, indicates that contour mapping and, possibly, grouping sites to provide regional values, could improve the accuracy of values given on the IACWD nationwide map (1982). Software is available for computerized mapping of skew coefficients and fitting contours through the data points. A New York State map with an equally spaced grid was plotted along with station skew coefficients for each study site, and GIS (geographic information systems) software was used to compute an unbiased skew coefficient for each node of the grid. A spatial "neighborhood" search of nearby stations was made to select a subset of skew coefficients to compute a skew coefficient for each node. As suggested by Landers and Wilson (1991), the node value was computed as a distance-weighted mean of the subset values, where the weights are based on distance from the grid node. The skew coefficient for each grid node was calculated from the following equation (Landers and Wilson, 1991), which also accounts for nonuniform variances and station distances from the grid node in question:

$$Z_i = \frac{\sum_{j=1}^{n_i} G_{s_j}(W_j)(1/d_j)}{\sum_{j=1}^{n_i} (W_j)(1/d_j)} \quad (6)$$

where:

Z_i = estimated skew coefficient at grid node i;
G_{s_j} = unbiased skew coefficient of station j;
n = number of stations selected to estimate Z_i;
d_j = distance from the grid node to centroid of drainage basin whose records define G_{s_j}; and
W_j = weight given to G_s at station j, as calculated from equation 5.

The station skew coefficients for the 226 selected stations were computed by using equation 1 and corrected for bias by using equation 2. An unbiased skew coefficient was then computed for each grid node by the weighting procedure of equation 6, and lines of equal skew coefficient were fitted through the grid-node coefficients by an automated GIS technique. Some manual smoothing of the contours was done to eliminate unrealistic irregularities in the electronic mapping. The final, generalized unbiased-skew-coefficient map is given in figure 1; locations of gaging stations used to develop this map are included. The attempt to provide refined skew coefficients through regionalization of the values, and an analysis of map errors, are given in the following sections.

Regionalization

Lumia (1991) separated New York into eight hydrologic regions to improve flood-frequency estimations. In the present study, the unbiased skew coefficients for 226 stations were plotted on maps of these eight regions to indicate which regions had somewhat uniform skew coefficients. Results indicated that a few of the regions could be combined; the result was five regions (fig. 2) with fairly uniform skew values. Skew coefficients were compared through statistical analyses to test for equality of mean skew values among regions.

Although data for several gaging stations in neighboring states were used in this study, the regional boundaries extend only to the New York border. Some statistics describing the stations' unbiased skew coefficients within each region, and in all five regions combined, are given in table 1 and plotted in figure 3. Differences in skew coefficients among the hydrologic regions are apparent in table 1 and figure 3 and were evaluated through several statistical tests

The statistical analyses indicated that the unbiased skew coefficients within each of the five regions were normally distributed, and standard parametric tests (such as ANOVA F-test) indicated that the mean skew values for each region were not equal. Multiple comparison tests were then used to identify which regions' means differed statistically from the others. Because the five sample sizes were unequal, ranging from 19 in Region E to 65 in Region B (table 1), simultaneous inference methods (SIM), including Fisher's Least Significant Difference (LSD) test (t-tests) and Tukey's multiple comparison test, were used (Helsel and Hirsch, 1992). Both tests yielded the following comparisons of means of unbiased skew coefficients among the five hydrologic regions:

For adjacent regions: $A < B > C < D > E < A > C = E$

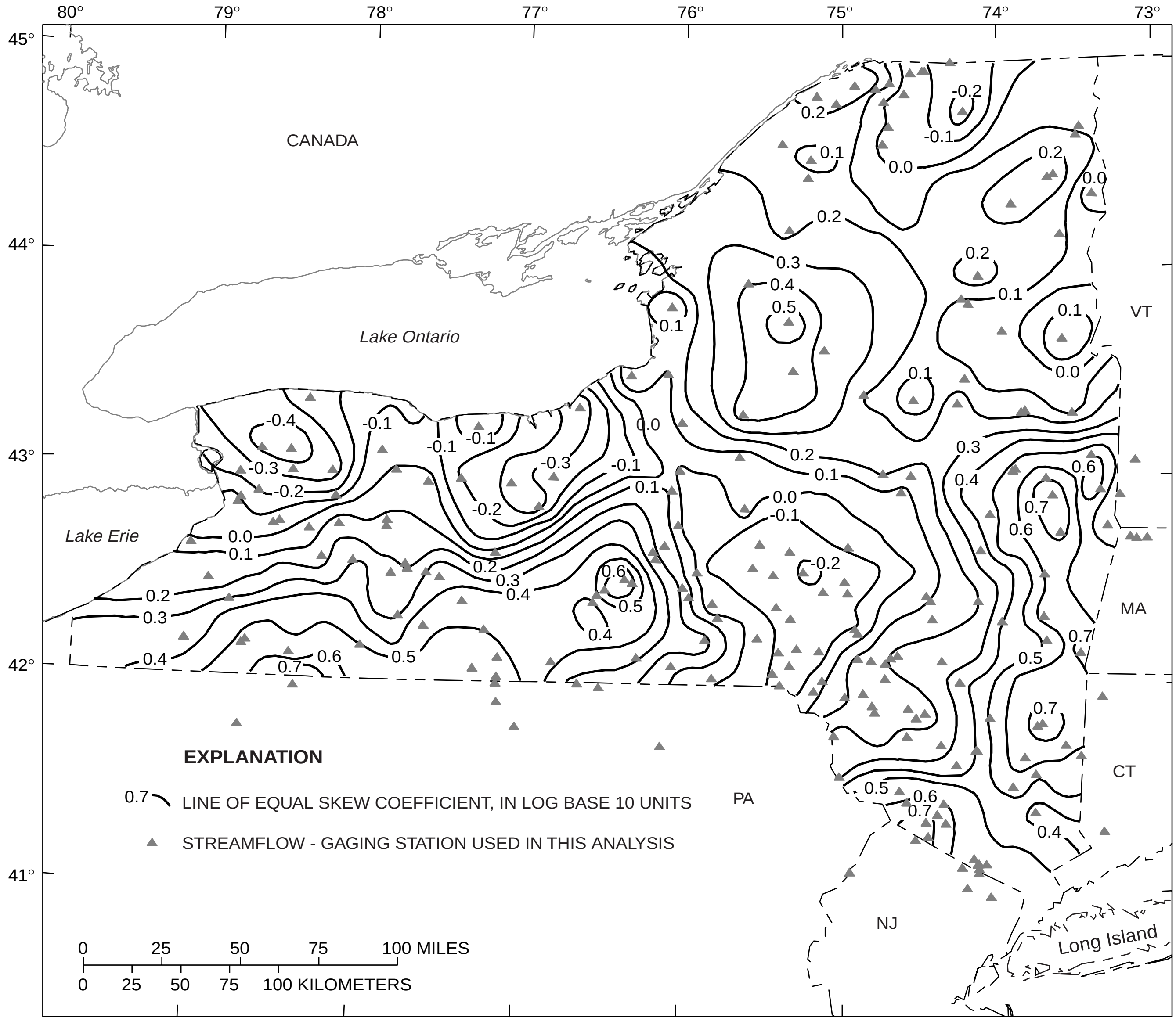
For nonadjacent regions: $A = D = B > E$

Only two adjacent regions (C and E, see fig. 2) had mean skew coefficients that did not differ statistically from each other, but other factors, such as topography, geology, and climate, indicate that delineation of these areas as separate hydrologic regions is justified.

Results of the statistical analyses and comparisons indicated that an evaluation of the unbiased skew coefficients within each of the five regions was warranted. A separate contour map of skew coefficients was developed for each region by the methods used for the statewide contour map. Results showed significant differences in skew contours along regional boundaries. The regional boundaries generally correspond to major basin boundaries, but no hydrologic reasons for the sharp differences along these boundaries were noted. An analysis of map errors (discussed in the next section) showed that the generalized skew coefficients on the contour maps of each region were no more accurate than those on the statewide contour map; therefore, the statewide map was adopted as the source of generalized skew coefficients for streamflow-gaging stations in New York, excluding Long Island. Although the five regional maps were not ultimately used for estimation of generalized skew coefficients, the five hydrologic areas were used to help evaluate the skew coefficients and their associated map errors, as discussed below.

Error Analysis

One of the methods suggested by IACWD (1982) for estimating generalized skew coefficients for annual peak discharges is to develop a prediction equation that relates skew coefficients to predictor or explanatory variables. An attempt was made to develop a multiple regression equation for each of the five hydrologic regions of New York and for all five regions combined by relating skew coefficients to topographic and climatic variables such as drainage area, slope, mean annual precipitation, and basin storage. No multiple regression equations with statistically significant explanatory variables could be developed, however; therefore, this method was not used in error analyses. Instead, statistical error analyses were used to evaluate other methods to identify which one would provide the most accurate



Base from U.S. Geological Survey digital data, 1:2,000,000, 1972

Figure 1. Generalized skew coefficients of New York, excluding Long Island

generalized skew coefficients. The methods evaluated were (1) the IACWD nationwide skew-coefficient map (1982), (2) the five regional skew maps prepared in the present study, (3) the statewide skew map prepared in this study, and (4) the weighted mean skew coefficients for each of the five hydrologic regions of New York. In method 4, weighted mean skew coefficients for each region were calculated from the unbiased skew coefficients for each station. The weighting factor was the number of years of annual peak-discharge record at each station, divided by the average number of recorded annual peak discharges at all stations within the region. The arithmetic unweighted means of the unbiased skew coefficients for stations within each region also were evaluated; the results are included in figure 4, which shows the mean generalized skew coefficients for each of the five regions. The highest mean is in region B (southeastern

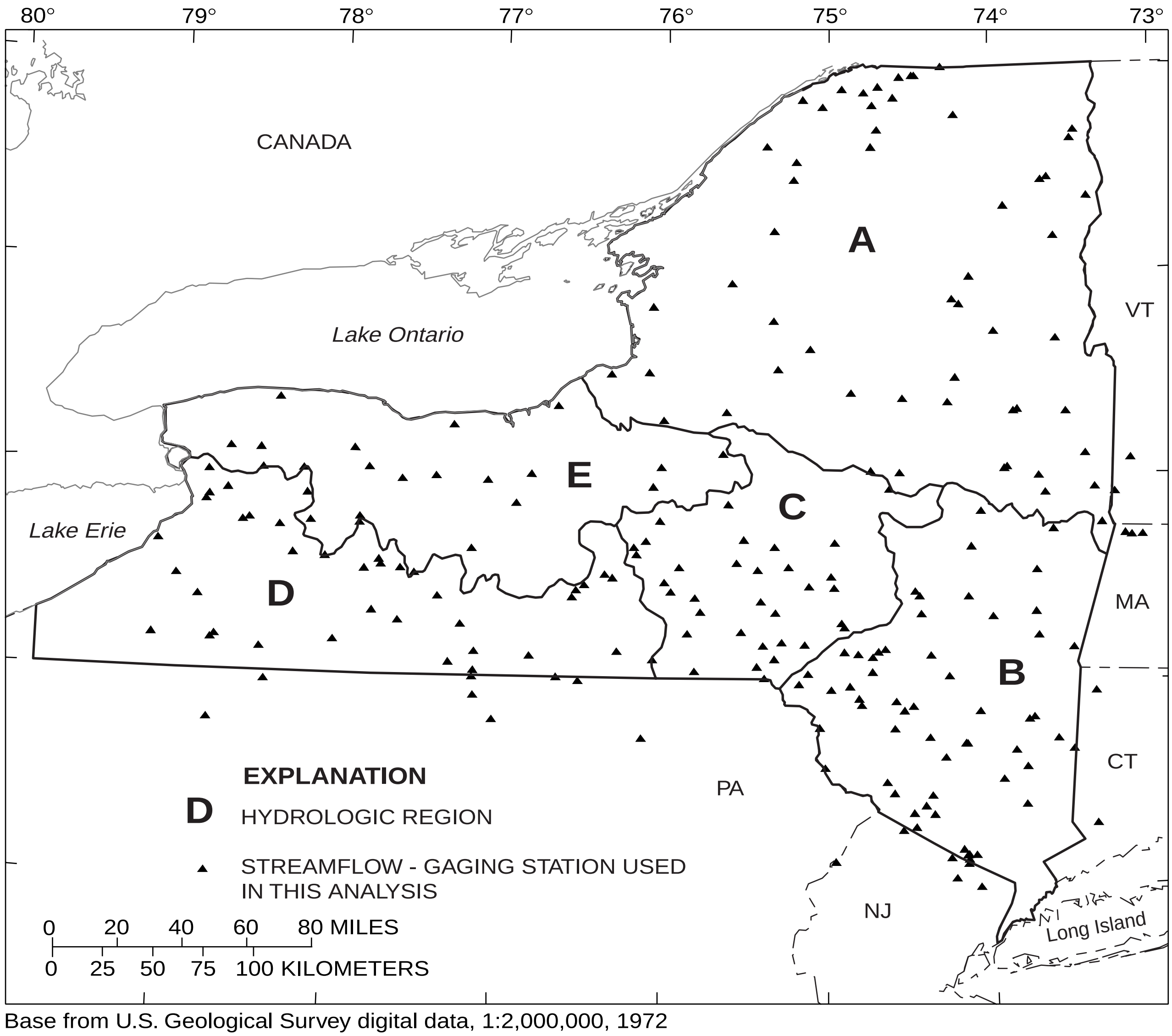


Figure 2. Five hydrologic regions of New York, excluding Long Island

New York), and the lowest in region E (western New York, south of Lake Ontario).

The MSE's (mean square errors) for each of the four methods of predicting generalized skew coefficients were computed within each of the five hydrologic regions and for all regions combined and are based on observed and predicted skew coefficients at gaging stations. Results are plotted in figure 5. Generally, the largest MSE's resulted from the first method (the IACWD nationwide map of 1982), and the second largest MSE's resulted from the fourth method (the weighted mean skews for each region). The smallest MSE's resulted from the third method (the statewide contour map) (fig. 1); and the second-smallest MSE's resulted from the second method (the five regional maps). The observed station skew

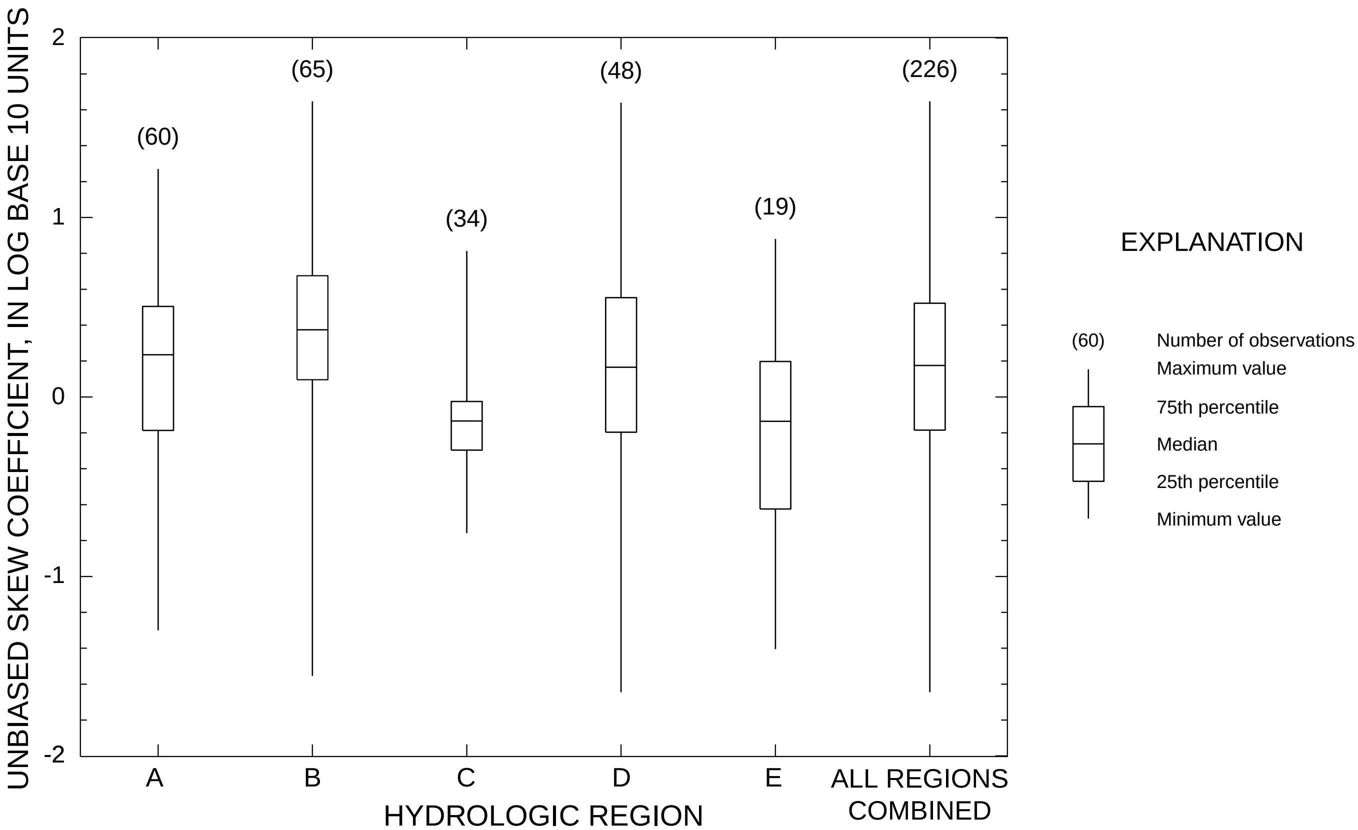


Figure 3. Unbiased skew coefficients at stations within five hydrologic regions of New York and in all regions combined. (Region locations are shown in fig. 2.)

Table 1. Selected statistics for unbiased skew coefficients for each of five hydrologic regions of New York, and for all five regions combined.

[Statistics based on annual peak-discharge data through 1996. Region locations are shown in fig. 2.]

Hydrologic region	Number of stations	Statistics for unbiased station skew data (log base 10 units)			
		Mean	Standard deviation	Minimum	Maximum
A	60	0.173	0.531	-1.300	1.270
B	65	0.361	0.512	-1.554	1.647
C	34	-0.092	0.370	-0.759	0.814
D	48	0.172	0.607	-1.645	1.640
E	19	-0.199	0.645	-1.405	0.881
All regions	226	0.156	0.558	-1.645	1.647

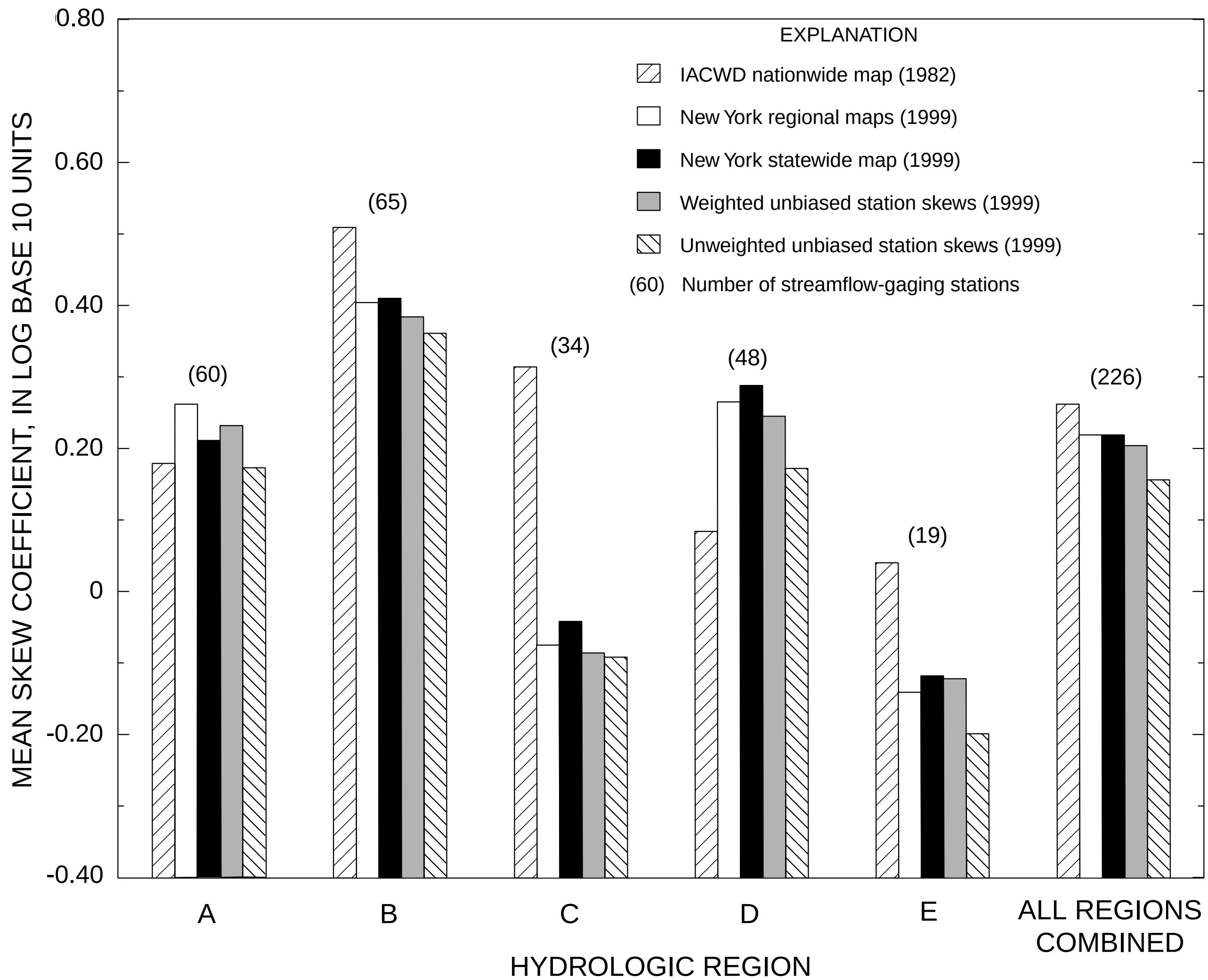


Figure 4. Mean generalized skew coefficients for the five hydrologic regions of New York and for the five regions combined, as calculated by five methods. (Region locations are shown in fig. 2.)

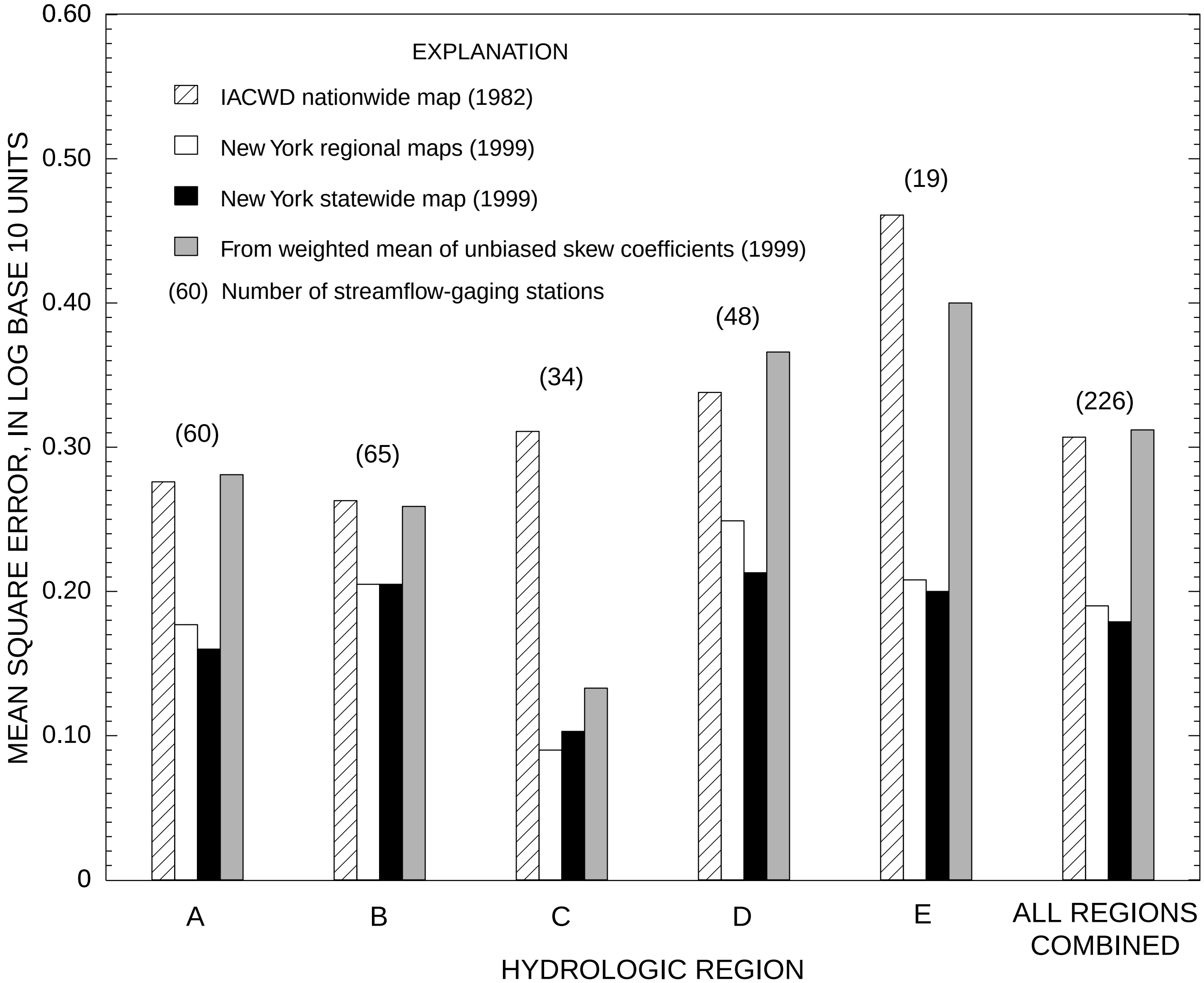


Figure 5. Mean square errors of generalized skew coefficients for five hydrologic regions of New York and for the five regions combined, as calculated by four methods. (Region locations are shown in fig. 2.)

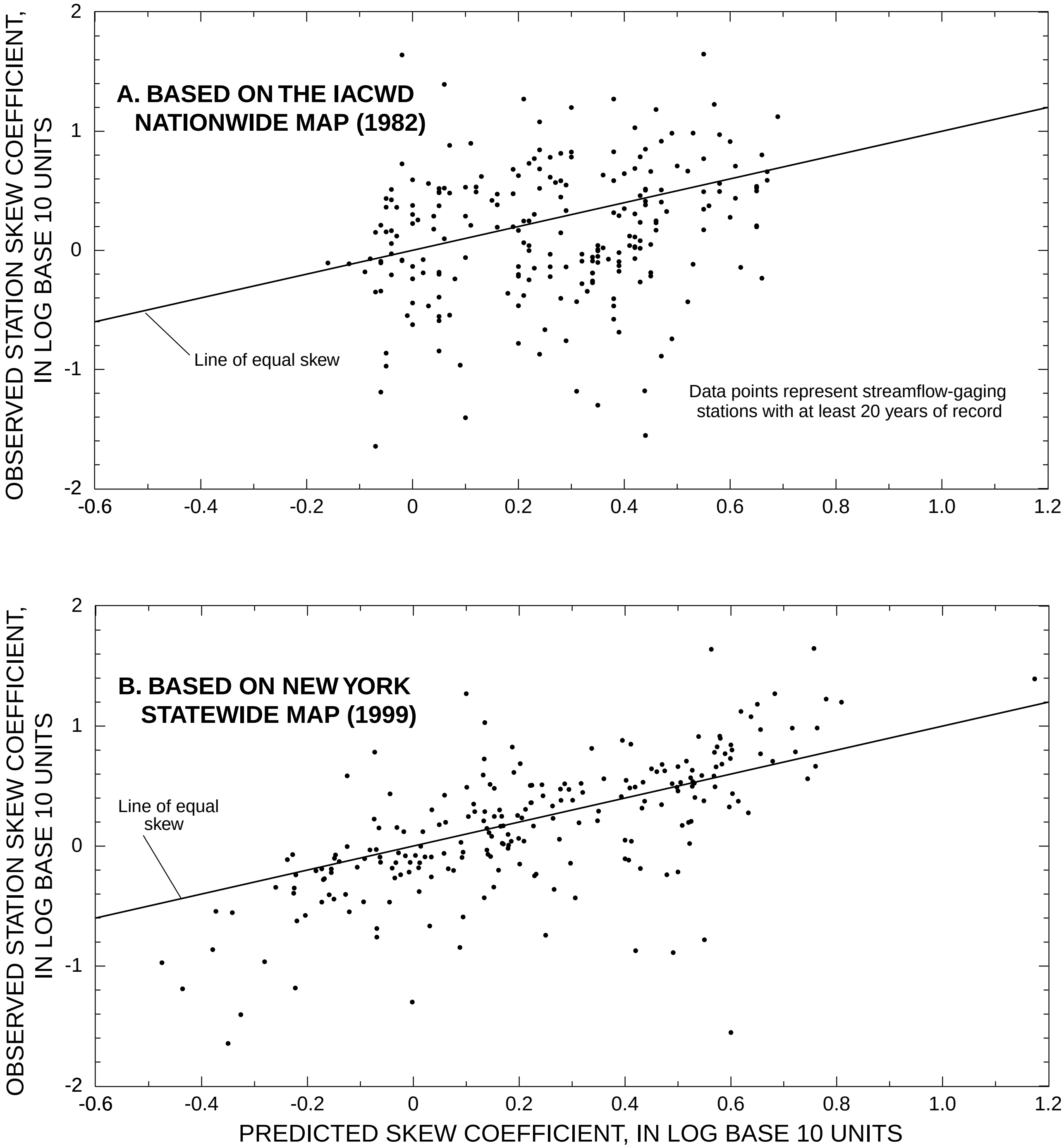


Figure 6. Observed station skew coefficients and skew coefficients predicted from (A) the IACWD nationwide map (1982) and (B) the New York statewide map from this study (1999).

coefficients (unbiased) are plotted with the skew coefficients predicted from the IACWD nationwide map (1982) in figure 6A, and with those predicted from the final statewide generalized skew map from this (1999) study in figure 6B. The results based on the statewide map show significantly less scatter from the line of equal skew than those based on the nationwide map. The wide scatter and poor accuracy obtained from the nationwide map was the impetus for this study. Thus, flood-frequency determinations based on generalized skew coefficients from the new statewide map should be more accurate than those based on skew coefficients from the 1982 nationwide map. Future flood-frequency computations will apply the MSE value derived from the statewide map to equation 3 when fitting the logarithms of annual peak discharges from a gaging station to a Pearson Type-III distribution.

SUMMARY

This report provides a contour map showing generalized skew coefficients for use in computing flood frequency at streamflow-gaging stations on rural, unregulated streams throughout New York (excluding Long Island). Skew coefficients were calculated for 226 gaging stations with at least 20 years of annual peak-discharge record and were corrected for bias. A weighting technique was used to generate a grid of skew coefficients to which map contours were fitted. An attempt was made to provide further refinement by grouping the skew coefficients into five hydrologic regions and developing a separate map of each. An error analysis showed that the resulting regional maps provided no greater accuracy than the statewide contour map; therefore, the statewide contour map was adopted as the final generalized skew-coefficient map of the State. The error analysis showed the statewide map to have a lower MSE than (1) the previously used nationwide map, (2) the five regional maps, or (3) the weighted mean skew coefficients for each of five hydrologic regions.

REFERENCES CITED

Fisher, R.A., 1931, The moments of the distribution of normal samples of measures of departure from normality: Proceedings of the Royal Society of London, v. 130, p. 16-28.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, 522 p.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency—Bulletin 17B of the Hydrology Subcommittee: U.S. Geological Survey, Office of Water-Data Coordination, 183 p.

Landers, M.N. and Wilson, K.V., Jr., 1991, Flood characteristics of Mississippi Streams: U.S. Geological Survey Water-Resources Investigations Report 91-4037, 82 p.

Lumia, Richard, 1991, Regionalization of flood discharges for rural, unregulated streams in New York, excluding Long Island: U.S. Geological Survey Water-Resources Investigations Report 90-4197, 119 p.

Tasker, G.D., 1978, Flood-frequency analysis with a generalized skew coefficient: Water Resources Research, v. 14, no. 2, p. 373-376.

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.