A National Streamflow Network Gap Analysis

Scientific Investigations Report 2013–5013

A National Streamflow Network Gap Analysis

By Julie E. Kiang, David W. Stewart, Stacey A. Archfield, Emily B. Osborne, and Ken Eng

Scientific Investigations Report 2013–5013

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

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

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

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

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

Acknowledgements

Contents

Figures

Tables

Conversion Factors and Datum

Inch/Pound to SI

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
mile, nautical (nmi)	1.852	kilometer (km)
yard (yd)	0.9144	meter (m)
Area		
square mile (mi^2)	2.590	square kilometer (km^2)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F=(1.8×°C)+32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C=(°F-32)/1.8

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

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

Elevation, as used in this report, refers to distance above the vertical datum.

Abbreviations

HUC   hydrologic unit code

LPIII   log Pearson type III statistical distribution

NSS   National Streamflow Statistics software

NWIS   National Water Information System

RMSE   root mean square error

SE   standard error

USGS   U.S. Geological Survey

WIE   Weighted Independent Estimates software

A National Streamflow Network Gap Analysis

By Julie E. Kiang, David W. Stewart, Stacey A. Archfield, Emily B. Osborne, and Ken Eng

Abstract

The U.S. Geological Survey (USGS) conducted a gap analysis to evaluate how well the USGS streamgage network meets a variety of needs, focusing on the ability to calculate various statistics at locations that have streamgages (gaged) and that do not have streamgages (ungaged). This report presents the results of analysis to determine where there are gaps in the network of gaged locations, how accurately desired statistics can be calculated with a given length of record, and whether the current network allows for estimation of these statistics at ungaged locations.

The analysis indicated that there is variability across the Nation's streamflow data-collection network in terms of the spatial and temporal coverage of streamgages. In general, the Eastern United States has better coverage than the Western United States. The arid Southwestern United States, Alaska, and Hawaii were observed to have the poorest spatial coverage, using the dataset assembled for this study. Except in Hawaii, these areas also tended to have short streamflow records. Differences in hydrology lead to differences in the uncertainty of statistics calculated in different regions of the country. Arid and semiarid areas of the Central and Southwestern United States generally exhibited the highest levels of interannual variability in flow, leading to larger uncertainty in flow statistics.

At ungaged locations, information can be transferred from nearby streamgages if there is sufficient similarity between the gaged watersheds and the ungaged watersheds of interest. Areas where streamgages exhibit high correlation are most likely to be suitable for this type of information transfer. The areas with the most highly correlated streamgages appear to coincide with mountainous areas of the United States. Lower correlations are found in the Central United States and coastal areas of the Southeastern United States. Information transfer from gaged basins to ungaged basins is also most likely to be successful when basin attributes show high similarity. At the scale of the analysis completed in this study, the attributes of basins upstream of USGS streamgages cover the full range of basin attributes observed at potential locations of interest fairly well. Some exceptions included very high or very low elevation areas and very arid areas.

Introduction

To understand streamflow in the context of ecological services, water availability, or interactions with water quality, streamflow information is needed for a wide variety of hydrologic conditions. Statistics providing information on the magnitude, timing, and frequency of streamflow events are needed at the location of interest.

Ideally, a gaged streamflow record is available at the location of interest, and statistics can be calculated directly from the streamflow time series. To estimate reliable statistics and identify trends in streamflow, long-term streamflow records are essential. The extent of the U.S. Geological Survey (USGS) streamflow network and the accuracy of statistics calculated from records of limited length are assessed in this report. Whereas other government agencies and private organizations may also collect streamflow information in some areas of the United States, this report focuses exclusively on the USGS streamgage network because USGS streamflow data are readily accessible from a single database and consistent quality control procedures have been applied to the entire dataset.

Because it is impossible to maintain a streamgage at every location of interest, methods have been developed to transfer streamflow information at gaged locations to estimate information at ungaged locations. These methods generally require that flow records at the gaged locations be relatively unaltered by upstream land-use change, flow regulation, discharges, or withdrawals. This report assesses the suitability of the USGS streamflow network for transferring information from gaged locations to ungaged locations.

For the purpose of estimating statistics, areas of the country that are not covered by the streamgage network and where there is low suitability for applying statistical methods for estimating statistics are priority areas for the addition of streamgages. Additional streamgages can be located strategically, if information about gaps is available. The goal of this report is to help identify where there are gaps in the streamflow network. Filling in these gaps with new streamgages will not only allow estimation of statistics directly from a new gaged record, but will also allow better estimation of streamflow statistics at ungaged locations.

Objectives of the Streamgage Gap Analysis

Streamgage gap analysis for gaged locations.—The objective of the streamgage gap analysis for gaged locations is to assess the spatial and temporal extent of the existing USGS

streamgage network. There are three main components to this part of the analysis:

- *spatial gap analysis* to characterize the spatial extent and coverage of the current streamgage network

- *temporal gap analysis* to characterize the change in the network over time and the coverage of long-term streamgages (50 years of record or more)

- *uncertainty analysis* to evaluate the accuracy of statistics calculated at gaged locations in different parts of the country

Streamgage gap analysis for ungaged locations.—The streamgage gap analysis for ungaged locations provides an assessment of the suitability of the USGS streamgage network for estimating statistics at ungaged locations. Two approaches are commonly taken for transferring information from the streamgage network to an ungaged location. The first involves the selection of one or more index streamgages whose upstream watershed is thought to be similar to the ungaged location. Streamflow characteristics calculated for the index streamgages are then adjusted and applied to the ungaged location. The second approach to information transfer utilizes regional relationships among streamflow characteristics and upstream basin characteristics. This part of the streamgage gap analysis considers the following approaches to assess the suitability of the USGS streamgage network for estimating statistics at ungaged locations:

- identify where an existing long-term continuous record index station is likely to be suitable or unsuitable for transfer of information using correlation analysis and basin attribute analysis

- evaluate the accuracy of regional regression analyses for estimation of streamflow statistics across the country

Study Area and Data

The network analysis examined the USGS streamgage network for the entire United States, including Alaska, Hawaii, and Caribbean (for purposes of this report, Puerto Rico), although some elements of the analysis were possible only for the conterminous United States. Only USGS streamgages included in the National Water Information System (NWIS) database were considered.

A total of 9,929 USGS streamgages were used in this study (highlight 1). The bulk of the dataset consisted of the 9,323 streamgages included in the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES) dataset, developed by Falcone and others (2010), as updated in March 2011. This 2011 update, referred to as GAGES–II-prelim, includes all streamgages active in water year[1] 2010 and all current and

---

[1]A water year is a 12-month period starting October 1 and continuing to the following September 30. The water year is designated by the calendar year in which it ends.

Highlight 1.   Datasets Used in the National Streamflow Network Gap Analysis

Complete dataset.—Consists of a total of 9,929 streamgages (fig. 1*A*), including reference- and non-reference-quality streamgages. Consists of GAGES–II-prelim streamgages and additional streamgages from Wieczorek (2011).

Reference-quality streamgages dataset.—A total of 1,630 streamgages are identified in GAGES–II-prelim as having reference-quality data and at least 20 years of record (fig. 1*B*). The reference-quality streamgages are judged to be largely free of human alterations to flow, whereas nonreference-quality streamgages may have substantial human alterations to flow (Falcone and others, 2010). Reference streamgages were used for statistical analyses (uncertainty of statistics and analysis of flow correlation) that require long-term records. In the basin similarity analysis, streamgages were included as long as they are identified to be of reference quality, regardless of the record length.

There are small differences between the streamgages identified as being reference quality in the final GAGES–II dataset (Falcone, 2011) and those identified as reference quality in the preliminary version used in this study. The GAGES–II dataset was not finalized until October 2011, after the completion of the analysis for this study. As a result of review of the final Falcone (2011) dataset by the U.S. Geological Survey, 19 streamgages classified as reference quality in the preliminary dataset were removed from that designation. In addition, 15 streamgages were reclassified as being of reference quality. The 19 reclassified streamgages that were included as reference-quality streamgages in the preliminary dataset and used in this analysis met all other criteria for assignment into the reference-quality classification. Because of this, and because of the small number of streamgages that were reclassified, use of the preliminary dataset is not believed to have a substantive effect on the results of this analysis. Appendix 1 lists the reference-quality streamgages with 20 or more years of record that were used in this study and notes the streamgages whose classification changed in the final GAGES–II dataset.

discontinued USGS streamgages with at least 20 years of record beginning after 1950. Similar to the original GAGES dataset, GAGES–II-prelim includes delineated drainage basins for the streamgages as well as values of a variety of basin characteristics calculated for the upstream watershed, such as mean annual precipitation, mean basin elevation, and land use descriptors. For this study, the GAGES–II-prelim dataset was supplemented by an additional 606 streamgages with 20 or more years of record (where record started before 1950) for which automated geographic information system (GIS) delineations (Wieczorek, 2011) match NWIS-reported drainage areas within 20 percent. This complete dataset was used for most of the GIS-based analysis to identify temporal and geographical gaps in the streamgage network. For most of this work, the complete dataset was further restricted to streamgages measuring flow for drainage areas less than or equal to 20,000 square miles (mi^2). The locations of the streamgages used in this study are shown in figure 1*A*.

Each of the streamgages included in the GAGES–II-prelim dataset are classified as being of either reference or nonreference quality. Reference-quality streamgages are those judged to be largely free of human alterations to flow, whereas nonreference quality streamgages may have substantial human alterations to flow (Falcone and others, 2010). All streamgages were screened and categorized as either reference quality or nonreference quality by calculating a hydrologic disturbance index (presence of dams, change in reservoir storage, number of canals, road density, proximity to major pollutant discharge site, estimates of water withdrawals, and fragmentation of undeveloped land), reviewing historical digital maps and imagery for evidence of hydrologic alteration and human activity, and reviewing comments in USGS annual water data reports for information on regulation or diversions. The streamgages that were identified in the GAGES–II-prelim dataset as of reference quality were the focus of many parts of the network analysis because these are the streamgages that facilitate estimation of natural flow statistics at gaged and ungaged locations. Daily streamflow data for 1,630 reference-quality streamgages with 20 or more years of record were downloaded from NWIS. These 1,630 streamgages are shown in figure 1*B* and listed in appendix 1.

The statistics that were considered in this analysis are listed in highlight 2. The flow statistics are divided into four categories: statistics describing long-term mean conditions, peak flows, low flows, and variability in flow. All the statistics chosen are commonly used by ecologists, flood plain managers, and water resource planners.

Many components of this study are conducted by or summarized by hydrologic unit code (HUC) 2-digit water resource regions (HUC regions; fig. 2). The Missouri River Basin (HUC region 10) is very large and has been separated into Upper Missouri (HUC region 10A) and Lower Missouri (HUC region 10B). Some parts of the analysis are also summarized by level II ecoregions of North America (ecoregions; fig. 3), as defined by the Commission for Environmental Cooperation (1997). Hawaii and Puerto Rico are not included in the ecoregions classification.

Highlight 2.   Statistics Used in Streamgage Gap Analysis

- *Long-term mean.—*
  - Mean annual flow
  - Mean monthly flows (mean daily flow throughout the month)
  - Maximum monthly flows (maximum daily flow during the month)
  - Minimum monthly flows (minimum daily flow during the month)
- *Peak flows.—*
  - 1 percent chance annual exceedance flood (100-year flood)
  - 2 percent chance annual exceedance flood (50-year flood)
  - 4 percent chance annual exceedance flood (25-year flood)
  - 10 percent chance annual exceedance flood (10-year flood)
- *Low flows.—*
  - 7Q2 (7-day average minimum flow experienced on average once in 2 years)
  - 7Q10 (7-day average minimum flow experienced on average once in 10 years)
- *Variability in flow.—*Coefficient of variation of flow (interannual variability)

Streamgage Gap Analysis for Gaged Locations

A streamgage network should have a wide spatial extent, with streamgages covering the variety of watershed conditions experienced in the network area. To accurately estimate flow statistics at gaged locations, long periods of record are generally preferable. This part of the analysis focuses on the following questions:

- Where are large geographical gaps located in the streamgage network (Spatial Gap Analysis section)?
- Where are large temporal gaps located in the streamgage network (Temporal Gap Analysis section)?
- How accurately can commonly used statistics be estimated at streamgage locations (Accuracy of Estimated Statistics section)?

A

EXPLANATION
- GAGES II-prelim gages
- Supplemental gages

Alaska

Hawaii

Puerto Rico

B

EXPLANATION
- Reference streamgage

Alaska

Hawaii

Puerto Rico
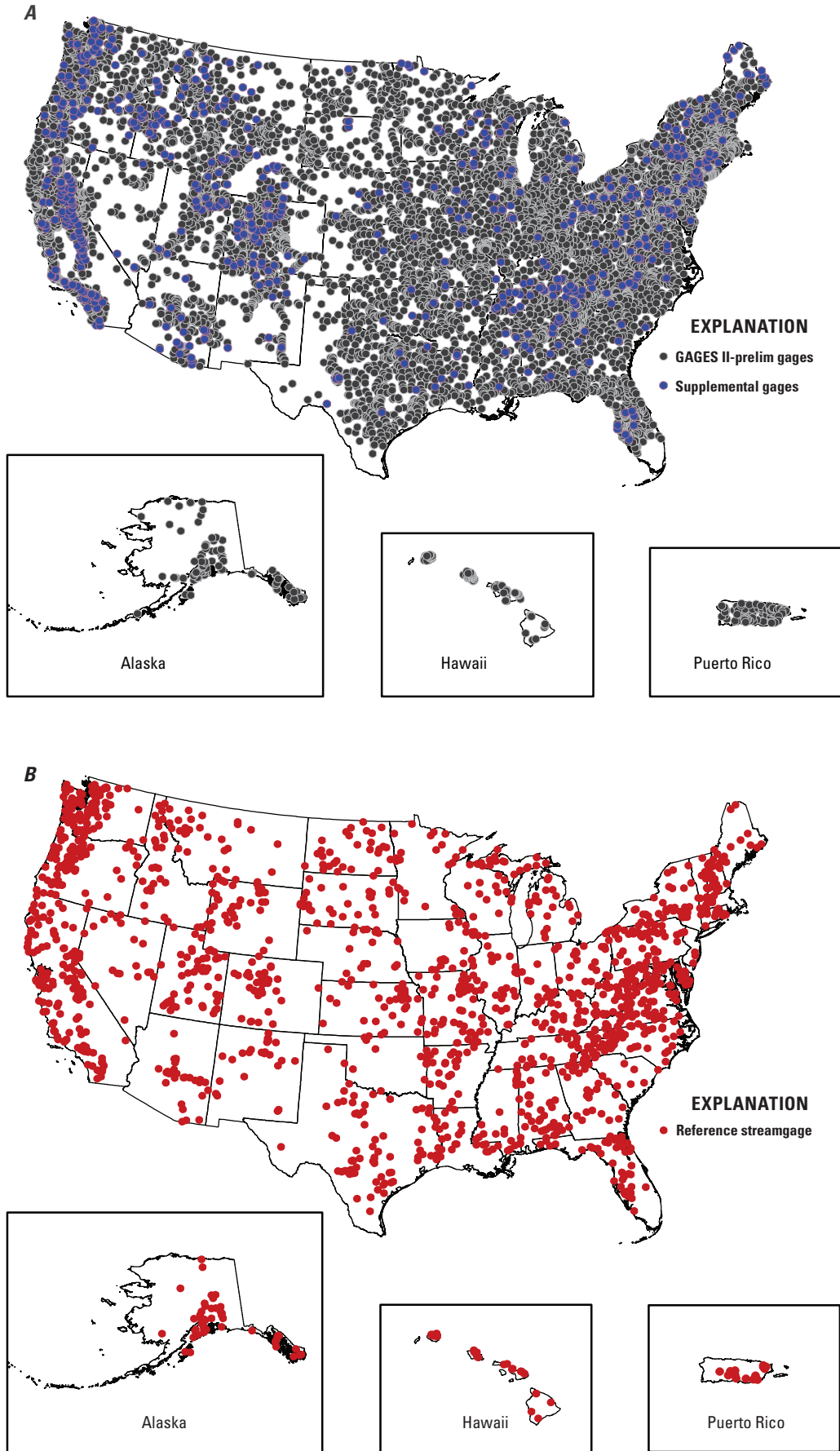
Figure 1.   Maps showing streamgages used in this analysis for the datasets that contain *A*, all (complete) and *B*, reference-quality streamgages.
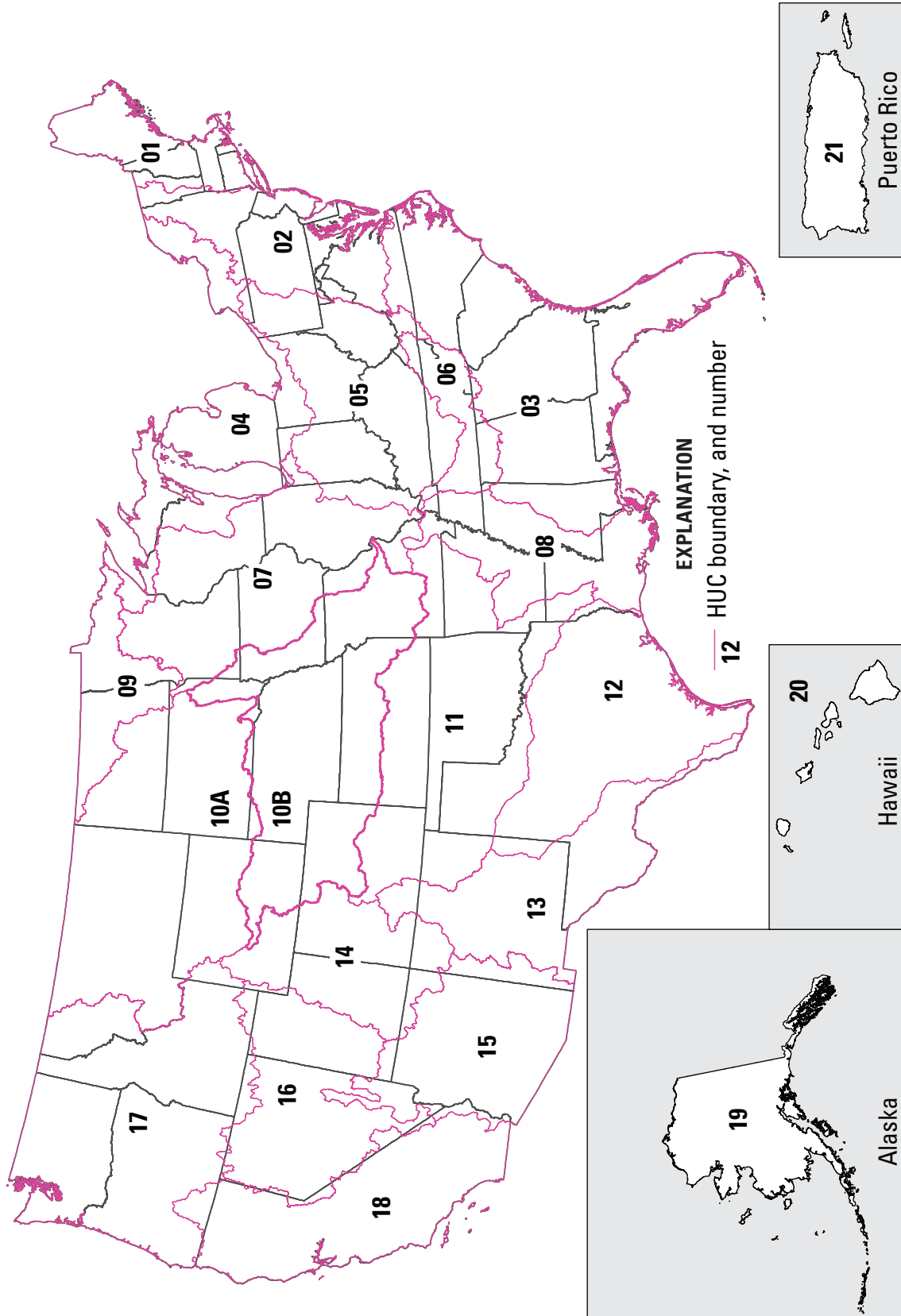
Figure 2. Map showing hydrologic unit code (HUC) regions for the United States. HUC regions shown in this map are indicated by two-digit codes.
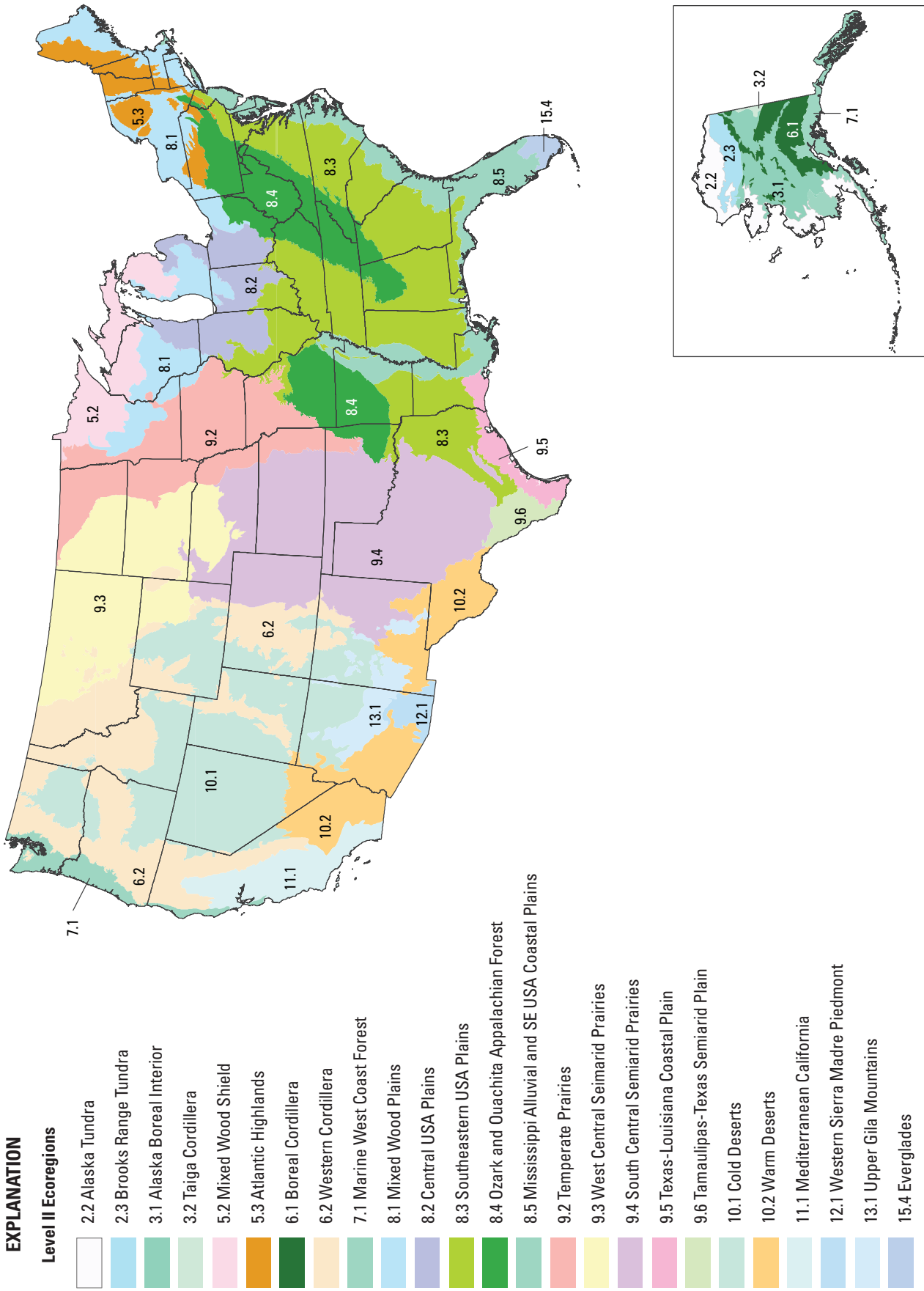
Figure 3.   Maps showing level II ecoregions of North America present in the United States; from Commission for Environmental Cooperation (1997).

EXPLANATION

Level II Ecoregions

- 2.2 Alaska Tundra
- 2.3 Brooks Range Tundra
- 3.1 Alaska Boreal Interior
- 3.2 Taiga Cordillera
- 5.2 Mixed Wood Shield
- 5.3 Atlantic Highlands
- 6.1 Boreal Cordillera
- 6.2 Western Cordillera
- 7.1 Marine West Coast Forest
- 8.1 Mixed Wood Plains
- 8.2 Central USA Plains
- 8.3 Southeastern USA Plains
- 8.4 Ozark and Ouachita Appalachian Forest
- 8.5 Mississippi Alluvial and SE USA Coastal Plains
- 9.2 Temperate Prairies
- 9.3 West Central Seimarid Prairies
- 9.4 South Central Semiarid Prairies
- 9.5 Texas-Louisiana Coastal Plain
- 9.6 Tamaulipas-Texas Semiarid Plain
- 10.1 Cold Deserts
- 10.2 Warm Deserts
- 11.1 Mediterranean California
- 12.1 Western Sierra Madre Piedmont
- 13.1 Upper Gila Mountains
- 15.4 Everglades

Spatial Gap Analysis

The extent of the existing streamgage network was assessed using GIS analysis. The drainage area covered by streamgages was mapped to highlight locations where there is sparse coverage by existing USGS streamgages. Differences in the geographical distribution of streamgages by region also were explored using GIS.

The size of a drainage basin affects scaling properties of hydrologic processes. For example, if upstream basin characteristics are similar, streamflow in a very small basin will tend to increase and decrease in response to precipitation events faster than in a large basin. Further, a large basin generally averages a hydrologic response over a large heterogeneous area; subbasins within the larger area may experience different amounts of precipitation, and different responses may occur. Consequently, a very large basin may not provide a great deal of information about local watershed response—even for a small basin contained within the large basin. Similarly sized watersheds are generally preferred when transferring information from a gaged location to an ungaged location. The reference-quality gages used in this analysis have upstream drainage areas ranging from less than 1 mi² to nearly 20,000 mi². To stay roughly within this range of drainage area, the spatial analysis of the USGS streamgage network focused on streamgages that measure drainage areas of less than 20,000 mi².

Figure 4 shows the spatial coverage of USGS streamgages in the complete dataset. When basins of drainage areas less than 20,000 mi² are considered, approximately 72 percent of the conterminous United States is gaged. The percentage of the land area that is gaged decreases to 45 percent for basins less than or equal to 2,000 mi², 33 percent for basins less than or equal to 1,000 mi², and 20 percent for basins less than or equal to 500 mi² in size. No attempt was made to assess the contribution of streamgages maintained by other agencies or organizations.

For the small basin sizes (fig. 4A, B), the streamgages in the complete dataset cover a larger percentage of the total area in the Eastern United States than in the Central and Western United States. In part, this is because the stream density is higher in the more humid Eastern United States than in many parts of the Western and Central United States, which are arid or semiarid. Areas with greater streamgage density also generally correspond to areas with high population. For large basin sizes (fig. 4C, D), the differences in spatial coverage of streamgages across the country are somewhat less apparent. The information on spatial coverage is summarized in table 1, which lists the percentage of each HUC region that is gaged by streamgages in the complete dataset by basin size. Some

Table 1.   Cumulative percentage of hydrologic unit code regions that are gaged.

[HUC, hydrologic unit code; DA, drainage area; ≤ , less than or equal to; mi², square mile]

	HUC region	Cumulative percentage gaged by all streamgages					Cumulative percentage gaged by reference gages				
		DA ≤500 mi²	DA ≤1,000 mi²	DA ≤1,500 mi²	DA ≤2,000 mi²	DA ≤20,000 mi²	DA ≤500 mi²	DA ≤1,000 mi²	DA ≤1,500 mi²	DA ≤2,000 mi²	DA ≤20,000 mi²
1	New England	36	44	56	60	74	6	9	13	13	13
2	Mid-Atlantic	45	55	59	61	78	10	12	12	12	12
3	South Atlantic-Gulf	26	39	47	50	79	7	9	10	11	11
4	Great Lakes	33	50	57	60	73	8	11	13	13	13
5	Ohio	38	52	58	63	86	9	10	11	11	11
6	Tennessee	34	46	51	56	69	10	13	13	13	13
7	Upper Mississippi	22	41	53	59	88	6	8	10	10	11
8	Lower Mississippi	15	30	38	43	63	3	6	6	6	6
9	Souris-Red-Rainy	14	32	42	55	65	5	11	15	23	25
10A	Upper Missouri	11	22	30	37	70	4	6	8	9	11
10B	Lower Missouri	17	31	43	49	84	3	5	7	9	16
11	Arkansas-White-Red	14	28	37	43	79	3	6	7	8	10
12	Texas-Gulf	18	29	35	40	79	5	8	9	9	10
13	Rio Grande	7	12	16	17	40	1	1	3	3	3
14	Upper Colorado	20	28	35	39	77	4	4	4	4	4
15	Lower Colorado	9	17	23	32	66	3	3	3	5	5
16	Great Basin	8	13	16	17	31	1	1	1	1	1
17	Pacific Northwest	23	37	44	48	76	8	10	11	11	13
18	California	24	35	40	41	62	5	8	8	8	8
19	Alaska	1	2	2	4	10	0	1	1	2	9
20	Hawaii	10	10	10	10	10	1	1	1	1	1
21	Caribbean	51	51	51	51	51	4	4	4	4	4
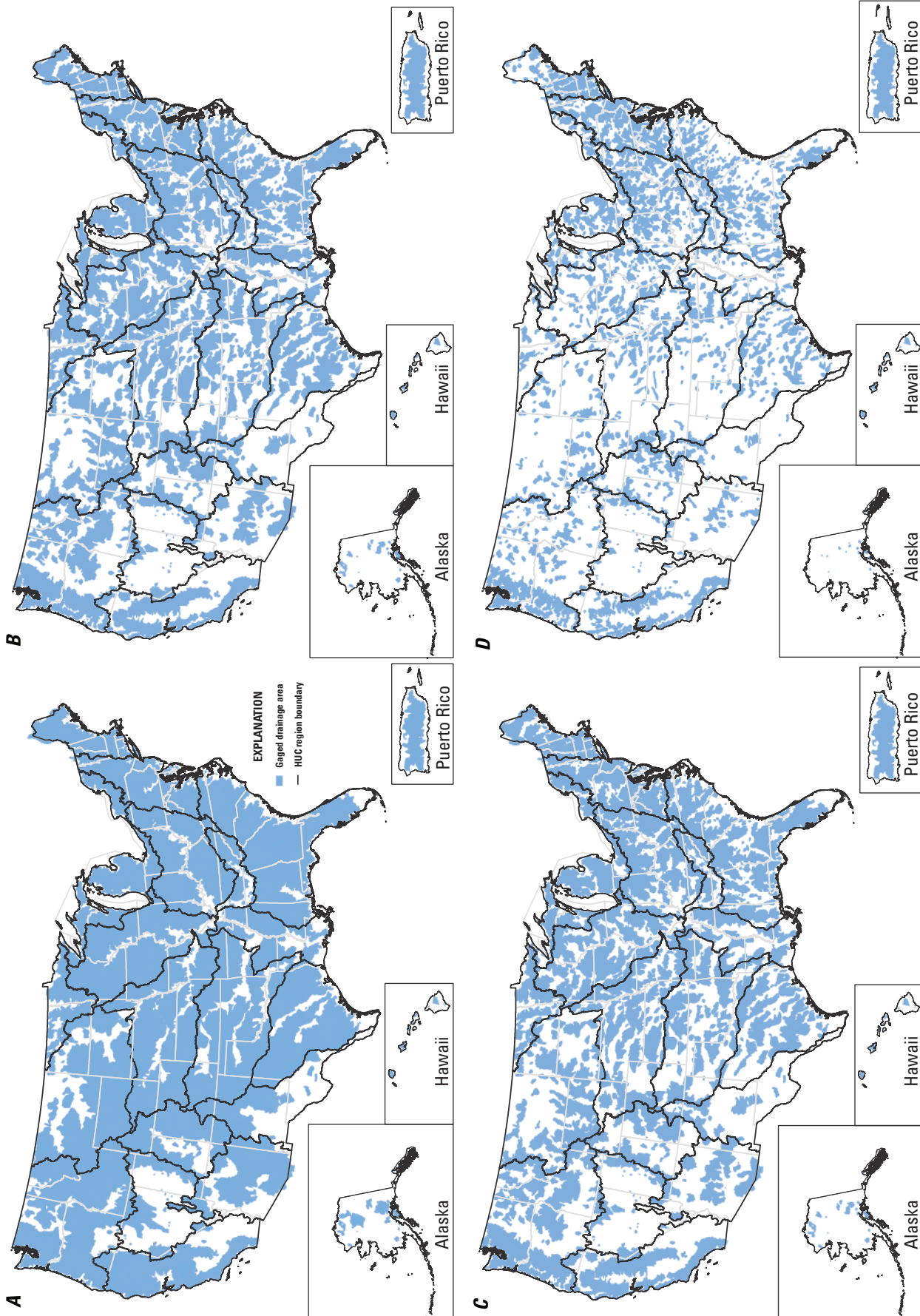
Figure 4. Maps showing the spatial coverage of streamgages in the complete dataset (GAGES–II-prelim streamgages with at least 20 years of record) that have upstream watersheds with drainage area A, less than 20,000 square miles (mi²); B, less than 2,000 mi²; C, less than 1,000 mi²; and D, less than 500 mi².

EXPLANATION

Gaged drainage area

HUC region boundary

HUC regions have low streamgage coverage across all basin sizes. In particular, the arid Rio Grande (HUC region 13) and Great Basin (HUC region 16) have only 30 to 40 percent areal coverage. Alaska (HUC region 19) and Hawaii (HUC region 20) have even poorer areal coverage, about 10 percent.

Figure 5 shows the spatial coverage of the gaged drainages in the reference-quality streamgage dataset. In the conterminous United States, 8.6 percent of the total land area is gaged by a streamgage in the reference-quality streamgage dataset. Figure 6 compares the spatial coverages of the complete dataset and the reference-quality streamgage dataset by HUC region. The bars in the figure are labeled with the number of streamgages in the HUC region. The regions with the greatest coverage in the complete dataset are not necessarily the same as the regions with the greatest coverage in the reference-quality streamgage dataset. For example, whereas the spatial coverage of streamgages in Souris-Red-Rainy (HUC region 9) was average for the complete dataset, the region has the highest spatial coverage for reference-quality streamgages (25 percent). In addition, whereas Alaska (HUC region 19) has one of the lowest percentages of

areal coverage in the complete dataset, most of the streamgages are of reference quality, so the percentage of area covered by reference-quality streamgages is more in line with the rest of the country (fig. 6). The arid southwest is seen to have low spatial coverage of streamgages, with Rio Grande (HUC region 13), Upper Colorado (HUC region 14), Lower Colorado (HUC region 15), and Great Basin (HUC region 16) all having spatial coverage of reference-quality streamgages less than 5 percent. Hawaii (HUC region 20) also shows very poor spatial coverage by streamgages in the reference-quality streamgage dataset at only 1 percent areal coverage. Table 1 lists, by basin size, the percentage of each HUC region that is covered by streamgages in the reference-quality streamgage dataset.

Differences in streamgage coverage by ecoregion are summarized in figure 7. Ecoregion 3.2 (Taiga Cordillera) does not include any streamgages from the complete dataset. There are no streamgages from the reference-quality streamgage dataset in the Brooks Range Tundra ecoregion (2.3), the Taiga Cordillera ecoregion (3.2), and the Everglades ecoregion (15.4). Because the ecoregions do not follow hydrologic divides, a streamgage

Table 2. Cumulative percentage of level II ecoregions in the United States that are gaged.

[Level II ecoregions are from Commission for Environmental Cooperation (1997). DA, drainage area; ≤ , less than or equal to; mi², square mile]

Level II ecoregion	Cumulative percentage gaged by all streamgages					Cumulative percentage gaged by reference gages				
	DA ≤500 mi²	DA ≤1,000 mi²	DA ≤1,500 mi²	DA ≤2,000 mi²	DA ≤20,000 mi²	DA ≤500 mi²	DA ≤1,000 mi²	DA ≤1,500 mi²	DA ≤2,000 mi²	DA ≤20,000 mi²
2.2　Alaska Tundra	0	1	1	4	6	0	0	0	0	2
2.3　Brooks Range Tundra	0	1	1	4	21	0	0	0	3	21
3.1　Alaska Boreal Interior	0	0	0	0	5	0	0	0	0	5
3.2　Taiga Cordillera	0	0	0	0	0	0	0	0	0	0
5.2　Mixed Wood Shield	21	41	52	62	77	7	13	15	19	19
5.3　Atlantic Highlands	49	64	75	81	97	11	14	17	17	17
6.1　Boreal Cordillera	1	5	5	7	22	1	4	4	7	21
6.2　Western Cordillera	36	55	63	68	93	9	12	13	13	15
7.1　Marine West Coast Forest	9	12	13	14	18	5	6	6	7	10
8.1　Mixed Wood Plains	32	46	54	58	80	5	6	8	8	8
8.2　Central USA Plains	39	57	64	68	91	6	6	8	8	8
8.3　Southeastern USA Plains	26	39	47	51	82	6	9	10	10	10
8.4　Ozark and Ouachita-Appalachian Forests	39	55	63	67	87	11	15	18	18	19
8.5　Mississippi Alluvial and SE USA Coastal Plains	13	23	28	32	45	4	4	6	6	7
9.2　Temperate Prairies	21	38	47	55	82	6	9	12	13	14
9.3　West Central Semiarid Prairies	7	17	26	33	64	1	4	5	8	16
9.4　South Central Semiarid Prairies	12	23	33	39	82	3	5	6	7	10
9.5　Texas-Louisiana Coastal Plain	11	15	19	19	21	1	2	2	2	2
9.6　Tamaulipas-Texas Semiarid Plain	7	13	15	16	65	5	6	6	6	6
10.1 Cold Deserts	7	13	18	21	57	1	1	2	2	2
10.2 Warm Deserts	2	5	7	10	20	0	0	0	0	0
11.1 Mediterranean California	32	41	46	48	70	6	6	6	6	6
12.1 Western Sierra Madre Piedmont	13	18	31	38	87	5	5	5	5	5
13.1 Upper Gila Mountains	21	35	42	49	81	6	9	10	15	15
15.4 Everglades	1	2	2	2	11	0	0	0	0	0
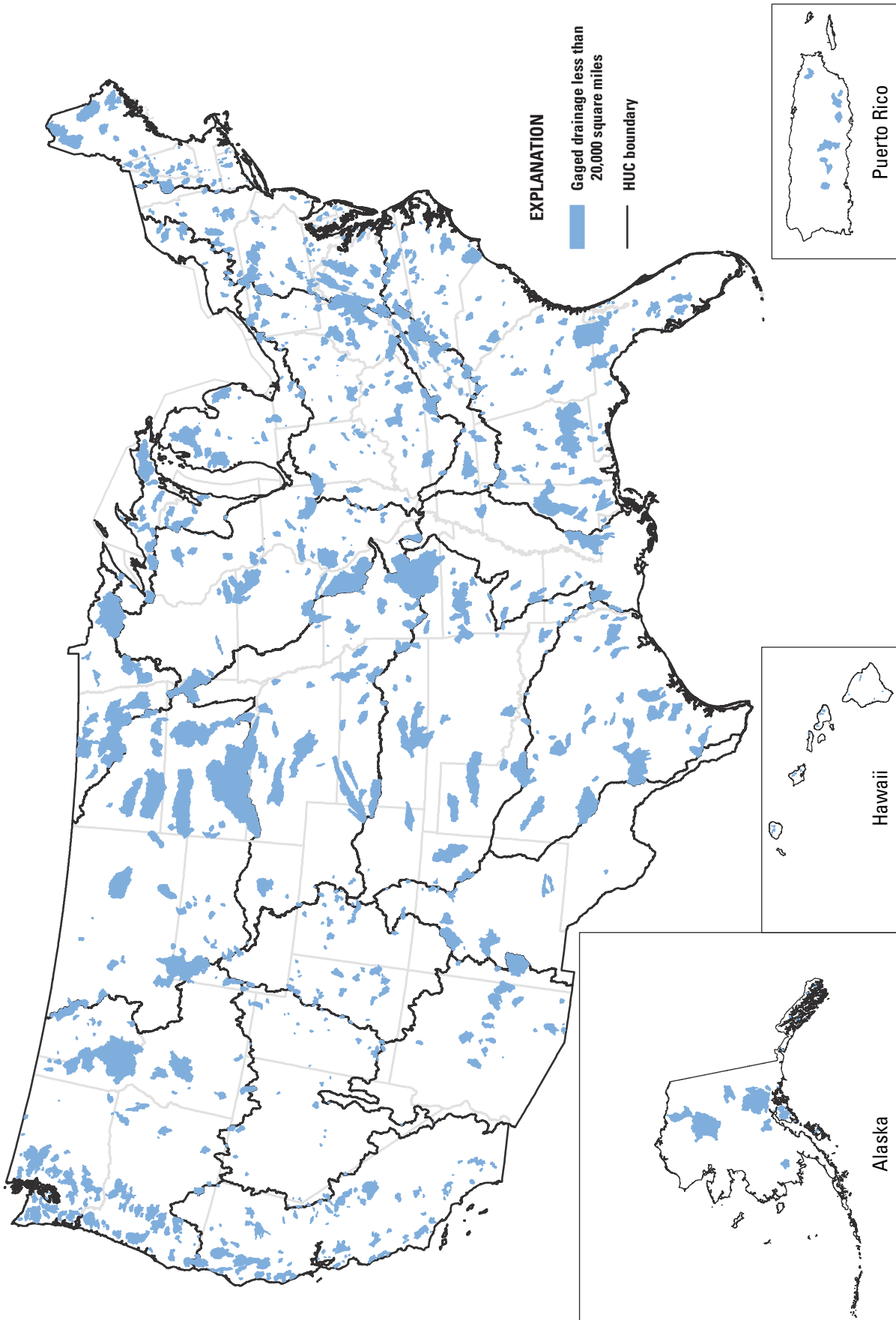
Figure 5.   Map showing spatial coverage of reference-quality streamgage drainages. All streamgages have an upstream drainage area of less than 20,000 square miles.
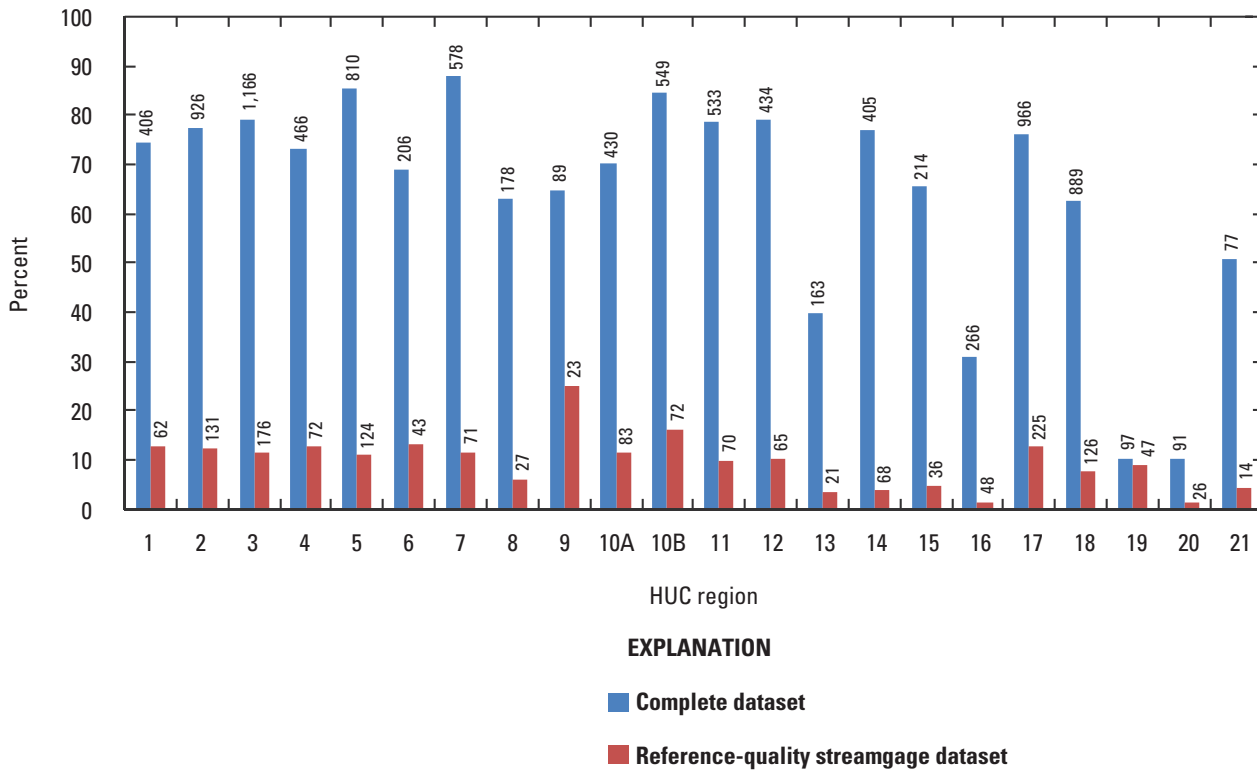
Figure 6. Graphs showing percentage of hydrologic unit code (HUC) regions gaged, complete dataset and reference-quality streamgage dataset. Each bar is labeled with the number of streamgages located in the HUC region. Figure 1 shows the location of each HUC region and table 1 lists the HUC regions.

may be located in one ecoregion, whereas the upstream drainage area can include area in another ecoregion. For example, whereas there are no reference-quality streamgages located within the Brooks Range Tundra, downstream streamgages outside of the ecoregion do measure flow from drainage areas that are partially contained in the ecoregion. The Everglades ecoregion includes a highly managed water system and has no reference-quality streamgages at all. Table 2 details the percentage of each ecoregion that is measured by streamgages in the reference-quality streamgage dataset, by varying drainage basin size.

Temporal Gap Analysis

Long-term streamflow records are essential to accurately calculate streamflow statistics. Long-term records are critical to tracking over time changes in flow due to land use changes, water management changes, or climate variability. Figure 8 shows the spatial coverage of streamgage drainages with more than 50 years of record. The spatial coverage of these long-term record streamgages is generally higher in the Eastern United States than in the Western United States. Alaska, Hawaii, and Puerto Rico are poorly covered by these long-term record streamgages. Overall, 47.9 percent of the conterminous United States is covered by streamgages with 50 or more years of record. The percentage coverage declines sharply when only streamgages measuring flow in smaller drainages are considered (fig. 8B).

Tables 3 and 4 list the percentage of area in each HUC region or in each ecoregion, respectively, where at least 10 years of continuous record are available in the two-decade period listed. The streamgage network grew rapidly in the early 1900s, and stabilized in the 1950s to 1960s. Since then, the number of streamgages and the area covered have been relatively stable, although many individual streamgages have been either discontinued or started since then.

Figure 9 shows the median number of years of record available at reference-quality streamgages in each of the HUC regions. This information should be considered in conjunction with the number of reference-quality streamgages and their spatial coverage (fig. 6). For example, although the median record length in Hawaii (HUC region 20) is quite long at 55 years, the areal coverage of reference-quality streamgages in Hawaii is only about 1.3 percent. The arid Southwest has the shortest median record length, and was also observed to have poor spatial coverage in the spatial gap analysis.

Calculation of Flow Statistics

Most of the flow statistics that describe the long-term characteristics of flow (highlight 2) were calculated for the set of streamgages in the reference-quality streamgage dataset using code written in the Matlab programming environment. For the annual statistics, the water year is used as the basis of calculation. A water year is a 12-month period from October 1
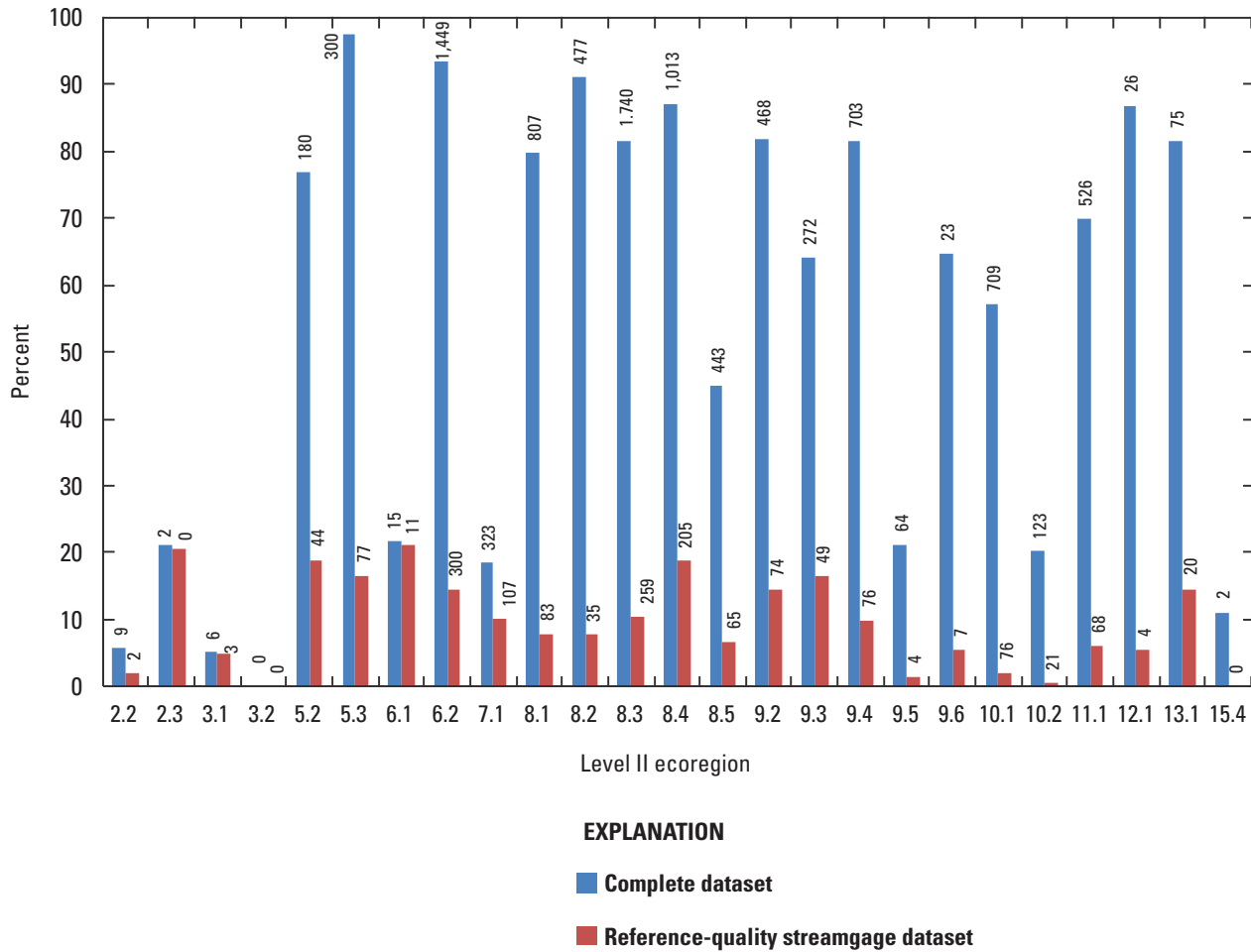
Figure 7. Graphs showing the percentages of ecoregions gaged for the complete dataset and the dataset that contains only reference-quality streamgages. Each bar is labeled with the number of streamgages in the ecoregion. A streamgage can measure flow in multiple ecoregions if part of its upstream drainage area is in another ecoregion. For example, about 20 percent of ecoregion 2.3 is measured by streamgages in the reference-quality streamgage dataset, but these streamgages are in a downstream ecoregion. Figure 2 shows the extent of each ecoregion region and table 2 lists the ecoregions.

of the previous year through September 30 of the water year and is designated by the calendar year in which it ends. For example, the year ending September 30, 2011, is the 2011 water year.

Statistics of peak flow frequency were obtained by querying the USGS National Streamflow Statistics (NSS) database, a national repository of streamflow statistics that are published in USGS reports. All peak flow statistics are based on the period of record that was used in the original report, as referenced in the NSS database. Statistics were not available for all the streamgages in the reference-quality streamgage dataset. Those streamgages that were not found in the NSS database are indicated in appendix 1.

Uncertainty of Estimated Statistics

The uncertainty of a statistic describes the potential difference between the estimated value of the statistic and the (unknown) true value of the statistic. The magnitude of

the uncertainty of a flow statistic is affected by the length of the record, the variability of the streamflow, and anthropogenic effects within the basin. Generally, the uncertainty of a hydrologic statistic is greater when a short-term record is used to estimate its value. Conversely, a long-term record typically decreases the uncertainty. Uncertainty is also affected by the variability of flow. Statistics calculated for a highly variable time series will generally be subject to more uncertainty than statistics calculated from a stable time series. Knowing the uncertainty of statistics calculated from records of varying temporal extent can be useful in determining whether longer term records would be beneficial in a particular region for the purpose of reducing uncertainty in a statistic. Because of differences in the variability of streamflow across the Nation, records of different lengths are required to decrease the level of uncertainty by even amounts in different parts of the country.

All the analyses in this section of the report use only the reference-quality streamgages to maintain as consistent a record as possible over the period of record. Changes in flow

Table 3. Percentage of hydrologic unit code regions with at least 10 years of gaged record collected at streamgages draining less than or equal to 2,000 square miles.

[HUC, hydrologic unit code]

HUC region	Percentage of area gaged				
	1910–1930	1931–1950	1951–1970	1971–1990	1991–2010
1 New England	26.5	52.2	55.3	53.6	48.9
2 Mid-Atlantic	24.0	51.0	56.4	56.3	56.6
3 South Atlantic-Gulf	3.6	31.7	38.1	38.8	42.0
4 Great Lakes	11.9	33.3	52.4	52.8	49.5
5 Ohio	16.1	53.2	58.9	57.9	52.5
6 Tennessee	37.7	52.1	52.9	49.9	39.5
7 Upper Mississippi	14.2	41.2	50.2	51.4	51.4
8 Lower Mississippi	0.0	24.2	34.3	33.1	29.3
9 Souris-Red-Rainy	0.0	29.9	51.5	51.0	45.8
10A Upper Missouri	4.4	21.2	41.8	38.8	29.8
10B Lower Missouri	3.4	8.8	27.3	25.8	22.7
11 Arkansas-White-Red	4.2	20.8	32.5	36.3	36.0
12 Texas-Gulf	0.0	22.6	31.0	35.5	28.5
13 Rio Grande	2.3	7.7	14.3	15.9	11.7
14 Upper Colorado	13.4	21.7	29.4	32.0	30.1
15 Lower Colorado	2.3	10.0	18.0	22.6	25.5
16 Great Basin	8.0	9.4	15.7	14.8	13.7
17 Pacific Northwest	16.1	33.2	40.1	40.5	37.3
18 California	19.0	30.7	38.4	39.1	34.7
19 Alaska	0.0	0.0	1.4	1.4	3.7
20 Hawaii	1.2	5.5	8.9	9.5	10.3
21 Caribbean	0.0	0.0	2.4	14.6	50.8

due to anthropogenic effects can confound estimates of variability in flow statistics. For example, as a result of urbanization, more impervious surfaces are created in a watershed, limiting infiltration and increasing surface runoff. Day-to-day variability in flow tends to increase. In addition, if flow tends to be lower before urbanization and greater after urbanization, this results in an increase in the range between the minimum and maximum observed flows. Using only the reference-quality streamgages minimizes these types of problems.

Mean Annual Flow

The mean annual flow is the average volume of water that flows past a streamgage or any selected point in a river reach in a year. The mean annual flow is equivalent to the mean daily flow multiplied by the number of days in a year. The annual flow past a continuous-record streamgage is computed for each year of record, producing an annual time series of streamflow. In this study, the mean annual flow is calculated as the mean of this annual time series. The mean is considered to be a sample mean because it is calculated from a sample of observations from the full population. Because of limited record lengths and natural variability in flow, there is

uncertainty in how well this sample mean represents the true long-term population mean at a streamgage.

The standard error of the mean is a measure of the uncertainty in the estimate of the mean for the entire population when the mean of the population is calculated from a finite sample. The standard error of the mean (SE) can be calculated as follows:

$$SE = \frac{s}{\sqrt{n}},\qquad(1)$$

where

s is the standard deviation of the annual mean flow time series and

n is the number of years of record.

In equation 1, the standard error is calculated in the same units as the time series (for example, cubic feet per second). The standard error can instead be expressed as a fraction of the mean annual flow (SE_{frac}) by dividing by the mean annual flow:

$$SE_{frac} = \frac{SE}{\bar{x}} = \frac{s}{\sqrt{n}}\frac{1}{\bar{x}} = \frac{CV}{\sqrt{n}},\qquad(2)$$

where

$\bar{x}$ is the mean of the time series and

CV is the coefficient of variation ($\frac{s}{\bar{x}}$), which quantifies the interannual variability observed in the time series.

When the actual length of observed record is used in equation 1 or equation 2, the standard error of the mean reflects the uncertainty in the mean for the observed record length. To calculate the standard error of the mean for a specified record length, different values of n can be inserted into equation 2. The coefficient of variation (fig. 10) is calculated using the full length of record to obtain the best possible estimate of the interannual variability in the time series.

Equation 2 can be rearranged to estimate the number of years of record needed to obtain a specified standard error, as follows:

$$n = \frac{CV^2}{SE_{frac}^{\;2}}.\qquad(3)$$

The standard error of the mean is most appropriate as a measure of the uncertainty of the mean when the time series is a series of independent and identically distributed random variables where the year-to-year variability in the time series is random and uncorrelated. This is not strictly the case for hydrologic time series, which tend to exhibit some persistence, often causing a series of years to be wetter than normal and another series of years to be drier than normal. If the long-term mean is estimated from only a few years of record, the estimate will tend to be high if the years included in the sample were wetter than normal. Conversely, the estimate will
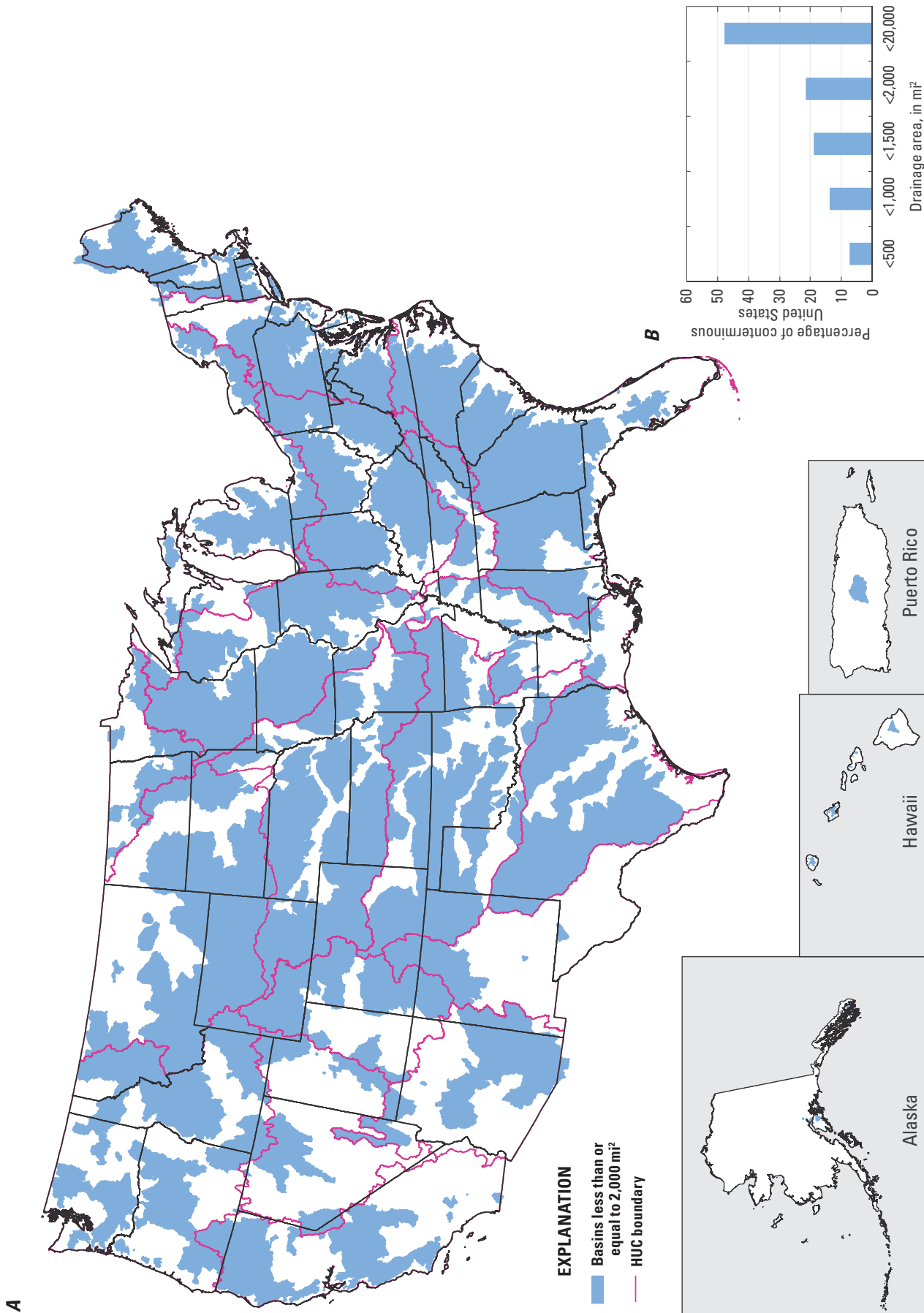
Figure 8.   *A,* Map showing streamgage drainage basins less than or equal to 2,000 square miles (mi²) with at least 50 years of record collected between 1900 and 2009 and *B,* graph showing percentages of conterminous United States with at least 50 years of record, by size of gaged drainage area. <, less than.
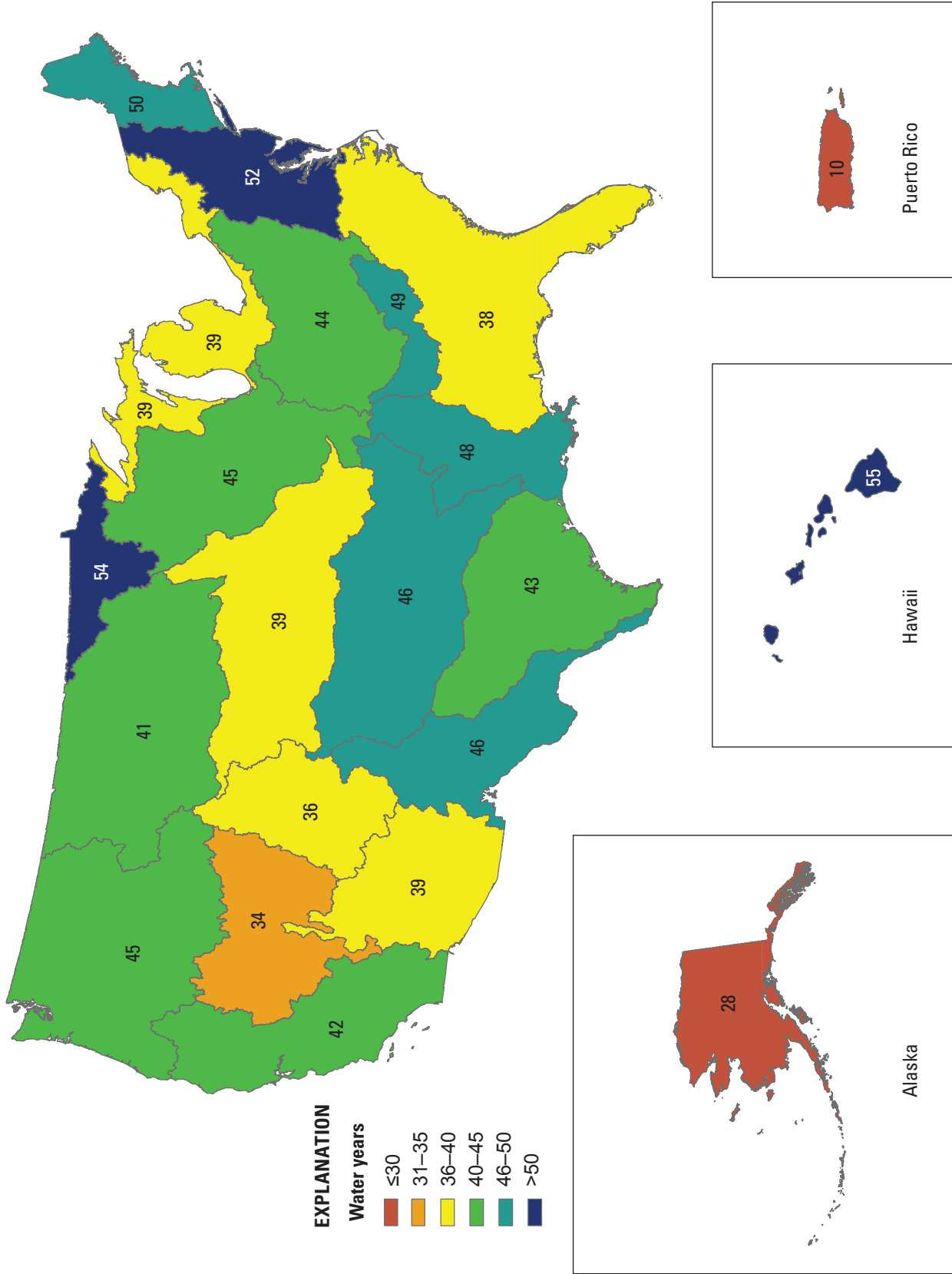
Figure 9. Map showing the median years of record for the reference-quality streamgages found in each hydrologic unit code (HUC) region. There is a limited number of reference-quality streamgages in all HUC regions. <, less than; >, greater than.

EXPLANATION

☐ HUC regions

Mean annual flow CV

- 0.00 to 0.25
- >0.25 to 0.50
- >0.50 to 0.75
- >0.75 to 1.00
- >1.00 to 1.25
- >1.25 to 1.50
- >1.50 to 1.75
- >1.75 to 2.00
- >2.00

Puerto Rico

Hawaii

Alaska

Figure 10.   Map showing the coefficient of variation of the mean annual flow. CV, coefficient of variation; HUC, hydrologic unit code; >, more than.

Table 4. Percentage of level II ecoregions with at least 10 years of gaged record collected at streamgages draining less than or equal to 2,000 square miles.

[Level II ecoregionas are from Commission for Environmental Cooperation (1997)]

Level II ecoregion		Percentage of area gaged during specified time period				
		1910–1930	1931–1950	1951–1970	1971–1990	1991–2010
2.2	Alaska Tundra	0.0	0.0	0.0	0.0	3.8
2.3	Brooks Range Tundra	0.0	0.0	0.0	0.0	4.0
3.1	Alaska Boreal Interior	0.0	0.0	0.0	0.1	0.3
3.2	Taiga Cordillera	0.0	0.0	0.0	0.0	0.0
5.2	Mixed Wood Shield	16.5	34.6	50.6	51.9	46.6
5.3	Atlantic Highlands	43.2	71.8	77.8	74.8	71.2
6.1	Boreal Cordillera	0.0	0.0	3.8	4.2	7.4
6.2	Western Cordillera	28.8	49.9	59.1	58.9	55.3
7.1	Marine West Coast Forest	0.7	6.8	11.1	10.8	12.2
8.1	Mixed Wood Plains	15.9	44.7	51.8	51.1	49.7
8.2	Central USA Plains	14.1	47.5	61.6	61.9	62.6
8.3	Southeastern USA Plains	5.6	34.3	42.1	41.7	40.6
8.4	Ozark and Ouachita-Appalachian Forests	18.9	54.6	60.6	57.4	54.7
8.5	Mississippi Alluvial and SE USA Coastal Plains	0.0	13.5	19.8	23.3	24.9
9.2	Temperate Prairies	3.8	26.9	47.4	47.7	45.1
9.3	West-Central Semiarid Prairies	2.4	5.5	24.2	24.5	18.5
9.4	South Central Semiarid Prairies	2.9	16.5	30.0	34.0	27.4
9.5	Texas-Louisiana Coastal Plain	0.0	5.8	10.7	13.0	11.9
9.6	Tamaulipas-Texas Semiarid Plain	0.0	3.9	4.3	8.3	8.3
10.1	Cold Deserts	4.9	7.5	15.9	15.9	14.2
10.2	Warm Deserts	0.2	1.7	4.8	6.5	7.1
11.1	Mediterranean California	15.5	35.6	44.7	44.7	38.0
12.1	Western Sierra Madre Piedmont	6.7	21.8	24.5	28.1	31.3
13.1	Upper Gila Mountains	2.0	22.0	38.5	43.7	39.7
15.4	Everglades	0.0	0.0	0.0	1.1	1.1

tend to be low if the years included in the sample were drier than normal. With a longer period of record, the estimate of the mean tends to stabilize. In addition, a persistent trend in a time series can increase the standard deviation of a time series even if the variability around the trend line remains the same. This increase is because a persistent trend tends to increase the difference between the smallest and largest values in a time series. The effect of trends in increasing the standard error of the mean is not considered in this study.

To restate, the standard error of the mean characterizes the uncertainty of the sample mean, taking into account the length of record and the overall variability of the time series, but not persistence that results in periods of wetter or drier than average flows, nor trends in flow. Both of these factors will tend to increase the uncertainty, because the variability is not randomly distributed from year to year. Despite its limitations, the standard error of the mean can be considered a good estimate of the minimum level of uncertainty that can be expected when using finite record lengths to calculate a mean. The standard error of the mean is used in this analysis to facilitate comparison of uncertainty in flow statistics across the United States.

The coefficient of variation for the streamgages in the reference-quality streamgage dataset is shown in figure 10.

The coefficient of variation of the mean annual flow is a measure of how much the flow varies from year to year and is calculated using the entire period of record available at each streamgage. Equation 2 is then used to calculate the standard error as a fraction of the mean annual flow at each streamgage. Results for an assumed record length of 10 years are mapped in figure 11. Figure 11*A* shows the standard error for individual stations, and figure 11*B* shows the median standard error for each HUC region. These are the standard errors that could be expected at these locations if only 10 years of record were available. Figure 12 is similar, but record lengths of 50 and 100 years were used; figure 12*A* shows the standard error that could be expected if 50 years of record were available, and figure 12*B* shows the standard error that could be expected if 100 years of record were available. In all cases, the coefficient of variation is calculated using the available length of record at each streamgage, to derive the best possible estimate of coefficient of variation at each streamgage.
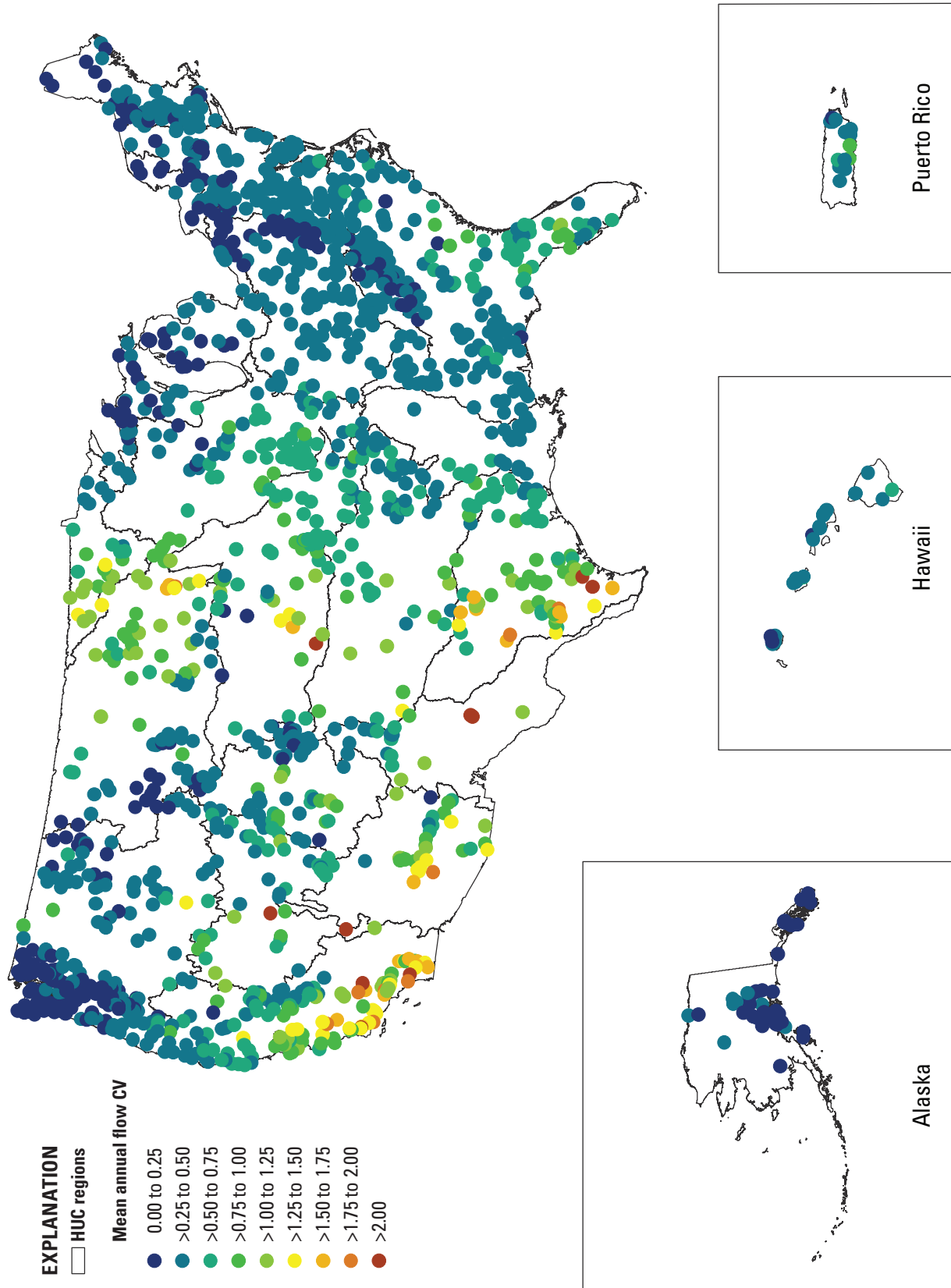
When only 10 years of record are available, there is considerable variability in the standard error of the mean annual flow across the country, with the greatest standard errors in the Central and Southwestern United States, where the interannual variability in annual flow is highest. There is a considerable

decrease in the standard error when 100 years of record are available, with most areas showing fractional standard errors less than 0.2 (20 percent). Figure 13 shows the number of years of record that are needed in each HUC region to obtain a standard error of 0.10 (10 percent, calculated using equation 3).

Figure 13 also suggests that the number of years of record needed for a standard error of 10 percent is quite low in some HUC regions. The uncertainty calculation based on standard error of the mean does not include the effects of persistence in hydrologic records. The effect of persistence is to increase the uncertainty, so the number of years of record needed to achieve 10 percent uncertainty would be higher if the effects of persistence were included. Also, the number of years of record needed shown in figure 13 apply only to estimates of the mean annual flow. Other statistics related to more extreme events require considerably longer records. The map is intended to show differences across the country.

The number of years of record needed in some HUC regions to achieve an uncertainty of 10 percent is very large (fig. 13). The number of years of record needed declines sharply if a slightly higher level of uncertainty is acceptable (see for example Souris-Red-Rainy in fig. 14). For each HUC region, table 5 shows the median coefficient of variation of annual mean flow at streamgages in the HUC region. These median coefficients of variation can be used in equation 3 to calculate the years of record needed for a desired level of uncertainty in estimating mean annual flow.

Table 5. Median coefficient of variation of annual mean flow of streamgages.

[HUC, hydrologic unit code; CV, coefficient of variation]

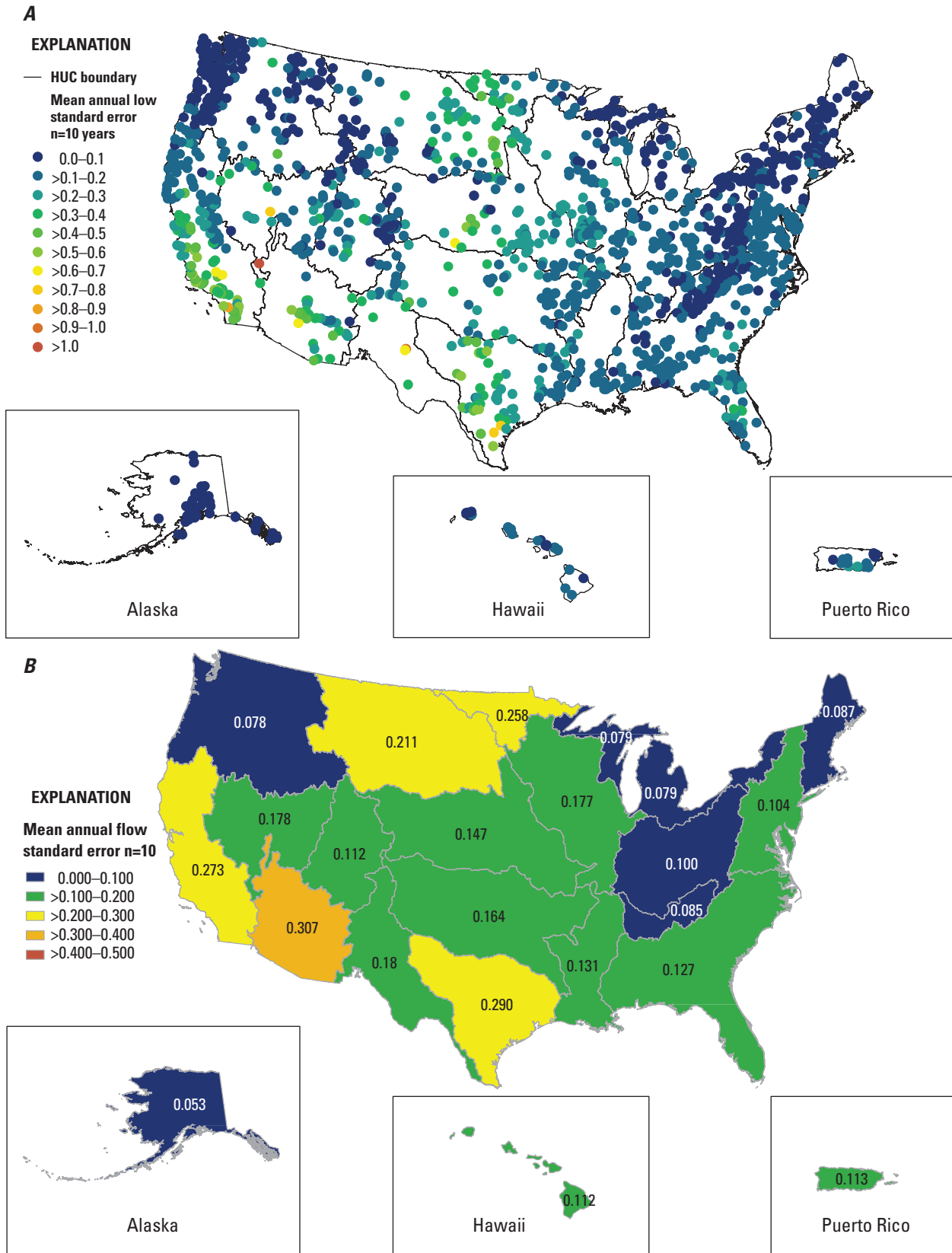| HUC region		CV
1	New England	0.27
2	Mid-Atlantic	0.33
3	South Atlantic-Gulf	0.40
4	Great Lakes	0.25
5	Ohio	0.32
6	Tennessee	0.27
7	Upper Mississippi	0.56
8	Lower Mississippi	0.41
9	Souris-Red-Rainy	0.82
10A	Upper Missouri	0.46
10B	Lower Missouri	0.67
11	Arkansas-White-Red	0.52
12	Texas-Gulf	0.92
13	Rio Grande	0.56
14	Upper Colorado	0.36
15	Lower Colorado	0.97
16	Great Basin	0.56
17	Pacific Northwest	0.25
18	California	0.86
19	Alaska	0.17
20	Hawaii	0.35
21	Caribbean	0.36

The number of years of record needed varies with the desired level of uncertainty within each HUC region. For example, in New England (HUC region 1), to achieve a fractional standard error of 0.10 (10 percent) for the mean annual flow, a record of less than 10 years is needed. As a comparison, in the Upper Mississippi Basin (HUC region 7), a record of more than 30 years is needed to obtain the same standard error of the mean. In the Lower Colorado Basin (HUC region 15), a record of 75 years is required to estimate the mean annual flow with an uncertainty of 10 percent. If the desired uncertainty were relaxed to 20 percent, then the number of years of record needed for Lower Colorado decreases to about 19 years.

Monthly Flow Statistics

The uncertainty of mean monthly flows, maximum monthly flows, and minimum monthly flows are discussed in this section. Uncertainty is directly related to the coefficient of variation of statistics. The data in table 6 show how the median coefficient of variation varies seasonally for mean monthly flow, minimum monthly flow, and maximum monthly flow in each HUC region. The data in table 7 show how the median coefficient of variation varies seasonally by level II ecoregion. The coefficient of variation cannot be calculated when the mean monthly flow statistic is equal to zero, so stations where this occurs are omitted from the calculations. The larger the coefficient of variation is, the larger the uncertainty of the long-term monthly flow statistics. Note that only a small number of streamgages were available in some HUC regions and ecoregions for these calculations. Estimated values of the coefficient of variation are more likely to be representative of the entire HUC region or ecoregion when many streamgages were included in the analysis.

In general, the spatial pattern of the coefficient of variation for monthly flows mirrors that seen for mean annual flow. The coefficient of variation is lowest in the Eastern United States and the Pacific Northwest and highest in the Central and Southwestern United States where the greatest year-to-year variability in the maximum monthly flow (the maximum daily flow that is seen in a particular month) occurs. Consequently, more years of record are required to obtain a stable estimate of the maximum monthly flow statistic than for the minimum monthly flow or mean monthly flow statistic. Minimum monthly flow generally has the lowest coefficient of variation; in other words, it has the smallest amount of variability from year to year. However, in some parts of the country and for some months, the coefficient of variation for minimum monthly flow is of similar magnitude as the coefficient of variation for mean monthly flow.

Using equation 2, the standard error of any of the monthly flow statistics can be estimated by applying the coefficient of variation for a HUC region or ecoregion and specifying the record length. Equation 3 allows the calculation of the number of years of record needed to achieve a desired level of uncertainty.

Figure 11. Maps showing *A*, standard error of mean annual flow for 10 years for each streamgage in the reference-quality streamgage dataset and *B*, median standard error of the mean annual flow for stations in each hydrologic unit code (HUC) region. In all HUC regions, a limited number of streamgages was available. >, more than or equal to.

A

B
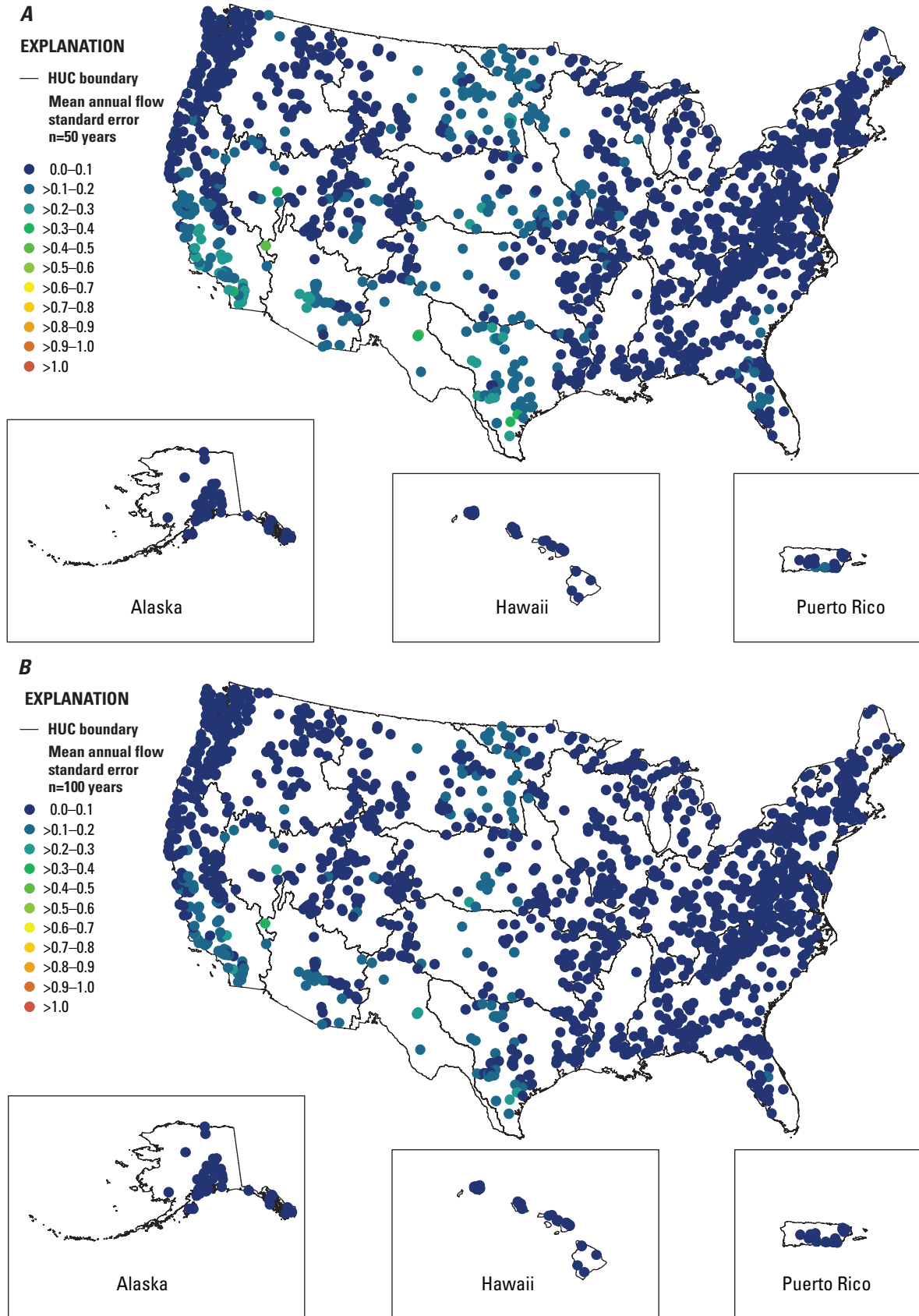
Figure 12.   Maps showing standard error of mean annual flow for *A*, 50 and *B*, 100 years. If 100 years of record were available, the standard error would be reduced to less than 0.2 for all but a few streamgages.

EXPLANATION

Water years

- Greater than 40
- 31 to 40
- 21 to 30
- 11 to 20
- 6 to 10
- Less than or equal to 5

Puerto Rico

Hawaii

Alaska
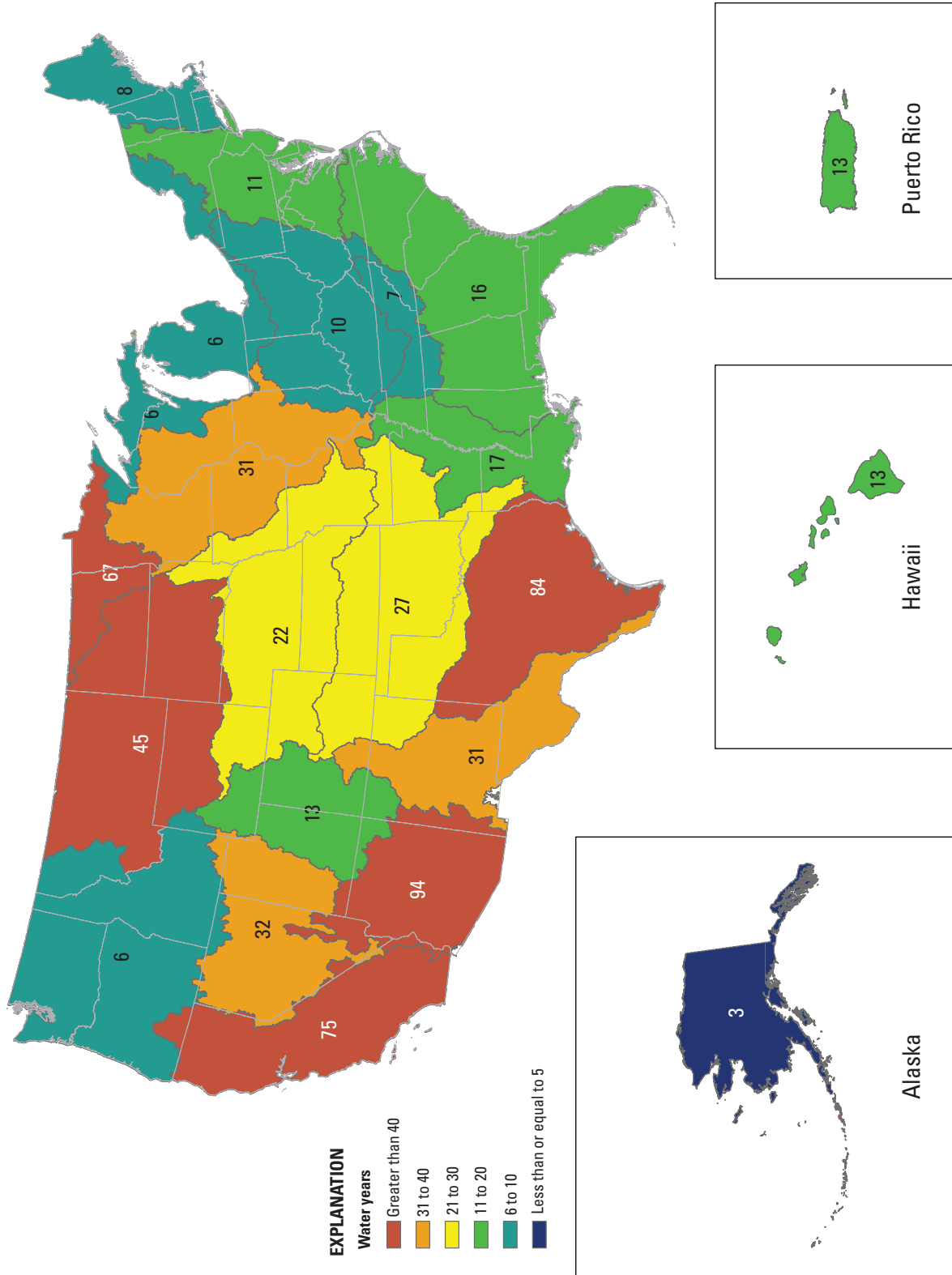
Figure 13. Map showing the number of years of record needed in each hydrologic unit code (HUC) region to estimate mean annual flow with a (fractional) standard error of 0.1 (10 percent). The number of years of record needed is calculated using the median coefficient of variation for each HUC region. The number of years of record needed at an individual location will vary. Only a limited number of streamgages was available in each HUC region. ≤, less than or equal to; ≥, more than or equal to.

Table 6. Median coefficient of variation for hydrologic unit code regions.

[HUC, hydrologic unit code; CV, coefficient of variation]

HUC region	Statistic	Median coefficient of variation											
		October	November	December	January	February	March	April	May	June	July	August	September
1 New England	Minimum	0.85	0.75	0.55	0.45	0.48	0.45	0.41	0.44	0.63	0.63	0.70	0.68
	Mean	1.06	0.65	0.63	0.63	0.67	0.55	0.40	0.47	0.70	0.84	1.04	1.06
	Maximum	1.30	0.80	0.91	0.98	0.94	0.74	0.54	0.66	1.13	1.21	1.40	1.49
2 Mid-Atlantic	Minimum	0.94	0.74	0.62	0.54	0.54	0.46	0.43	0.47	0.58	0.64	0.77	0.86
	Mean	1.13	0.82	0.67	0.65	0.60	0.48	0.52	0.56	0.94	0.95	1.22	1.45
	Maximum	1.53	1.14	1.01	1.04	0.96	0.77	0.80	0.96	1.63	1.40	1.81	2.07
3 South Atlantic-Gulf	Minimum	0.79	0.66	0.60	0.51	0.47	0.43	0.49	0.58	0.67	0.76	0.82	0.76
	Mean	1.11	1.10	0.81	0.66	0.60	0.60	0.72	0.90	1.01	1.00	0.94	1.14
	Maximum	1.56	1.44	1.11	0.98	0.95	0.94	1.10	1.24	1.40	1.29	1.42	1.62
4 Great Lakes	Minimum	0.57	0.47	0.38	0.38	0.35	0.42	0.43	0.43	0.51	0.45	0.42	0.51
	Mean	0.73	0.53	0.45	0.46	0.49	0.50	0.37	0.54	0.60	0.68	0.73	0.79
	Maximum	0.99	0.72	0.73	0.70	0.79	0.68	0.52	0.74	0.89	1.04	1.14	1.24
5 Ohio	Minimum	1.11	1.00	0.82	0.64	0.54	0.44	0.44	0.60	0.76	0.86	1.00	1.09
	Mean	1.44	1.00	0.78	0.65	0.58	0.53	0.53	0.76	1.03	1.11	1.22	1.55
	Maximum	1.72	1.24	0.92	0.85	0.80	0.78	0.82	1.00	1.31	1.38	1.48	1.83
6 Tennessee	Minimum	0.47	0.47	0.52	0.46	0.39	0.31	0.32	0.33	0.37	0.41	0.40	0.42
	Mean	0.74	0.80	0.62	0.53	0.46	0.47	0.44	0.50	0.58	0.63	0.66	0.84
	Maximum	1.26	1.21	0.94	0.89	0.77	0.83	0.81	1.04	1.16	1.17	1.31	1.54
7 Upper Mississippi	Minimum	1.59	1.33	1.16	1.03	0.99	0.76	0.68	0.75	0.91	1.28	1.25	1.54
	Mean	1.72	1.51	1.27	1.12	0.93	0.85	0.87	1.00	1.07	1.42	1.56	1.93
	Maximum	1.88	1.62	1.58	1.47	1.20	1.01	1.13	1.19	1.33	1.58	1.78	2.07
8 Lower Mississippi	Minimum	0.34	0.44	0.60	0.53	0.51	0.46	0.51	0.50	0.51	0.50	0.39	0.39
	Mean	1.47	1.23	0.93	0.83	0.72	0.67	0.88	1.14	1.05	1.17	1.17	1.66
	Maximum	2.20	1.69	1.22	1.11	0.99	0.99	1.25	1.45	1.83	2.06	1.79	2.33
9 Souris-Red-Rainy	Minimum	1.35	1.40	1.15	1.27	1.17	1.80	1.44	1.10	1.26	1.58	1.86	1.63
	Mean	1.49	1.53	1.26	1.15	1.74	1.41	1.01	1.25	1.39	1.49	2.00	1.75
	Maximum	2.03	1.72	1.36	1.09	2.89	1.57	1.14	1.44	1.59	1.96	2.23	1.96
10A Upper Missouri	Minimum	0.56	0.48	0.59	0.68	0.55	0.61	0.63	0.79	0.75	0.66	0.56	0.64
	Mean	0.58	0.44	0.45	0.61	0.67	0.67	0.90	0.82	0.96	0.88	0.83	0.80
	Maximum	0.70	0.61	0.54	0.69	1.18	1.21	1.21	1.27	1.24	1.17	1.02	1.09
10B Lower Missouri	Minimum	1.51	1.16	1.06	1.06	1.02	0.86	0.81	0.92	1.08	1.34	1.27	1.46
	Mean	1.80	1.43	1.23	1.21	1.06	1.11	0.99	1.10	1.22	1.80	1.70	1.81
	Maximum	2.21	1.59	1.54	1.62	1.45	1.33	1.29	1.26	1.35	1.72	1.89	1.90

Table 6. Median coefficient of variation for hydrologic unit code regions.—Continued

[HUC, hydrologic unit code; CV, coefficient of variation]

HUC region	Statistic	Median coefficient of variation											
		October	November	December	January	February	March	April	May	June	July	August	September
11 Arkansas-White-Red	Minimum	1.45	1.23	0.90	0.75	0.66	0.63	0.67	0.80	0.94	1.23	1.43	1.70
	Mean	1.85	1.32	1.07	0.92	0.81	0.80	0.86	0.98	1.27	1.39	1.65	1.81
	Maximum	2.13	1.78	1.52	1.40	1.35	1.23	1.29	1.25	1.61	1.74	2.17	2.23
12 Texas-Gulf	Minimum	1.64	1.58	1.15	1.23	1.11	1.07	1.06	1.15	1.61	1.79	2.12	1.91
	Mean	2.24	2.04	1.86	1.45	1.76	1.36	1.45	1.43	1.99	2.37	2.37	2.17
	Maximum	2.62	2.59	3.12	1.91	2.14	1.97	2.06	1.76	2.12	3.03	3.04	2.57
13 Rio Grande	Minimum	0.43	0.46	0.40	0.36	0.31	0.35	0.51	0.62	0.76	0.56	0.56	0.48
	Mean	0.67	0.57	0.37	0.32	0.35	0.54	0.81	0.92	0.92	0.72	0.80	0.70
	Maximum	1.08	0.84	0.55	0.47	0.54	0.69	0.88	0.98	0.94	0.84	0.93	1.02
14 Upper Colorado	Minimum	0.50	0.42	0.37	0.36	0.34	0.38	0.60	0.66	0.61	0.68	0.53	0.44
	Mean	0.50	0.38	0.37	0.36	0.36	0.44	0.54	0.45	0.58	0.71	0.56	0.45
	Maximum	0.59	0.42	0.40	0.39	0.49	0.67	0.70	0.48	0.59	0.77	0.65	0.70
15 Lower Colorado	Minimum	0.70	0.98	0.93	1.05	1.30	1.13	1.29	1.23	1.36	0.89	0.88	0.67
	Mean	2.61	1.83	1.87	2.05	1.56	1.15	1.38	1.34	1.63	1.64	1.40	1.75
	Maximum	3.08	2.41	2.36	2.26	1.91	1.73	1.49	1.57	1.51	2.02	1.79	2.54
16 Great Basin	Minimum	0.47	0.41	0.42	0.45	0.46	0.43	0.54	0.71	0.85	0.77	0.59	0.51
	Mean	0.43	0.37	0.46	0.50	0.43	0.47	0.57	0.72	0.83	0.77	0.66	0.54
	Maximum	0.53	0.54	0.69	0.66	0.76	0.72	0.66	0.80	0.91	0.93	0.75	0.66
17 Pacific Northwest	Minimum	0.40	0.51	0.45	0.44	0.42	0.36	0.35	0.36	0.42	0.36	0.29	0.31
	Mean	0.60	0.57	0.53	0.49	0.49	0.44	0.35	0.36	0.48	0.45	0.37	0.46
	Maximum	0.87	0.80	0.75	0.70	0.76	0.64	0.58	0.46	0.62	0.59	0.57	0.75
18 California	Minimum	0.91	0.95	1.11	1.11	1.08	1.03	0.97	0.91	1.07	1.13	1.01	0.97
	Mean	1.47	1.56	1.45	1.26	1.19	1.07	1.14	1.00	1.02	1.10	1.12	1.18
	Maximum	2.36	2.09	2.01	1.57	1.50	1.26	1.52	1.36	1.11	1.15	1.35	2.23
19 Alaska	Minimum	0.39	0.41	0.37	0.42	0.40	0.41	0.39	0.65	0.38	0.34	0.34	0.40
	Mean	0.41	0.45	0.53	0.44	0.51	0.43	0.39	0.41	0.31	0.30	0.36	0.42
	Maximum	0.56	0.56	0.60	0.64	0.64	0.58	0.68	0.45	0.38	0.42	0.55	0.57
20 Hawaii	Minimum	0.63	0.62	0.58	0.56	0.61	0.61	0.56	0.56	0.64	0.53	0.64	0.67
	Mean	0.67	0.73	0.70	0.82	0.87	0.89	0.75	0.73	0.68	0.62	0.70	0.72
	Maximum	0.94	0.95	0.99	1.08	1.07	0.99	0.97	1.01	0.94	0.89	0.99	1.00
21 Caribbean	Minimum	0.48	0.46	0.41	0.39	0.33	0.34	0.40	0.44	0.48	0.50	0.62	0.58
	Mean	0.70	0.77	0.65	0.58	0.52	0.51	0.69	0.80	0.65	0.61	0.66	0.85
	Maximum	1.10	1.23	1.30	1.54	1.22	0.99	1.14	0.94	0.90	1.02	0.95	1.45

Table 7. Median coefficient of variation for level II ecoregions in the United States.

[Level II ecoregions are from Commission for Environmental Cooperation (1997). CV, coefficient of variation; HUC, hydrologic unit code; n/a, not available]

HUC region	Statistic	October	November	December	January	February	March	April	May	June	July	August	September
2.2 Alaska Tundra	Minimum	0.79	1.05	1.94	2.37	2.52	2.54	2.54	2.42	0.60	0.41	0.65	0.59
	Mean	0.74	0.90	1.24	2.21	2.47	2.53	2.46	1.10	0.41	0.47	0.60	0.60
	Maximum	0.99	0.81	1.10	1.99	2.41	2.52	2.55	1.04	0.50	0.74	1.02	0.80
2.3 Brooks Range Tundra	Minimum	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Mean	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Maximum	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
3.1 Alaska Boreal Interior	Minimum	0.54	0.46	0.45	0.39	0.41	0.42	0.44	0.72	0.50	0.43	0.38	0.46
	Mean	0.52	0.50	0.49	0.39	0.41	0.42	0.69	0.45	0.52	0.44	0.65	0.54
	Maximum	0.78	0.55	0.49	0.44	0.41	0.42	1.07	0.55	0.61	0.71	1.53	0.73
3.2 Taiga Cordillera	Minimum	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Mean	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	Maximum	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
5.2 Mixed Wood Shield	Minimum	0.49	0.44	0.34	0.32	0.30	0.38	0.52	0.42	0.49	0.47	0.37	0.51
	Mean	0.72	0.53	0.44	0.33	0.36	0.62	0.34	0.54	0.57	0.63	0.67	0.76
	Maximum	0.99	0.66	0.63	0.47	0.57	0.81	0.47	0.67	0.76	0.91	1.04	1.03
5.3 Atlantic Highlands	Minimum	0.85	0.71	0.51	0.48	0.50	0.47	0.41	0.45	0.59	0.59	0.72	0.70
	Mean	1.05	0.63	0.60	0.64	0.65	0.50	0.41	0.49	0.76	0.90	1.10	1.17
	Maximum	1.34	0.86	0.96	1.01	0.96	0.73	0.55	0.75	1.22	1.29	1.52	1.59
6.1 Boreal Cordillera	Minimum	0.39	0.34	0.34	0.31	0.28	0.28	0.30	0.75	0.37	0.23	0.27	0.40
	Mean	0.37	0.37	0.34	0.33	0.29	0.29	0.37	0.43	0.33	0.22	0.25	0.41
	Maximum	0.52	0.43	0.38	0.35	0.34	0.31	0.60	0.47	0.39	0.27	0.43	0.58
6.2 Western Cordillera	Minimum	0.37	0.44	0.43	0.41	0.38	0.37	0.42	0.49	0.53	0.47	0.36	0.34
	Mean	0.50	0.53	0.50	0.47	0.48	0.44	0.40	0.40	0.53	0.54	0.43	0.42
	Maximum	0.79	0.75	0.75	0.68	0.74	0.65	0.58	0.45	0.59	0.66	0.57	0.68
7.1 Marine West Coast Forest	Minimum	0.50	0.59	0.48	0.46	0.46	0.37	0.35	0.38	0.36	0.35	0.32	0.41
	Mean	0.71	0.58	0.52	0.49	0.49	0.45	0.40	0.41	0.44	0.41	0.42	0.54
	Maximum	0.89	0.75	0.67	0.64	0.69	0.67	0.71	0.62	0.76	0.60	0.78	0.84
8.1 Mixed Wood Plains	Minimum	0.85	0.73	0.53	0.51	0.48	0.48	0.40	0.51	0.60	0.59	0.67	0.74
	Mean	1.02	0.66	0.58	0.59	0.60	0.44	0.47	0.53	0.86	0.89	1.05	1.20
	Maximum	1.36	0.86	0.86	0.92	0.87	0.68	0.63	0.84	1.35	1.32	1.43	1.65
8.2 Central USA Plains	Minimum	1.33	1.21	0.84	0.80	0.66	0.60	0.49	0.56	0.76	0.96	0.97	1.30
	Mean	1.69	1.31	0.94	1.03	0.79	0.67	0.64	0.88	1.06	1.30	1.57	1.97
	Maximum	1.98	1.37	1.21	1.18	1.09	0.91	0.87	1.18	1.42	1.51	2.03	2.11
8.3 Southeastern USA Plains	Minimum	0.91	0.75	0.67	0.54	0.49	0.43	0.49	0.58	0.67	0.79	0.87	0.86
	Mean	1.38	1.09	0.83	0.68	0.63	0.60	0.70	0.93	1.03	1.14	1.28	1.41
	Maximum	1.83	1.48	1.11	1.03	0.96	0.94	1.04	1.19	1.43	1.48	1.72	1.99
8.4 Ozark and Ouaçguta-Appalachian Forests	Minimum	0.86	0.79	0.69	0.57	0.52	0.42	0.44	0.51	0.58	0.61	0.68	0.77
	Mean	1.21	0.97	0.76	0.65	0.57	0.53	0.54	0.65	0.94	0.94	1.05	1.42
	Maximum	1.66	1.36	1.02	0.98	0.90	0.83	0.89	1.02	1.41	1.37	1.56	2.02

Table 7. Median coefficient of variation for level II ecoregions in the United States.—Continued

[Level II ecoregions are from Commission for Environmental Cooperation (1997). CV, coefficient of variation; HUC, hydrologic unit code]

HUC region	Statistic	Median coefficient of variation											
		October	November	December	January	February	March	April	May	June	July	August	September
8.5 Mississippi Alluvial and SE USA Coastal Plains	Minimum	1.10	0.96	1.02	0.88	0.71	0.71	0.93	0.98	1.36	1.23	1.15	1.13
	Mean	1.31	1.29	1.26	0.91	0.89	0.90	1.07	1.28	1.34	1.13	1.14	1.30
	Maximum	1.46	1.53	1.49	1.03	1.14	1.01	1.23	1.47	1.51	1.26	1.31	1.47
9.2 Temperate Prairies	Minimum	1.79	1.45	1.28	1.27	1.26	1.29	0.99	1.03	1.24	1.71	1.90	1.95
	Mean	1.98	1.82	1.49	1.45	1.34	1.22	1.22	1.20	1.36	1.96	1.99	1.98
	Maximum	2.23	1.98	1.81	1.80	1.84	1.33	1.33	1.38	1.47	1.99	2.12	2.07
9.3 West Central Semiarid Prairies	Minimum	0.76	0.64	0.71	0.82	1.03	0.87	0.81	0.93	0.86	0.90	0.96	0.83
	Mean	1.24	0.66	0.62	0.93	1.27	1.01	0.99	1.25	1.12	1.27	1.21	1.19
	Maximum	1.49	0.87	0.72	1.20	1.80	1.33	1.39	1.50	1.34	1.58	1.54	1.86
9.4 South Central Semiarid Prairies	Minimum	1.68	1.37	1.12	1.23	1.11	1.06	1.03	1.13	1.42	1.45	1.92	1.99
	Mean	2.18	1.68	1.53	1.36	1.42	1.36	1.32	1.40	1.62	2.04	1.93	1.86
	Maximum	2.63	2.45	2.58	1.91	1.95	1.99	2.02	1.81	1.95	2.48	2.31	2.30
9.5 Texas-Louisiana Coastal Plain	Minimum	1.33	1.16	1.30	1.15	1.29	0.97	1.03	1.31	1.51	1.95	1.40	1.31
	Mean	2.10	2.42	2.10	1.54	2.03	2.12	2.20	1.59	2.01	2.62	2.34	2.57
	Maximum	2.53	2.69	2.69	1.53	2.03	2.00	2.15	1.62	1.95	2.97	3.09	2.58
9.6 Tamaulipas-Texas Semiarid Plain	Minimum	3.36	2.36	3.03	3.32	3.88	3.37	3.35	2.68	4.93	3.50	4.67	3.83
	Mean	3.04	3.50	4.22	4.00	3.92	3.22	2.87	2.41	2.74	3.15	3.81	2.91
	Maximum	3.44	4.01	4.98	4.69	4.88	3.23	3.76	3.34	3.16	3.65	3.96	3.50
10.1 Cold Deserts	Minimum	0.43	0.41	0.49	0.51	0.47	0.56	0.60	0.72	0.77	0.70	0.59	0.50
	Mean	0.50	0.47	0.56	0.62	0.55	0.60	0.65	0.74	0.80	0.76	0.66	0.55
	Maximum	0.63	0.64	0.91	0.89	0.88	0.75	0.85	0.86	0.89	0.80	0.77	0.71
10.2 Warm Deserts	Minimum	2.51	2.09	1.97	1.69	1.71	1.86	1.85	2.47	2.68	2.72	2.70	1.95
	Mean	2.95	3.12	2.38	2.59	2.55	1.91	1.67	2.11	2.77	2.39	2.33	2.66
	Maximum	2.97	3.67	2.77	2.88	2.54	2.64	2.65	2.07	3.02	2.68	2.69	3.00
11.1 Mediterranean California	Minimum	1.34	0.99	1.24	1.30	1.38	1.24	1.25	1.15	1.35	1.49	1.56	1.55
	Mean	1.86	2.20	1.76	1.44	1.48	1.37	1.44	1.17	1.18	1.39	1.54	1.49
	Maximum	2.63	2.64	2.20	1.80	1.73	1.76	1.90	1.55	1.22	1.45	1.74	2.87
12.1 Western Sierra Madre Piedmont	Minimum	1.24	1.50	1.10	1.25	1.21	1.23	1.28	1.46	1.67	1.81	1.42	1.23
	Mean	3.03	1.43	1.40	3.01	1.71	1.36	1.24	1.29	1.52	1.89	1.47	1.64
	Maximum	4.25	2.75	3.05	3.96	3.09	1.96	1.56	1.20	2.16	2.12	1.69	2.02
13.1 Upper Gila Mountains	Minimum	0.65	0.97	0.82	0.99	1.17	1.01	1.20	1.19	0.98	0.57	0.76	0.54
	Mean	2.39	1.65	1.78	1.80	1.44	1.06	1.31	1.19	1.23	0.83	1.08	1.48
	Maximum	3.01	2.31	2.28	2.13	1.80	1.56	1.13	1.36	1.22	1.82	1.70	2.54
15.4 Everglades	Minimum	1.06	1.39	1.25	1.42	1.33	2.02	1.15	0.81	1.36	1.42	0.92	0.93
	Mean	0.92	1.24	1.36	1.25	1.58	1.70	2.22	1.35	1.52	0.97	0.70	0.82
	Maximum	0.97	1.15	1.41	1.30	1.61	1.51	2.80	2.45	1.57	0.87	0.74	0.98
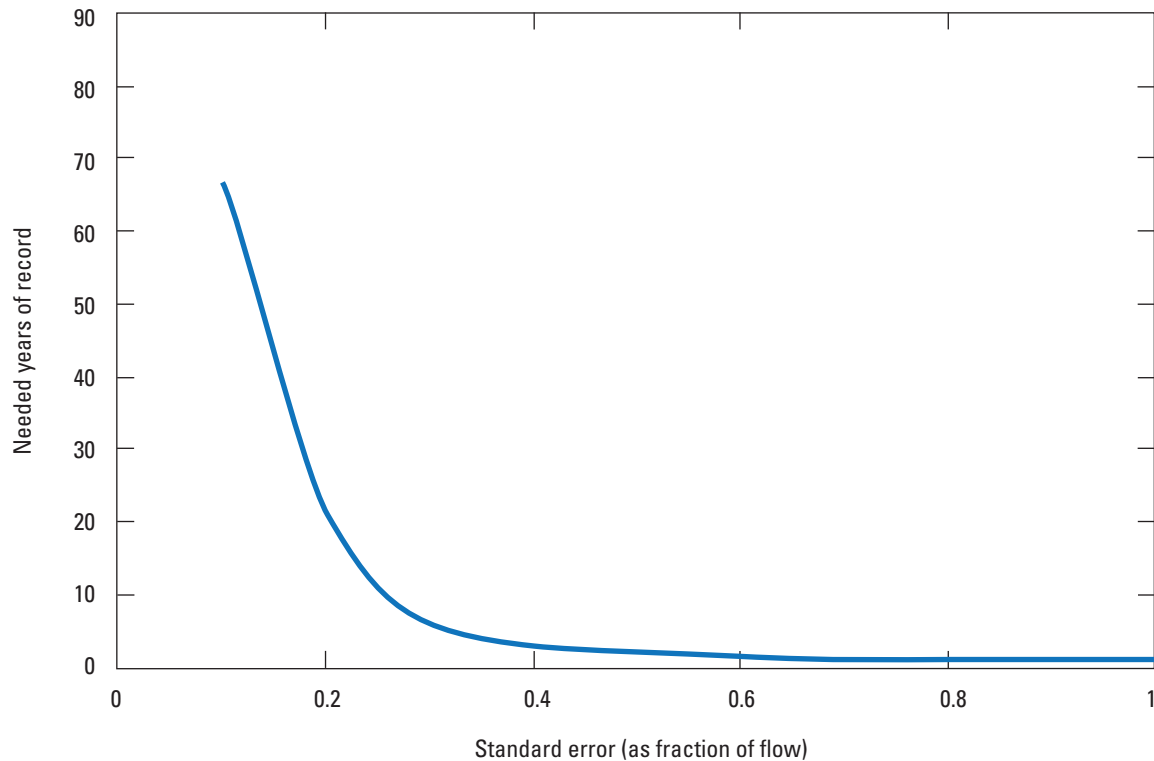
Figure 14.   Graph showing the number of years of streamflow record needed in hydrologic unit code region 9 (Souris-Red-Rainy) to estimate mean annual flow with a specified standard error.

Peak Flow Statistics

In the United States, the frequency of peak flows is calculated according to guidelines specified in Guidelines for Determining Flood Flow Frequency (Bulletin 17B; U.S. Interagency Advisory Committee on Water Data, 1982). According to Bulletin 17B, annual time series of peak flows are fit to a log-Pearson type III (LPIII) distribution with adjustments for historical information and low outliers. Most peak flow frequency studies conducted by the USGS have been done on a State-by-State basis, in cooperation with local agencies. Results of these studies are entered into the USGS National Streamflow Statistics (NSS) database (Ries, 2007).

The LPIII statistics for 1,200 stations were obtained by querying the NSS database and retrieving LPIII statistics (mean, standard deviation, and skew) for all the stations in the reference-quality streamgage dataset where this information was available. Information was not available for all streamgages in the reference-quality streamgage dataset, and the streamgages that were not included in this analysis are noted in appendix 1.

The standard error of various flood quantiles were obtained by entering the LPIII statistics into the Weighted Independent Estimates (WIE) software that calculates the standard error of flood frequency quantiles (Cohn and others, 2012). WIE

estimates the standard error assuming that the LPIII statistics pulled from NSS are valid for a systematic record with the specified record length. These estimates should be considered minimum uncertainties for these statistics, as they account for only one source of uncertainty.

Figure 15A shows the standard error of the 10 percent annual exceedance flood (10-year flood) for an assumed record length of 10 years where data were available from reference-quality streamgage locations, and figure 15B shows the same for an assumed record length of 100 years. With the longer record length, uncertainty is considerably diminished, and nearly all the United States shows an uncertainty of less than 20 percent. In general, the same spatial patterns of uncertainty that were observed for mean annual flows were also observed for peak flows. Areas with high interannual variability in flows, such as the Central and Southwestern United States, generally have the highest uncertainties. For peak flows, there is also a high uncertainty area in the Mid-Atlantic. Information on uncertainty of peak flows was not available for some States in the Mid-Atlantic.

Figure 16 shows the standard error of various exceedance level floods in all the HUC regions, all assuming that 10 years of record are available. The more extreme the flood event (lower exceedance probability), the greater the uncertainty in all regions of the country, although this is most pronounced in areas with large interannual variability in flood events.
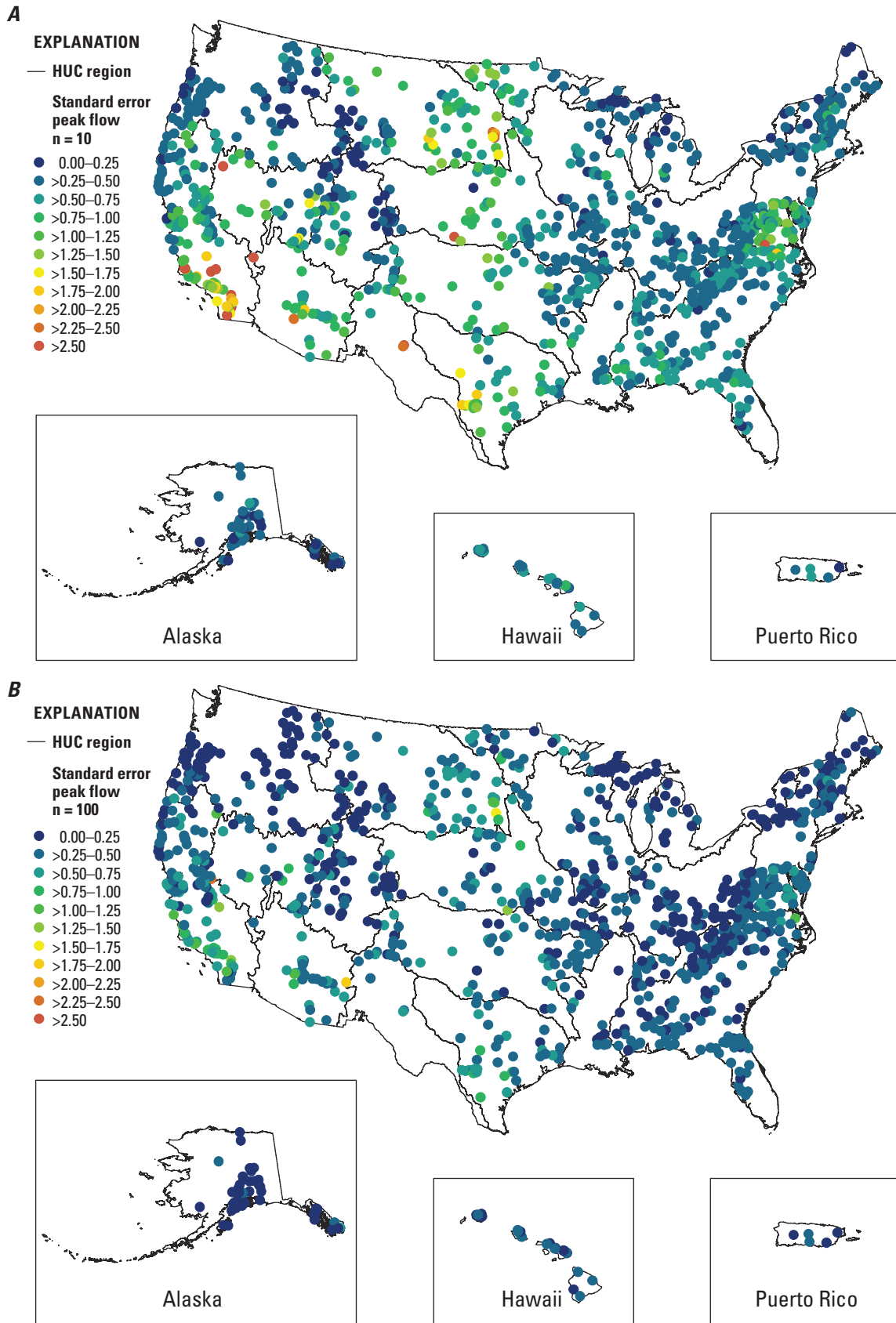
Figure 15. Maps showing standard error of 10 percent annual exceedance probability flood (10-year flood) for *A*, a record length of 10 years and *B*, for a record length of 100 years. >, more than.
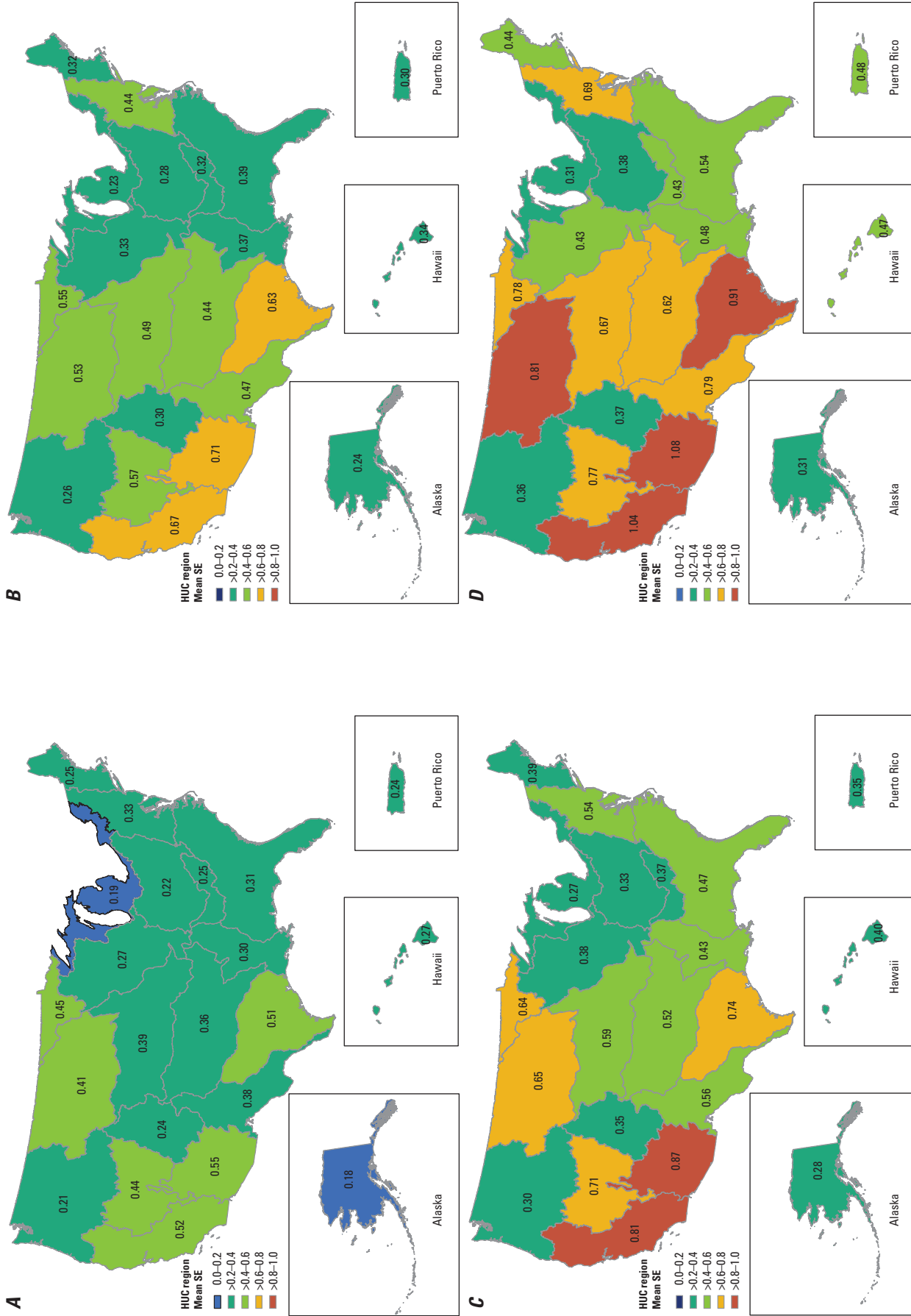
Figure 16. Maps showing standard error of selected peak flow quantiles by hydrologic unit code (HUC) region, all assuming a record length of 10 years and the same log-Pearson type III statistical distribution statistics. Annual exceedance probabilities equal *A*, 0.10 (10-year flood); *B*, 0.04 (25-year flood); *C*, 0.02 (50-year flood); and *D*, 0.01 (100-year flood). These estimates are based on the limited number of streamgages available in each HUC region. >, more than.

Low-Flow Statistics

Similar to peak flow studies, studies designed to calculate low-flow frequency statistics have been conducted by the USGS primarily on a State-by-State basis, in cooperation with local agencies. The most commonly calculated low-flow frequency statistics are the annual minimum 7-day average flow expected to occur every 10 and 2 years (7Q10 and the 7Q2, respectively). Both are calculated by fitting a time series of annual minimum 7-day average flows to a probability distribution. USGS studies most commonly use the LPIII distribution, as implemented in the Surface Water Statistics (SWSTAT) software.

Whereas it was possible to calculate the 7-day annual time series and estimate the 7Q10 and 7Q2 using estimated moments of the time series, a standard method for estimating the variance of the resulting estimate does not exist when a conditional probability adjustment is applied to account for zero flows. Nearly one-third of the reference-quality streamgages (530 of 1,630) experienced zero flow for at least one consecutive 7-day period during their period of record. Consequently, variance estimates were unavailable for nearly one-third of the reference-quality streamgages used in this study, particularly in the semiarid Central and arid Southwestern United States, but also sporadically throughout most of the country. As a result of the inability to calculate meaningful values of uncertainty for many streamgages, the uncertainty of low-flow statistics was omitted from this study.

Streamgage Gap Analysis for Ungaged Locations

At ungaged locations or locations with only limited data, it may not be possible to estimate statistics directly from the gaged record. At these locations, streamflow statistics can sometimes be calculated by transferring information from a continuous flow streamgage (index streamgage) with similar characteristics or by conducting a regional analysis of stream-flow statistics using many streamgages. Factors that affect the use of these methods include the length of record available at continuous record streamgages in the region and the hydro-logic similarity between the ungaged or limited-information watershed and other gaged watersheds in the region. This part of the study focuses on the following questions:

- Where has regional analysis been successful (Previous Regional Analyses section)?

- Are there long-term continuous record streamgages with sufficient record length and needed hydrologic similarity available (Availability of Index Streamgages section)?

Previous Regional Analyses

Regional regression equations to estimate statistics have been developed for many regions of the country, but they vary in terms of the flow statistics considered and the accuracy of the equations. The regression equations use more easily calculated basin characteristics to estimate the value of flow statistics. Table 8 lists the basin attributes that have been commonly used in previous regression studies.

The National Streamflow Statistics (NSS) database was queried in November 2010 to obtain regional regression equations that are applicable in rural locations for peak flows (1 percent and 10 percent annual exceedance probability flows), low flows (7Q10), and average flows (mean and median annual flows). Whereas peak flow equations were available in NSS for nearly all States in the United States (exceptions were Kansas, Oklahoma, and Louisiana), low and average flow equations were available for only about one-third of the States. The available equations were analyzed to summarize the most commonly used basin characteristics and the variability in equation accuracy across the country.

Drainage area was always included as an explanatory variable in the regression equations obtained from NSS, as was also found by Benson and Carter (1973). Measures of precipitation, elevation, and slope were also commonly included as significant variables for all types of flow. Measures of precipitation included mean annual precipitation as well as measures of seasonal precipitation or the intensity of precipitation events. Less commonly used but also appearing in many equations were variables describing the soils in the basin (soil permeability or soil type), land cover (impervious area, forest cover, or grass cover), and the amount of water storage available in the basin. In addition, information on temperature or snow was often used in colder climates. Other variables that were sometimes used included measures of the basin shape, indicators of geology, and flow indices (such as a recession index or variability index). Table 9 categorizes the frequency with which variables were used for regression of different types of statistics.

The accuracy of the regional regression equations varies by flow statistic and by region. Two different metrics were commonly reported in previous regression studies, the standard error of the estimate and the standard error of prediction. The standard error of the estimate is a measure of how well the regression model fits the observed data; it is a measure of the goodness of fit of the regression model. The standard error of prediction is a measure of how well the regression equation might be expected to predict future observations; it is calculated using both a measure of the goodness of fit of the regression model and a term representing uncertainty in the model coefficients. Whereas both metrics are a measure of the uncertainty in model results, the two measures cannot be directly compared and so are mapped separately.

Figures 17–20 show the standard error of prediction in parts *A* and the standard error of the estimate in parts *B* for previous regressions of mean annual flow (fig. 17), 1 percent annual exceedance probability (100-year flood; fig. 18), 10 percent annual exceedance probability (10-year flood; fig. 19), and the 7-day average, 10-year minimum flow (7Q10; fig. 20). Uncertainties were generally lowest in the Eastern United States and highest in the Central and Southwestern United

Table 8.   Basin attributes commonly used in regression studies.

Attribute	Explanation
Drainage area	The drainage area is the upstream watershed area that drains to a streamgage. Some studies use the contributing drainage area, which excludes any upstream areas that form a closed basin and so do not contribute flow to the streamgage.
Elevation	Different studies may use maximum elevation in the basin, average elevation in the basin, or other similar variables.
Evapotranspiration	Water withdrawn from a land area by evaporation from water surfaces and moist soil plus plant transpiration. Different studies may use mean annual evapotranspiration or seasonal evapotranspiration or other measures.
Flow index	A numerical index calculated from streamflow records, such as an index of the rate of recession or variability in streamflow. A flow index must be estimated at an ungaged location.
Geology	Typically, the percentage of the basin that is underlain by a particular geology.
Impervious area/ development index	A variable that is intended to characterize the extent of urbanization. Most commonly, this is the percentage of the basin that is covered by impervious surfaces but other measures of development are also used.
Land cover (natural)	Typically, the percentage of the basin that is covered by a particular land cover type, such as forest or grassland.
Latitude	The latitude of the streamgage or basin centroid.
Longitude	The longitude of the streamgage or basin centroid.
Precipitation	A measure of precipitation in the basin. Different studies may use annual maximum precipitation, average annual precipitation, different measures of seasonal precipitation, or an index of precipitation intensity.
Shape	A measure of the shape of the basin, such as length divided by width.
Slope	A measure of the slope, or steepness, of the basin. This may be the average slope, slope of the main channel, the percentage of the basin where slope exceeds a threshold, or other permutations.
Snow	A measure of the amount of snowfall or the percentage of precipitation falling as snow.
Soils	A variable that characterizes the soils or soil permeability. This may be the percentage of the basin with a specified soil type, the average soil permeability, or other similar measures.
Stream density	A measure of the stream density, such as the total length of streams in a basin divided by the basin area.
Water storage	A measure of the upstream water storage in the basin. This may include the percentage of the basin covered by lakes, wetlands, or other water bodies.
Winter temperature	Average minimum temperature, average January temperature, or similar. It is used to characterize whether snow and ice are a factor in the basin.

Table 9.   Basin attributes appearing in regression equations developed by the USGS and included in the National Streamflow Statistics (NSS) database.

Frequency	Peak flow equations	Low flow equations	Average (mean and median) flow equations
Frequently used	Drainage area[1]	Drainage area[1]	Drainage area[1]
	Precipitation	Precipitation	Precipitation
	Slope		Elevation
Moderately used	Elevation	Slope	Soils
	Impervious area/development index	Elevation	Slope
	Water storage	Soils	Land cover (natural)
		Other	Latitude
Occasionally used or rarely used	Shape	Winter temperature	Flow Index
	Soils	Geology	Winter temperature
	Geology	Flow index	Water storage
	Land cover (natural)	Land cover (natural)	Impervious area/development index
	Snow	Impervious area/development index	Shape
	Winter temperature	Stream density	Longitude
	Evapotranspiration	Water storage	Geology
	Latitude	Shape	Stream density
	Longitude	Longitude	Other
	Other		

[1]Drainage area was used in all regression equations.

States. Differences in the available data when the studies were completed, the methods used, and other factors are likely to have contributed to differences in the accuracy of the regression equation. In general, the results of peak flow equations had the lowest uncertainty, followed by mean annual flow statistics. Low-flow statistics generally had the largest uncertainties and were sometimes extremely large. For example, the standard error of prediction of the low-flow regression equations in Colorado and Arkansas exceeded 200 percent (fig. 20*A*).

Availability of Index Streamgages

At locations where measured streamflow information is limited or nonexistent, estimates of streamflow statistics can sometimes be made by transferring information from streamgages with a long-term record. The watershed upstream of the long-term streamgage, however, must exhibit similar hydrologic behavior to the watershed upstream of the location of interest.

Selection of an appropriate index streamgage is a key to the success of information transfer techniques. One method of selecting an index streamgage relies on finding streamgages where the streamflow is highly correlated with flow at the location of interest. If there is a short streamflow record available, the short time series can be used to estimate correlations between the short record site and nearby long-term sites. When the location of interest is ungaged, Archfield and Vogel (2010) showed that correlation maps can be effective in identifying possible index streamgages. Correlation maps have been developed for all the United States and can indicate which streamgages are the least correlated to others in the streamgage network. Because the least correlated streamgages are most unlike the others, these streamgages are important for the independent information they provide. Locations where the analysis shows that there are no highly correlated streamgages could be considered for addition of new streamgages, should resources become available.

Another method for selecting index streamgages is the hydrologic similarity between the watershed of interest and nearby gaged watersheds. The more similar they are in terms of basin attributes—such as drainage area size, mean annual precipitation, elevation, geology, or land use—the more likely they are to be useful as an index site. Regionalization of statistics is typically accomplished by relating values of easily measurable basin attributes to the values of flow statistics of interest and then assessing the range of basin attributes covered by the streamgage network. Ideally, the streamgage network would cover the full range of known basin attributes.

Correlation Analysis

Streamflow is the expression of the complex interaction between climate, geology, and land surface conditions in its contributing catchment. Correlation between two streamflow time series could be interpreted as a measure of how similarly two catchments (pair of streamgages) respond to these interactions and therefore can be used to understand similarities and uniqueness in the hydrologic response of the network. Furthermore, methods to estimate daily streamflow at ungaged locations commonly require transferring the timing of streamflows from an index streamgage to an ungaged location, making correlation an important factor in deciding which index streamgage to select. For ungaged locations, Archfield and Vogel (2010) introduced the map-correlation method, which can be used to select the index streamgage estimated to have the highest correlated daily streamflows with the ungaged location.

Correlations between streamflow time series at reference-quality streamgage locations within each major HUC region (fig. 1*B*) were evaluated to answer the following questions:

- What is the distribution of correlation across the streamgage network for daily streamflow and high- and low-flow events (Estimation of Correlation section)?

- Can the streamgage network provide highly correlated index streamgages suitable for estimating daily streamflow and high- and low-flow events across the entire network (Distribution of Correlations Across the United States section)?

- Are there regions of the United States where distance can be used as a surrogate for the selection of the highest correlated streamgage (Relation Between Distance and Correlation section)?

- If distance is not a good surrogate for correlation, how well is the map-correlation method able to estimate correlation between a gaged and ungaged location (Map Correlation Across the United States section)?

Estimation of Correlation

The correlation of daily streamflow between each possible pairing of the 1,630 study streamgages was estimated by HUC region. Correlations were estimated using the coincident, observed streamflow at each pair of streamgages. The map-correlation method was then applied to determine the utility of the method to estimate correlations at ungaged locations. To examine the availability of highly correlated index streamgages so that daily streamflow in sparsely measured locations could be estimated, correlations were computed for three coincident periods—the entire coincident daily streamflow record, the daily streamflow record for a high-flow period (the period when streamflows at both streamgages were greater than the 0.1 exceedance probability), and the daily streamflow record for a low-flow period (the period when streamflows at both streamgages were less than the 0.9 exceedance probability).

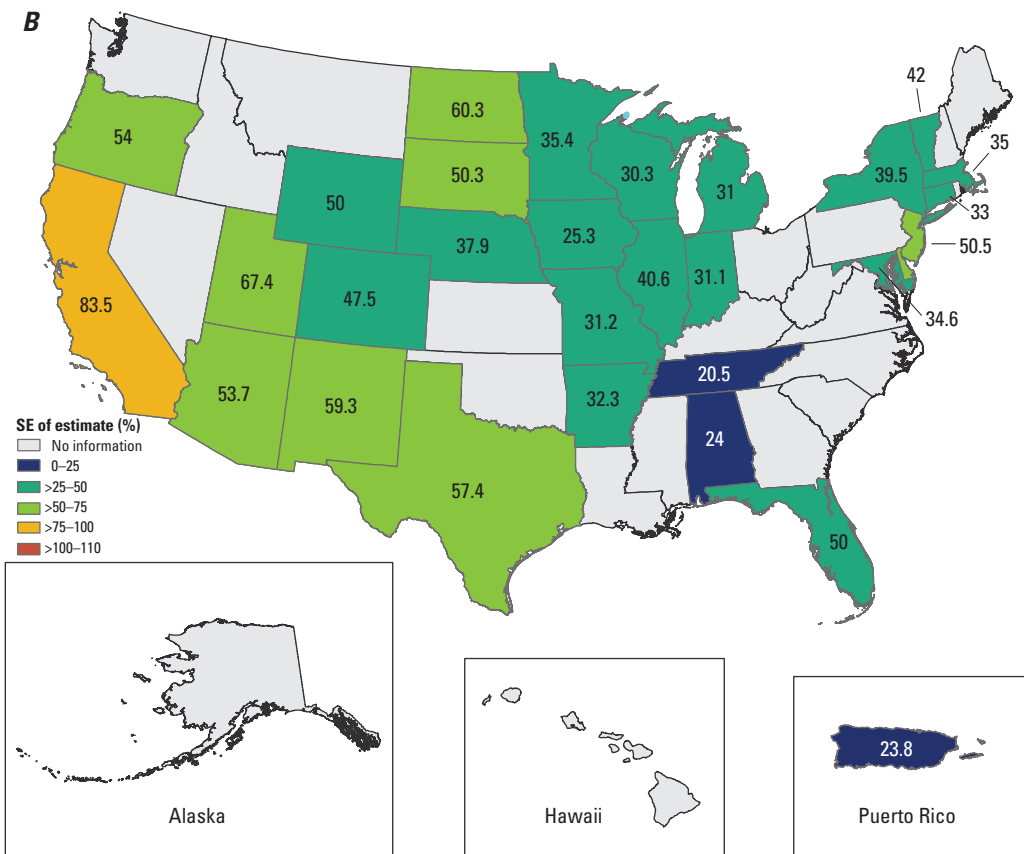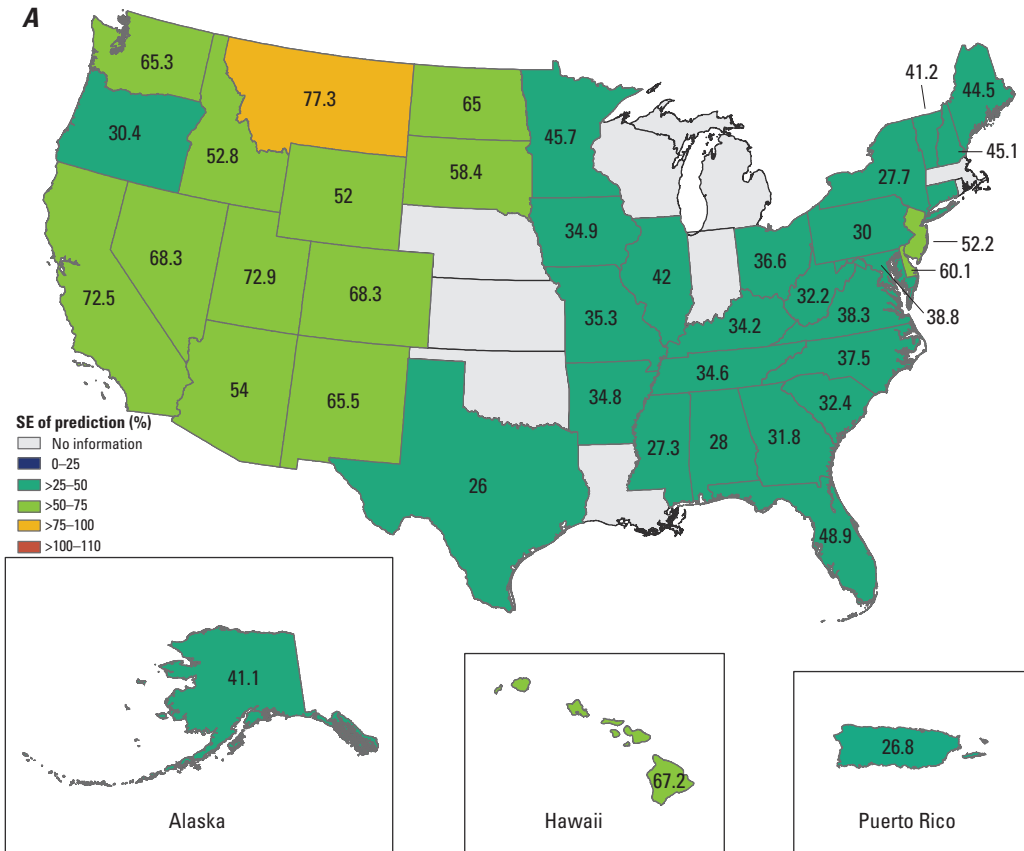### Estimation of Correlation Using Observed Streamflow

For each HUC region in the United States, correlations between coincident streamflows at each pair of streamgages

A

SE of prediction (%)

- No information
- 0–25
- >25–50
- >50–75
- >75–100
- >100–110

B

SE of estimate (%)

- No information
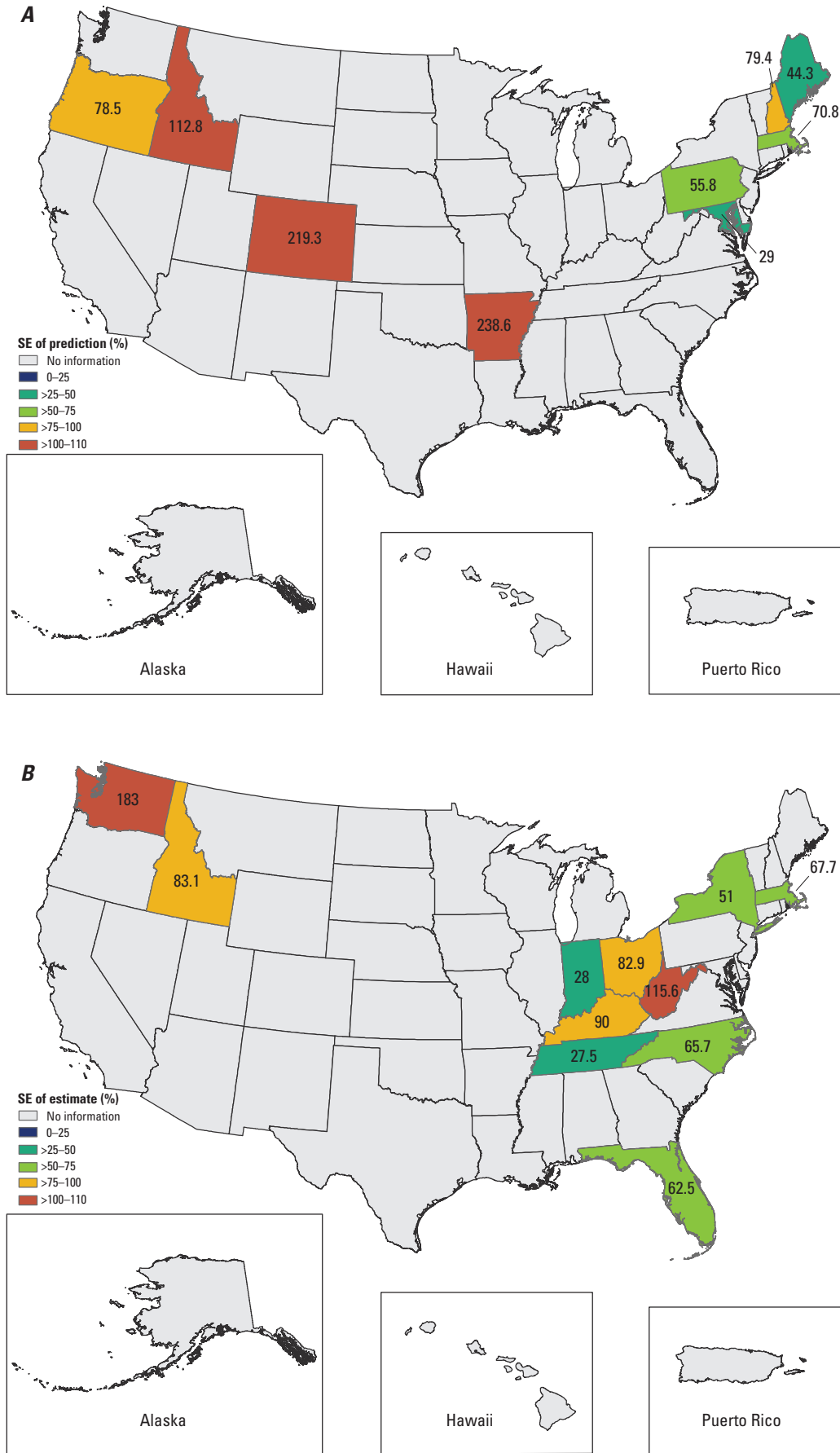- 0–25
- >25–50
- >50–75
- >75–100
- >100–110
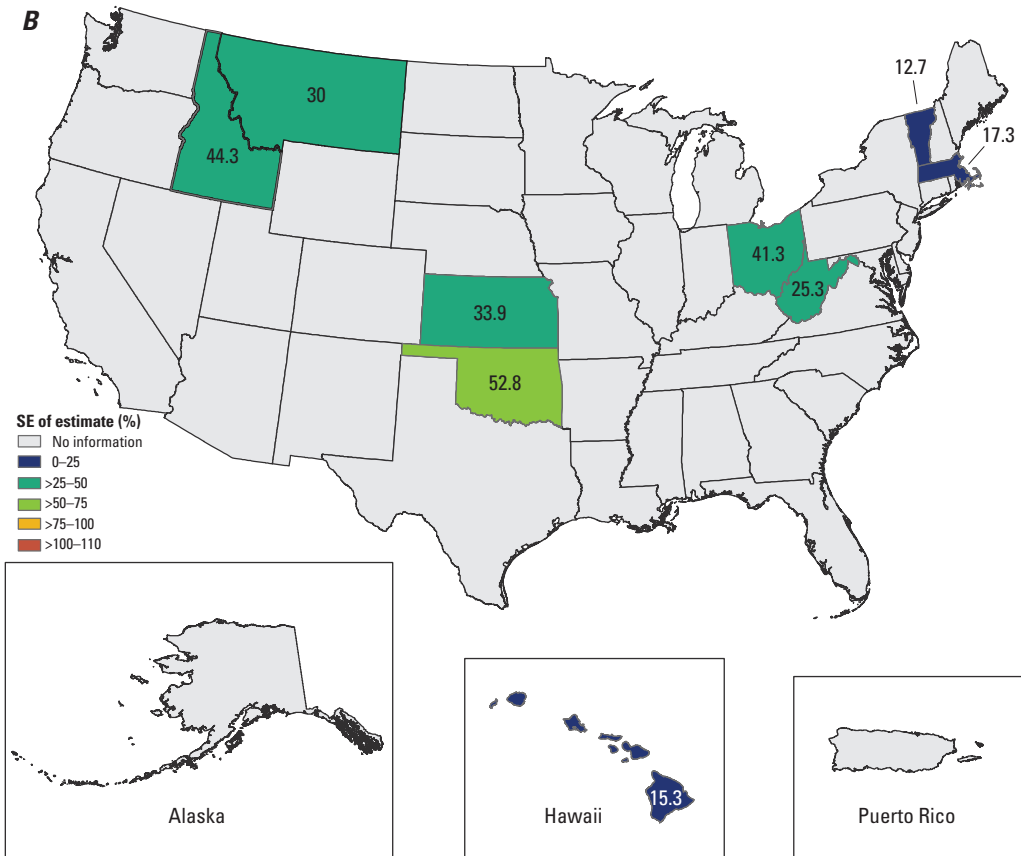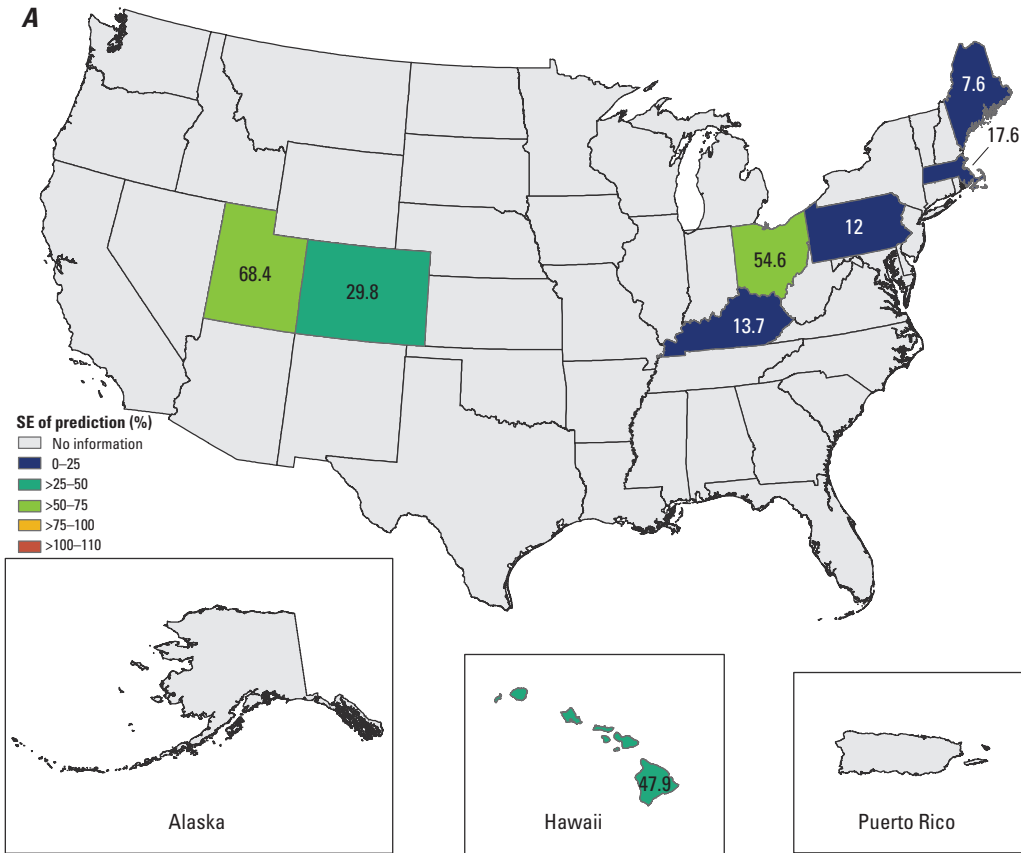
Figure 17.   Maps showing the standard error (SE) of *A*, prediction and *B*, the estimate for regression equations used to predict mean annual flow. %, percent; >, more than.
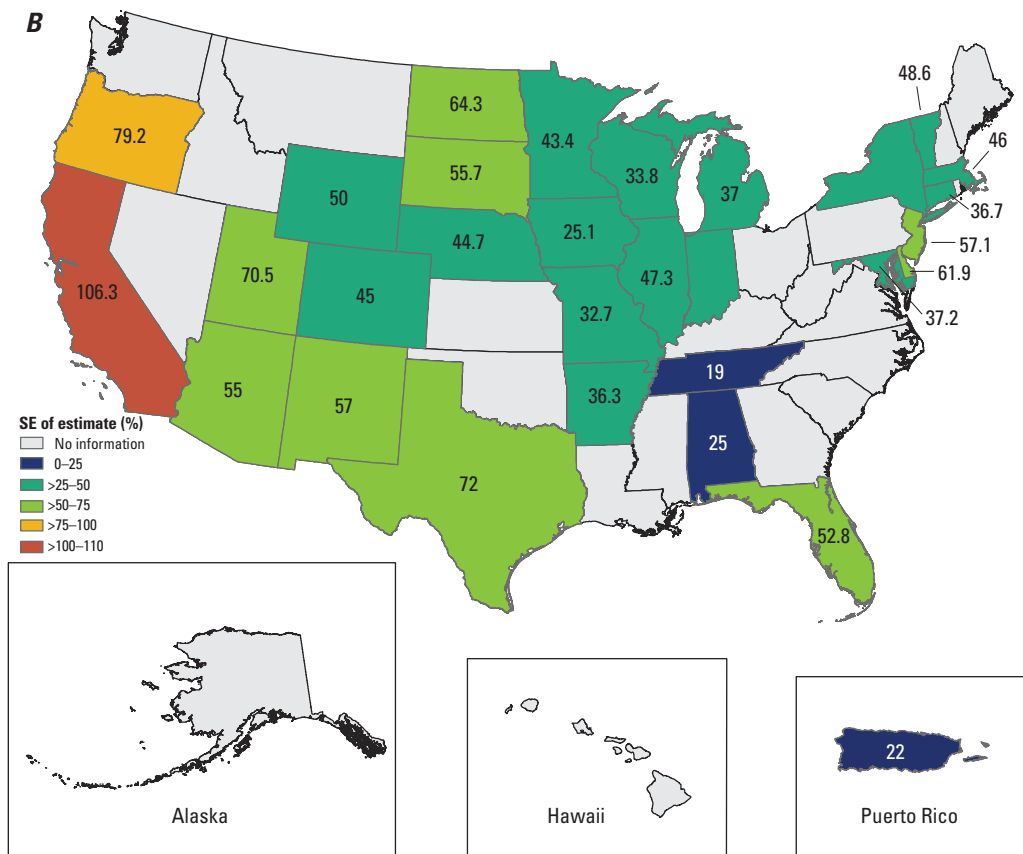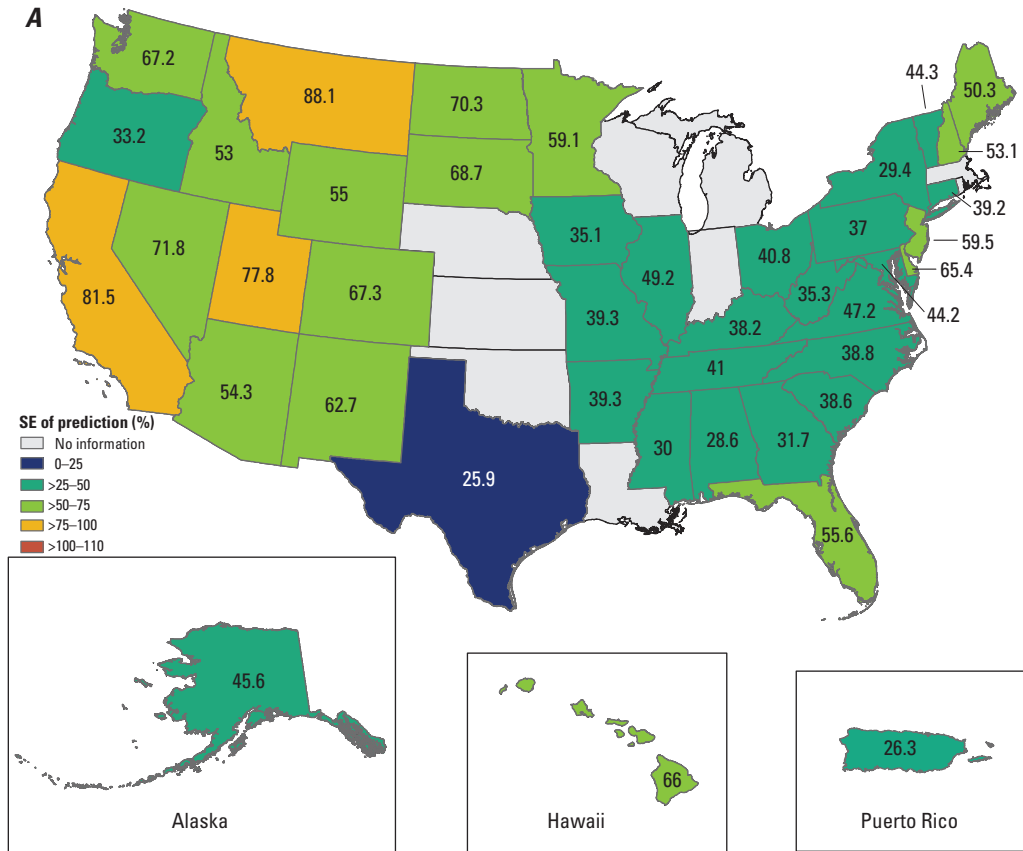
Figure 18. Maps showing the standard error (SE) of A, prediction and B, the estimate for regression equations used to predict the 1 percent annual exceedance probability flood (100-year flood). %, percent; >, more than.

A

SE of prediction (%)
- No information
- 0–25
- >25–50
- >50–75
- >75–100
- >100–110

Alaska

Hawaii

Puerto Rico

B

SE of estimate (%)
- No information
- 0–25
- >25–50
- >50–75
- >75–100
- >100–110

Alaska

Hawaii

Puerto Rico

Figure 19.   Maps showing the standard error (SE) of *A*, prediction and *B*, the estimate for regression equations used to predict the 10 percent annual exceedance probability flood (10-year flood). %, percent; >, more than.

A

78.5

112.8

219.3

238.6

79.4

44.3

70.8

55.8

29

SE of prediction (%)
No information
0–25
>25–50
>50–75
>75–100
>100–110

Alaska

Hawaii

Puerto Rico

B

183

83.1

51

67.7

28

82.9

115.6

90

27.5

65.7

62.5

SE of estimate (%)
No information
0–25
>25–50
>50–75
>75–100
>100–110

Alaska

Hawaii

Puerto Rico

Figure 20. Graphs showing the standard error (SE) of *A*, prediction and *B*, the estimate for regression equations used to predict the 7-day average, 10-year minimum flow (7Q10). %, percent; >, more than.

were estimated using Pearson's correlation coefficient (r), which measures the linear correlation between two variables (Helsel and Hirsch, 1992); the correlation coefficient was selected because of its common use in the selection of index streamgages for record extension and patching. With this method, the natural logarithms of the coincident streamflows are computed to linearize the relation between the two time series before the correlation was estimated.

Pearson r values can range from -1 to 1, with a value of 1 indicating perfect correlation and a value of -1 indicating perfect negative correlation. A value of zero indicates that the two variables exhibit no correlation. There are limitations to the use of r to estimate correlation, most notably that the natural logarithm cannot be computed if streamflow is zero. For this analysis, zero-valued observed streamflows were removed from the record before the correlation was estimated. As a result, for some pairs of streamgages, the number of coincident streamflows may be too small to result in a reliable estimate of r. For this reason, only r values statistically significant at the 90 percent confidence level were reported.

To estimate the coincident period between streamflows for the high- and low-flow periods, the streamflows at the 0.1 and 0.9 exceedance probability were first estimated at each streamgage using the Weibull plotting position and a single-order statistic (as presented in Vogel and Fennessey, 1994, equation 2). The streamflow record was then truncated to contain only the periods of high or low flow. The dates for which each pair of streamgages were in a high- or low-flow period were recorded, and only the corresponding streamflows on those dates were used to compute the r value between the high- and low-flow periods, respectively.

Mapping of Correlation for Ungaged Locations

Using the methods described in Archfield and Vogel (2010), the correlation between a streamgage and an ungaged location was estimated using map correlation for each HUC region in the United States. Map correlation has been applied previously to regions in the Eastern United States (Archfield and Vogel, 2010); however, the utility of the method for estimating correlations at ungaged sites across the United States is unknown.

Map correlation is an application of ordinary kriging and is summarized here from Archfield and Vogel (2010). A spatial model, termed a variogram, is fit to the correlation values between each streamgage and the other streamgages in the study area, resulting in one spatial model per streamgage in the study area. This spatial model can then be used to estimate the correlation between the streamgage and an ungaged location. When there is more than one potential index streamgage, the spatial models can be used to discern which index streamgage results in the highest estimated correlation with the ungaged location.

To evaluate the utility of map correlation to estimate the correlation between an ungaged location and a particular streamgage, a leave-one-out cross-validation experiment was used. The streamgages used to develop the spatial model were removed—one at a time—and the parameters of the spatial model were estimated again without the removed site. The spatial model was then used to estimate the correlation at the removed site. Estimated correlations were then compared with the observed correlations using the root mean square error (RMSE) to determine the goodness-of-fit of the spatial model.

Distribution of Correlations Across the United States

Correlations were computed between each streamgage and all other streamgages in the same HUC region, resulting in a set of r values for each streamgage in the HUC region for all coincident streamflows (fig. 21), coincident high-flow events (fig. 22), and coincident low-flow events (fig. 23). The figures show box plots for the correlation between each streamgage and all other streamgages in the HUC region. The streamgages are shown in order of ascending station ID (not labeled on figs. 21–23). For all coincident streamflows, the distribution of correlations tend to be higher and less variable for HUC regions located in the Eastern United States, with the exception of the Southeastern United States (HUC region 3; fig. 21). By contrast, HUC regions in the Central United States (HUC regions 10 and 11) show substantially lower correlations between streamgages (fig. 21). Differences in the distribution of correlations across HUC regions do not appear to be dependent upon streamgage density.

Correlations between streamflows for coincident high-flow (fig. 22) and low-flow (fig. 23) periods tended to be lower and more variable than correlations for all coincident streamflows. For streamgages shown as a discontinuous line on figures 22 and 23, coincident high- or low-flow periods were not experienced at all or not frequently enough to produce a significant correlation. This finding suggests that there are streamgages in the network that respond uniquely to high- or low-flow conditions. This situation occurs more frequently for low flows (fig. 23) than for high flows (fig. 22), potentially suggesting that streamgage pairs may respond similarly to a high-flow event but not a low-flow event.

The maximum correlation at each streamgage was also mapped (fig. 24). Mapping the maximum correlation is the equivalent to selecting the highest correlation in each box plot and mapping that value. For correlations computed from coincident daily streamflow, most streamgages have streamflows highly correlated with at least one other streamgage; this is particularly true for streamgages located in the Eastern and Western United States with the exception of coastal locations in the Southeastern United States. For the Central and Southwestern United States, particularly the Midwest and Texas, there are very few highly correlated streamgages. For this region of the United States, the use of an index to transfer timing to an ungaged location may not yield reliable time-series estimates. Areas of high maximum correlations appear to follow a pattern closely related to mountainous regions of

the United States, whereas regions with lower elevations tend to have lower maximum correlations (fig. 24).

For each streamgage, the number of other streamgages with a correlation value greater than 0.8 was examined (fig. 25). If a streamgage was highly correlated with many other streamgages, then several streamgages would be appropriate for use as an index streamgage for that particular streamgage. Mountainous regions of the United States and southern Alaska appear to have many more streamgages that correlate well with other streamgages in their respective HUC regions. In contrast a large part of the Central United States, coastal areas in the Southeastern United States, Hawaii, and Puerto Rico have at least one highly correlated streamgage (fig. 24) but few other streamgages considered suitable as an index streamgage (fig. 25). For these regions, methods to select the index streamgage need to be precise because there is a greater chance of not choosing a highly correlated index streamgage.

For correlations between coincident streamflows for high- and low-flow periods, higher maximum correlation values are observed for high-flow periods compared with low-flow periods (figs. 26 and 27, respectively). Correlations for coincident high-flow periods behave similarly to correlations computed from all coincident streamflow values, with higher maximum correlations along the Eastern (excluding Southeastern) and Western United States and lower maximum correlations in the Interior and Southeastern United States. Correlations estimated from coincident low-flow periods suggest that there is more unique behavior in the response of the streamgage network to low-flow periods than to high-flow periods.

Relation Between Distance and Correlation

Given that most areas of the United States have at least one and sometimes many suitable streamgages for information transfer, the relation between distance and correlation is examined. If there are areas of the country where the nearest streamgage can provide a surrogate for the highest correlated index streamgage, then more sophisticated and data-intensive techniques to select the index streamgage need not be applied.

To determine if distance between streamgages could be used to predict correlation, a linear regression equation was fit between the distances, and r values computed between each pair of streamgages by HUC region. For example, there are 131 streamgages in New England (HUC region 1). Assuming that each of the other 130 streamgages has enough coincident record to yield a statistically significant r value, there are 130 other streamgages for each streamgage in New England to measure the distance between paired streamgage outlets and the correlation between coincident streamflow time series. The coefficient of determination (r^2) of the regression between distance and r was determined for all coincident streamflows (fig. 28), coincident high-flow events (fig. 29), and coincident low-flow events (fig. 30). A subset of relations between distance and r was selected and plotted to verify that the assumptions of linear regression were met, most notably that the relation between distance and r was linear.

Whereas distance and correlation are highly related for some streamgages, in general, distance was seen to be a poor predictor of correlation. The relatively weak ability of distance to act as a surrogate for correlation seen nationally confirms the findings of Archfield and Vogel (2010), which applied only to New England. For correlations determined from all coincident streamflows, areas with high r^2 values are located primarily in the Northeastern United States and along the southwestern part of the Mississippi River Basin (fig. 28). These geographic patterns do not appear related to streamgage density or to the number of highly correlated sites at the streamgage; if the geographic patterns adhered to these two criteria, then the Western United States also would have high r^2 values. For correlations determined from coincident high and low flows (figs. 29 and 30, respectively), there are either weak or no significant relations between distance and correlation, with the exception of a few streamgages in scattered locations across the country.

Map Correlation Across the United States

Distance, although sometimes highly related to correlation, does not appear to be a reliable way to select the most correlated index streamgage. Therefore, the utility of the map-correlation method was investigated to determine how well the correlation between a gaged and ungaged location could be estimated. Using the methods described in Archfield and Vogel (2010) and cross validation described in this report, RMSEs were determined from observed and map-correlation-estimated correlation values for coincident daily streamflows at each streamgage, coincident streamflows for high-flow periods, and coincident streamflows for low-flow periods. The RMSEs between observed and estimated correlations determined for the coincident daily streamflows show that the map-correlation method is able to provide good estimates of correlation at ungaged locations in the Northeastern and Mid-Atlantic United States and Puerto Rico and to a lesser extent in the Southeastern United States and the Great Lakes region (fig. 31). For other regions of the United States, the RMSEs are much larger, making the estimates unreliable for estimating correlation at ungaged locations (fig. 31). High RMSEs are also found nationwide for correlations determined for high- and low-flow periods (figs. 32 and 33, respectively).

There may be several reasons for the poor performance of the map-correlation method across much of the United States. The map-correlation method, as applied in this study, uses the variogram model form and binning strategy developed for applications of the map-correlation method to areas in the Eastern and Mid-Atlantic United States. Another spatial model form may be appropriate for streamgages in the Central and Western United States and in Alaska and Hawaii. Furthermore, map correlation has not been applied previously to correlations determined for only high- and low-flow periods, and other spatial model forms may be appropriate for this reason.
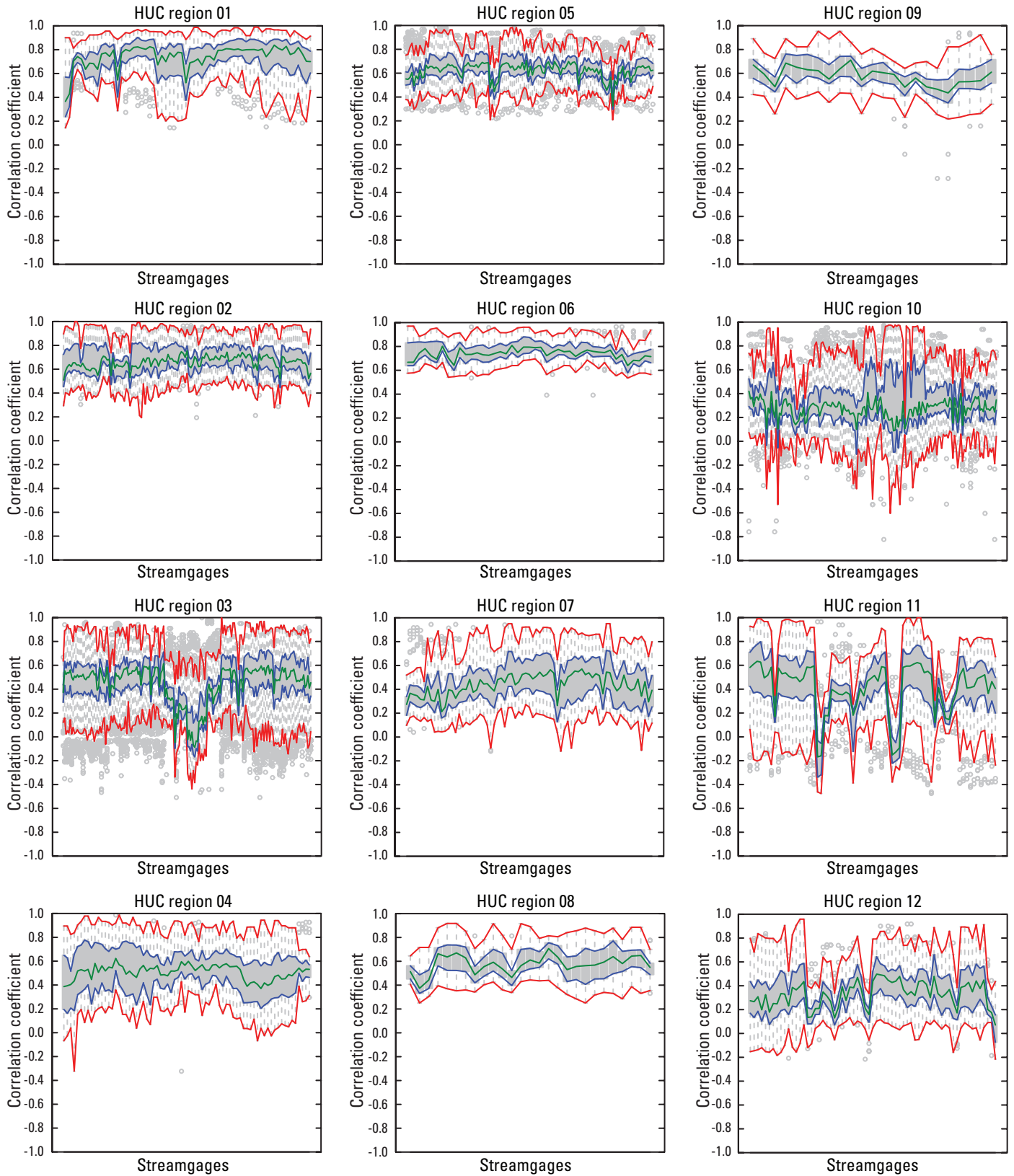
Figure 21.   Graphs showing the distribution of Pearson's correlation coefficient between coincident daily streamflows at each streamgage and all other streamgages in each hydrologic code (HUC) region. The streamgages within each HUC are plotted in ascending order by streamgage identification number (appendix 1).
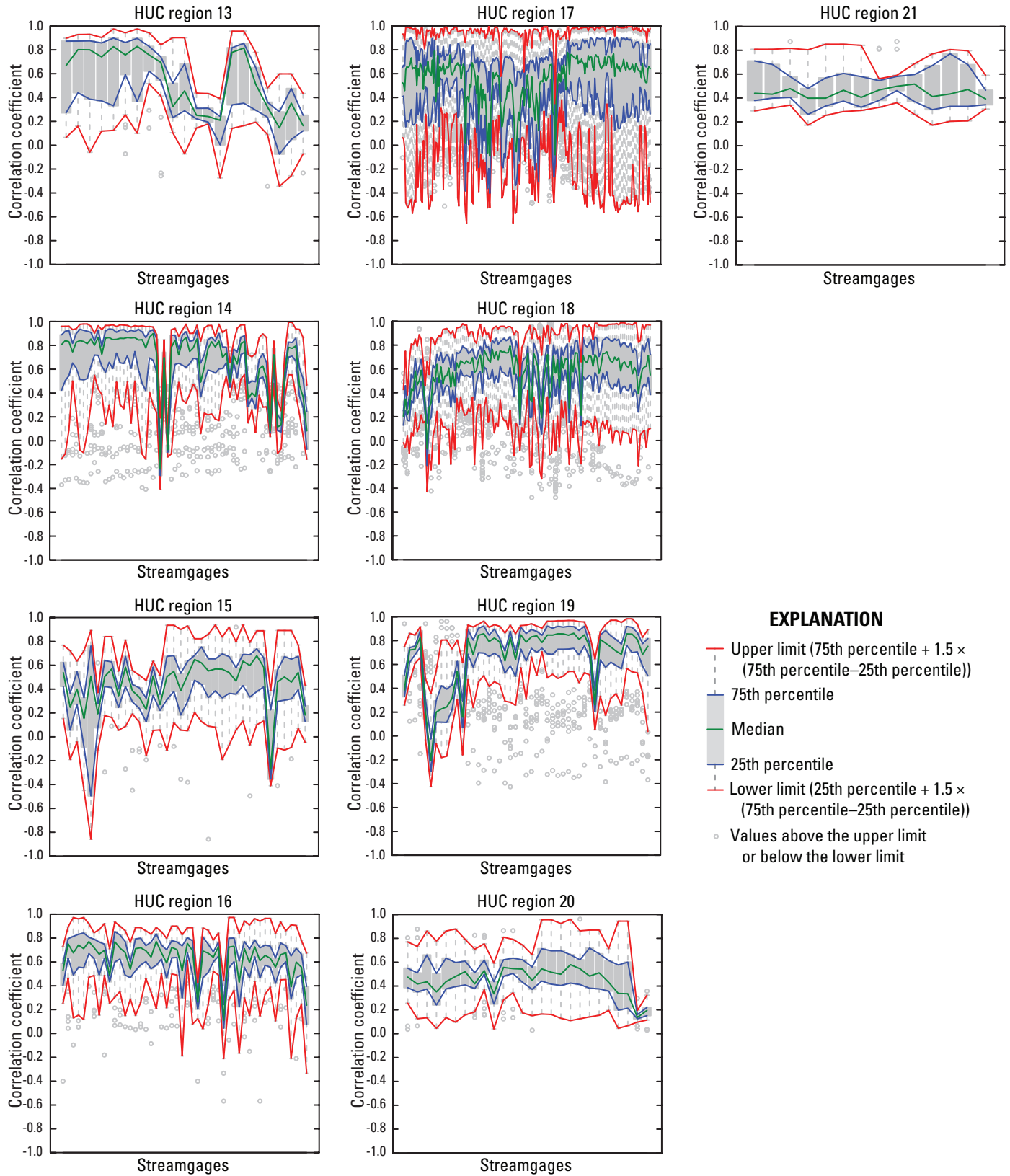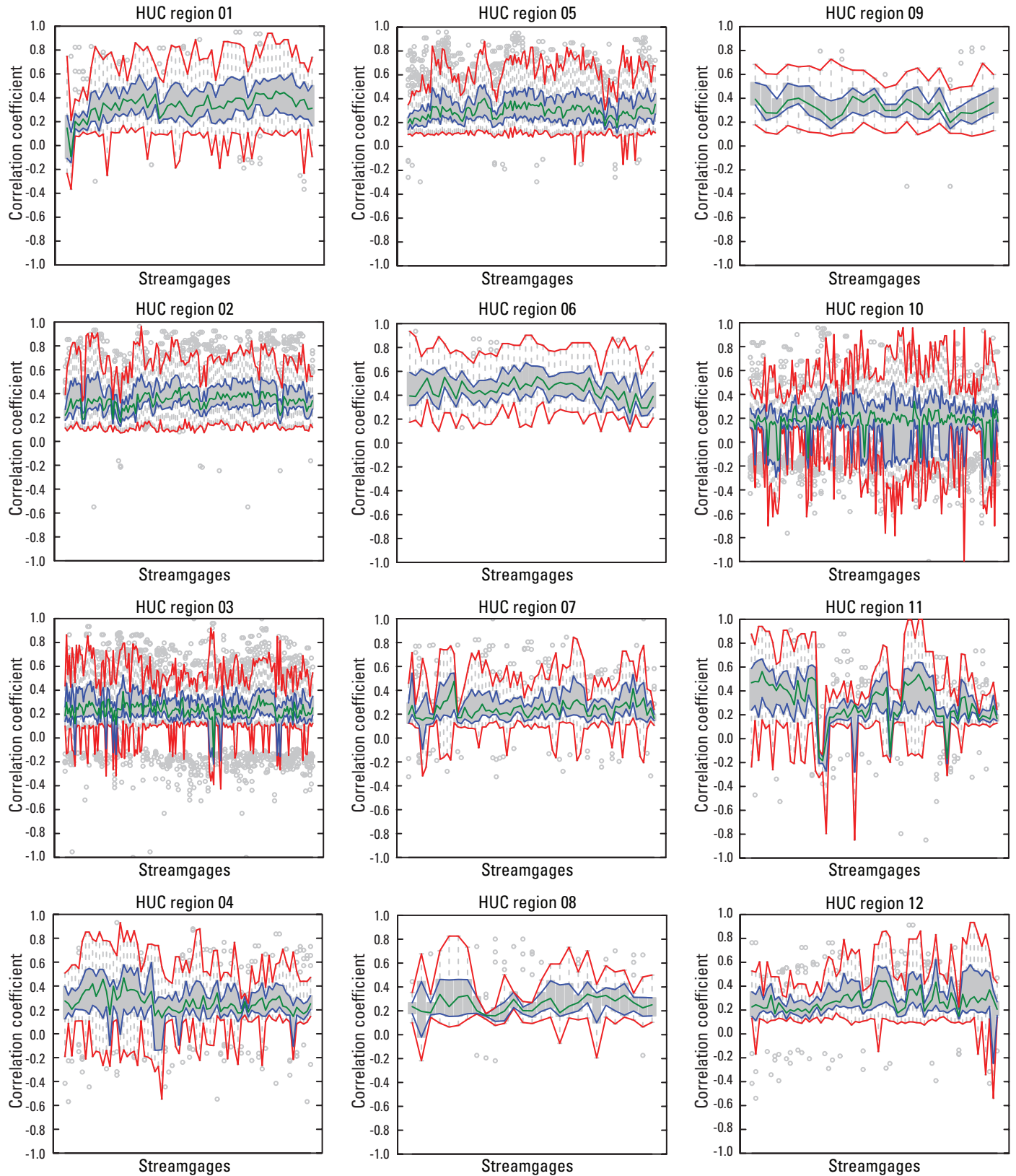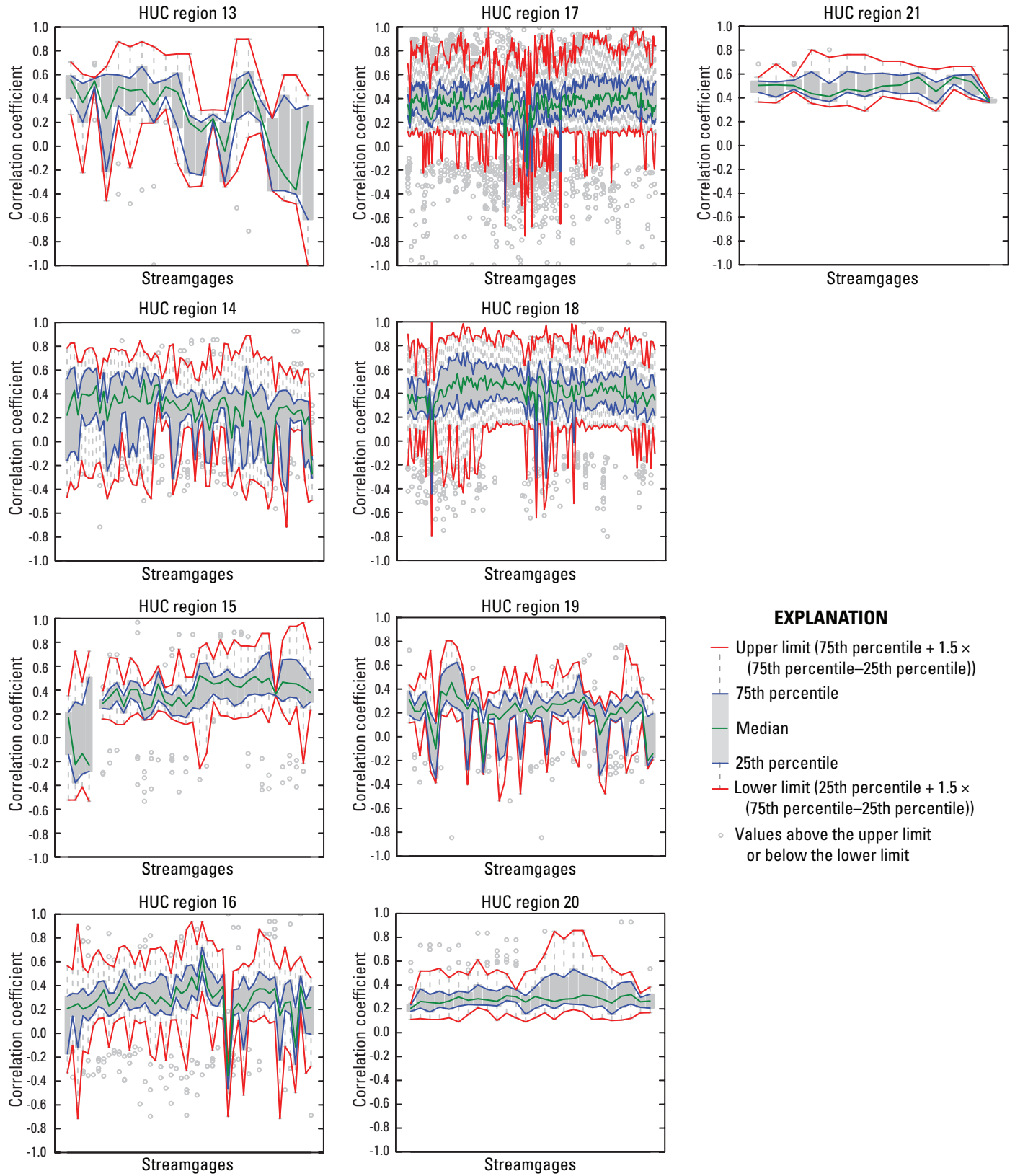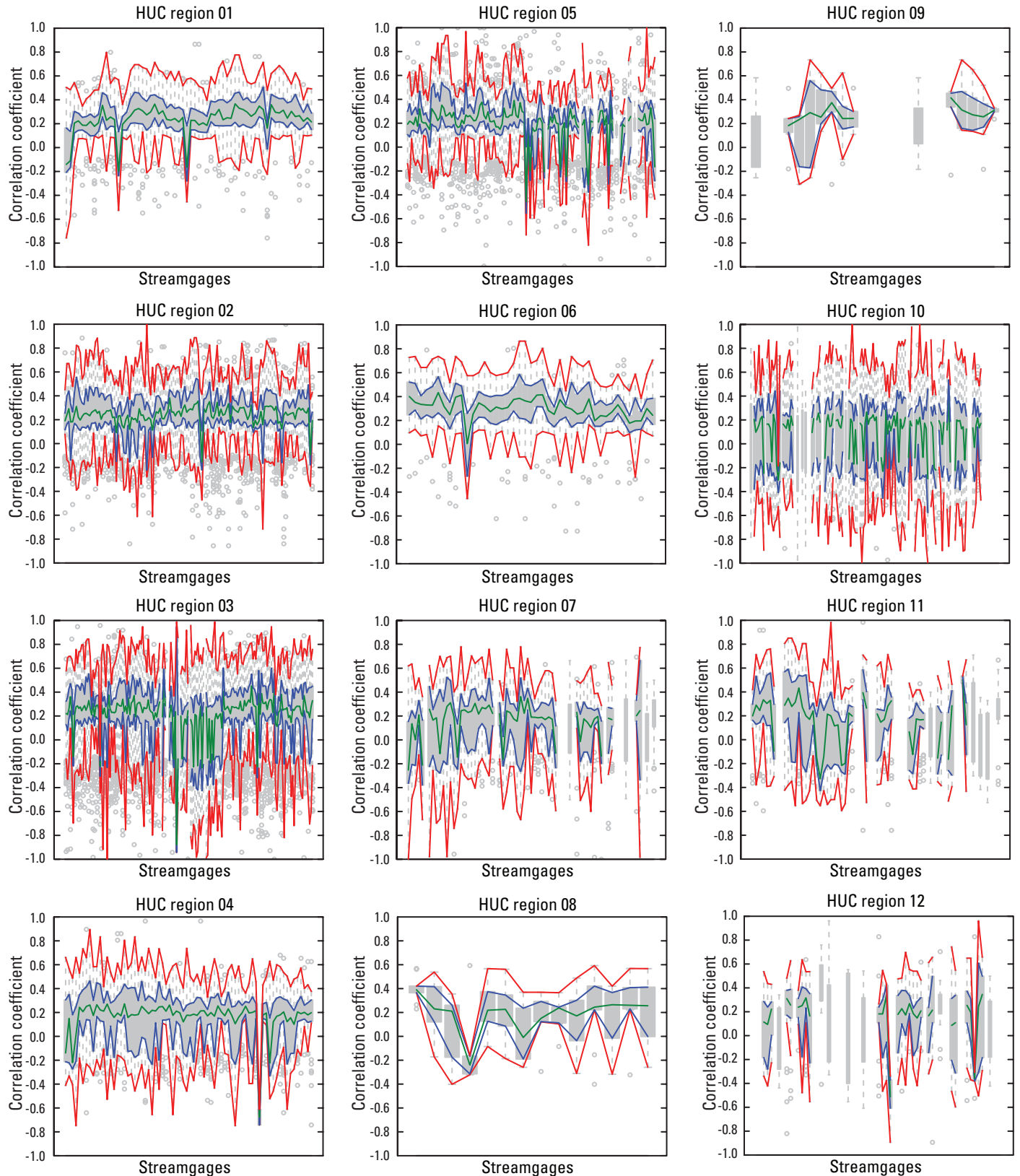
Figure 21.   Continued.

Figure 22.   Graphs showing the distribution of Pearson's correlation coefficient between coincident streamflows greater than the 0.1 exceedance probability (high flows) at each streamgage and all other streamgages in each hydrologic code (HUC) region. The streamgages within each HUC are plotted in ascending order by streamgage identification number (appendix 1). Discontinuous lines indicate periods where coincident flows were experienced either not at all or not frequently enough to produce a significant correlation.

Figure 22. Continued.

Figure 23. Graphs showing the distribution of Pearson's correlation coefficient between coincident streamflows less than the 0.9 exceedance probability (low flows) at each streamgage and all other streamgages in each hydrologic code (HUC) region. The streamgages within each HUC are plotted in ascending order by streamgage identification number (appendix 1). Discontinuous lines indicate periods where coincident flows were experienced either not at all or not frequently enough to produce a significant correlation.
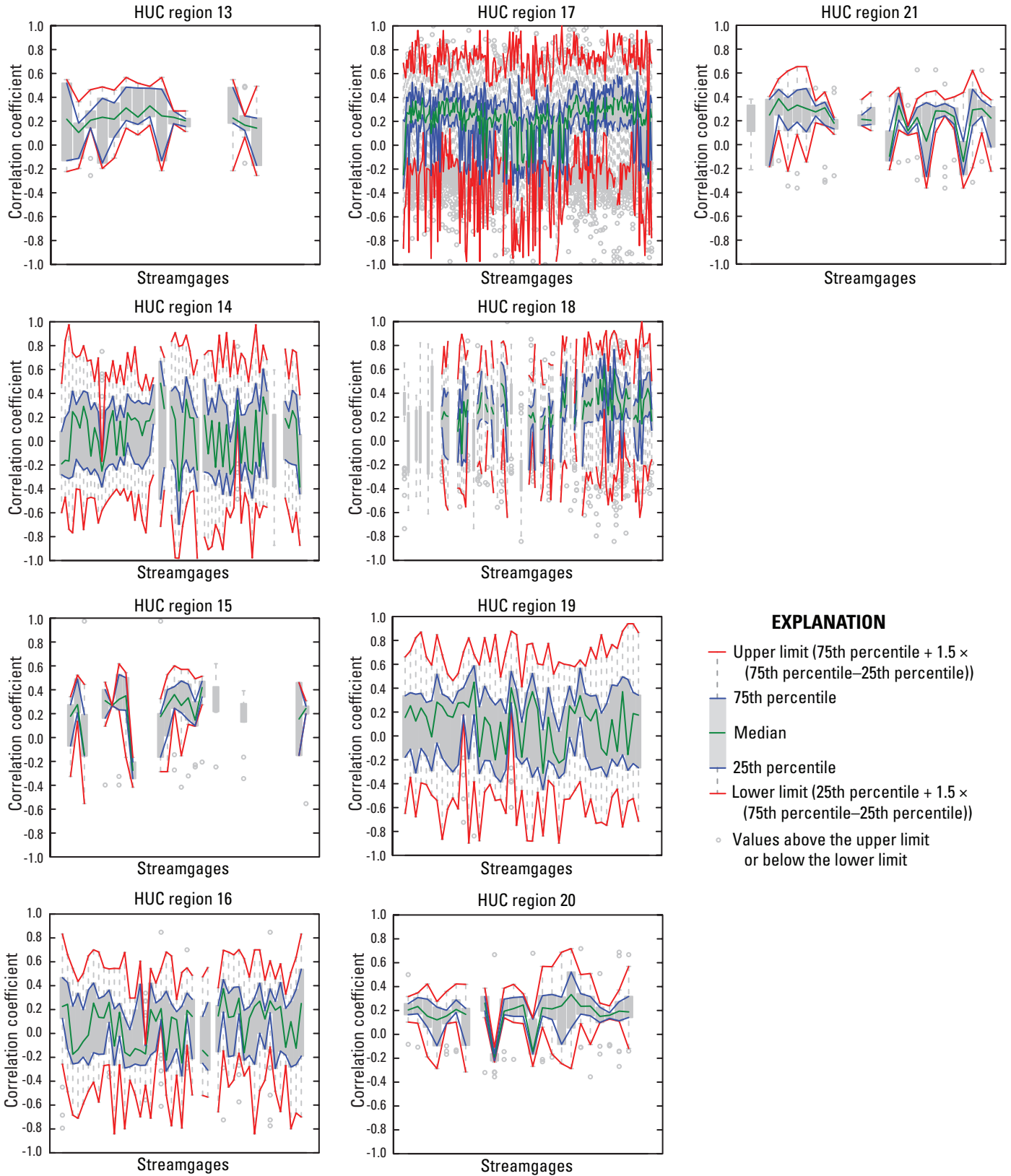
Figure 23. Continued.
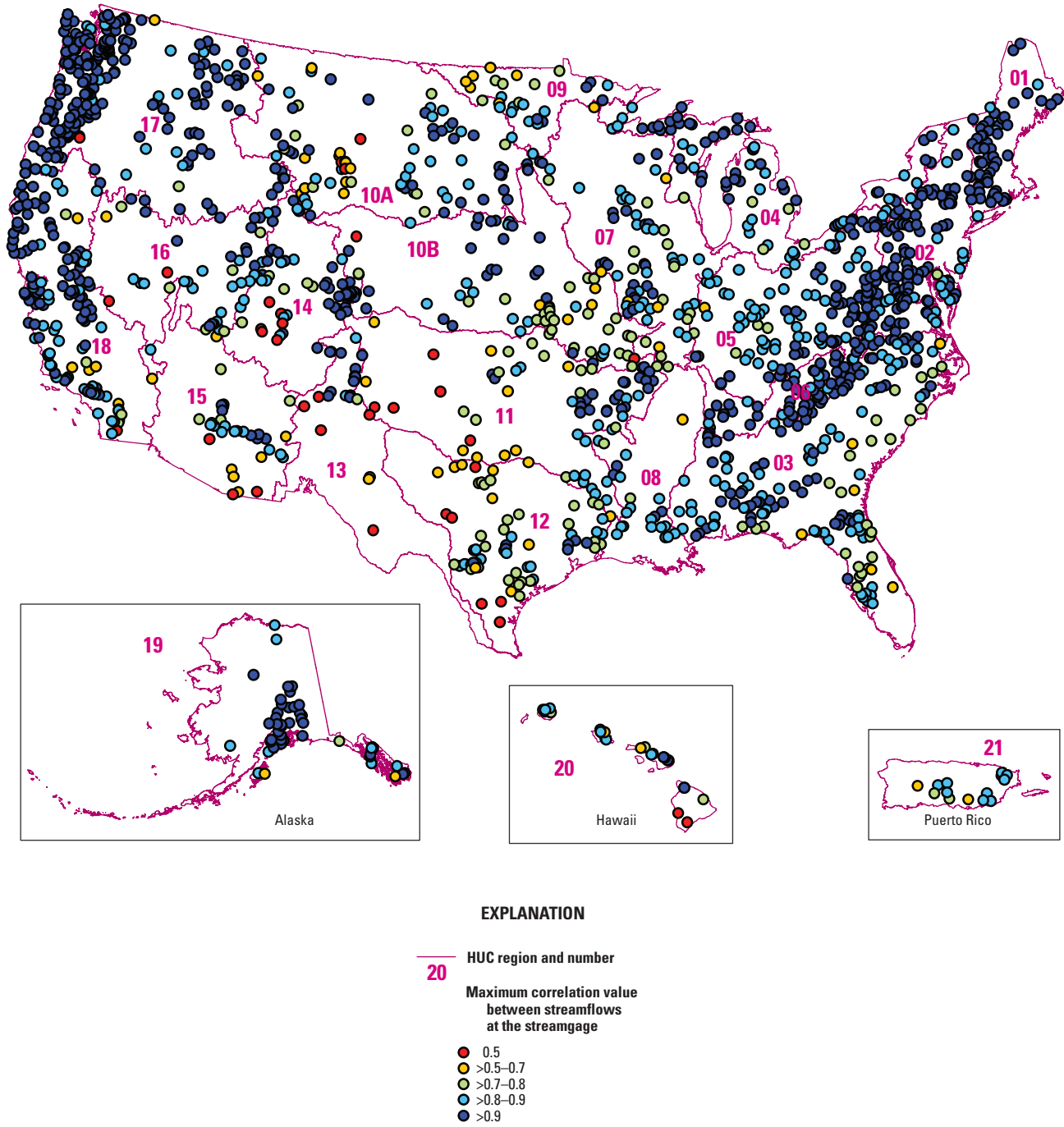
Figure 24.   Map showing the maximum values of Pearson's correlation coefficient for each hydrologic unit code (HUC) region between coincident daily streamflows at each streamgage and all other streamgages. <, less than; >, more than.
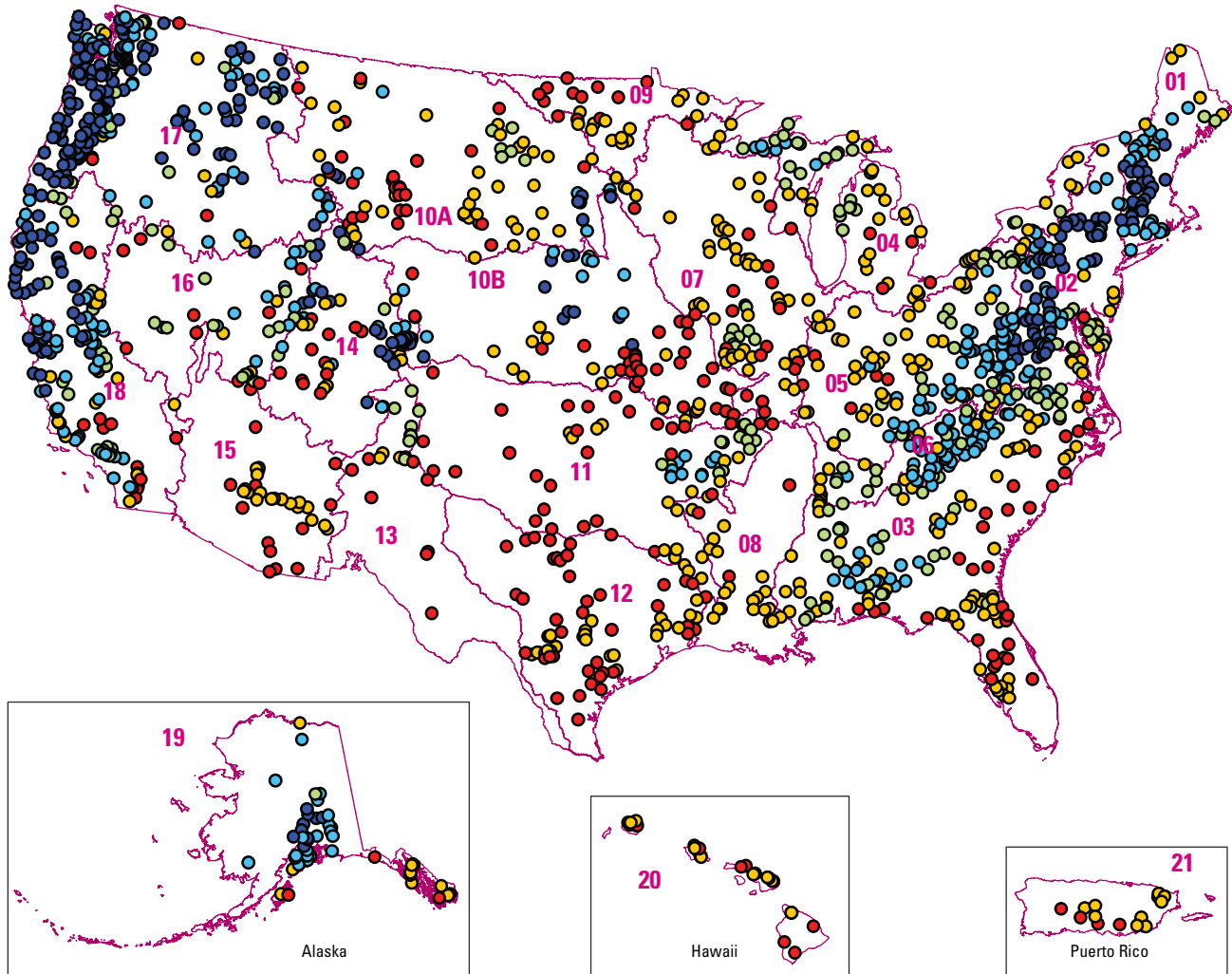
EXPLANATION

— **HUC region and number**
20

**Number of streamgages with
    correlations greater than 0.8**

- 🔴 0
- 🟡 >0–5
- 🟢 >5–10
- 🔵 >10–25
- 🔵 >25

Figure 25. Map showing number of streamgages with correlated daily streamflows greater than 0.8 (where 1.0 is the highest correlation possible) in each hydrologic unit code region. <, less than; >, more than.
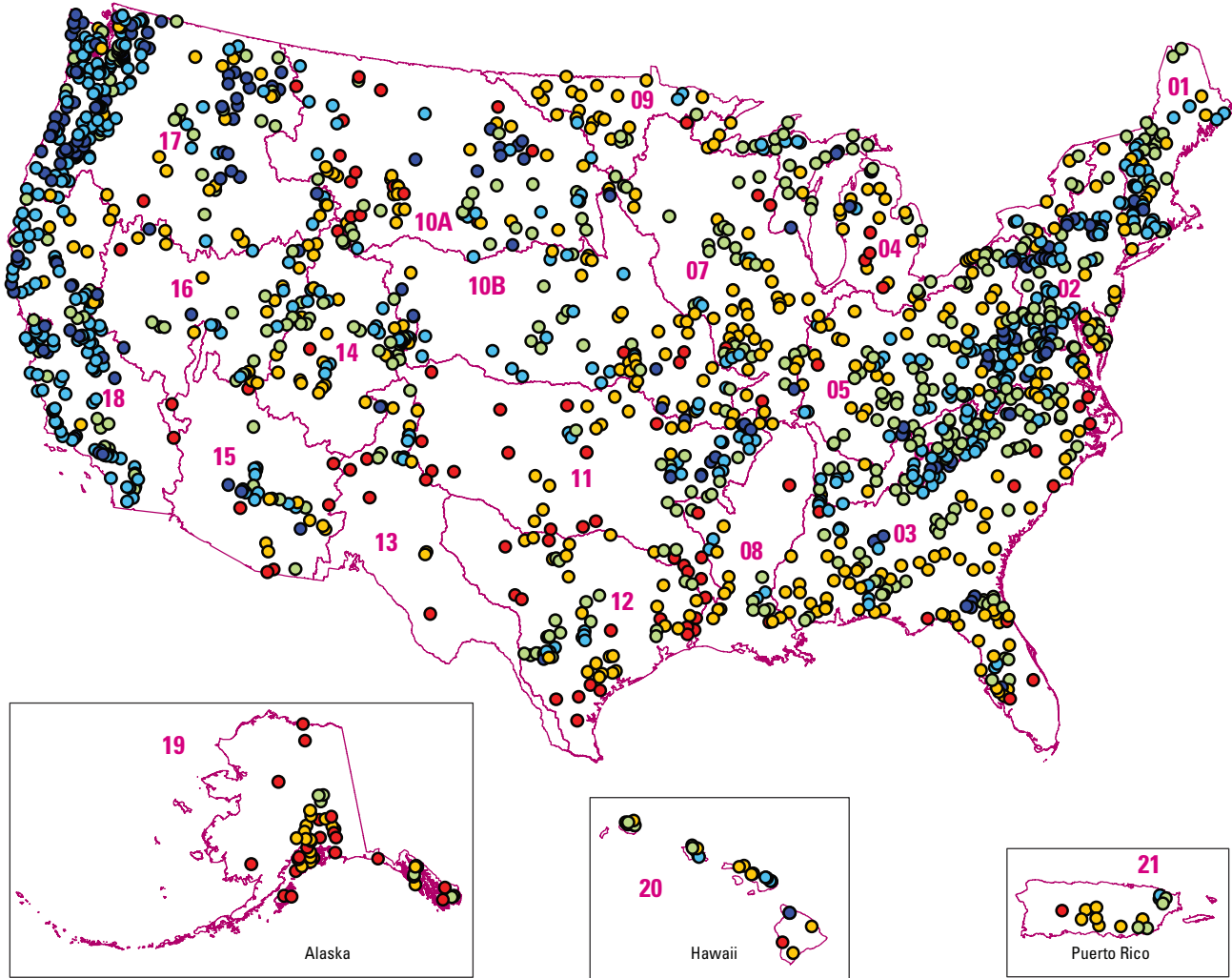
EXPLANATION

—— **HUC region and number**
20

Maximum correlation value between streamflows
at the streamgage

- ● < 0.5
- ● 0.5–0.7
- ● >0.7–0.8
- ● >0.8–0.9
- ● >0.9

Figure 26.   Map showing the maximum values of Pearson's correlation coefficient between coincident streamflows greater than the 0.1 exceedance probability (high flows) at each streamgage and all other streamgages in each hydrologic unit code region. <, less than; >, more than.
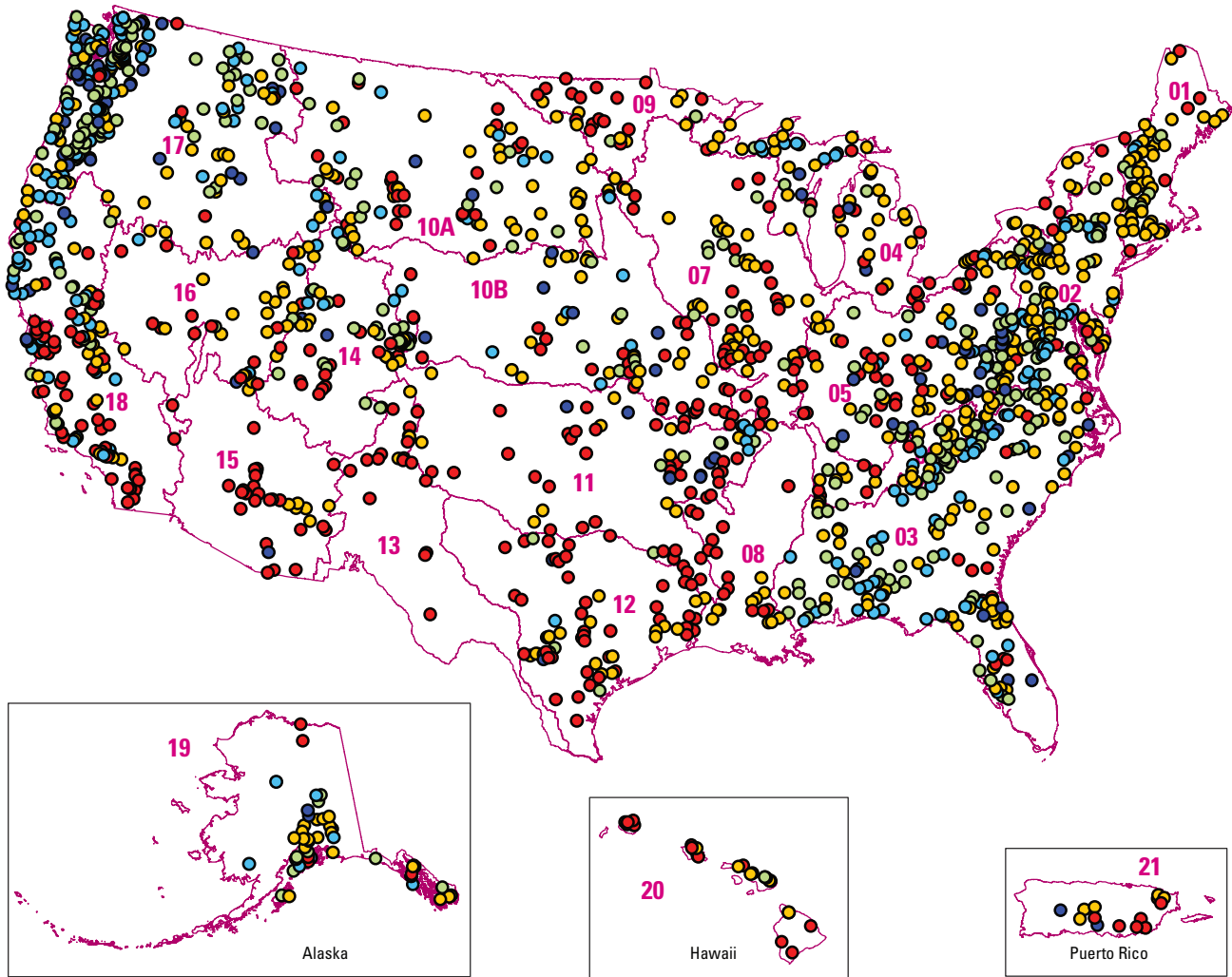
Figure 27. Map showing the maximum values of Pearson's correlation coefficient between coincident streamflows less than the 0.9 exceedance probability (low flows) at each streamgage and all other streamgages in each hydrologic unit code region. <, less than; >, more than.

EXPLANATION

— **HUC region and number**

20

Maximum correlation value between streamflows at the streamgage

- < 0.5
- 0.5–0.7
- >0.7–0.8
- >0.8–0.9
- >0.9

Figure 28.   Map showing coefficient of determination (R-squared value) obtained by regressing distance against correlation between distance and correlation between coincident daily streamflows at each streamgage and all other streamgages in each hydrologic code unit region. Regression not significant means that a statistically significant equation relating distance to correlation could not be determined at the 90-percent confidence level for that streamgage. <, less than; >, more than.
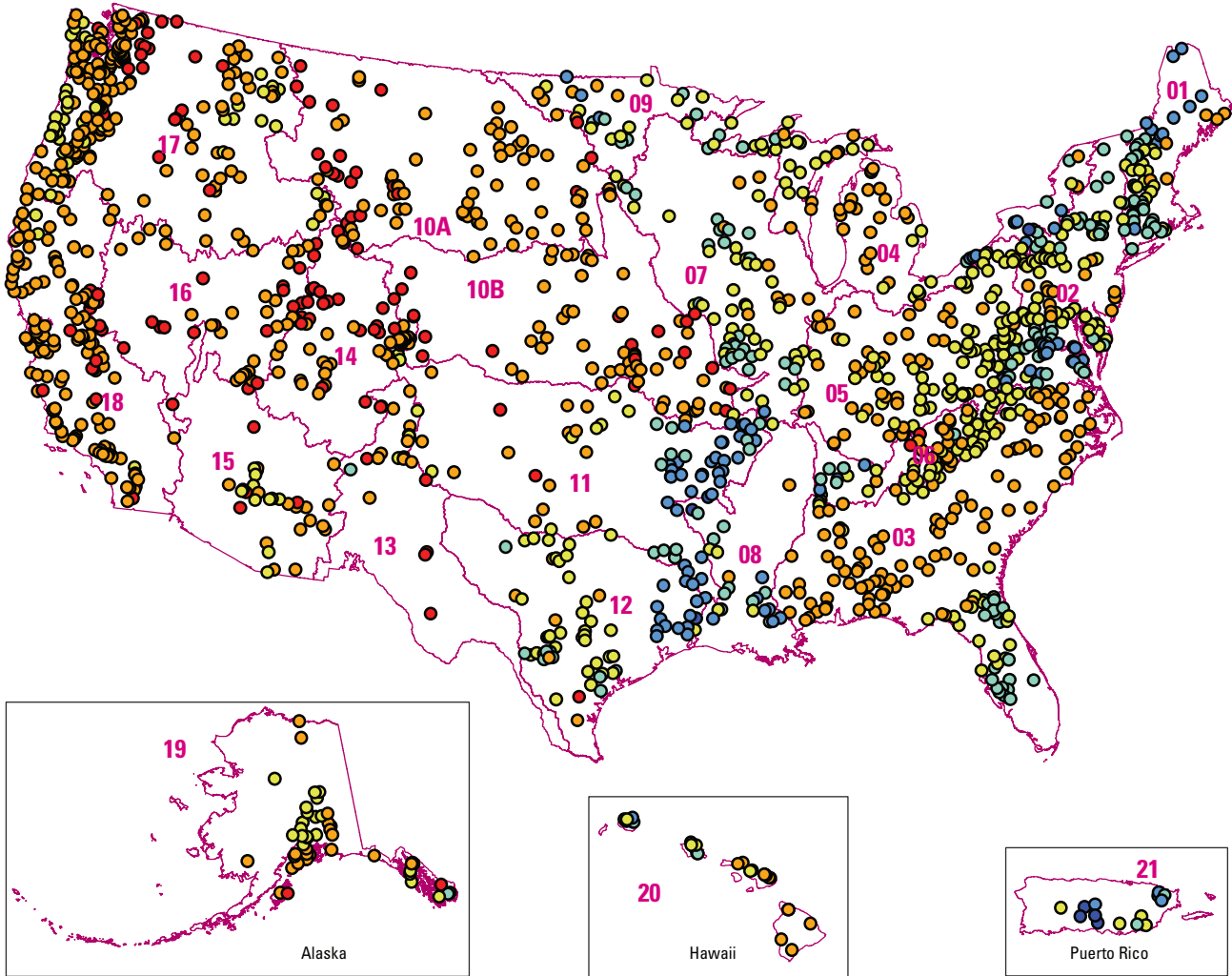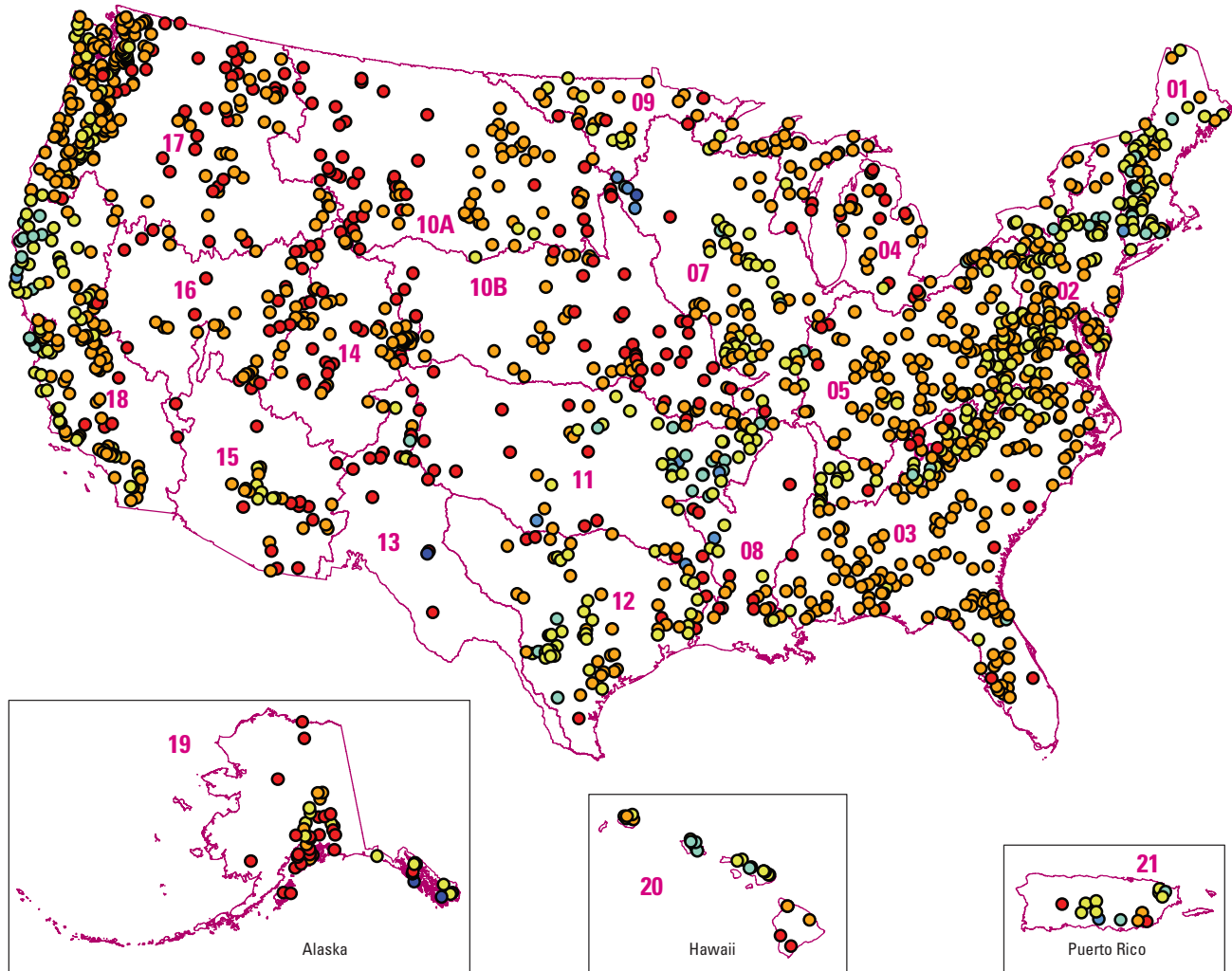
EXPLANATION

—— **HUC region and number**
20

R-squared value obtained by fitting a regression between distance and correlation at the streamgage

- ● Regression not significant
- ● < 0.5
- ● 0.5–0.7
- ● >0.7–0.8
- ● >0.8–0.9
- ● >0.9

Figure 29.   Map showing coefficient of determination (R-squared value) obtained by regressing distance against correlation between coincident streamflows greater than the 10-percent exceedence probability (high flows) at each streamgage and all other streamgages in each hydrologic code unit region. Regression not significant means that a statistically significant equation relating distance to correlation could not be determined at the 90-percent confidence level for that streamgauge. <, less than; >, more than.
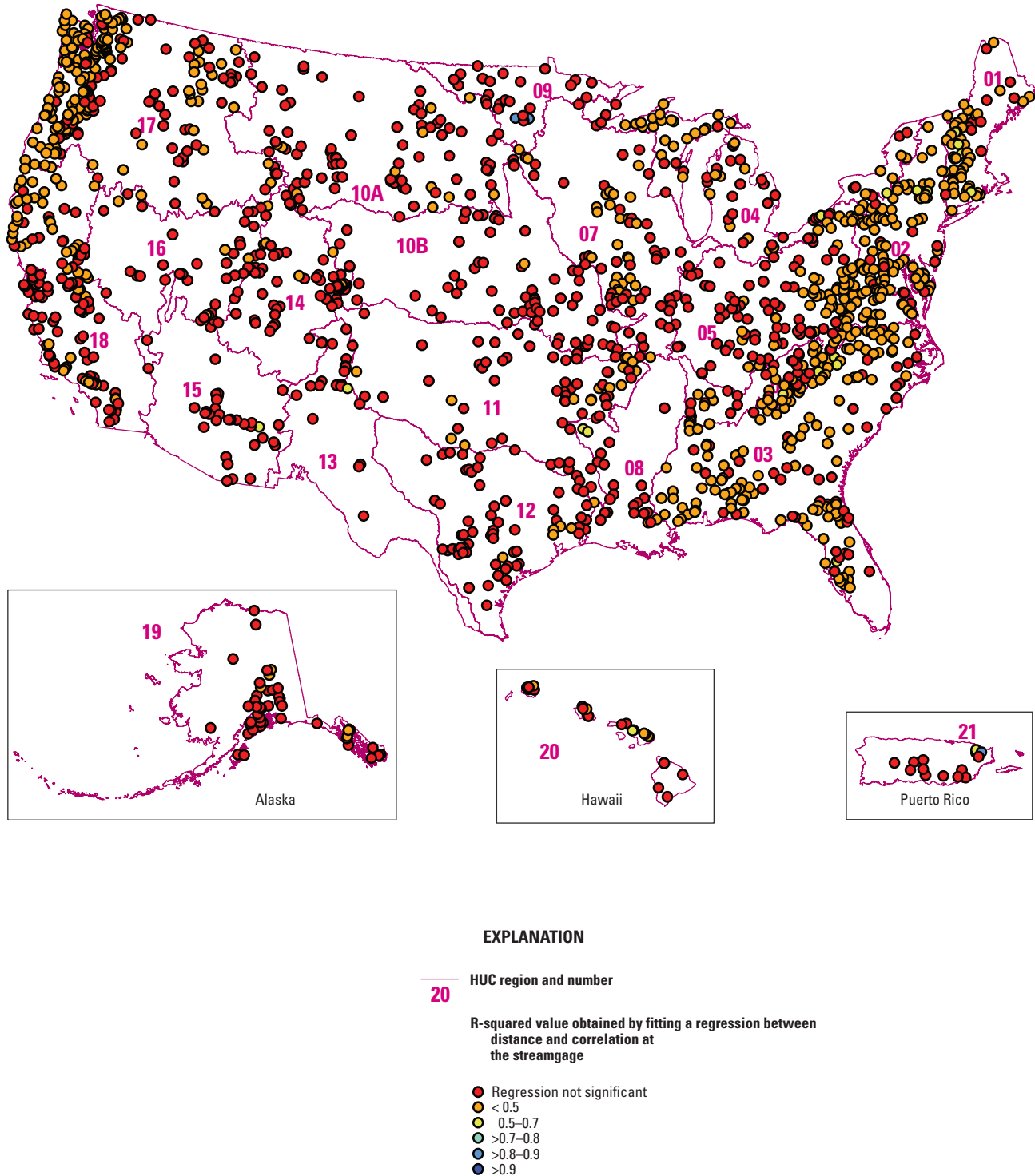
Figure 30. Map showing coefficient of determination (R-squared value) obtained by regressing distance against correlation between coincident streamflows less than the 90-percent exceedence probability (low flows) at each streamgage and all other streamgages in each hydrologic code unit region. Regression not significant means that a statistically significant equation relating distance to correlation could not be determined at the 90-percent confidence level for that streamgage. <, less than; >, more than.

EXPLANATION

—— **HUC region and number**
20

Root-mean-square error between observed and estimated correlations
○ Not enough data for fit variogram
● <0.025
● 0.025–0.05
● >0.05–0.1
● >0.1

Figure 31. Map showing root mean square error computed from observed and map-correlation-estimated correlation between coincident daily streamflows at each streamgage and all other streamgages in each hydrologic code unit region. <, less than; >, more than.
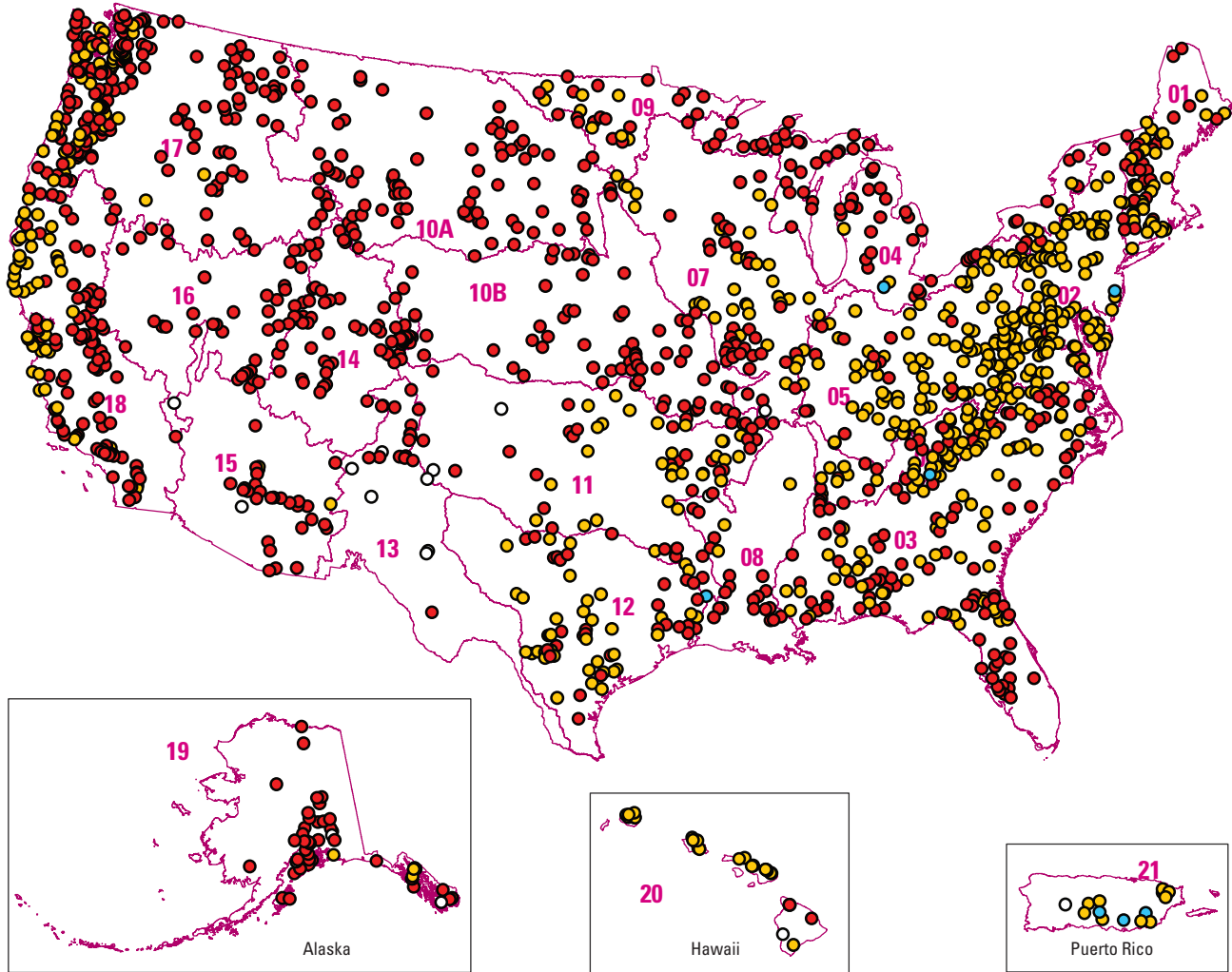
Figure 32.   Map showing root mean square error computed from observed and map-correlation-estimated correlation between coincident streamflows greater than the 10-percent exceedence probability (high flows) at each streamgage and all other streamgages in each hydrologic code unit region. <, less than; >, more than.
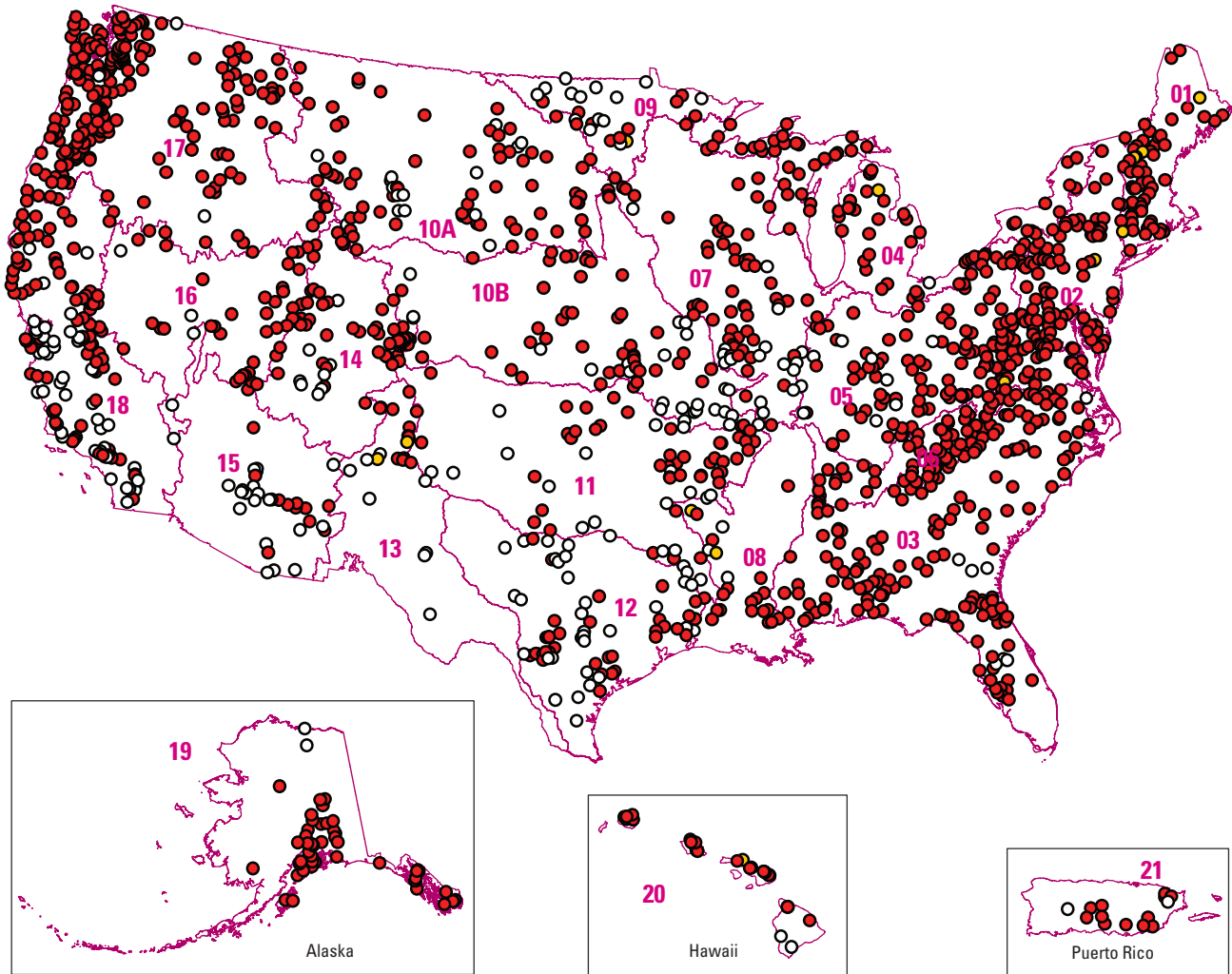
EXPLANATION

—— **2-digit hydrologic unit boundary, and number**
20

Root-mean-square error
 between observed and
 estimated correlations
○ Not enough data for
   fit variogram
● <0.025
● >0.025–0.05
● >0.05–0.1
● >0.1

Figure 33. Map showing root mean square error computed from observed and map-correlation-estimated correlation between coincident streamflows less than the 90-percent exceedence probability (low flows) at each streamgage and all other streamgages in each hydrologic code unit region. <, less than; >, more than.

Basin Attribute Coverage

Basin attribute similarity is another commonly used method for selecting an index streamgage. Basin attributes are also used to develop regional regression equations for estimating streamflow statistics at ungaged locations. In both cases, a streamgage network that covers the full range of basin attributes that is observed at ungaged locations is desirable. This part of the analysis assesses how well the streamgages in the reference-quality streamgage dataset cover the full range of basin attributes observed at ungaged locations in the continental United States. Alaska, Hawaii, and Puerto Rico were not included because the GIS layers used to calculate basin attributes within the continental United States were not available for Alaska, Hawaii, and Puerto Rico.

Basin attributes commonly used in regional regression studies are listed in table 9 in this report. With the exception of latitude, those attributes that were noted as "frequently used" and "moderately used" were used in this analysis of basin similarity, as listed below:

- drainage area

- average basin elevation

- average basin slope

- forest cover

- water storage

- impervious area

- average annual precipitation

- soil permeability

- snow, as a percentage of annual precipitation

- average annual temperature

The first eight attributes listed are the list of "frequently used" and "moderately used" attributes from published regional regression equations. Although latitude was also "moderately used" in low-flow studies, it was not included in this analysis of basin similarity. Snow and average annual temperature are included because they are important in cold and high elevation areas.

The analysis also includes annual average temperature and the percentage of precipitation that falls as snow, two variables that were often used in colder climates. Values for these variables were calculated by Falcone and others (2010) for reference-quality streamgages in GAGES–II-prelim.

The same underlying GIS layers were used to calculate values of the basin attributes in "gageable watersheds". The gageable watersheds are defined in this study as HUC 10-digit watersheds (level-5 HUCs) that are also headwater basins. Ideally, additional gageable watersheds would have been defined that were smaller than the headwaters of HUC 10-digit

watersheds, but additional GIS processing beyond the scope of this study was needed to achieve this result. The gageable watersheds have a mean drainage area of about 180 mi^2.

The distribution of the attribute values in gaged watersheds is compared with the distribution of these attribute values in the gageable watersheds to assess whether the full range of basin characteristics observed at gageable watersheds is being covered by streamgages (appendix 2). Where the full range in values for gageable watersheds is covered by the gaged watersheds (reference quality), the streamgage network is doing a good job representing the range of basin characteristics that are actually observed.

The drainage area at the reference-quality streamgages is the NWIS reported value; if there is no NWIS value reported, then the GIS-derived drainage area is used. Generally, there is fairly good coverage of the full range of drainage area observed at gageable watersheds by the reference-quality streamgages.

The streamgages in the reference-quality streamgage network appear to be doing a fairly good job of representing the range of basin characteristics for most of the attributes listed above. In part, this may be because the average size of the gageable watersheds is fairly large and characteristics are averaged over these large areas. More extreme values of basin attributes that may occur over limited spatial areas are not seen in attributes calculated for the gageable watersheds based on HUC 10-digit watersheds. Using smaller basins would give a more complete picture of whether the full range of basin attributes observed at ungaged locations is being covered by the streamgage network.

Overall, the reference-quality streamgages cover the range of attributes observed at gageable watersheds fairly well, but there are some exceptions. Analysis indicated that the streamgage network does not adequately represent areas with low forest cover or high impervious cover. This finding is expected because the reference-quality streamgages were selected so that they do not measure flow at basins with substantial land-cover changes, such as deforestation or urbanization. The streamgages in the complete dataset (not just the reference-quality streamgage dataset) are likely to cover these types of attributes more fully. Reference-quality streamgages were the focus of this analysis because these are the streamgages that are most likely to be of value in estimating statistics at ungaged locations using existing regression and index site transfer methods.

There are also instances where specific basin attributes are not covered well in some HUC regions. For example, the highest elevations are not covered well in Souris-Red-Rainy (HUC region 9), Arkansas-White-Red (HUC region 11), or Texas-Gulf (HUC region 12). Conversely, the lowest elevations are not being covered well in the Rio Grande (HUC region 13), Lower Colorado (HUC region 15), or Great Basin (HUC region 16). These arid and semiarid HUC regions seem

to have a variety of mismatches in gageable and gaged basin attributes. For example, very low precipitation values are seen at the gageable watersheds but not in the watersheds draining to reference-quality streamgages. Permeability, snow, temperature, and water cover are other variables that have gaps in their coverages in these regions.

Summary and Conclusions

The U.S. Geological Survey (USGS) conducted a gap analysis to evaluate how well the USGS streamgage network meets a variety of needs, focusing on the ability to calculate various statistics at locations that have streamgages (gaged) and that do not have streamgages (ungaged), and to determine where there are gaps in the network of gaged locations, how accurately desired statistics can be calculated with a given length of record, and if the current network allows for estimation of these statistics at ungaged locations. The results of this analysis are summarized in table 10.

Spatial Gap Analysis

USGS streamgages in the complete dataset measure flow at drainage areas of 20,000 square miles (mi^2) or less. Flow is measured for about 72 percent of the land area in the conterminous United States. Alaska and Hawaii have the lowest density of streamgages in the United States, with large parts of these States containing no USGS streamgages. The spatial distribution of streamgages in the complete dataset is uneven within the conterminous United States as well, with the most streamgages concentrated in the most populated areas of the country. The density of USGS streamgages tends to be low in much of the arid Southwest. Regional differences in the spatial coverage of streamgages are especially pronounced for streamgages that measure flow at drainage areas smaller than 1,000 mi^2.

The overall spatial coverage of streamgages in the reference-quality streamgage dataset is much lower, at 8.6 percent for the conterminous United States. The spatial coverage of reference-quality streamgages in Alaska is similar (8.9 percent). The lowest reference-quality streamgage coverage is observed in the Southwestern United States, Hawaii, and Puerto Rico.

Temporal Gap Analysis

The USGS streamgage network has been fairly stable in terms of number of streamgages since the 1950s and 1960s. The median record length for all streamgages in the complete dataset is 42 years, The median record length for all streamgages within a hydrologic unit code (HUC) region is lowest for Alaska (HUC region 19) and Puerto Rico (HUC region 21). In an effort to consider record length in conjunction with spatial coverage, the areal coverage of streamgages with record lengths of 50 years or greater was determined. The spatial coverage of these long-term record streamgages is lowest outside the conterminous United States (Alaska, Hawaii, and Puerto Rico) and in arid and semiarid areas of the Central and Southwestern United States.

Uncertainty Analysis

A minimum level of uncertainty in various streamflow statistics was estimated based on the interannual variability in the streamflow time series used to calculate the statistics. As calculated in this study, differences in uncertainty of flow statistics across the country are the result of differences in the interannual variability of flow. For each of the statistics used in this analysis, the greatest variability is generally observed in the arid and semiarid Central and Southwestern United States. A greater number of years of record are required in these areas to achieve comparable levels of uncertainty throughout the country for most statistics.

This analysis is limited by the use of some simplifying assumptions that cause the estimated uncertainty to be lower than would be expected if all sources of uncertainty were considered. In particular, persistence in hydrologic records was not considered in this study. The estimates of uncertainty may or may not apply in areas that are ungaged. If the interannual variability at ungaged sites is similar to that for gaged areas, the uncertainty will be similar. However, if the interannual variability is different, then the uncertainty may also be quite different.

Previous Regression Studies

Regional regression equations developed in previous studies to estimate peak flows, low flows, and annual flows use many of the same basin attributes as explanatory variables. Drainage area, precipitation, elevation, and slope were the most commonly used basin attributes in previous USGS studies. Peak flow equations were the most widely available across the country and are generally the most accurate. Equations to estimate low-flow frequency statistics are the least accurate. Equations were generally most accurate in the Eastern United States and lowest in the Central and Southwestern United States, but exceptions abound. However, due to the different times when the studies were completed and the methods used, not all the differences in regression equation accuracy may be directly attributable to differences in hydrology or data availability.

Correlation

Correlations between reference-quality streamgages for all coincident streamflow and high- and low-flow periods were evaluated to determine similarities and uniqueness in the streamgage network as well as to understand the suitability of the network to provide index streamgages for information transfer to an ungaged location. Most highly correlated streamgages are in the Eastern and Western United States and appear to coincide with mountainous regions of the United States. Areas with highly correlated streamgages provide index streamgages with the potential for good transfer of information to an ungaged location.

Lower correlations are found in the Central United States and coastal areas in the Southeastern United States, suggesting that few or no streamgages in these areas would yield index streamgages suitable for information transfer, regardless of the index streamgage selection method used. The use of distance as a surrogate to select a highly correlated index streamgage appears to work in some areas although not reliably. The use of the map-correlation method to estimate correlation between a streamgage and ungaged location showed promise for the Eastern United States but less so for other areas of the Nation. Correlations estimated between coincident high- and low-flow periods were lower than correlations estimated from the entire coincident periods at each streamgage, and there appears to be more similarity among streamgages for coincident high-flow periods than for coincident low-flow periods, suggesting there is more dissimilarity amongst streamgages during low-flow periods than high-flow periods.

Basin Attribute Similarity

The USGS streamgages in the reference-quality streamgage dataset were evaluated to determine if the basin attributes observed at upstream watersheds covered the range observed at gageable watersheds. The gageable watersheds are defined in this study as watershed-level HUC (10-digit) codes that are headwater basins. Any gaps in the coverage of the full range of basin attributes is a potential problem for transferring streamflow information from gaged to ungaged areas. Basin attributes that were commonly used in previous USGS studies to develop regional regression equations were included in this analysis.

The USGS streamgage network of reference-quality streamgages defined for this study covers the range of basin attributes observed at gageable watersheds in most parts of the country fairly well, for most attributes. Some notable exceptions include often remote areas such as those with very high

elevation or extreme aridity. In some cases areas with very low elevations were also poorly covered. The analysis also indicates that the streamgage network does not adequately represent areas with low forest cover or high impervious cover. The streamgages in the complete dataset, not only those in the reference-quality streamgage dataset, are likely to cover these types of attributes more fully. Reference-quality streamgages were the focus of this analysis because these are the streamgages that are most likely to be of value in estimating statistics at ungaged locations when using existing regression and index site transfer methods.

This analysis is limited by the use of only the set of gageable watersheds. A more complete analysis of basin attribute similarity would require the definition of additional gageable watersheds at finer scales than were available for use in this study.

Further Work

Further work to expand parts of this network gap analysis would increase the usefulness and confidence in these results. The uncertainty analysis included in this report calculates a minimum level of uncertainty and does not consider the effects of persistence in hydrologic time series, which is expected to increase uncertainty above the estimates reported here. To better understand the full uncertainty of streamflow statistics, additional time-series analysis could be undertaken to estimate the effects of persistence on uncertainty. The uncertainty of low-flow frequency statistics was not undertaken in this report because of issues in estimating the variance of the frequency statistics when the conditional probability adjustment is used to accommodate zero flow conditions. Additional work to develop, test, and implement an estimator could provide these estimates of low-flow uncertainty.

Alaska, Hawaii, and Puerto Rico were omitted from the basin similarity analysis because the GIS layers used in the conterminous United States were not available. To complete a basin similarity analysis outside the conterminous United States, alternative datasets would need to be identified or developed. Headwater basins at the smaller scale of subwatersheds (HUC 12-digit codes) could be added to the set of gageable watersheds to broaden the view of what basin characteristics need to be covered by the streamgage network. In addition to subwatersheds, other watersheds could be defined at every outlet of basins as defined in the U.S. Department of Agriculture Natural Resources Conservation Service (2012) Watershed Boundary Dataset to develop an even more complete set of gageable watersheds.

Table 10. Summary of key findings by hydrologic unit code (HUC) region.

[HUC, hydrologic unit code]

HUC region	HUC region name	Spatial coverage, percentage of area that is measured by streamgages	Temporal coverage, median years of record	Uncertainty of statistics, qualitative assessment	Maximum correlation between reference data-set streamgages, median for all flows
1	New England	Complete dataset: 74 percent Reference dataset: 13 percent	50 years	Low	0.95
2	Mid-Atlantic	Complete dataset: 78 percent Reference dataset: 12 percent	52 years	Low to moderate	0.95
3	South Atlantic-Gulf	Complete dataset: 79 percent Reference dataset: 11 percent	38 years	Low to moderate	0.89
4	Great Lakes	Complete dataset: 73 percent Reference dataset: 13 percent	39 years	Low	0.91
5	Ohio	Complete dataset: 86 percent Reference dataset: 11 percent	44 years	Low	0.91
6	Tennessee	Complete dataset: 69 percent Reference dataset: 13 percent	49 years	Low	0.93
7	Upper Mississippi	Complete dataset: 88 percent Reference dataset: 11 percent	45 years	Moderate	0.86
8	Lower Mississippi	Complete dataset: 63 percent Reference dataset: 6 percent	48 years	Low to moderate	0.84
9	Souris-Red-Rainy	Complete dataset: 65 percent Reference dataset: 25 percent	54 years	High	0.82
10A	Upper Missouri	Complete dataset: 70 percent Reference dataset: 11 percent	39 years	Moderate	0.85*
10B	Lower Missouri	Complete dataset: 84 percent Reference dataset: 16 percent	41 years	Moderate	0.85*
11	Arkansas-White-Red	Complete dataset: 79 percent Reference dataset: 10 percent	46 years	Moderate	0.83
12	Texas-Gulf	Complete dataset: 79 percent Reference dataset: 10 percent	43 years	High	0.79
13	Rio Grande	Complete dataset: 40 percent Reference dataset: 3 percent	46 years	Moderate	0.90
14	Upper Colorado	Complete dataset: 77 percent Reference dataset: 4 percent	36 years	Low to moderate	0.94
15	Lower Colorado	Complete dataset: 66 percent Reference dataset: 5 percent	39 years	High	0.83
16	Great Basin	Complete dataset: 31 percent Reference dataset: 1 percent	34 years	Moderate	0.89
17	Pacific Northwest	Complete dataset: 76 percent Reference dataset: 13 percent	45 years	Low	0.97
18	California	Complete dataset: 62 percent Reference dataset: 8 percent	42 years	High	0.94
19	Alaska	Complete dataset: 10 percent Reference dataset: 9 percent	28 years	Low	0.93
20	Hawaii	Complete dataset: 10 percent Reference dataset: 1 percent	55 years	Low to moderate	0.86
21	Caribbean	Complete dataset: 51 percent Reference dataset: 4 percent	29 years	Low to moderate	0.81

*Separate values were not calculated for the Upper and Lower Missouri River Basin. This value represents the median for the entire Missouri region (HUC region 10).

References Cited

Archfield, S.A., and Vogel, R.M., 2010, Map correlation method—Selection of a reference streamgage to estimate daily streamflow at ungaged catchments: Water Resources Research, v. 46, W10513, 15 p., accessed [date], at http://www.agu.org/pubs/crossref/2010/2009WR008481.shtml.

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

Cohn, T.A., Berenbrock, Charles, Kiang, J.E., and Mason, R.R., Jr., 2012, Calculating weighted estimates of peak streamflow statistics: U.S. Geological Survey Fact Sheet 2012–3038, 4 p., available at http://pubs.usgs.gov/fs/2012/3038/.

Commission for Environmental Cooperation, 1997, Ecological regions of North America—Toward a common perspective: Montreal, Canada, Commission for Environmental Cooperation, 71 p. (Accessed September 24, 2012, at http://www.cec.org/Storage/42/3484_eco-eng_EN.pdf.)

Falcone, J.A., Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010, GAGES—A streamgage database for evaluating natural and altered flow conditions in the conterminous United States: Ecology, v. 91, p. 621.

Falcone, J.A., 2011, GAGES–II—Geospatial attributes of gages for evaluating streamflow: U.S. Geological Survey, accessed September 24, 2012, at http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Amsterdam, Elsevier Studies in Environmental Science 49, 522 p.

Natural Resources Conservation Service, 2012, Watershed boundary dataset: U.S. Department of Agriculture Nature Resources Conservation Service, accessed October 10, 2012, at http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/water/watersheds/dataset.

Ries, K.G., III, 2007, The national streamflow statistics program—A computer program for estimating streamflow statistics for ungaged sites: U.S. Geological Survey Techniques and Methods, book 4, chap. A6, 37 p. (Also available at http://pubs.usgs.gov/tm/2006/tm4a6/.)

U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: U.S. Geological Survey Bulletin 17B, 28 p., 14 appendixes. (Also available at http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.)

Vogel, R.M., and Fennessey, N.M., 1994, Flow duration curves I—A new interpretation and confidence intervals: Journal of Water Resources Planning and Management, v. 120, no. 4, p. 485–504.

Wieczorek, Michael, 2011, USGS streamgage NHDPlus version 1 basins 2011: U.S. Geological Survey, accessed November 29, 2012, at http://water.usgs.gov/GIS/metadata/usgswrd/XML/streamgagebasins.xml

Appendix 1.   Gages Used in the Reference-Quality Streamgage Dataset

Reference-quality gages used in a national streamflow gap analysis study are available as a separate Microsoft Excel file (click here to download).

Appendix 2. Basin Attributes at Gageable Watersheds and Gaged Basins for Each Hydrologic Unit Code Region
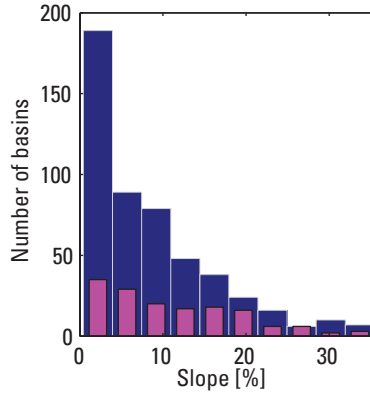
HUC Region 1: New England
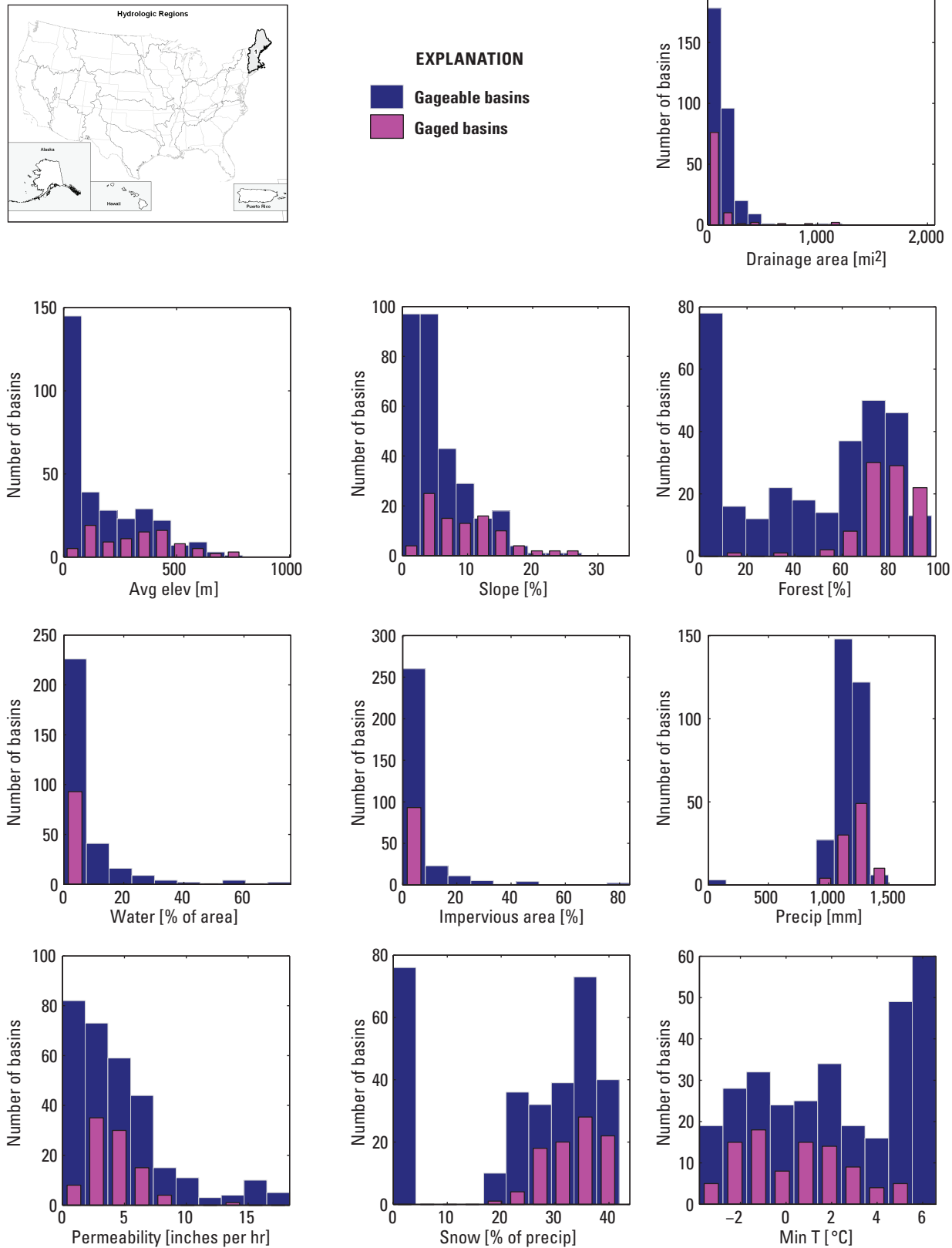
Figure 2–1.   Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in New England (hydrologic unit code (HUC) region 1). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

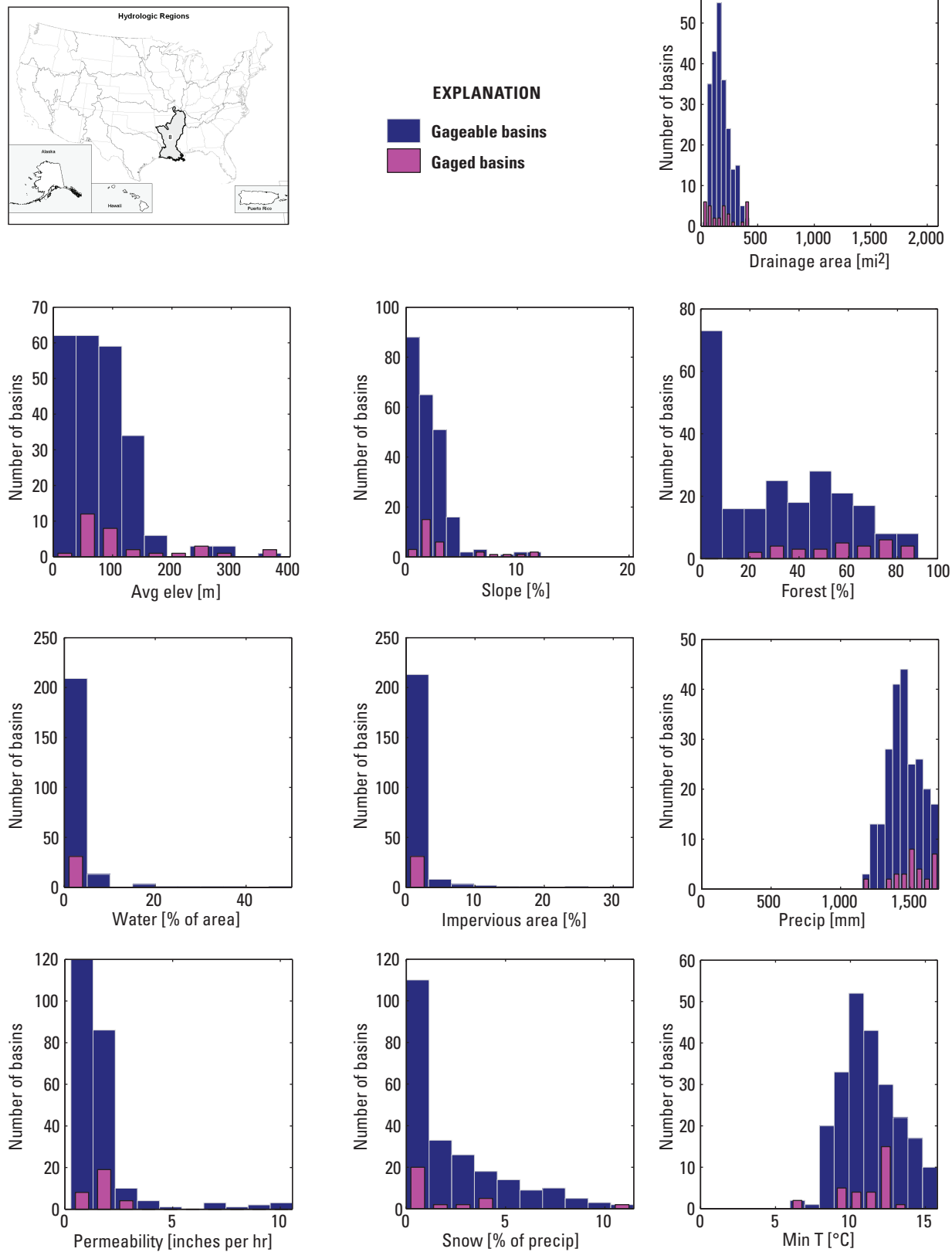HUC Region 2: Mid-Atlantic
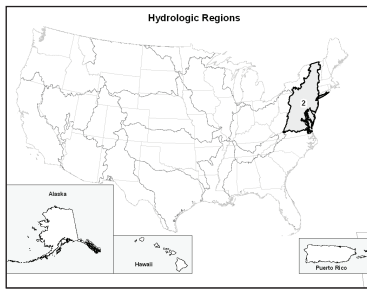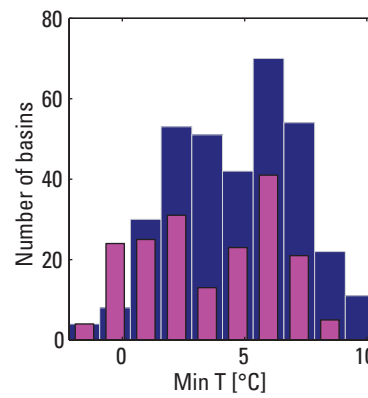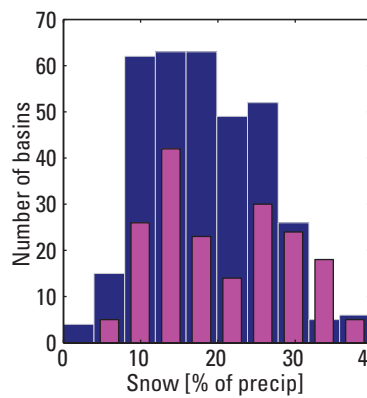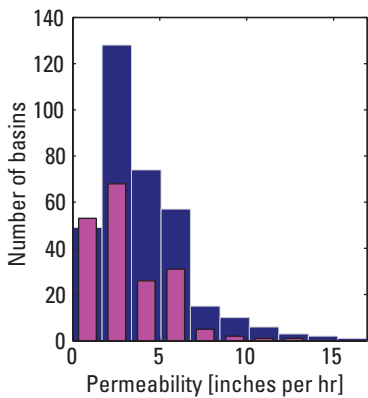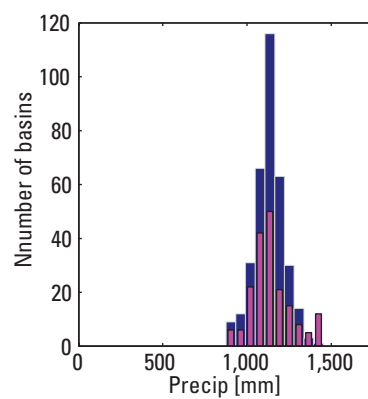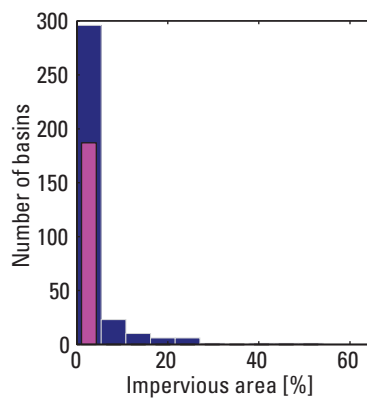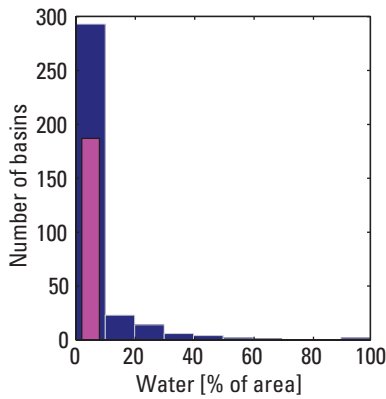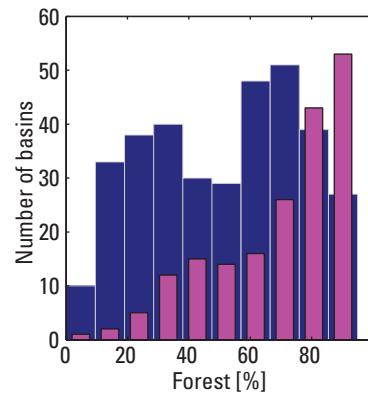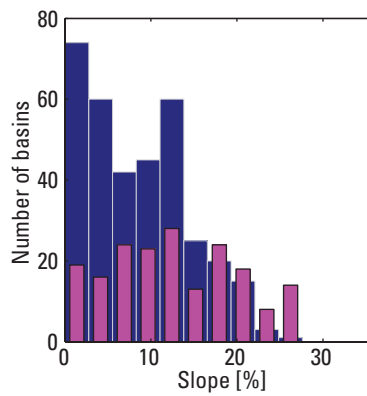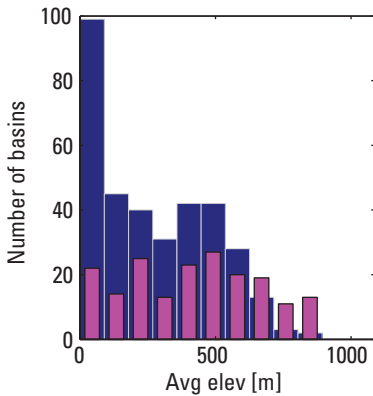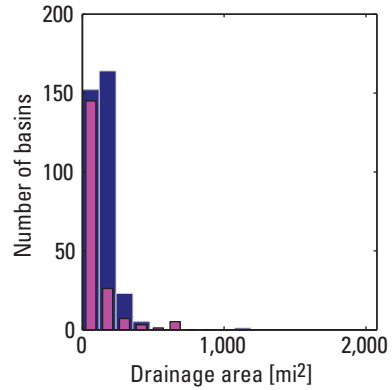
Figure 2–2. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Mid-Atlantic (hydrologic unit code (HUC) region 2). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
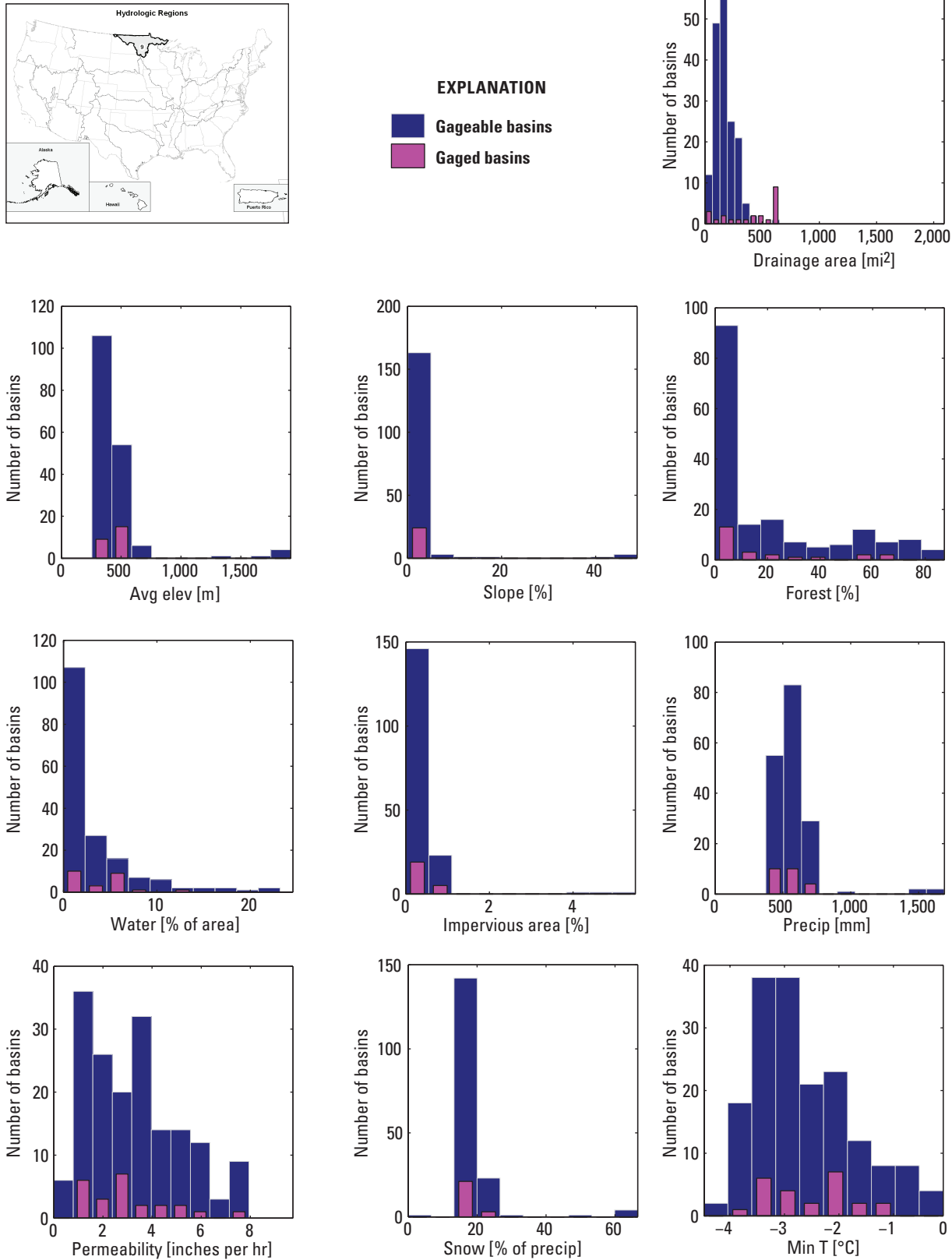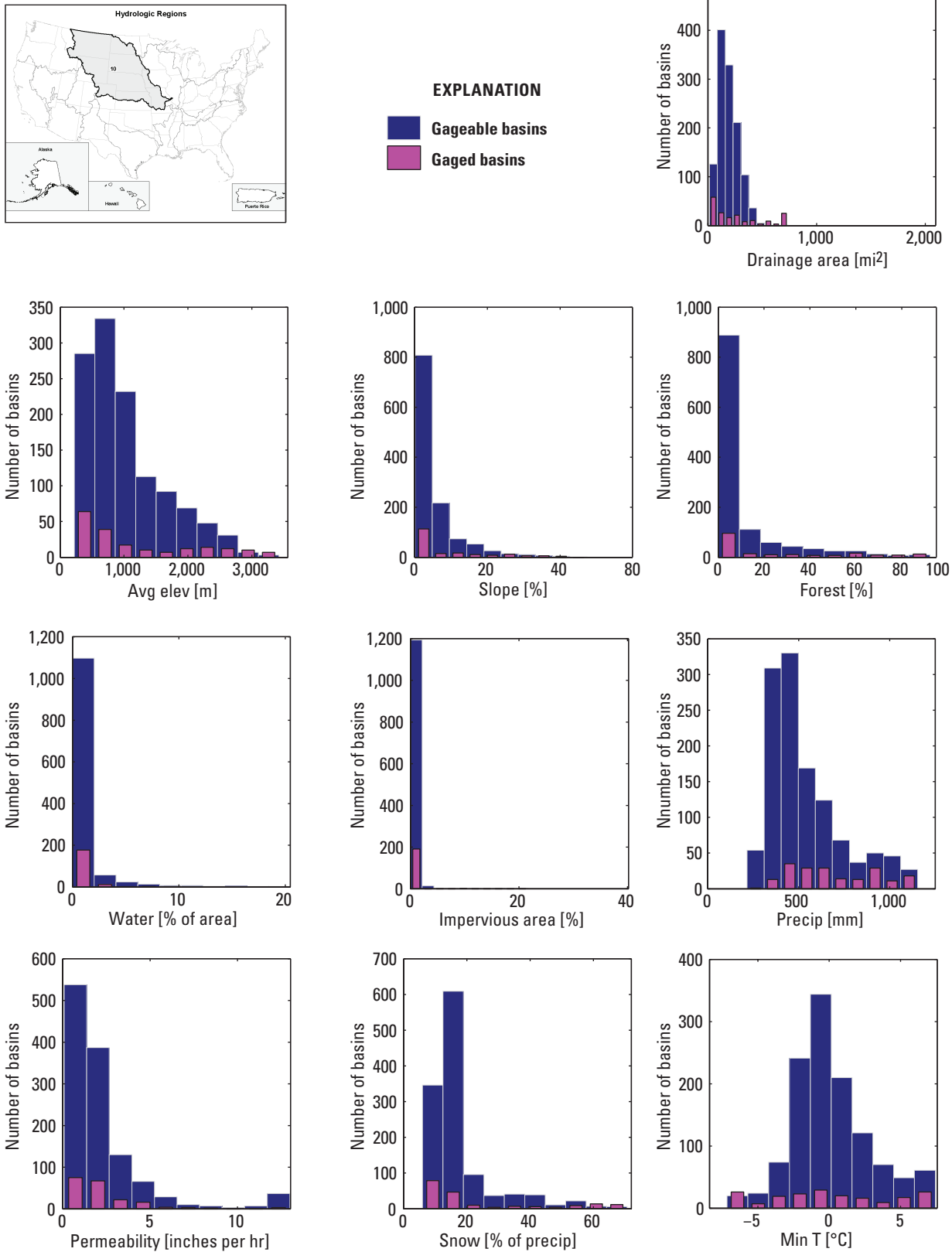
HUC Region 3: Southeast
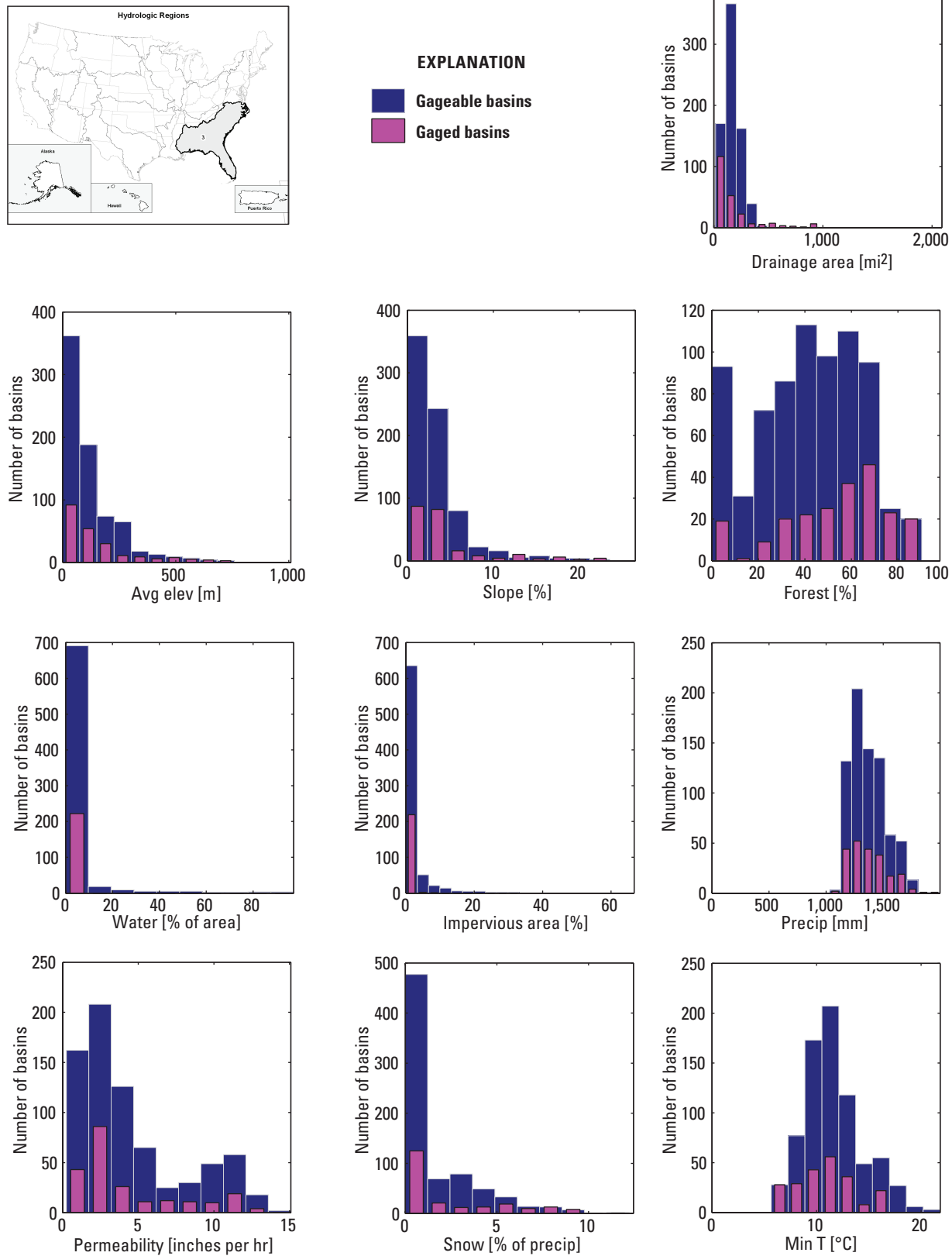
Figure 2–3.   Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Southeast (hydrologic unit code (HUC) region 3). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 4: Great Lakes
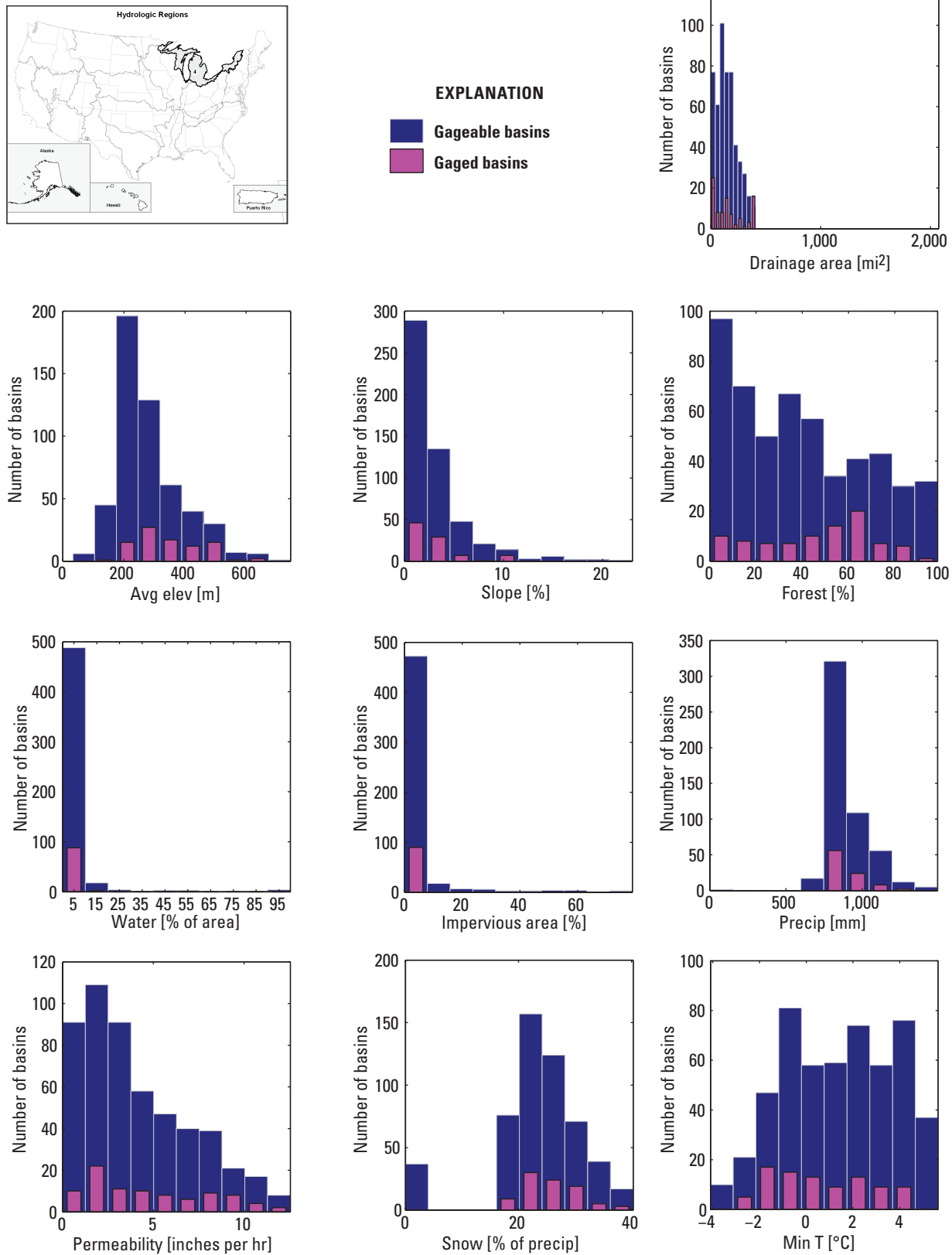
Figure 2–4. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Great Lakes (hydrologic unit code (HUC) region 4). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
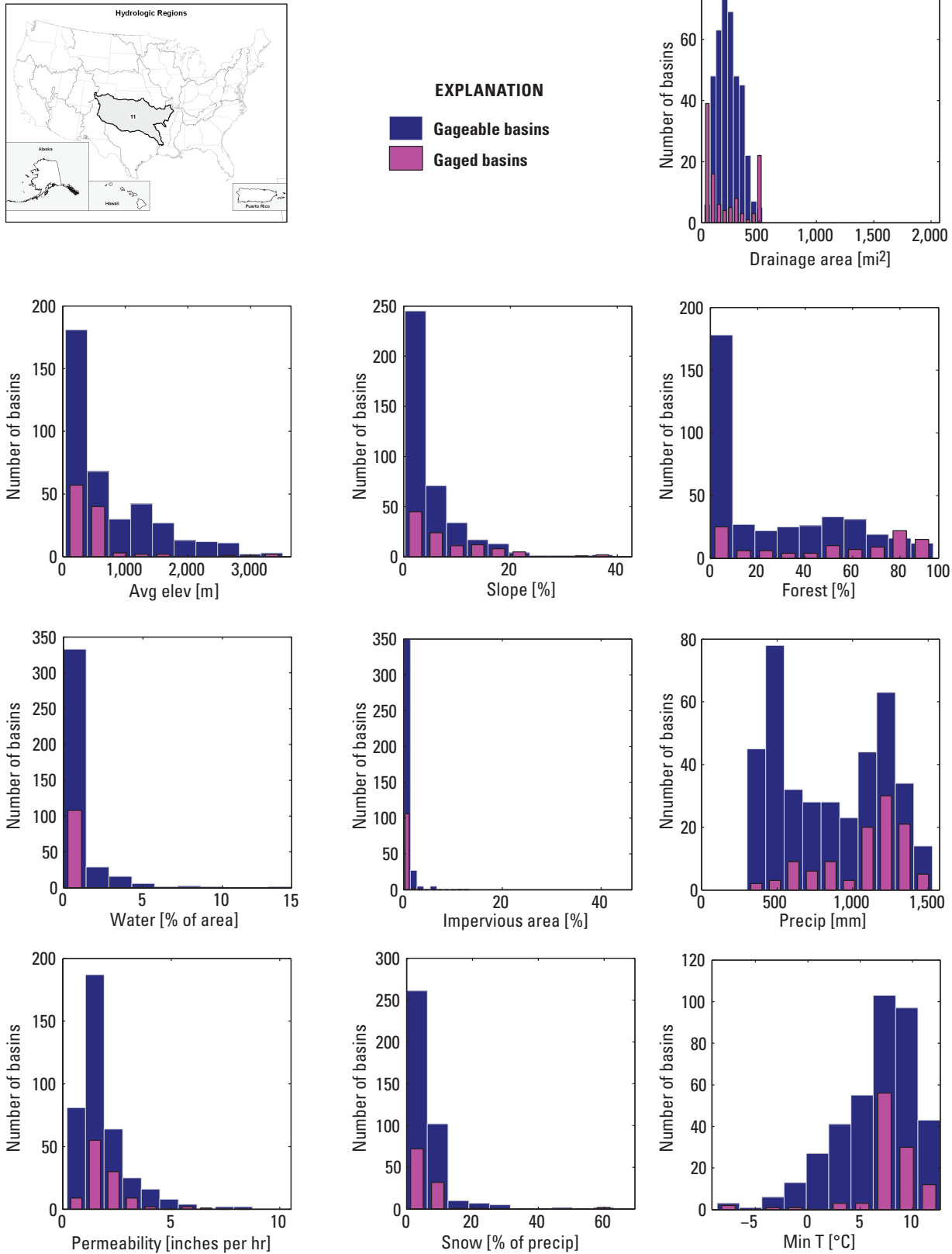
HUC Region 5: Ohio
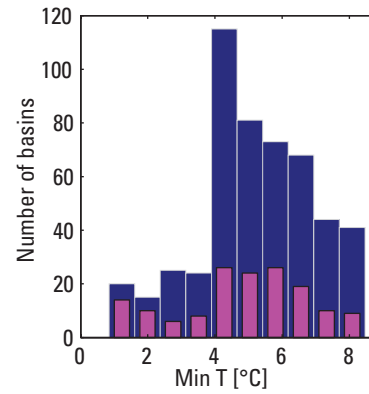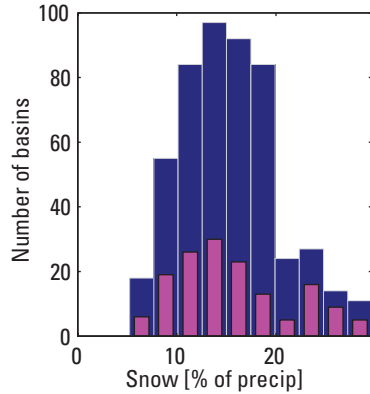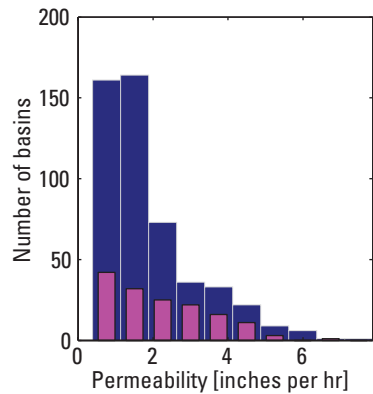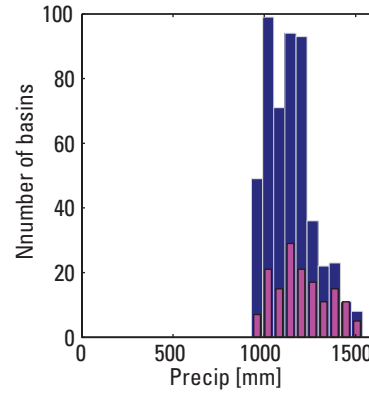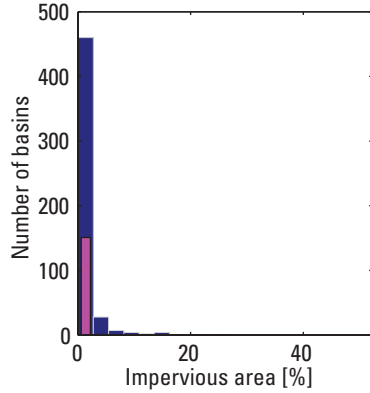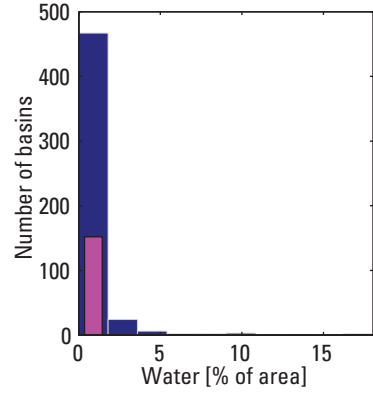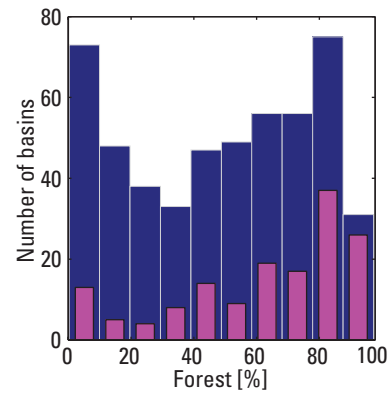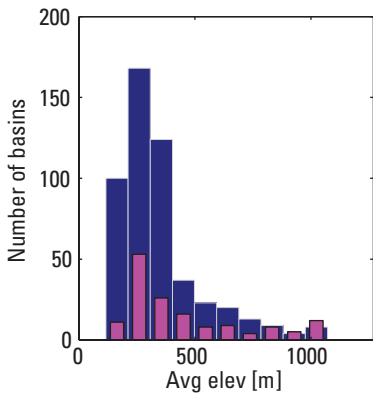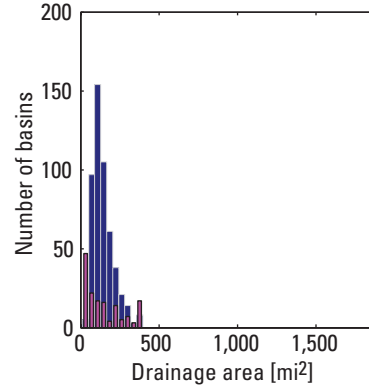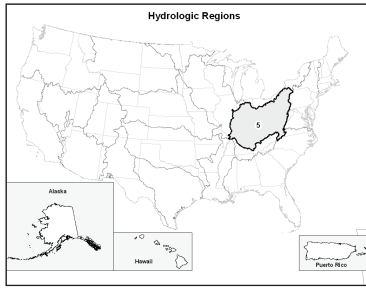
Figure 2–5. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Ohio (hydrologic unit code (HUC) region 5). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 6: Tennessee
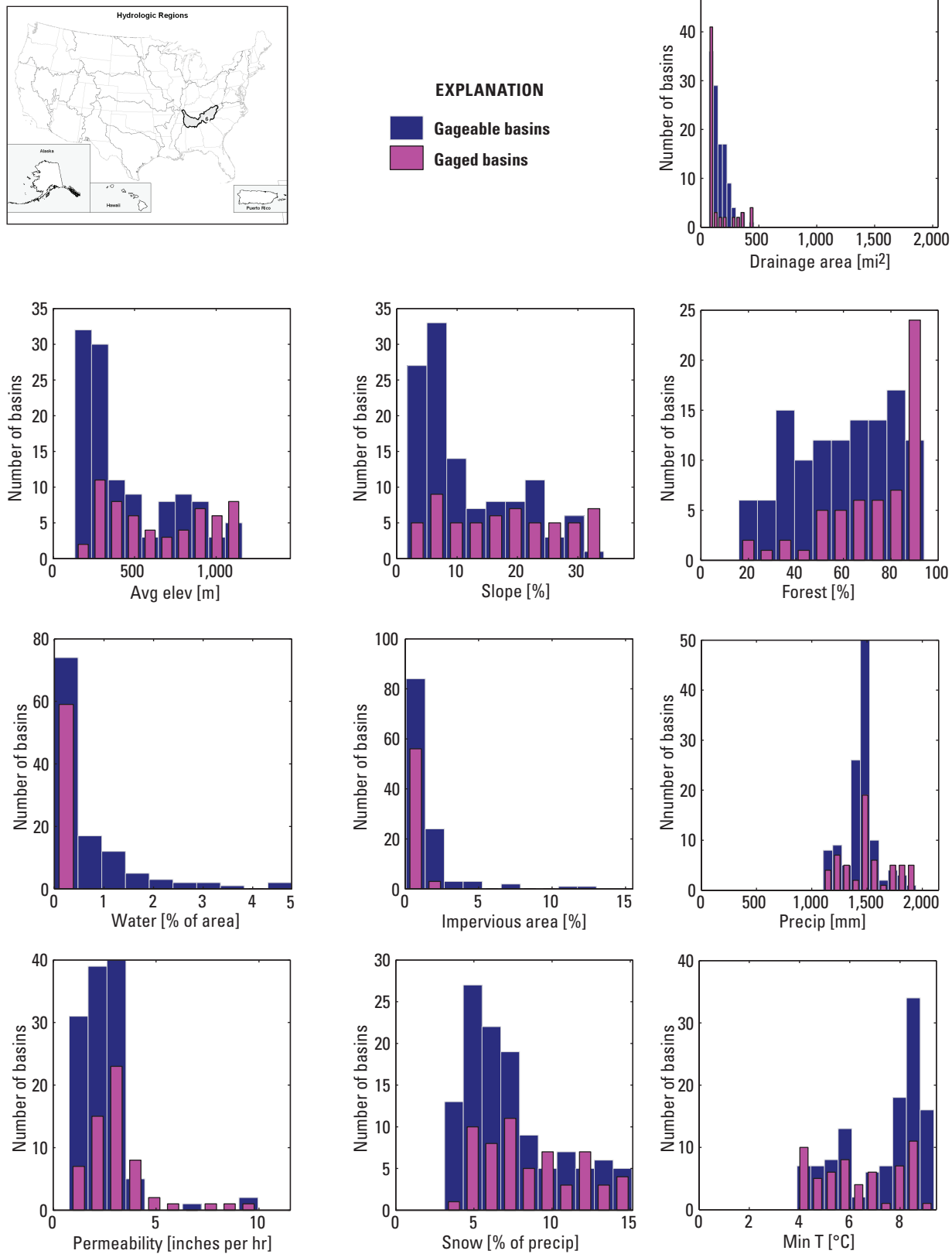
Figure 2–6.   Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Tennessee (hydrologic unit code (HUC) region 6). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
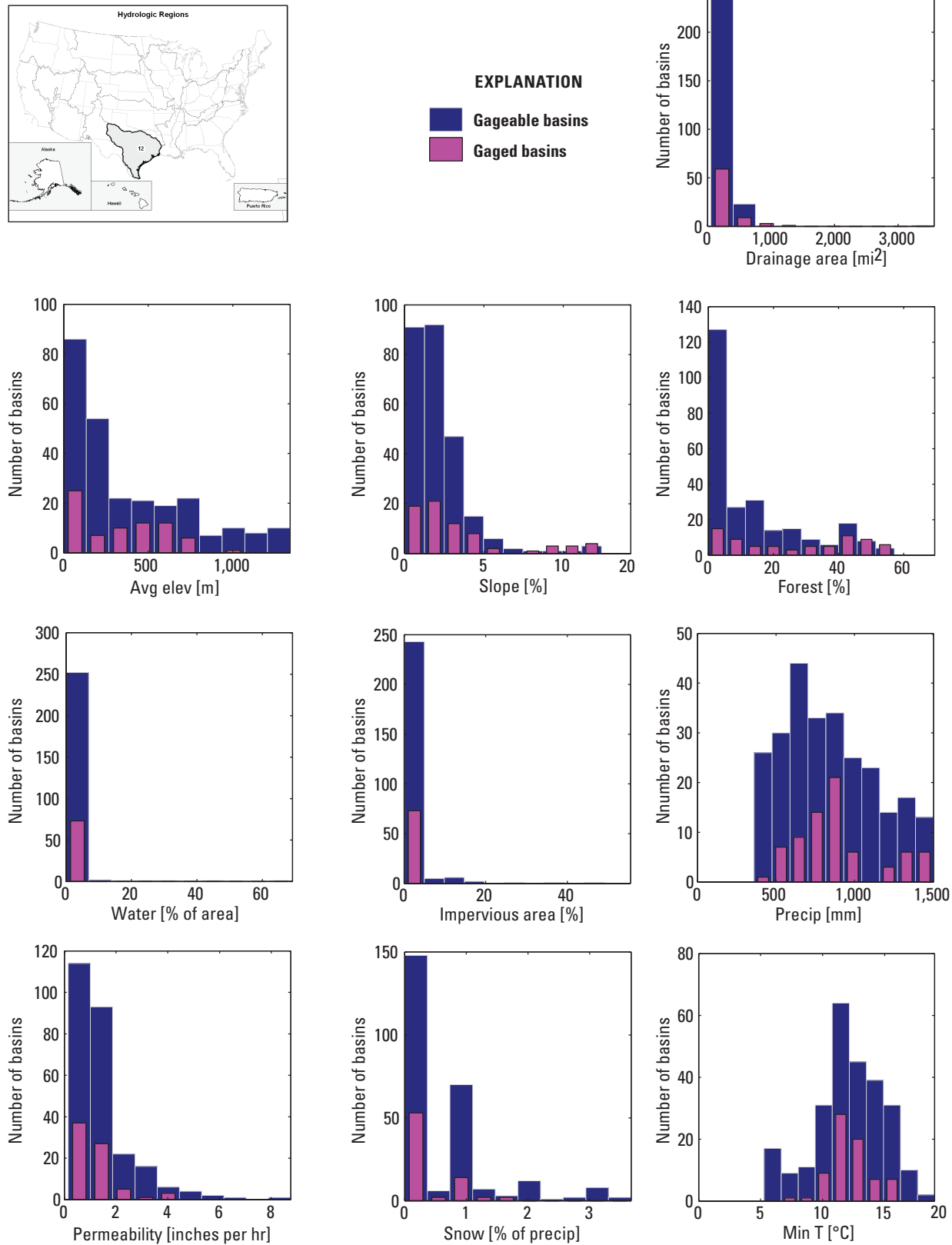
HUC Region 7: Upper Mississippi
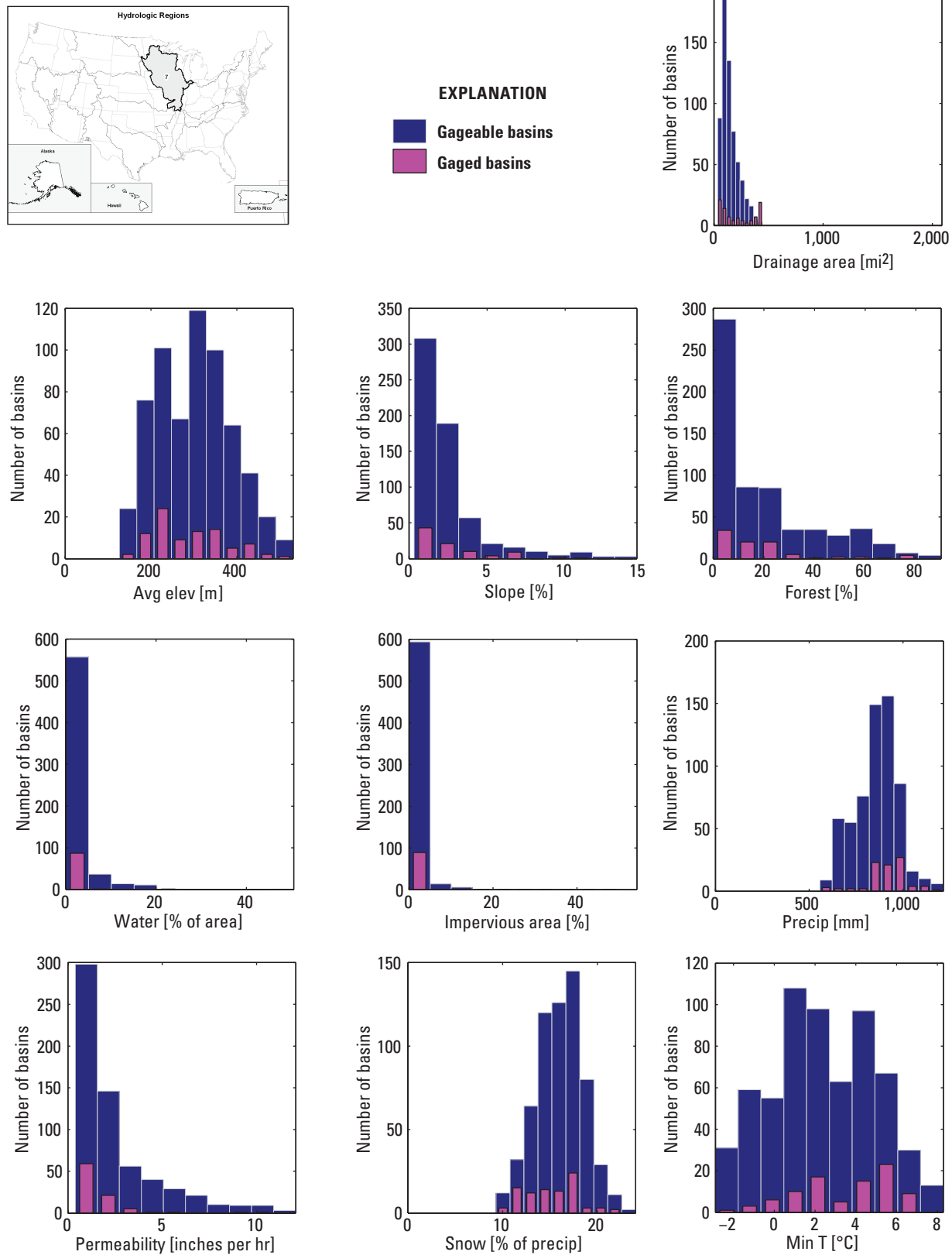
Figure 2–7. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Upper Mississippi (hydrologic unit code (HUC) region 7). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
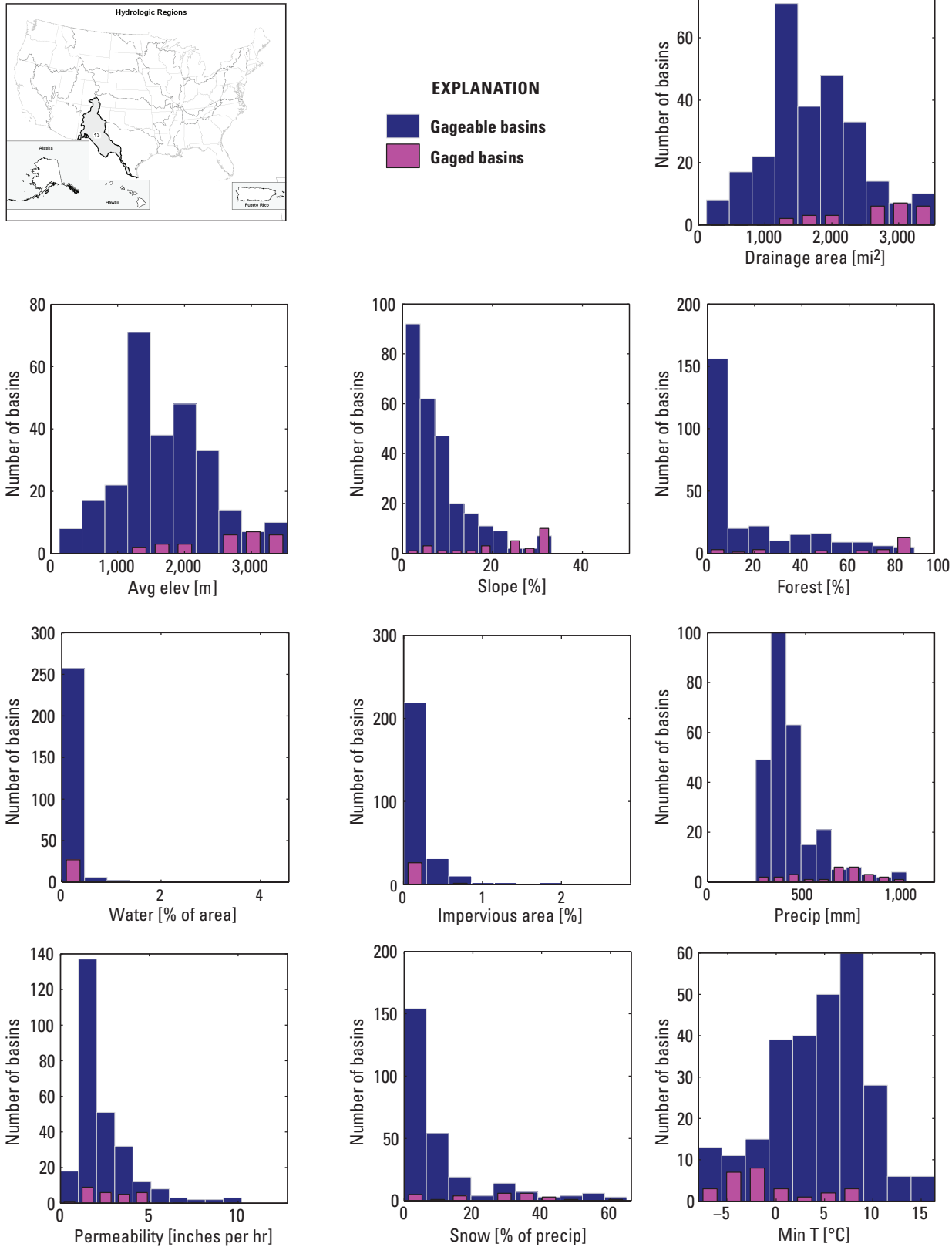
HUC Region 8: Lower Mississippi

Figure 2–8. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Lower Mississippi (hydrologic unit code (HUC) region 8). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 9: Souris-Red-Rainy
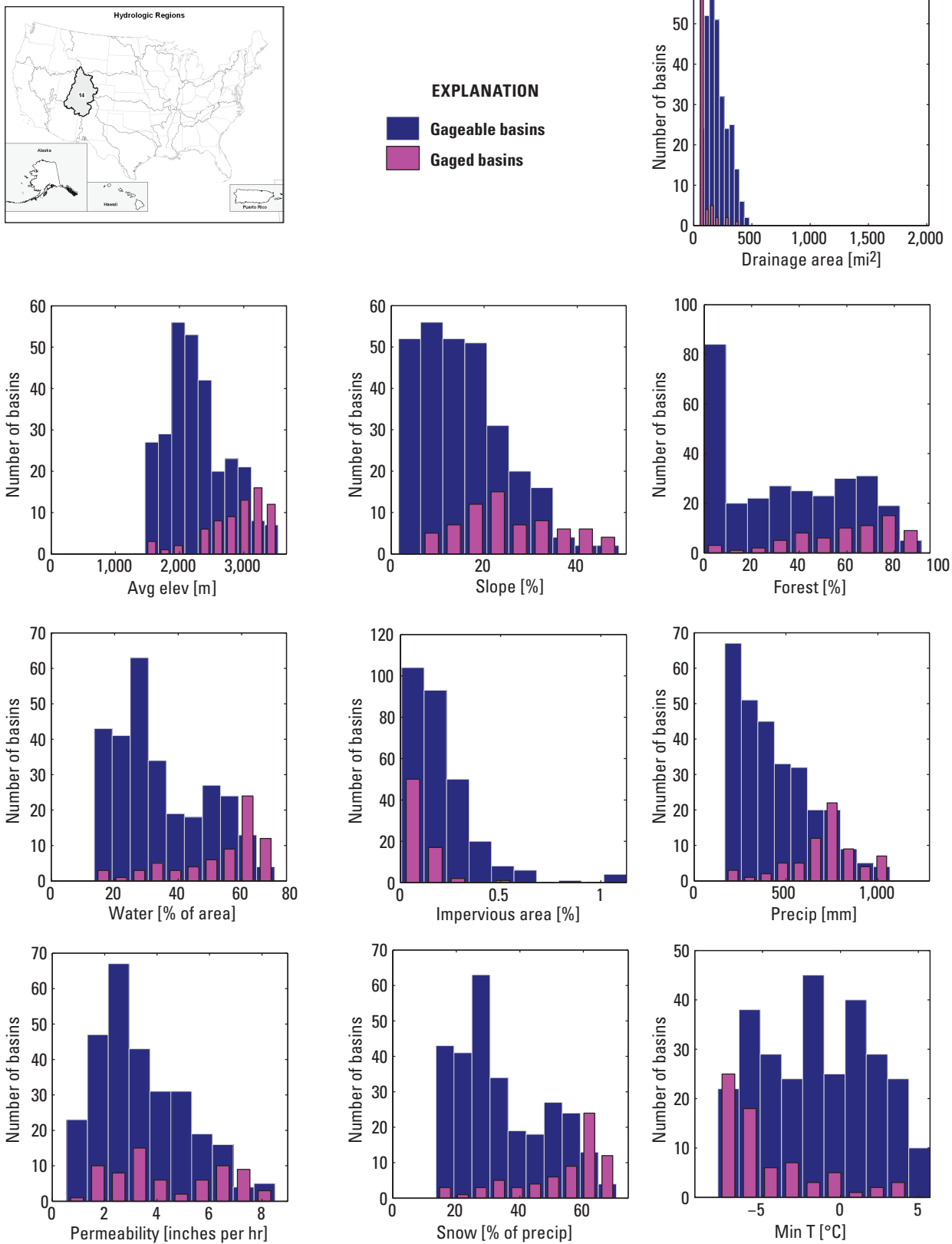

Figure 2–9.   Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Souris-Red-Rainy (hydrologic unit code (HUC) region 9). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 10: Missouri

Figure 2–10. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Missouri (hydrologic unit code (HUC) region 10). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
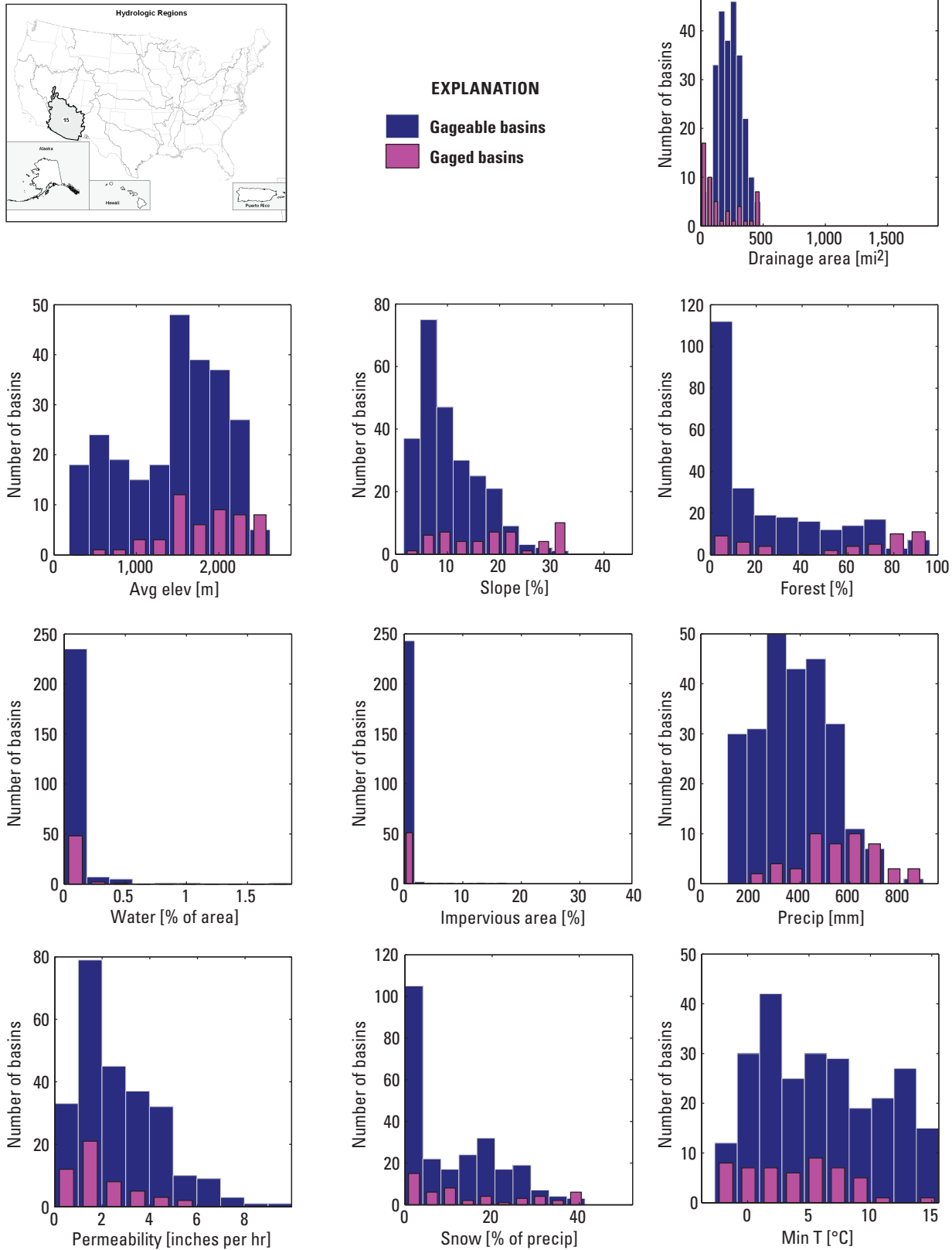
HUC Region 11: Arkansas-White-Red

Figure 2–11. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Arkansas-White-Red Basin (hydrologic unit code (HUC) region 11). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 12: Texas-Gulf

Figure 2–12. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Texas-Gulf Basin (hydrologic unit code (HUC) region 12). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
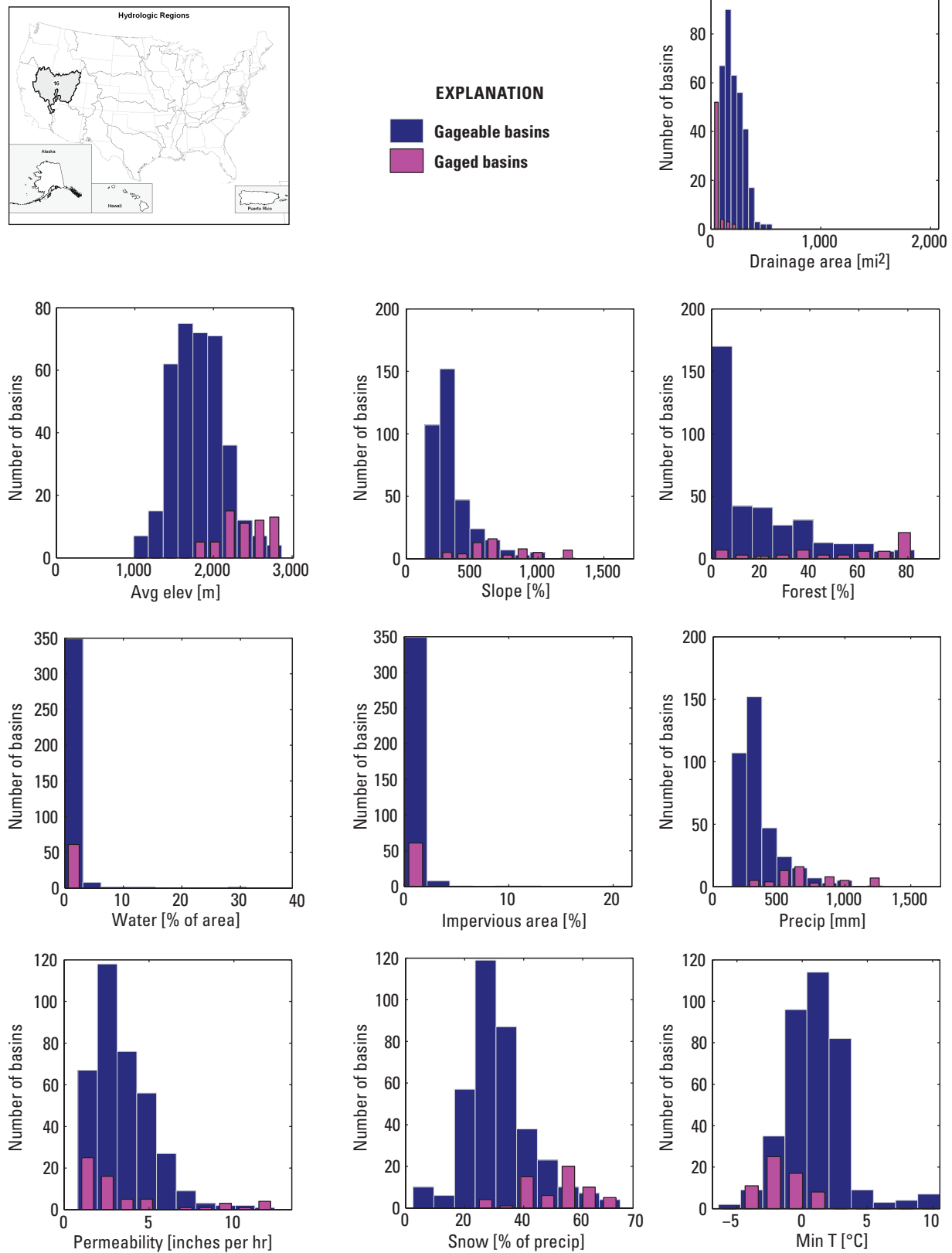
HUC Region 13: Rio Grande

Figure 2–13. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Rio Grande (hydrologic unit code (HUC) region 13). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 14: Upper Colorado

Figure 2–14. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Upper Colorado (hydrologic unit code (HUC) region 14). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
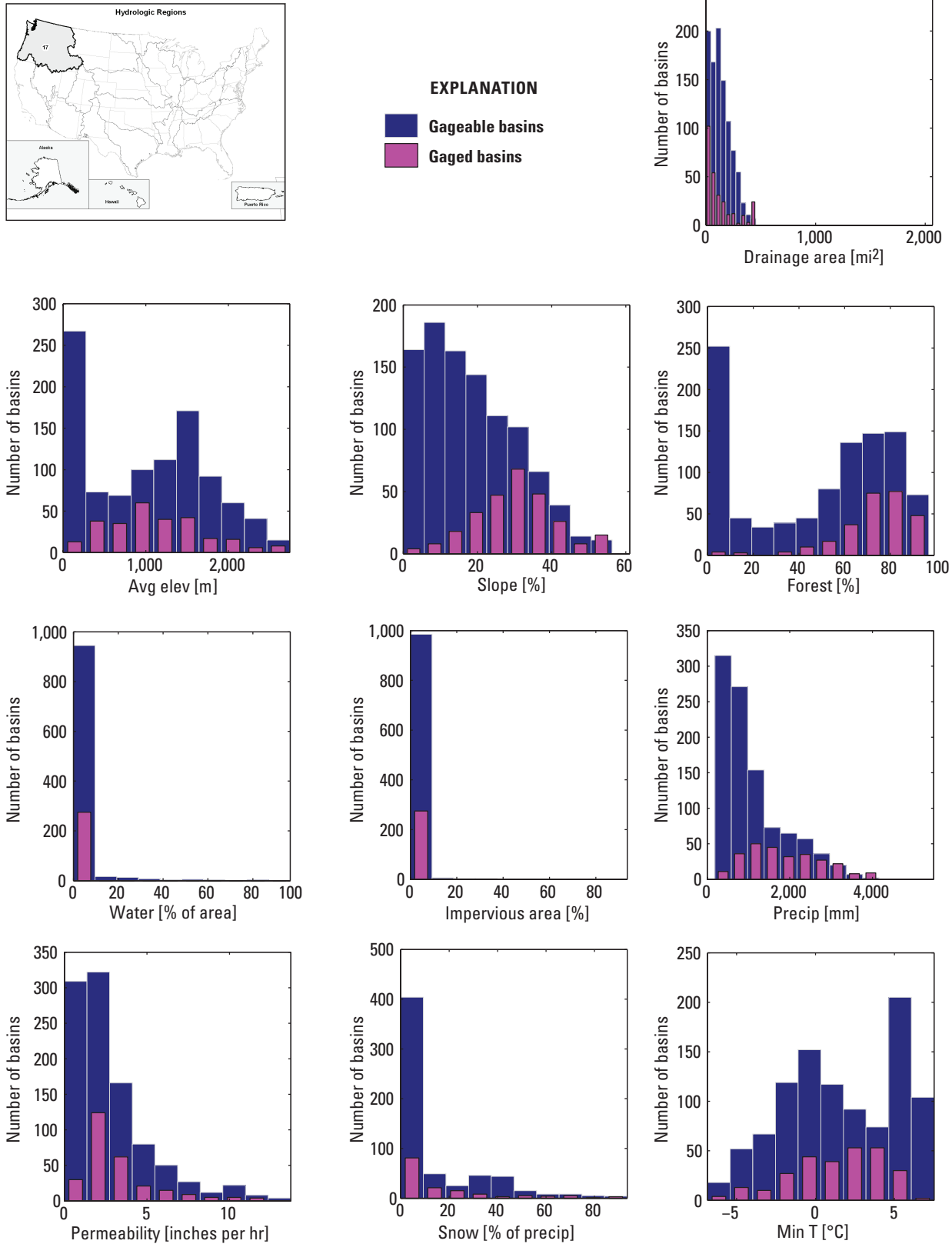
HUC Region 15: Lower Colorado

Figure 2–15. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in Lower Colorado (hydrologic unit code (HUC) region 15). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.
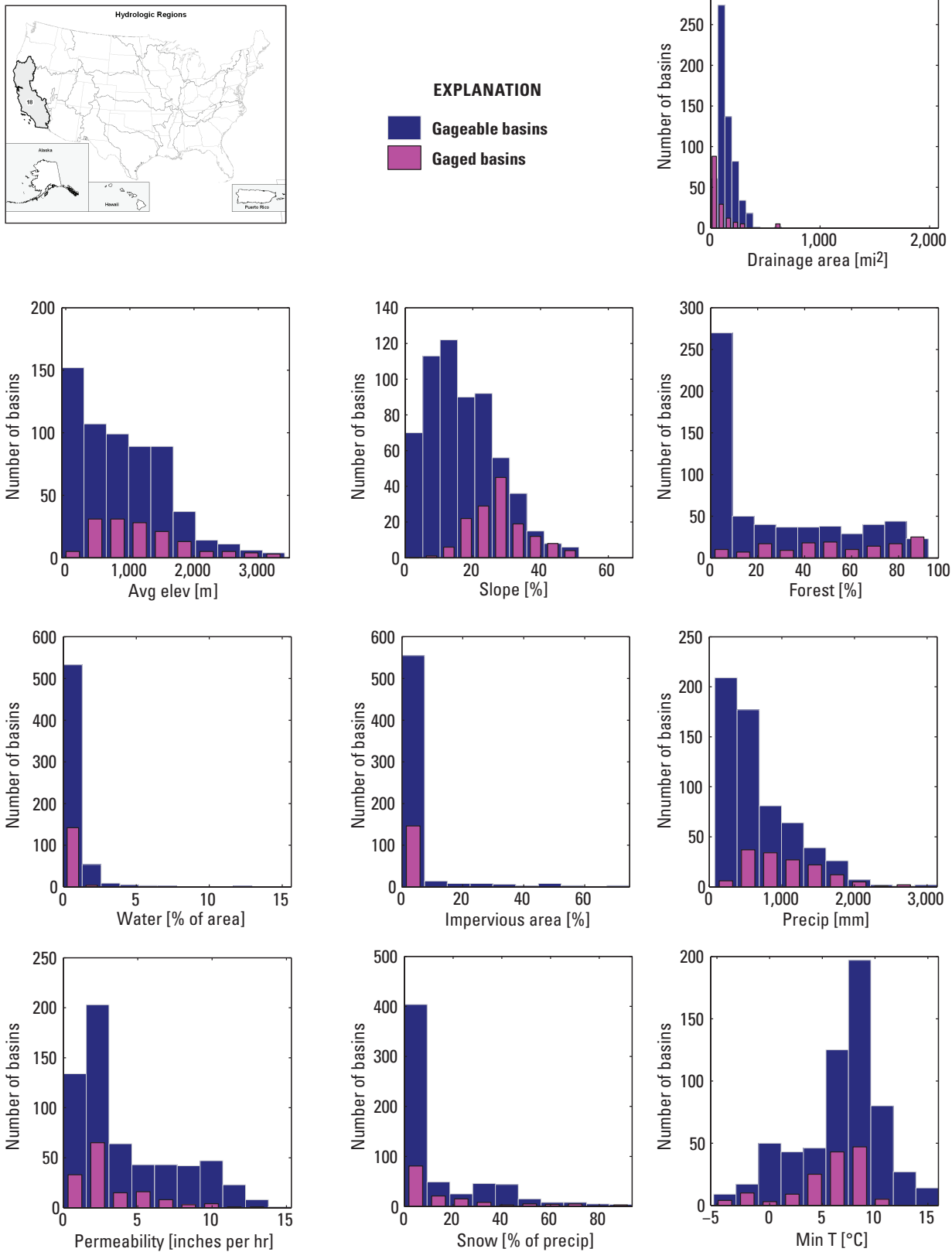
HUC Region 16: Great Basin

Figure 2–16. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Great Basin (hydrologic unit code (HUC) region 16). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 17: Pacific Northwest

Figure 2–17. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in the Pacific Northwest (hydrologic unit code (HUC) region 17). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.

HUC Region 18: California

Figure 2–18. Histograms of basin attributes at gageable watersheds and gaged basins (reference-quality streamgage dataset) in California (hydrologic unit code (HUC) region 18). m, meter; mi², square mile; mm, millimeter; in/hr, inches per hour; %, percent; °C, degrees Celsius.