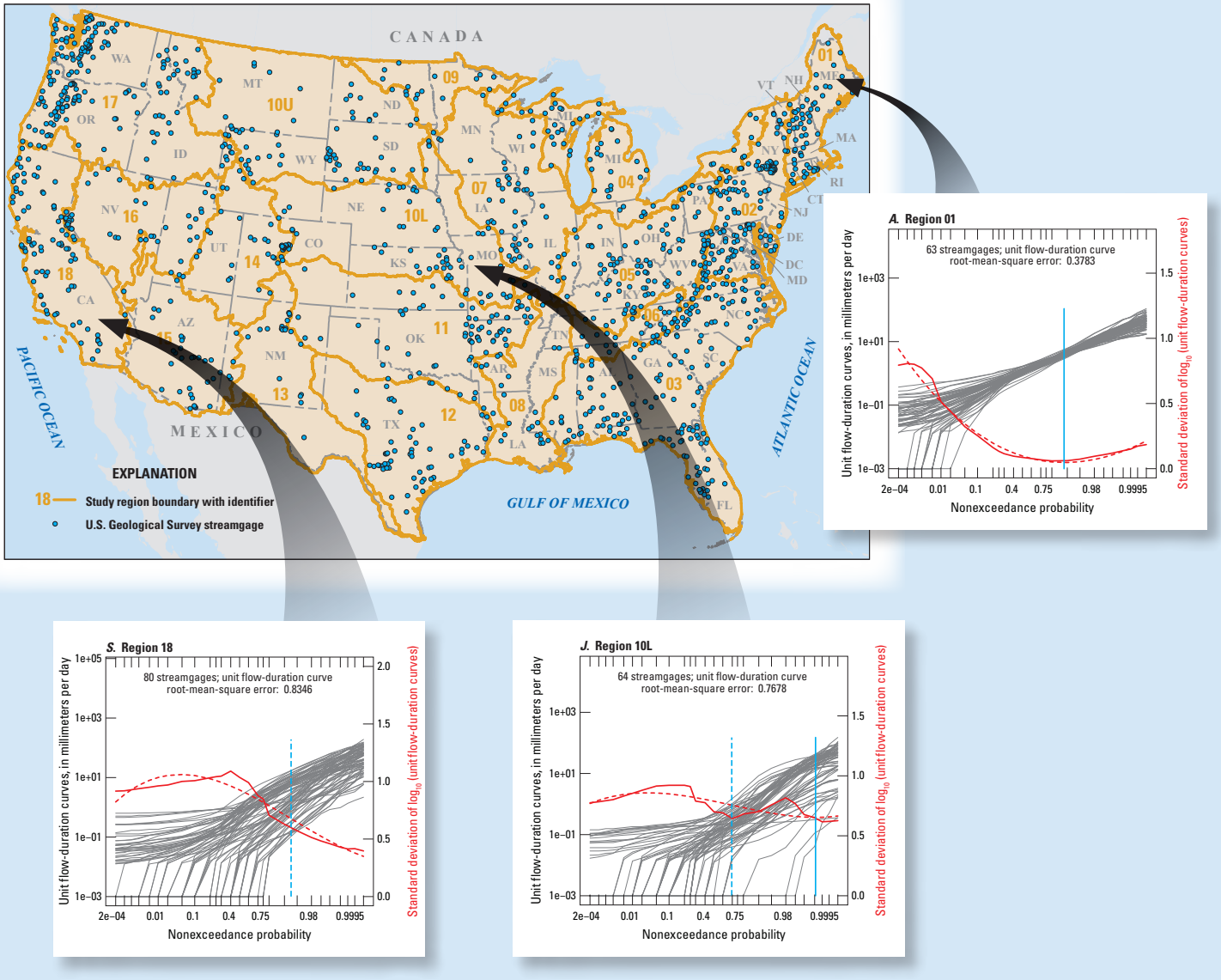
USGS
science for a changing world

Water Availability and Use Science Program

Refinement of a Regression-Based Method for Prediction of Flow-Duration Curves of Daily Streamflow in the Conterminous United States



Scientific Investigations Report 2018–5072

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

Cover. Map of conterminous United States showing study regions and streamgages, and sample graphs showing unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from streamgages in the region.

Refinement of a Regression-Based Method for Prediction of Flow-Duration Curves of Daily Streamflow in the Conterminous United States

By Thomas M. Over, William H. Farmer, and Amy M. Russell

Water Availability and Use Science Program

Scientific Investigations Report 2018–5072

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

Suggested citation:
Over, T.M., Farmer, W.H., and Russell, A.M., 2018, Refinement of a regression-based method for prediction of flow-duration curves of daily streamflow in the conterminous United States: U.S. Geological Survey Scientific Investigations Report 2018–5072, 34 p., https://doi.org/10.3133/sir20185072.

ISSN 2328-0328 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

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)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch per hour (in/h)	0.0254	meter per hour (m/h)

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square kilometer (km^2)	0.3861	square mile (mi^2)
	Flow rate	
millimeter per year (mm/yr)	0.03937	inch per year (in/yr)

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

$$°F = (1.8 \times °C) + 32.$$

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

$$°C = (°F - 32) / 1.8.$$

Datum

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

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

Abbreviations

AIC	Akaike Information Criterion
CONUS	conterminous United States
FDC	flow-duration curve
GAGES–II	Geospatial Attributes of Gages for Evaluating Streamflow, version II
GOF	goodness-of-fit
IQR	interquartile range
LOOCV	leave-one-out cross-validation
MAP	mean annual precipitation
PM	performance measure
pseudo-R^2	pseudo coefficient of determination
RMSE	root-mean-square error
USGS	U.S. Geological Survey
WY	water year (A water year is the 12-month period October 1 through September 30 designated by the calendar year in which it ends.)

Refinement of a Regression-Based Method for Prediction of Flow-Duration Curves of Daily Streamflow in the Conterminous United States

By Thomas M. Over, William H. Farmer, and Amy M. Russell

Abstract

Regional regression is a common tool used to estimate daily flow-duration curves (FDCs) at ungaged locations. In this report, several refinements to a particular implementation of the regional regression method for estimating FDCs are evaluated by consideration of different methodological options through a leave-one-out cross-validation procedure in the 19 major river basins of the conterminous United States. Regression analyses in this report are based on streamflow data from water years 1981–2013 (October 1, 1980 to September 30, 2013) from 1,378 mostly undisturbed watersheds. Linear regression using selected basin characteristics at 27 quantiles with nonexceedance probabilities ranging from 0.02 to 99.98 percent was applied. The regression computations were primarily by weighted least squares, with left-censored Gaussian regression solved by maximum likelihood in the presence of zero-valued quantiles.

The regional regression method as applied to the FDC estimation problem includes several methodological options that require determination of the better of two or more choices. The options considered in this report include (1) the setting of the maximum number of basin characteristics considered in the regression models for each region, (2) the method of placing the quantiles into groups ("flow regimes") having the same basin characteristics used as independent variables, (3) the maximum number of candidate models retained from regressions at the single-quantile level that are retained for testing of the best model at the flow-regime scale, and (4) whether drainage area should be forced into the models. In all, 5 binary options were considered for most regions, resulting in 32 methodological combinations. Leave-one-out cross-validation predictions of FDC quantiles at each streamgage used in the study were used to evaluate compared options. Various performance measures were computed based on the predicted quantiles; these were combined by region and the methods were ranked for each measure.

Based on examination of the ranked methods compared across the measures, the following treatments produced the more accurate results: (1) using fewer basin characteristics (of the two options considered), (2) utilizing a variance of the unit FDC-based method of determining the flow regimes rather than fixed regimes, (3) retaining more models from the quantile-level regressions regime-wide consideration, and (4) forcing drainage area into the regression models. Results of analyses also indicate that performance varies more by region than by methodological option, with FDCs in arid regions and those with a large value of a measure of intraregional FDC heterogeneity being harder to predict, particularly with respect to the low-flow quantiles.

Introduction

One of the primary goals of the Water Census Project of the Water Availability and Use Science Program of the U.S. Geological Survey (USGS) is to compute consistent nationwide estimates of current and future water flows and storage. As part of its contribution to this effort, the USGS is developing methods to estimate natural streamflow at ungaged locations. One method being investigated to estimate natural streamflow, the flow-duration curve (FDC) transfer method, requires FDCs to be estimated in ungaged locations. This report is a contribution to that investigation.

A daily FDC is the probability distribution of daily streamflow, and it specifies how often the observed daily streamflow is above or below a given value. Information from FDCs is useful for a variety of purposes in water resources management (Vogel and Fennessey, 1995), hydrologic modeling (Westerberg and others, 2011), and catchment classification and process identification (Ley and others, 2011; Yokoo and Sivapalan, 2011; Cheng and others, 2012). When combined with timing information in a sequence of nonexceedance probabilities, an FDC also can be used as the basis for the estimation of daily streamflow (Fennessey, 1994; Mohamoud, 2008; Archfield and others, 2010; Farmer and others, 2014).

Castellarin and others (2013) identify four classes of methods to predict FDCs in ungaged basins: (1) regression methods, (2) index flow methods (including parametric methods), (3) geostatistical methods, and (4) estimation from short records. Regression-based methods have been developed

and applied in several recent USGS publications (Ries and Friesz, 2000; Flynn, 2003; Perry and others, 2004; Ahearn, 2010; Archfield and others, 2010; Esralew and Smith, 2010; Linhart and others, 2012; Bent and others, 2014; Farmer and others, 2014; Over and others, 2014; Stuckey and others, 2014; Gazoorian, 2015; Stuckey, 2016; Ziegeweid and others, 2015).

In the regression method, the FDC is characterized by a monotonic sequence of selected discrete quantiles. Characterizing FDCs in this way avoids the need to select a probability distribution for the FDCs, a need that has proven to be challenging (Castellarin and others, 2004; Archfield, 2009; Blum and others, 2017). Basin characteristics, which often must be transformed because of deviations from normality, are used as explanatory variables in regression equations to predict the quantiles and may be selected separately for each quantile or by using one set of characteristics for each contiguous group of quantiles, such as high-, medium-, and low-flow quantiles. Challenges for the regression method include the following: (1) making a choice about how many quantiles to predict and how to define the contiguous groups of quantiles for which a common set of basin characteristics is used for prediction, if any; and (2) avoiding inconsistently predicted quantiles, where a quantile at a lower nonexceedance probability is predicted to be larger than a quantile at a higher nonexceedance probability (Archfield and others, 2010; Poncelet and others, 2017). Predicting a dense sequence of quantiles reduces the importance of the accuracy of the method used to interpolate between the quantiles, if such interpolates are needed, but it increases the number of prediction equations needed. Similarly, selecting basin characteristics separately for each quantile also increases the number of independent equations, but it avoids the need to define quantile groups. An additional challenge to the regression method is that, like most FDC-prediction methods, it is based on an assumption of homogeneous regions and stationary streamflow distributions.

Purpose and Scope

This report presents an analysis to address the need for refinement of a regression-based method for the prediction of period-of-record FDCs at ungaged locations based on daily streamflow data for natural conditions. Different methodological options of a form of the regression method previously investigated by Farmer and others (2014) were evaluated. Linear regression models developed using basin characteristics for explanatory variables were used to estimate 27 FDC quantiles with nonexceedance probabilities ranging from 0.02 to 99.98 percent. The quantiles of the FDC were divided into as many as three "flow regimes" (low, medium, high) with the same set of basin characteristics used for regression models of each quantile in each given regime. Observed FDC quantiles were computed from daily streamflow during water years (WYs) 1981 to 2013 from 1,378 relatively undisturbed "reference" quality streamgages as defined in the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES–II) dataset (Falcone, 2011). A WY is the 12-month period from October 1 through September 30 designated by the calendar year in which it ends. Sets of regression equations were developed for each of the 19 major river basins of the conterminous United States (CONUS) (fig. 1).

Five binary methodological options were investigated: (1) whether or not to force drainage area into the set of basin characteristics used to develop the regression equations, (2) whether flow regimes were defined by preselected sets of quantiles or by an automated procedure based on the similarity of the unit FDCs (that is, FDCs divided by drainage area) within each region, (3) two options on the maximum number of basin characteristics to use in the regression models, (4) two values for the maximum number of candidate models retained from regressions at the single-quantile level for testing of the best model at the flow-regime level, and (5) whether the best model for a flow regime is determined using the adjusted coefficient of determination or using the Akaike Information Criterion (AIC). Statistics characterizing the differences between the predicted and observed FDCs were computed for each combination of methodological options and were used to evaluate the different methodological options.

Description of Study Area

The domain of the analysis presented in this report is the CONUS. For this report, the CONUS is divided into the 19 regions used in the GAGES–II dataset (Falcone, 2011) based on major river basins and similar to 2-digit hydrologic unit code watersheds (fig. 1). The CONUS includes a wide range of hydrologic and climatic conditions (table 1). Climatically, regional median mean annual precipitation (MAP) of the basins ranges from 1,678 millimeters (mm) in region 17 (Pacific Northwest), which also includes the basin with the largest MAP of 4,606 mm, to 507 mm in region 09 (Souris-Red-Rainy) and 517 mm in region 15 (Lower Colorado). The regional median of the aridity index computed by dividing mean annual potential evapotranspiration by MAP is highest in regions 12 (Texas-Gulf), 09, and 15 (1.180, 1.143, and 1.131, respectively), whereas it is lowest in region 17 (0.312).

The highly variable climatic conditions in the study regions interact with other factors to create an even wider range, proportionally, of streamflow conditions as measured by unit mean discharge (that is, discharge per unit area): unit mean discharge ranges from a minimum of 0.11 millimeter per year at a streamgage in region 15 (Lower Colorado) to a maximum of 4,028 millimeters per year at a streamgage in region 17 (Pacific Northwest) (Russell and others, 2018). For all regions except region 01 (New England), the ratio of maximum to minimum unit mean discharge is much larger than the corresponding ratio of maximum to minimum MAP. Streamflow conditions, as measured by the runoff ratio (unit mean discharge per unit MAP), also vary widely and vary within and among regions, similar to the unit mean discharge values.

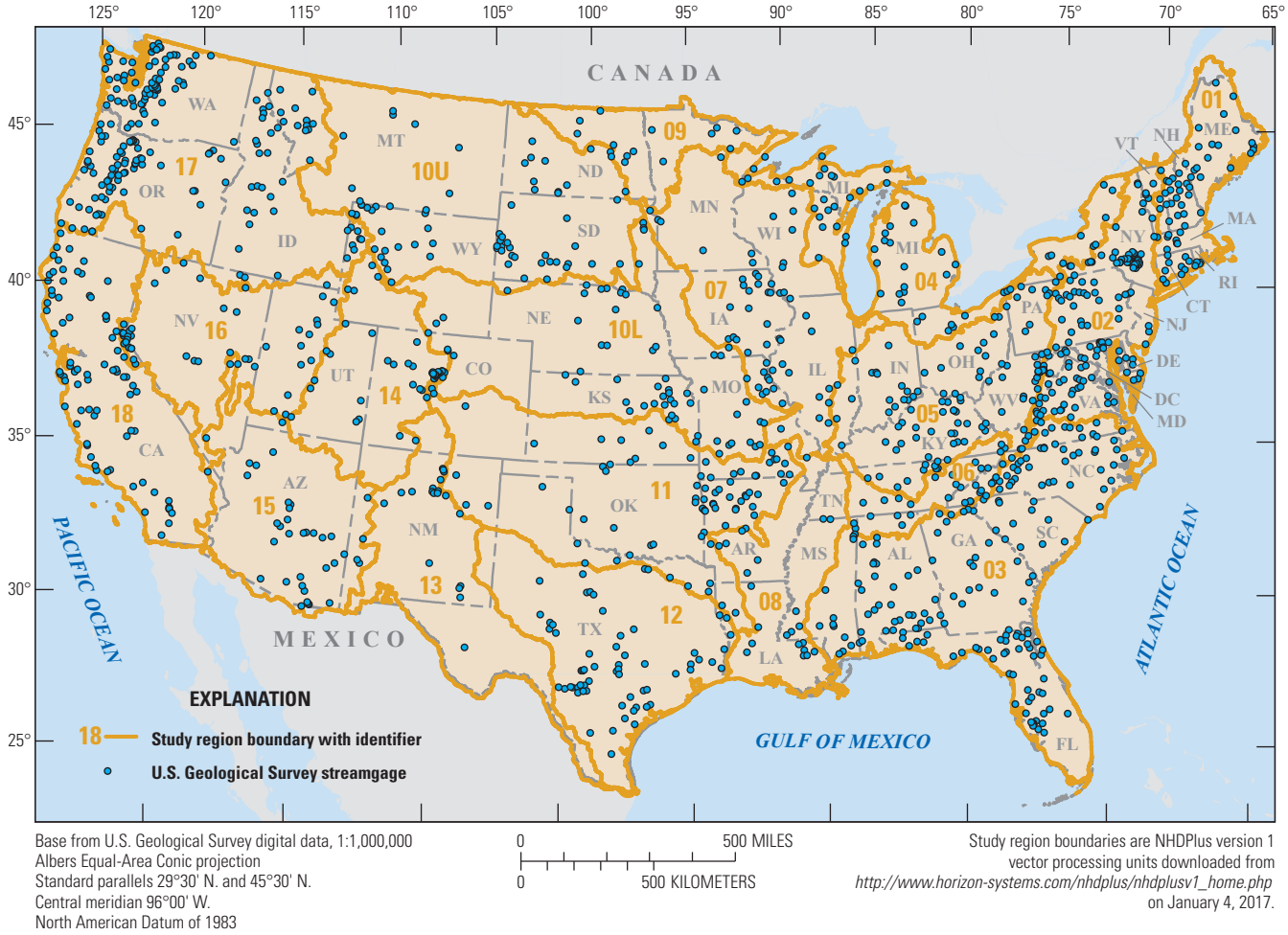


Figure 1. Map of conterminous United States showing study regions and streamgages used in this study.

Densities of the streamgages selected for use in this report also vary widely across the CONUS. The largest density is in region 02 (Mid-Atlantic) where 0.514 streamgages per 100 square kilometers (km^2) were used in this report; at the other extreme, in region 13 (Rio Grande), only 0.043 streamgages per 100 km^2 were used, or 19 streamgages in a region with an area of 437,065 km^2. For additional information regarding densities of reference-qaulity streamgages across the CONUS, see Kiang and others (2013).

Redefining regional boundaries to increase the hydrologic homogenity of the regions could be useful for improving regional FDC predictions (Castellarin and others, 2013) but was not considered in this report.

Methods of Study

In this report, FDCs are represented with 27 quantiles with nonexceedance probabilities of 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 98, 99, 99.5, 99.8, 99.9, 99.95, and 99.98 percent. Considering each of the 19 study regions (fig. 1) individually, these FDCs were estimated using a leave-one-out cross-validation (LOOCV) scheme of linear regression. In each region, USGS streamgages that were identified as being of "reference" quality (as defined by Falcone and others [2010] and Falcone [2011]) and had at least 10 complete WYs of daily streamflow data during the study period from WY 1981 through WY 2013 (that is, October 1, 1980, through September 30, 2013) were selected. The criterion of 10 complete WYs of daily streamflow data was adopted based on the recommendation of the same criterion for flood flow frequency analysis by the Interagency Advisory Committee on Water Data (1982, p. 2). Daily streamflow data were downloaded from the USGS National Water Information System (USGS, 2016), and all available complete WYs of daily streamflow data during the study period from the selected streamgages (Russell and others, 2018) were used to compute the empirical FDC quantiles to which regression equations were fitted. Positive quantiles were transformed with the base–10 logarithm. When nonexceedance quantiles had associated flow values of zero, a censored regression approach was applied. In total, 1,378 streamgages with 35,053 WYs of daily streamflow data, or approximately 12.8 million daily streamflow values, were used in this analysis.

Table 1. Basic properties of streamgage records and gaged watersheds used in this study by study region.

[km², square kilometer; mm/yr, millimeter per year; mm, millimeter; m, meter; PET, potential evapotranspiration; RMSE, root-mean-square error; FDC, flow-duration curve; NA, not applicable. Shading with "cooler" colors indicates numerically smaller values, whereas shading with "warmer" colors indicates numerically larger values, where the cool-to-warm color order is green, yellow, orange, and red, except for unit mean discharge, runoff ratio, basin precipitation, and basin elevation values, where the ordering is opposite to that. Drainage area, basin elevation, basin precipitation, and basin potential evapotranspiration values are from the Geospatial Attributes of Gages for Evaluating Streamflow, version II, dataset (Falcone, 2011). Basin precipitation is based on data from 1971 to 2000, and basin potential evapotranspiration is based on data from 1961 to 1990]

Region	Region name	Area (km²)	Number of streamgages	Streamgage density (per 100 km²)	RMSE of unit FDC quantiles	Nonexceedance probability of low regime break when all streamgages are included	Nonexceedance probability of high regime break when all streamgages are included	Mean number of complete water years (1981–2013)	Drainage area (km²)			Unit mean discharge (mm/year)				Runoff ratio (unit mean discharge/mean annual basin precipitation, dimensionless)			
									Minimum	Median	Maximum	Minimum	Mean	Maximum	Maximum/minimum	Minimum	Median	Maximum	Maximum/Minimum
01	New England	191,019	63	0.330	0.378	NA	0.883	25.48	1.6	90.8	3,676	537	728	1,136	2.12	0.469	0.583	0.730	1.56
02	Mid-Atlantic	276,189	142	0.514	0.452	NA	0.954	26.02	1.7	149.5	1,778	276	490	1,295	4.69	0.253	0.463	0.887	3.51
03	S. Atlantic-Gulf	724,336	176	0.243	0.650	0.352	0.978	26.63	3.1	270.8	6,622	109	388	1,308	12.0	0.084	0.288	0.909	10.79
04	Great Lakes	617,056	67	0.109	0.530	0.032	0.934	24.06	6.2	341.6	2,946	222	384	808	3.64	0.265	0.415	0.680	2.56
05	Ohio	421,961	105	0.249	0.558	0.369	0.983	24.12	4.3	295	2,953	235	499	1,031	4.39	0.179	0.411	0.715	3.99
06	Tennessee	105,950	34	0.321	0.487	0.185	0.981	27.12	4.4	178	2,082	326	687	1,318	4.04	0.273	0.453	0.681	2.50
07	Upper Miss.	492,027	58	0.118	0.651	0.408	0.982	26.34	65.4	692.3	3,858	76	284	482	6.38	0.127	0.306	0.442	3.47
08	Lower Miss.	276,482	21	0.076	0.571	0.344	0.981	26.05	69.8	542.9	2,073	398	487	741	1.86	0.242	0.308	0.459	1.90
09	Souris-Red-Rainy	227,845	13	0.057	0.742	0.956	NA	29.46	92.8	2,358	7,582	15.5	29.3	235	15.1	0.0352	0.0616	0.329	9.34
10L	Lower Mo.	540,005	64	0.119	0.768	0.695	0.998	25.13	4	669.4	8,265	0.53	204	1,010	1,915	0.0011	0.214	0.881	831
10U	Upper Mo.	809,225	70	0.087	0.918	0.970	NA	25.70	20.8	521.5	25,790	4.40	47.6	737	168	0.0132	0.081	0.753	57.1
11	Arkansas-White-Red	642,284	80	0.125	0.822	NA	NA	23.00	9.6	494	6,885	0.97	361	755	777	0.0021	0.304	0.525	248
12	Texas-Gulf	471,091	62	0.132	0.794	NA	NA	27.66	31.8	452.8	6,665	0.39	91.8	442	1,139	0.0008	0.113	0.297	385
13	Rio Grande	437,065	19	0.043	1.012	NA	NA	22.89	29.7	163.3	3,547	1.26	96.5	411	325	0.0042	0.141	0.541	128
14	Upper Colorado	293,569	41	0.140	0.635	0.572	0.9996	25.41	7.6	63.4	460	18.1	419	780	43.0	0.0374	0.533	0.964	25.8
15	Lower Colorado	424,349	39	0.092	0.798	NA	NA	24.69	2.2	213.4	7,782	0.11	38.2	227	2,040	0.00021	0.066	0.251	1,194
16	Great Basin	367,049	48	0.131	0.669	0.254	0.995	24.29	1.5	44.85	716	12.1	181	1,134	93.6	0.0388	0.260	1.25	32.2
17	Pacific Northwest	783,066	196	0.250	0.548	0.286	0.982	24.81	1.7	225.7	14,270	8.26	1,085	4,028	487	0.0271	0.652	1.17	43.0
18	California	436,631	80	0.183	0.835	0.929	NA	26.39	5.4	145.3	1,980	18.7	386	2,064	110	0.0351	0.354	0.896	25.5

Table 1. Basic properties of streamgage records and gaged watersheds used in this study by study region.—Continued

[km², square kilometer; mm/yr, millimeter per year; mm, millimeter; m, meter; PET, potential evapotranspiration; RMSE, root-mean-square error; FDC, flow-duration curve; NA, not applicable. Shading with "cooler" colors indicates numerically smaller values, whereas shading with "warmer" colors indicates numerically larger values, where the cool-to-warm color order is green, yellow, orange, and red, except for unit mean discharge, runoff ratio, basin precipitation, and basin elevation values, where the ordering is opposite to that. Drainage area, basin elevation, basin precipitation, and basin potential evapotranspiration values are from the Geospatial Attributes of Gages for Evaluating Streamflow, version II, dataset (Falcone, 2011). Basin precipitation is based on data from 1971 to 2000, and basin potential evapotranspiration is based on data from 1961 to 1990]

Region	Mean annual basin precipitation (mm)				Fraction of zero discharges			Median basin elevation (m)			Mean annual basin potential evapotranspiration (PET) (mm)			Aridity index: Mean annual basin PET/mean annual basin precipitation		
	Minimum	Median	Maximum	Maximum/Minimum	Minimum	Mean	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum
01	974	1,238	1,883	1.93	0	0.0003	0.0084	54	334	980	446	553	691	0.237	0.461	0.552
02	843	1,128	1,748	2.07	0	0.0010	0.0357	8	466	1,039	483	607	826	0.277	0.575	0.727
03	1,061	1,334	1,944	1.83	0	0.0147	0.3350	5	90	851	655	953	1,162	0.363	0.689	0.899
04	777	860	1,242	1.60	0	0.0021	0.0855	130	320	643	510	566	679	0.424	0.668	0.832
05	965	1,189	1,530	1.59	0	0.0139	0.1344	154	324	1,151	545	720	823	0.386	0.617	0.752
06	1,146	1,513	2,072	1.81	0	0.0014	0.0357	160	796.5	1,490	578	692	875	0.302	0.487	0.659
07	565	919	1,201	2.13	0	0.0148	0.1653	148	288	501	538	696	805	0.654	0.759	1.146
08	1,369	1,550	1,690	1.23	0	0.0162	0.1510	35	86	396	819	975	1,016	0.501	0.608	0.693
09	441	507	722	1.64	0	0.1180	0.6314	341	454	545	528	574	602	0.739	1.143	1.364
10L	498	903	1,195	2.40	0	0.0452	0.7786	233	399	3,428	339	745	842	0.284	0.867	1.413
10U	274	551	980	3.58	0	0.0606	0.8091	432	1,413	3,270	307	558	700	0.379	1.059	2.280
11	416	1,164	1,576	3.79	0	0.0758	0.9839	66	387	3,680	306	826	1,005	0.350	0.695	1.778
12	461	854	1,488	3.23	0	0.2355	0.9929	12	357	938	872	1,009	1,191	0.685	1.180	2.074
13	298	674	898	3.02	0	0.3089	0.9905	1,288	2,698	3,314	370	465	811	0.424	0.676	2.057
14	485	732	1,106	2.28	0	0.0136	0.3385	2,285	3,205	3,601	307	354	524	0.312	0.481	1.029
15	223	517	908	4.08	0	0.2228	0.9965	523	1,965	2,759	451	698	1,231	0.497	1.131	5.531
16	310	667	1,726	5.56	0	0.0146	0.2999	1,759	2,366	3,015	402	455	614	0.259	0.657	1.966
17	305	1,678	4,606	15.1	0	0.0063	0.4966	101	1,122	2,753	362	518	656	0.118	0.312	1.947
18	378	1,108	2,826	7.48	0	0.1033	0.7219	249	976	3,290	313	668	894	0.218	0.607	2.153

For each region, the sequence of modeled FDC quantiles was divided into as many as three groups of contiguous quantiles called flow regimes. A common set of basin characteristics was selected for the regression models for each flow regime in each region. Basin characteristics were selected from the GAGES–II dataset (Falcone and others, 2010; Falcone, 2011) following the approach of Farmer and others (2014). Selected candidate basin characteristics were first transformed to obtain approximately Gaussian distributions. The basin characteristics were selected for inclusion in the ultimate regional regression by an exhaustive search algorithm with the option of forcing drainage area into the regression models. The maximum number of basin characteristics in the models was set with a formula that depends on the number of streamgages in the model.

Data Preparation

The data used to test the methodological options for FDC estimation by the regional regression method in this report consist of the empirical FDC quantiles, which are the dependent variables in the regressions, and the basin characteristics, which are the independent variables. This section describes the computation of the FDC quantiles and the selection and processing of the basin characteristics.

Computation of Flow-Duration Curves

The reported daily streamflow data used in this study were rounded to the nearest 0.01 cubic foot per second (ft^3/s). In particular, discharges with magnitudes less than 0.005 ft^3/s were censored (set to zero). At some streamgages, a large fraction of the observations had zero discharges and were censored (table 1). In addition, in region 03 (S. Atlantic-Gulf), the region with the lowest basin elevations (table 1), two of the qualifying streamgages, USGS station numbers 02313700 and 02322800, have negative values because of tidal and wind effects 18.2 and 0.2 percent of the time, respectively, during the study period. Because a preliminary investigation indicated that it would be difficult to determine when using the results of this study for predicting which locations of interest might also have such effects, these stations were retained. The general approach taken in this study assumes streamflow is positive; however, for purposes of this study, these negative values were censored by setting them to zero.

Because of the censored data values, two versions of the FDC quantiles from streamgage records were computed and are archived in Russell and others (2018). In the first version, quantiles were estimated from the original data with the `quantile` function of the R **`stats`** package (R Core Team, 2016), giving zeroes when the quantile value is less than 0.01. In the second version, termed "filled quantiles," the censored quantile values were filled with estimated positive values by using the function `quantile.lcens` from the R package **`smwrQW`** (D.L. Lorenz, U.S. Geological Survey, written commun., 2016) with the method parameter set to "log MLE," which applies a maximum likelihood estimation assuming the data are lognormally distributed (Helsel, 2012, p. 64–65). In both versions, for the noncensored values, the Blom (1958) plotting position formula was used (Helsel and Hirsch, 2002, p. 23–24):

$$p_i = \frac{i - a}{n + 1 - 2a},\ i = 1, 2, ..., n, \tag{1}$$

where

- p_i is the nonexceedance probability of the corresponding ordered daily discharge value Q_i,
- i is its rank (smallest to largest),
- a is a constant which determines the particular plotting position formula (a=3/8 for the Blom formula), and
- n is the number of daily discharge values.

The plotting position formula was implemented by selecting type 9 of the `quantile` function of the R **`stats`** package. After assignment of the plotting positions, quantile estimates were obtained by linear interpolation between the ordered discharges and their corresponding nonexceedance probabilities p_i.

In the determination of the filled quantiles, occasionally the data values estimated for the largest censored values were larger than the smallest noncensored values. As a result, sometimes quantile values greater than the censoring level were increased, and some of the noncensored filled flow quantile values are greater than the corresponding noncensored unfilled flow quantile values. Overall, among the 27 quantiles estimated for each of the 1,378 streamgages used in this study, 3,071 or 8.3 percent of the quantiles and 380 or 27.6 percent of the streamgages, had zero values that were filled. Of the noncensored quantiles, 223 or 0.65 percent were increased as a result of the filling process, affecting the quantiles of 168 streamgages.

Selection and Processing of Basin Characteristics

Potential basin characteristics were obtained from the GAGES–II dataset (Falcone and others, 2010; Falcone, 2011). Only selected characteristics that align with the goal of predicting natural streamflow were retained. The selected potential characteristics include those characterizing (1) basin morphometry (such as drainage area, slope, and elevation), (2) basin-average soil and land-use properties, (3) basin averages of long-term average climatic properties (precipitation and temperature values and quantities derived from them, such as relative humidity, potential evapotranspiration, dates of first and last freezes), and (4) estimated hydrologic properties (such as monthly runoff) derived from applications of uncalibrated hydrologic models. To reduce the incidence of potential

basin characteristic variables with high cross-correlation, some related morphometric, soil, and land-use variables were deleted, using judgment to decide on the most physically meaningful variables to retain, and monthly climatic and hydrologic variables were analyzed to create new variables characterizing the annual cycle. Two other variables, WATERWETNLCD06 and ELEV_RANGE_M_BASIN, were added by means of computations on existing GAGES–II variables (appendix 1, table 1.1). WATERWETNLCD06 was created by summing the water and wetland landcover fraction variables because both have many zeroes. ELEV_RANGE_M_BASIN is the difference of the maximum and minimum basin elevations and provides elevation range information independent of median basin elevation. Quantitative geologic basin characteristics describing subsurface geology were not used in this study because such variables are not included in the GAGES–II dataset. A complete annotated list of basin characteristic variables retained and developed for use in this analysis is provided in appendix 1, table 1.1.

Additional processing of selected basin characteristics was done on a region-by-region basis, beginning with deletion of variables with a large (greater than 50 percent) fraction across all sites in the region having the same value (usually zero). The basin characteristics considered for use in each region are indicated in appendix 1, table 1.2. The values of remaining characteristics were transformed by a series of steps: (1) centering the value if the coefficient of variation among the basin values was less than 0.10, (2) power-law or logarithmic transformation to minimize the absolute value of the skewness, (3) recentering the value if the coefficient of variation among the transformed basin values was less than 0.10, and (4) dividing by the maximum absolute value to obtain numerical values of comparable magnitude to prevent numerical issues during the regression analysis. Tables of the original and final (transformed) basin characteristic values used in the FDC regressions and the details of the transformations applied are provided in a separate data release (Russell and others, 2018).

Determination of Flow Regimes

Basin characteristics selected for use in regression models for the FDC quantiles should vary between neighboring quantiles as little as possible to account for the correlations among neighboring FDC quantiles and reduce the incidence of inconsistent FDC estimates (Archfield and others, 2010; Castellarin and others, 2013; Poncelet and others, 2017). At the same time, different basin characteristics are expected to be useful predictors for different sections of the FDC, such as low and high flow, based on the presence of different process controls (Castellarin and others, 2013). To balance these two factors, the FDC quantiles in this study were divided into flow regimes; the number of flow regimes depends on the heterogeneity of unit FDCs in the region of interest. Within each regime, the same set of basin characteristics was used to develop regional regression equations with quantile-dependent coefficients. This flow-regime approach has been used previously by Over and others (2014) and Farmer and others (2014).

In this report, two methods of setting the flow regimes are compared. The first method uses a default set of flow regimes regardless of the properties of the FDCs in the region being studied. This default set designates the low-flow regime as extending across nonexceedance probabilities from 0.02 to 10 percent, the medium-flow regime from 20 to 90 percent, and the high-flow regime from 95 to 99.98 percent, similar to those used by Farmer and others (2014) for the southeastern United States. The selected default flow regimes were adjusted slightly from those used in Farmer and others (2014) to a slightly skewed form with a longer low-flow tail. The longer low-flow tail was selected in recognition of the skewed nature of typical streamflow distributions, where, for example, the nonexceedance probability of the mean flow is typically 70 to 80 percent (not shown).

The second method was developed for this study and uses the variance among the observed unit FDCs (FDCs divided by drainage area) to define as many as three custom flow regimes for a given region. Standard deviations of the log-transformed unit FDCs among the streamgages in the region are computed as a function of probability. Then, a third-order polynomial is fit to computed standard deviations. A typical shape of this polynomial in a moderately humid region includes a maximum for low flows, then a decrease to a minimum at moderately high flows (around a nonexceedance probability of 95 to 98), and finally a rise from the minimum for the extreme high flows (fig. 2*C–H*). As many as two regime breaks are placed according to the shape of this polynomial—one at the minimum standard deviation to the right of the maximum, and the other at the inflection point between them.

Not all regions have two regime breaks. For some of the most humid regions, such as regions 01 and 02 (fig. 2*A–B*), the inflection point between the maximum and minimum standard deviations is not observed, indicating that there is no low-flow regime. Less humid to arid regions fall into one of two categories: (1) either they lack the minimum standard deviation in medium-level nonexceedance probabilities and have only the low-medium breakpoint at the inflection point, in which case they lack a high-flow regime (for example, fig. 2*S*), or (2) they have the low-flow maximum but lack both breakpoints, in which case they have only a low-flow regime (for example, fig. 2*N* and *P*). The lack of identified medium- and high-flow regimes in some regions does not mean there are no medium and high flows in the streams of the region, but that the method does not identify such regimes. These regimes are not identified in part because the regime is relatively narrow, and because the smoothing induced by the use of a third-order polynomial obscures the inflection point between the regimes. For example, in figure 2*N*, an inflection point and a minimum to the right of the maximum can be observed in the

unsmoothed standard deviation values indicated by the solid red line.

The custom flow-regime methodology presented in this report was used for two reasons. First, the minimum in unit FDC variance that typically appears near moderately high flows in humid regions seems to be a good way to define the lower bound of the high-flow regime. Near this minimum, Over and others (2014) observed that basin characteristic regression coefficients values crossed through zero in a direction that depended on whether the characteristic would be expected to be larger at high or low flow. Second, typically low flows are harder to predict and cover many more quantiles than high flows. Therefore, the degree of freedom was added to split the low-flow regime arising from the break at the minimum standard deviation into low- and medium-flow regimes. A visual inspection of the regional unit FDC curves confirms the reasonableness of this method of defining the low- and medium-flow regimes (fig. 2).

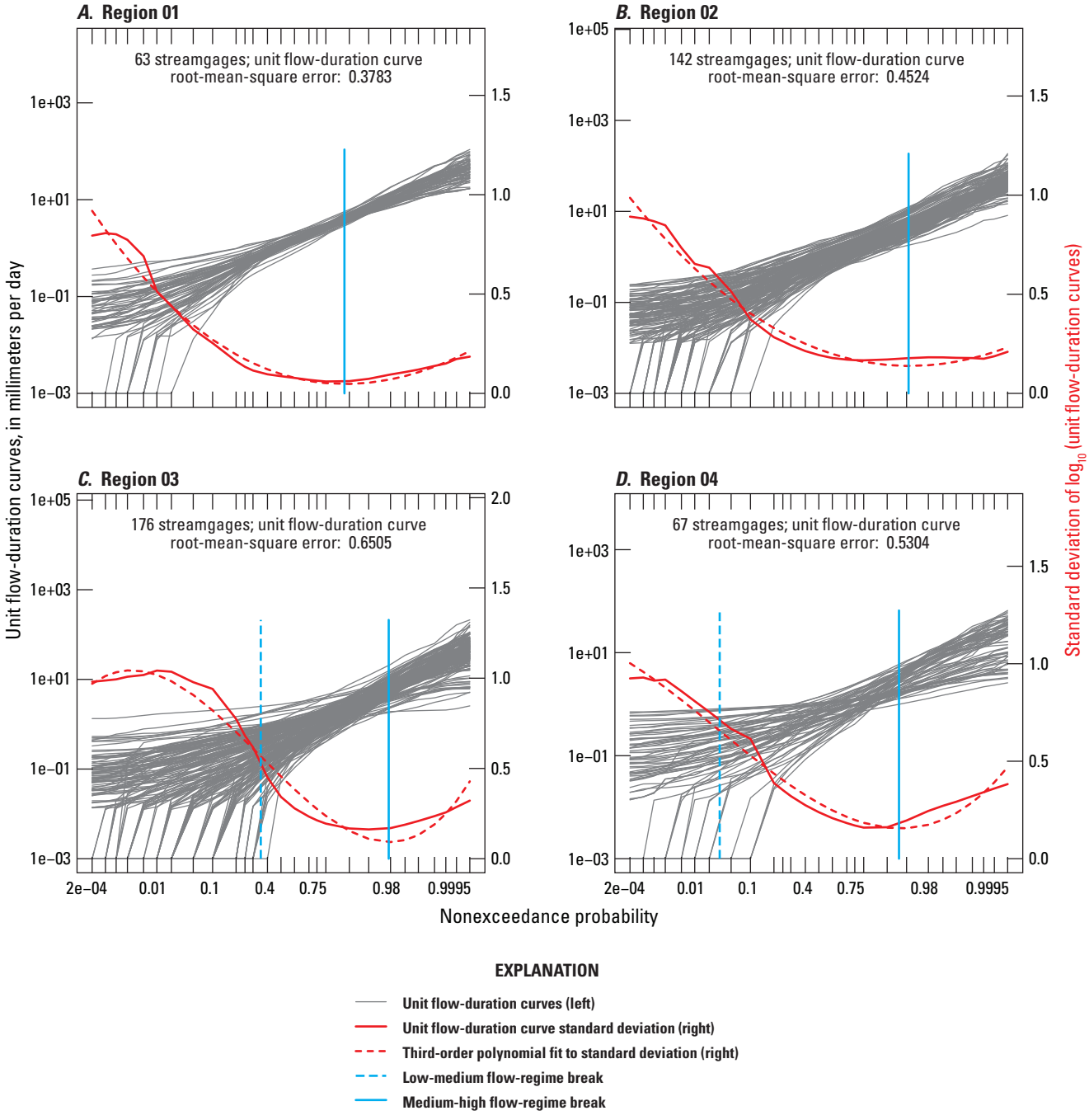


Figure 2. Unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from all qualifying streamgages in the regions used in this study.

Computation of Regression Models

Regression models were developed for each flow regime of each region by using a four-step process. In the first step, the best *nbest* models for each quantile and each value of *nv* from 1 to *nvmax*, where *nbest* is the number of regression models retained from this initial search for further analysis, *nv* is the number of basin characteristic variables, and *nvmax* is the largest number of basin characteristics considered, were determined by an exhaustive search with the function `regsubsets` from the **`leaps`** R package (Lumley, 2012). Because the `regsubsets` function operates within a least-squares framework, this search was performed with the $\log_{10}$-transformed "filled" observed FDC quantile values (that is, those having values estimated for the censored quantiles) as the dependent variable and weights proportional to the length of record. These *nbest* models were determined with and without forcing drainage area into the models.

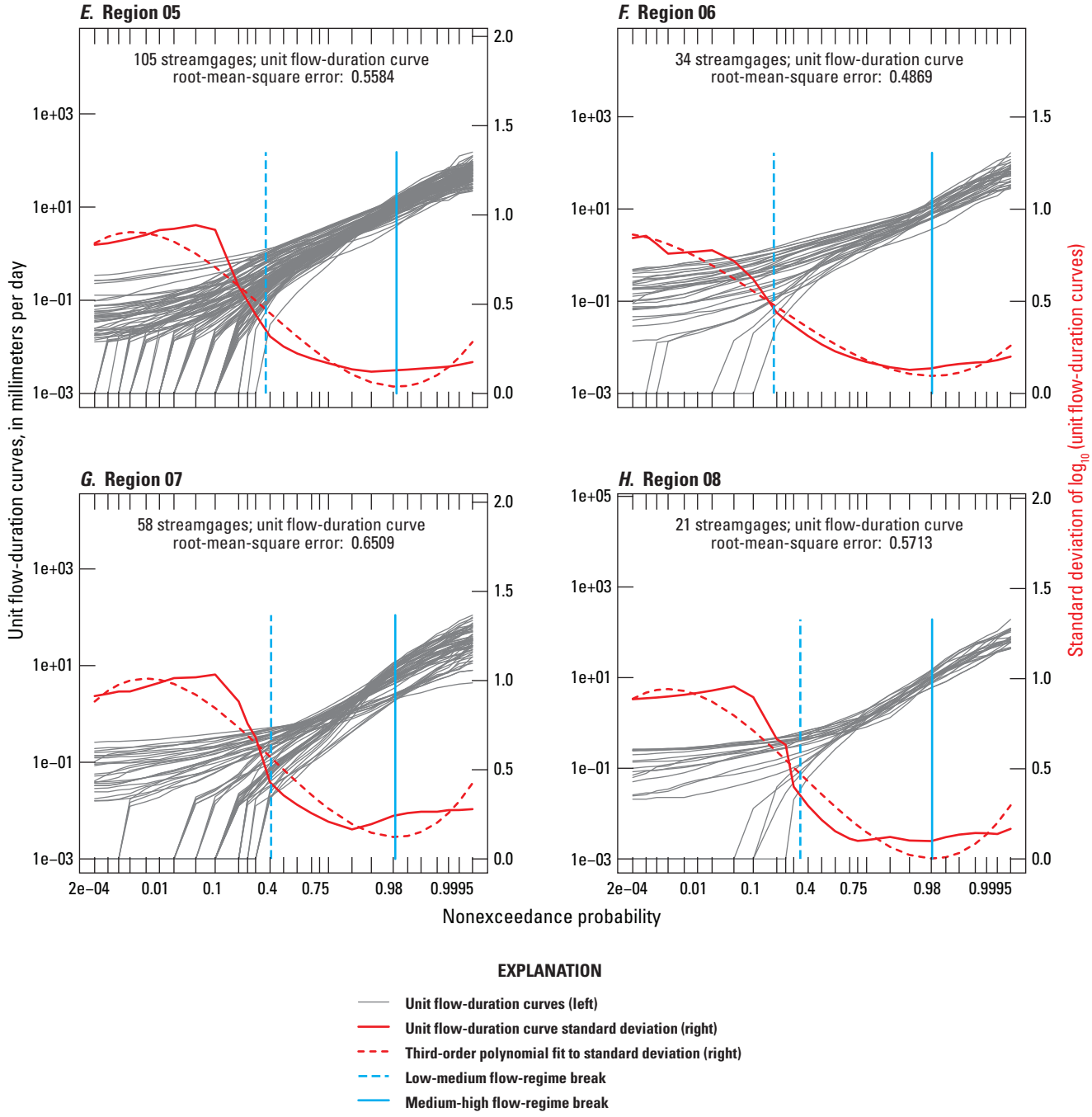


Figure 2. Unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from all qualifying streamgages in the regions used in this study.—Continued

The parameter *nvmax*, indicating the largest number of basin characteristics considered, was selected according to the following equation:

$$nvmax = \min\left(\max\left(\text{ceiling}\left(\frac{ngages}{nvmax.divisor}\right), 2\right), 6\right), \quad (2)$$

where

min returns the numerical minimum of its arguments,

max returns the numerical maximum of its arguments,

ngages is the number of streamgages in the region,

nvmax.divisor is a free parameter indicating the number of basin characteristics allowed per streamgage, and

ceiling( ) rounds the argument to the next largest integer.

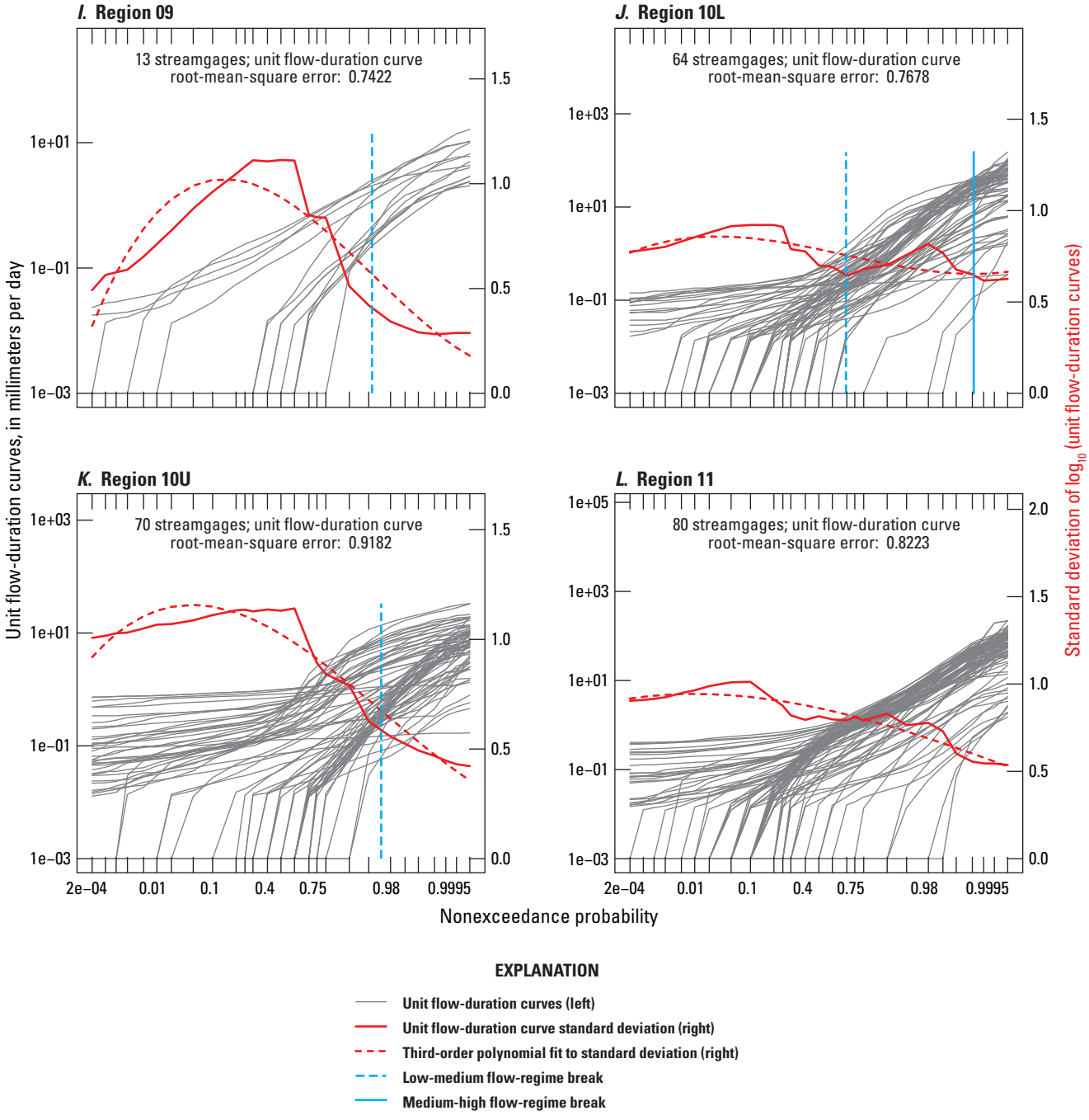


Figure 2. Unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from all qualifying streamgages in the regions used in this study.—Continued

According to equation 2, the minimum number of variables to be considered in regression models is 2 and the maximum is 6. The maximum is set at 6 because larger numbers of variables make regsubsets very slow. The minimum was set at 2 because using a single characteristic was determined to give poor results in preliminary testing.

In the second step, censored linear regression models for the best *nbest* models identified by regsubsets for each quantile were computed by using maximum likelihood estimation with left-censored Gaussian regression as implemented in the **`censReg`** function from the R package **`smwrQW`** (D.L. Lorenz, U.S. Geological Survey, written commun., 2016). Log_{10}-transformed censored (unfilled) FDC quantiles were the dependent variables in regression equations. Models having maximum variance-inflation-factor values exceeding 10 were eliminated.

In the third step, the best models for each flow regime and number of basin characteristics *nv* were determined.

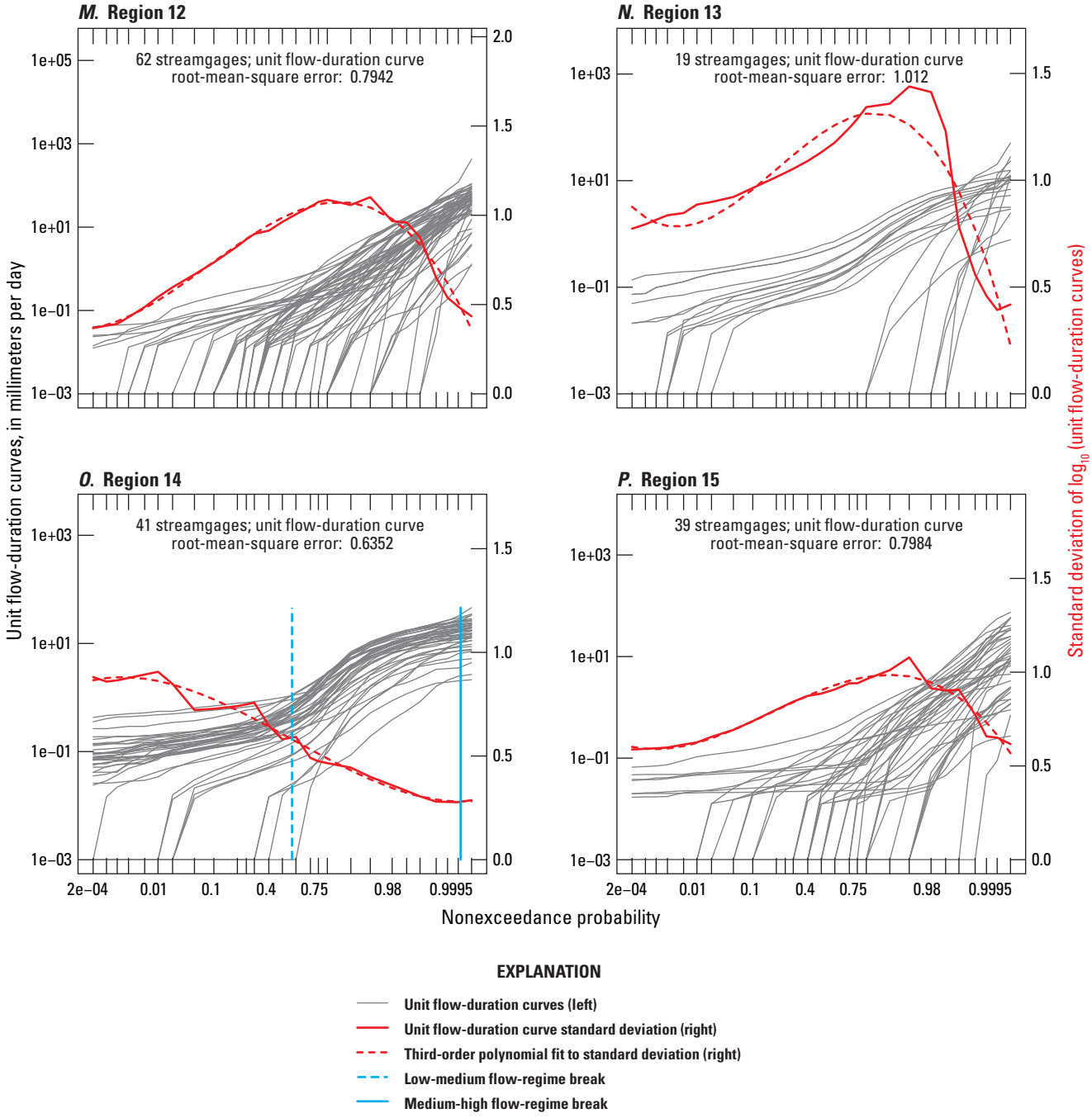


Figure 2. Unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from all qualifying streamgages in the regions used in this study.—Continued

First, values of a goodness-of-fit (GOF) measure termed the *pseudo-R^2* that is analogous to the usual coefficient of determination (R^2) but is defined in a maximum likelihood regression context (UCLA—Statistical Consulting Group, 2011) were computed from the censored linear regression fits in step two with the `summary.censReg` function from the R package **`smwrQW`** (D.L. Lorenz, U.S. Geological Survey, written commun., 2016), which uses the method presented in McKelvey and Zavoina (1975). Then models for each quantile were ranked according to their *pseudo-R^2* values, and the ranks of each model were summed across the quantiles in the flow regime to obtain a ranked set of the best models across the flow regime. Of these ranked models, the best *ntop* of them were selected for further analysis.

In the fourth and final step, the best *ntop* models for each flow regime and number of basin characteristics *nv* were compared by computing averages across the quantiles in the flow regime of two measures of goodness-of-fit that account for the effects of differences in the number of degrees of freedom: the adjusted *pseudo-R^2* and the AIC. Adjusted *pseudo-R^2* values were computed the same way as a usual adjusted R^2 is computed from the usual R^2 (Faraway, 2005, p. 127):

$$\overline{R}^2_{adj} = 1 - \left(1 - \overline{R}^2\right)\frac{n-1}{n-p}, \tag{3}$$

where

$\overline{R}^2_{adj}$ is the adjusted *pseudo-R^2*,

$\overline{R}^2$ is the *pseudo-R^2*,

n is the number of streamgages in the regression model, and

p is the number of coefficients in the regression model, including the intercept.

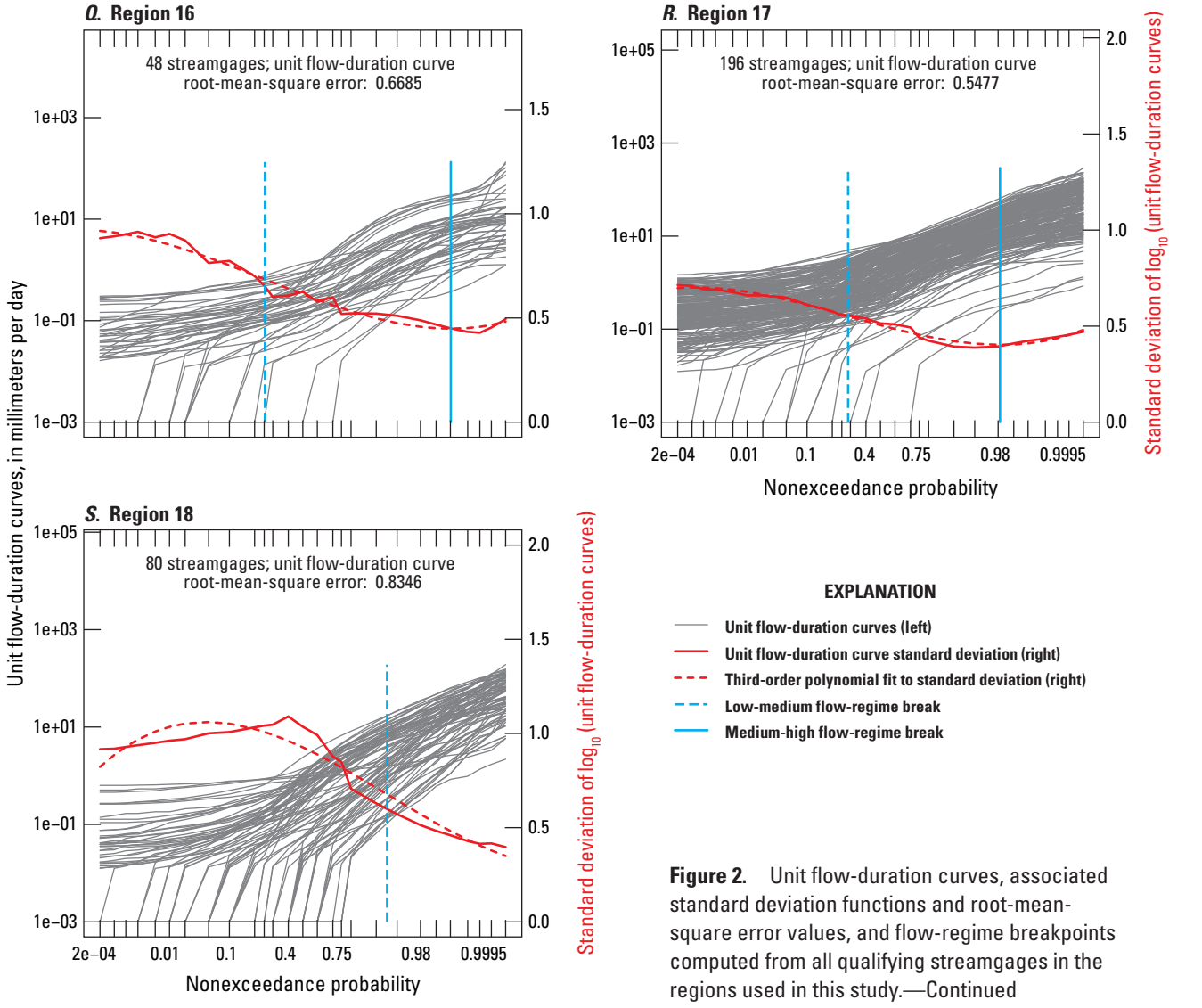


Figure 2. Unit flow-duration curves, associated standard deviation functions and root-mean-square error values, and flow-regime breakpoints computed from all qualifying streamgages in the regions used in this study.—Continued

The AIC is a likelihood-based GOF measure with adjustment for the number of degrees of freedom and is computed as $-2\log(L) + 2p$, where L is the likelihood of the regression model and p is the number of coefficients (Faraway, 2005, p. 126). The models with the highest average adjusted *pseudo-*R^2 and the lowest average AIC values across the flow regime, for any number of basin characteristics *nv*, were deemed to be the best models for the flow regime.

Different versions of certain steps in the regression model development process were tested to determine the sensitivity of the results to these aspects of the method and to find an approximately optimal method across the regions: (1) regimes were selected using (a) the default ranges of quantiles or (b) those obtained from the variance of unit FDC-based flow-regime determination method; (2) the maximum number of basin characteristics was chosen using equation 2 with *nvmax.divisor* set to 10 or 20 (that is, roughly, at most one basin characteristic per 10 or 20 streamgages); (3) drainage area was or was not forced into the regression models; (4) the numbers of regression models retained for further analysis at the quantile and regime levels (*nbest* and *ntop*) both were taken to be (a) 3 or (b) 100; and (5) the determination of the best model across the regime was made using average adjusted *pseudo-*R^2 or average AIC.

Because these methodological options are binary and there are five of them, there are as many as $2^5 = 32$ different methodological combinations (table 2). However, because the number of basin characteristics per region has a maximum of 6 and a minimum of 2, for regions with 101 or more streamgages (regions 02, 03, 05, and 17), the maximum number of basin characteristics *nvmax* was 6 regardless of whether *nvmax.divisor* takes the value 10 or 20; similarly, for regions with 20 or fewer streamgages (regions 09 and 13), the maximum number of basin characteristics was 2 regardless of whether the value of *nvmax.divisor* was 10 or 20. Therefore, for these six regions only, $2^4 = 16$ method combinations were tested.

Estimates of FDC quantiles for each region for these 16 or 32 combinations were computed in an LOOCV experiment. In other words, for each region and combination, the flow regimes and regression models were determined once for each streamgage in that region, where the FDC of the streamgage was predicted using the collection of streamgages in the region, and the streamgage being predicted was left out. The selected regression models and the associated predicted FDC quantiles resulting from the LOOCV experiment are archived in Russell and others (2018).

Testing and Evaluating Predicted Flow-Duration Curves

Several performance measures (PMs; table 3) were computed for use in testing and evaluating the predicted FDCs for each of the regions and methodological combinations. Because uncensored predicted quantile values less than the censoring level (0.01 ft^3/s) of the observed daily streamflow are available, for PMs that compare the predicted FDCs to the observed FDCs, the filled observed FDCs (which also include estimated values less than the censoring level) were used. Three classes of PMs were considered: (1) error measures, (2) bias measures, and (3) smoothness measures. The error measures are divided into three subclasses: (a) moment ratios, such as the ratio of predicted to observed mean discharge; (b) all-quantile error measures, such as root-mean-square error (RMSE), the mean absolute error of predicted quantiles relative to the observed, and distributional GOF measures (the Kolmogorov-Smirnov and Kuiper's statistics; Press and others, 1992), all computed on the $\log_{10}$-transformed quantiles; and (c) quantile ratios, which are ratios of predicted to observed quantile values at particular probabilities. The distributional error measures were included in recognition that the FDC is a representation of a probability distribution.

Except for the distributional GOF measures, the all-quantile and quantile ratio error measures were computed in two versions, one using censored and the other noncensored predicted and observed, filled FDC quantiles (table 3). For the PMs, a "censored" quantile is one where quantile values less than the censoring level of the daily discharge of 0.01 ft^3/s were set to selected values: when the quantile value was less than 0.005 ft^3/s, the corresponding censored quantile was set to 0.001 ft^3/s, and when the filled quantile value was between 0.005 and 0.01 ft^3/s, the censored quantile value was set to 0.01 ft^3/s. The censored version of the selected PMs was included to provide PM values that do not depend on the details of the method used to fill the observed FDCs.

The predicted moments for the moment ratio PMs were obtained by computing implied daily streamflow values, one for each day of the observed streamflow record, using the plotting position formula (eq. 1) and the predicted FDCs. The moments of the observed moments were computed directly from the observed daily streamflow during the complete WYs of data considered in this study. Before computing the moments, the same censoring adjustment was applied to the daily streamflow values as was used for computing the censored quantiles in the quantile-based error measures.

Table 2. Methodological options for flow-duration curve quantile regressions considered in this study.

[*nbest*, the number of regression models retained from initial search for further analysis; *ntop*, the number of regression models retained for determination of the best models across flow regimes; FDC, flow-duration curve; q_p, flow-duration curve quantile at nonexceedance probability p; *nvmax.divisor*, divisor of the number of streamgages in formula for computing *nvmax* (eq. 2); *nvmax,* maximum number of basin characteristics considered in regression models; R^2, coefficient of determination; &, and; |, or; AIC, Akaike Information Criterion; ≤, less than or equal; >, greater than; <, less than]

Decision	Option A		Option B	
	Definition	Abbreviation	Definition	Abbreviation
Number of regression models retained at (a) quantile and (b) flow-regime levels	(a) *nbest*=3; (b) *ntop*=3.	nb=3	(a) *nbest*=100; (b) *ntop*=100.	nb=100
Is drainage area forced into regression equations?	No	forcDA=N	Yes	forcDA=Y
Flow-regime definition	Default: low flow: All FDC quantiles q_p such that $p \leq 0.1$; medium: all *qp* such that $0.2 < p \leq 0.9$; high: all q_p such that $p > 0.9$.	defReg	Custom: Up to three, computed from the shape of variance among regional unit FDCs.	custReg
Choice of value of *nvmax.divisor* in formula for maximum number of basin characteristics *nvmax* (eq. 2)	*nvmax.divisor*=10 implies at most one basin characteristic per 10 streamgages, with *nvmax* having a minimum of 2 and a maximum of 6	nvmax=orig[a]	*nvmax.divisor*=20 implies at most one basin characteristic per 20 streamgages, with *nvmax* having a minimum of 2 and a maximum of 6	nvmax=reduc[a]
Goodness-of-fit (GOF) criterion for selecting the best regression model	adjusted R^2 (adjR2)	GOF=adjR2	Akaike Information Criterion (AIC)	GOF=AIC
Examples of methodological combinations				
Notation	Meaning			
nb=100	Union of all methods[b] where nb=100 (opposite and complement of "nb=3")			
nb=3	Union of all methods where nb=3 (opposite and complement of "nb=100")			
nb=100&forcDA=Y	Union of all methods where nb=100 and forcDA=Y (opposite of nb=3&forcDA=N")			
nb=3&forcDA=N	Union of all methods where nb=3 and forcDA=N (opposite of nb=100&forcDA=Y")			
nb3|forcDA=N	Union of all methods where nb=100 or forcDA=N (complement of "nb=100&forcDA=Y")			
nb=100&forcDA=Y&custReg&nvmax=reduc	Union of all methods where nb=100, forcDA=Y, custom flow regimes are used, and nvmax=reduc.			
nb=100&forcDA=Y&custReg&nvmax=reduc&GOF=adjR2	The method where nb=100, forcDA=Y, custom flow regimes are used, nvmax=reduc, and GOF=adjR2.			

[a]Because of the constraints on the value of *nvmax* (that is, $1 < nvmax < 7$) for regions with 20 or fewer streamgages, two basin characteristics are used regardless of the value of *nvmax.divisor*, and for regions with more than 100 streamgages, at most 6 basin characteristics are used regardless of the value of *nvmax.divisor*. Therefore, the "nvmax=orig" and "nvmax=reduc" cases coincide and the notation "nvmax=orig&reduc" is used.

[b]In this report, in the context of methodological options and combinations, the term "method" is used to designate methodological combinations for which all five options are specified.

Table 3. Definitions of performance measures used to evaluate predicted flow-duration curves in this study.

[Q, daily streamflow; ft^3/s, cubic foot per second; q_p, flow-duration curve quantile at nonexceedance probability p; FDC, flow-duration curve; <, less than; ≤, less than or equal to]

Name	Definition	Dimensions	Treatment of censoring (applied to both observed and predicted quantile and streamflow values except as noted)
	Moment ratio error measures		
mean_rat	Ratio of predicted to observed mean discharge	streamgage, region, method	Q<0.005 ft^3/s set to 0.001 ft^3/s and 0.005≤Q<0.01 set to 0.01 ft^3/s
var_rat	Ratio of predicted to observed discharge variance	streamgage, region, method	Q<0.005 ft^3/s set to 0.001 ft^3/s and 0.005≤Q<0.01 set to 0.01 ft^3/s
log_mean_diff	Difference of predicted to observed mean of $\log_{10}$(discharge)	streamgage, region, method	Q<0.005 ft^3/s set to 0.001 ft^3/s and 0.005≤Q<0.01 set to 0.01 ft^3/s
log_var_diff	Difference of predicted to observed variance of $\log_{10}$(discharge)	streamgage, region, method	Q<0.005 ft^3/s set to 0.001 ft^3/s and 0.005≤Q<0.01 set to 0.01 ft^3/s
	All-quantile error measures		
sqerr.comb	Squared errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	region, method	Used filled, observed quantiles
sqerr.comb.cens	Squared errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
abserr.comb	Absolute values of errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	region, method	Used filled, observed quantiles
abserr.comb.cens	Absolute values of errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
RMSE.statn	Root-mean-square errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	streamgage, region, method	Used filled, observed quantiles
RMSE.statn.cens	Root-mean-square errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
MAE.statn	Mean of absolute values of errors of $\log_{10}$-transformed quantiles, across stations (streamgages) and quantiles	streamgage, region, method	Used filled, observed quantiles
MAE.statn.cens	Mean of absolute values of errors of log10-transformed quantiles, across stations (streamgages) and quantiles	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
RMSE.quant	Root-mean-square errors of log10-transformed quantiles, by quantile across stations (streamgages)	quantile, region, method	Used filled, observed quantiles
RMSE.quant.cens	Root-mean-square errors of log10-transformed quantiles, by quantile across stations (streamgages)	quantile, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
MAE.quant	Mean of absolute values of errors of $\log_{10}$-transformed quantiles, by quantile across stations (streamgages)	quantile, region, method	Used filled, observed quantiles

Table 3. Definitions of performance measures used to evaluate predicted flow-duration curves in this study.—Continued

[Q, daily streamflow; ft^3/s, cubic foot per second; q_p, flow-duration curve quantile at nonexceedance probability p; FDC, flow-duration curve; <, less than; ≤, less than or equal to]

Name	Definition	Dimensions	Treatment of censoring (applied to both observed and predicted quantile and streamflow values except as noted)
	Moment ratio error measures—Continued		
MAE.quant.cens	Mean of absolute values of errors of $\log_{10}$-transformed quantiles, by quantile across stations (streamgages)	quantile, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
KS.stats	Kolmogorov-Smirnov statistics[a]: maximum absolute difference between predicted and observed streamflow probability distributions	streamgage, region, method	Used filled, observed quantiles
Kup.stats	Kuiper's statistics[a]: Sum of maximum difference of predicted and observed probability distributions and maximum difference of observed and predicted probability distributions	streamgage, region, method	Used filled, observed quantiles
	Quantile-specific error measures (quantile ratios)		
qrat001	Ratio of predicted to observed quantiles q_p with p=0.001	streamgage, region, method	Used filled, observed quantiles
qrat01	Ratio of predicted to observed quantiles q_p with p=0.01	streamgage, region, method	Used filled, observed quantiles
qrat10	Ratio of predicted to observed quantiles q_p with p=0.1	streamgage, region, method	Used filled, observed quantiles
qrat50	Ratio of predicted to observed quantiles q_p with p=0.5	streamgage, region, method	Used filled, observed quantiles
qrat90	Ratio of predicted to observed quantiles q_p with p=0.9	streamgage, region, method	Used filled, observed quantiles
qrat99	Ratio of predicted to observed quantiles q_p with p=0.99	streamgage, region, method	Used filled, observed quantiles
qrat999	Ratio of predicted to observed quantiles q_p with p=0.999	streamgage, region, method	Used filled, observed quantiles
qrat001.cens	Ratio of predicted to observed quantiles q_p with p=0.001	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat01.cens	Ratio of predicted to observed quantiles q_p with p=0.01	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat10.cens	Ratio of predicted to observed quantiles q_p with p=0.1	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat50.cens	Ratio of predicted to observed quantiles q_p with p=0.5	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat90.cens	Ratio of predicted to observed quantiles q_p with p=0.9	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat99.cens	Ratio of predicted to observed quantiles q_p with p=0.99	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
qrat999.cens	Ratio of predicted to observed quantiles q_p with p=0.999	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
	Bias measures		
bias.comb	Differences (predicted - observed) of $\log_{10}$-transformed quantiles	region, method	Used filled, observed quantiles

Table 3. Definitions of performance measures used to evaluate predicted flow-duration curves in this study.—Continued

[Q, daily streamflow; ft^3/s, cubic foot per second; q_p, flow-duration curve quantile at nonexceedance probability p; FDC, flow-duration curve; <, less than; ≤, less than or equal to]

Name	Definition	Dimensions	Treatment of censoring (applied to both observed and predicted quantile and streamflow values except as noted)
	Bias measures—Continued		
bias.comb.cens	Differences (predicted - observed) of $\log_{10}$-transformed quantiles	region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
bias.statn	Differences (predicted - observed) of $\log_{10}$-transformed quantiles by station (streamgage), across quantiles	streamgage, region, method	Used filled, observed quantiles
bias.statn.cens	Differences (predicted - observed) of $\log_{10}$-transformed quantiles by station (streamgage), across quantiles	streamgage, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
bias.quant	Differences (predicted - observed) of $\log_{10}$-transformed quantiles by quantile, across streamgages	quantile, region, method	Used filled, observed quantiles
bias.quant.cens	Differences (predicted - observed) of $\log_{10}$-transformed quantiles by quantile, across streamgages	quantile, region, method	q_p<0.005 ft^3/s set to 0.001 ft^3/s, 0.005≤q_p<0.01 set to 0.01
	Smoothness measures		
FracIncQuantsAvgs	Fraction of increasing quantiles across streamgages and quantiles	streamgage, region, method	Used filled, observed quantiles
Fracs.IncQuantsByStatn	Fraction of increasing quantiles by station (streamgage), across quantiles	streamgage, region, method	Used filled, observed quantiles
Fracs.IncQuantsByQuant	Fraction of increasing quantiles by quantile, across streamgages	streamgage, region, method	Used filled, observed quantiles
PredSSRavgs	Sum, across streamgages and quantiles, of sum of squares of residuals of regression fit of predicted FDC to quadratic polynomial of z-scores of exceedance probabilities	region, method	Used filled, observed quantiles
PredSSRs.byStatn	Sum, by station (streamgage), across quantiles, of sum of squares of residuals of regression fit of predicted FDC to quadratic polynomial of z-scores of exceedance probabilities	streamgage, region, method	Used filled, observed quantiles
PredSSRs.byQuant	Sum, by quantile, across streamgages, of sum of squares of residuals of regression fit of predicted FDC to quadratic polynomial of z-scores of exceedance probabilities	quantile, region, method	Used filled, observed quantiles
PredSSRs.byStatn.ratio	Ratio of sum, by station (streamgage), across quantiles, of sum of squares of residuals of regression fit of predicted FDC to quadratic polynomial of z-scores of exceedance probabilities to same quantity computed for the observed FDCs.	streamgage, region, method	Used filled, observed quantiles
PredSSRs.byQuant.ratio	Ratio of sum, by quantile, across streamgages, of sum of squares of residuals of regression fit of predicted FDC to quadratic polynomial of z-scores of exceedance probabilities to same quantity computed for the observed FDCs.	quantile, region, method	Used filled, observed quantiles

[a]Press and others (1992).

Bias measures were also computed using the $\log_{10}$-transformed quantiles. These measures, like most of the error measures, include both censored and uncensored versions.

The smoothness measures assess the overall smoothness of the predicted FDCs by focusing on two properties: (1) the fraction of predicted quantiles that increase for increasing nonexceedance probability (as required for consistency with the meaning of FDCs as representations of probability distributions) and (2) a more general measure of smoothness: the sum of squared residuals from a quadratic fit to the predicted $\log_{10}$-transformed FDC quantiles a function of the associated unit Gaussian quantile, that is,

$$SSR = \sum_{i,p} \varepsilon_{i,p}^2, \tag{4}$$

where

SSR is the sum of squared residuals and

$\varepsilon_{i,p}$, where i is the streamgage index and p is the nonexceedance probability, is a residual from the quadratic regression:

$$q_{i,p} = a_{i,0} + a_{i,1} z_p + a_{i,2} z_p^{\ 2} + \varepsilon_{i,p}, \tag{5}$$

where

$q_{i,p}$ is the predicted $\log_{10}$-transformed quantile at nonexceedance probability p of the ith streamgage,

z_p is the unit Gaussian quantile at nonexceedance probability p, that is, $p = (Z \le z_p)$, where Z is a unit Gaussian random variable, and

$a_{i,0}$, $a_{i,1}$, and $a_{i,2}$ are the fitted coefficients.

In the LOOCV approach there is one FDC prediction per method and streamgage (the streamgage left out of the fitting and being predicted by the others). Furthermore, the streamgages are grouped into regions. Therefore, there are four dimensions to the PM dataset: PM, method, region, and streamgage. To extract meaningful information from this dataset, the values were summarized to allow comparisons among the methods across the dataset. The following summarization steps were applied:

1. *Summarizing the PM values within the regions*:

   A. *Computed PM summary statistic values*: For each PM, summary statistics (mean, median, and interquartile range [IQR]) were computed across the streamgages for each region and method and are given in appendix 2, table 2.1. For example, the mean, median, and IQRs across streamgages and quantiles for each region and method of the PM designated as "abserr.comb" (the absolute values of the differences between the observed and predicted $\log_{10}$-transformed quantiles) are shown in figure 3.

   B. *Ranked PM summary statistic values*: To facilitate subsequent comparisons among regions and PMs, the methods were ranked by summary statistic for each PM and region on a scale of 0 to 10 with ranks defined using a plotting position approach, where better values are associated with smaller ranks. In particular, the rank $r(i)$ of ith best method is given by $r(i) = (i - 0.5)/(n/10)$, where n is the number of methods. For example, for regions with 16 methods, the method with the best summary statistic value was assigned the rank $r(1) = (1 - 0.5)/(16/10) = 0.3125$, whereas the method with the worst was assigned the rank $r(16) = (16 - 0.5)/(16/10) = 9.6875$. The interpretation of a summary statistic depends on the statistic: for example, for the "mean_rat" PM (the ratio of predicted to observed mean discharge) and other moment and quantile ratio PMs (table 3), mean and median summary statistic values are better the closer they are to 1, whereas the mean and median values of RMSE and other all-quantile error PMs are better the smaller they are. For IQR summary statistics, smaller is always better. A graphical example is given in figure 4 where the "abserr.comb" results given in figure 3 converted to ranks are presented. The ranked PM summary statistic values by method and region for all PMs are given in appendix 2, table 2.2.

2. *Summarizing the PMs across the regions*: The same three summary statistics (mean, median, IQR) were computed on the ranks of each PM summary statistic across selected groups of regions for selected methodological combinations. By comparing the summary statistics of the ranks, the effect on the FDC-prediction accuracy, as measured by the various PMs of the different choices on the methodological options that constitute the different methodological combinations, can be seen. Using this approach, each region is given equal weight, which is different than giving each streamgage equal weight, because the regions have different numbers of streamgages. The median summary statistics for three sets of regions are given in appendix 3, and an example plot showing the median rank value across all the regions for the summary statistics of all PMs for a given methodological combination is given in figure 5.

3. *Summarizing across the PMs*: Each statistic summarizing the PM summary statistic ranks across the regions for selected methodological combinations as obtained in step 2 was further summarized by computing the same set of summary statistics (mean, median, and IQR) across the PMs, for all PMs, and for the different classes of PMs. The results of this step are a set of vectors giving the summary statistics across the PMs, one for each selected methodological combination, which can be examined directly to draw conclusions about the

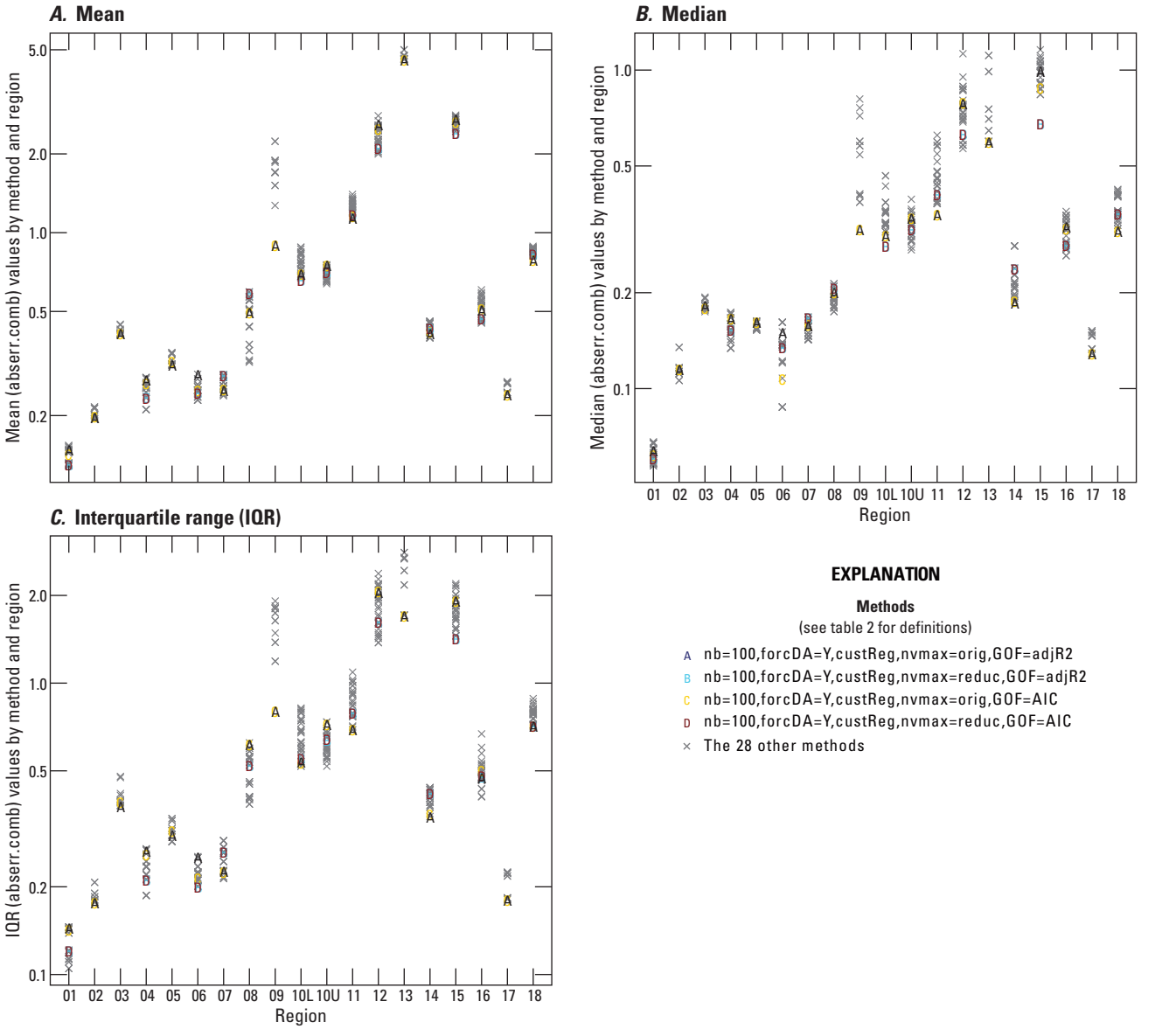


Figure 3. Absolute values of prediction errors of $\log_{10}$-transformed quantiles across streamgages and quantiles (performance measure "abserr.comb") by method and region, summarized by: *A*, mean; *B*, median; and *C*, interquartile range (IQR).

effects of the changes in parameters that constitute the different methodological options. The vector of median values summarizing the median ranks across all PMs and regions for selected methodological combinations is given in column B of each table in appendix 3. An example plot of these values is given in figure 6, which shows the medians across all summary statistics of all performance measures of median rank values across all regions. The methodological combination selected for the plot in figure 5 is the sixth from the right in figure 6.

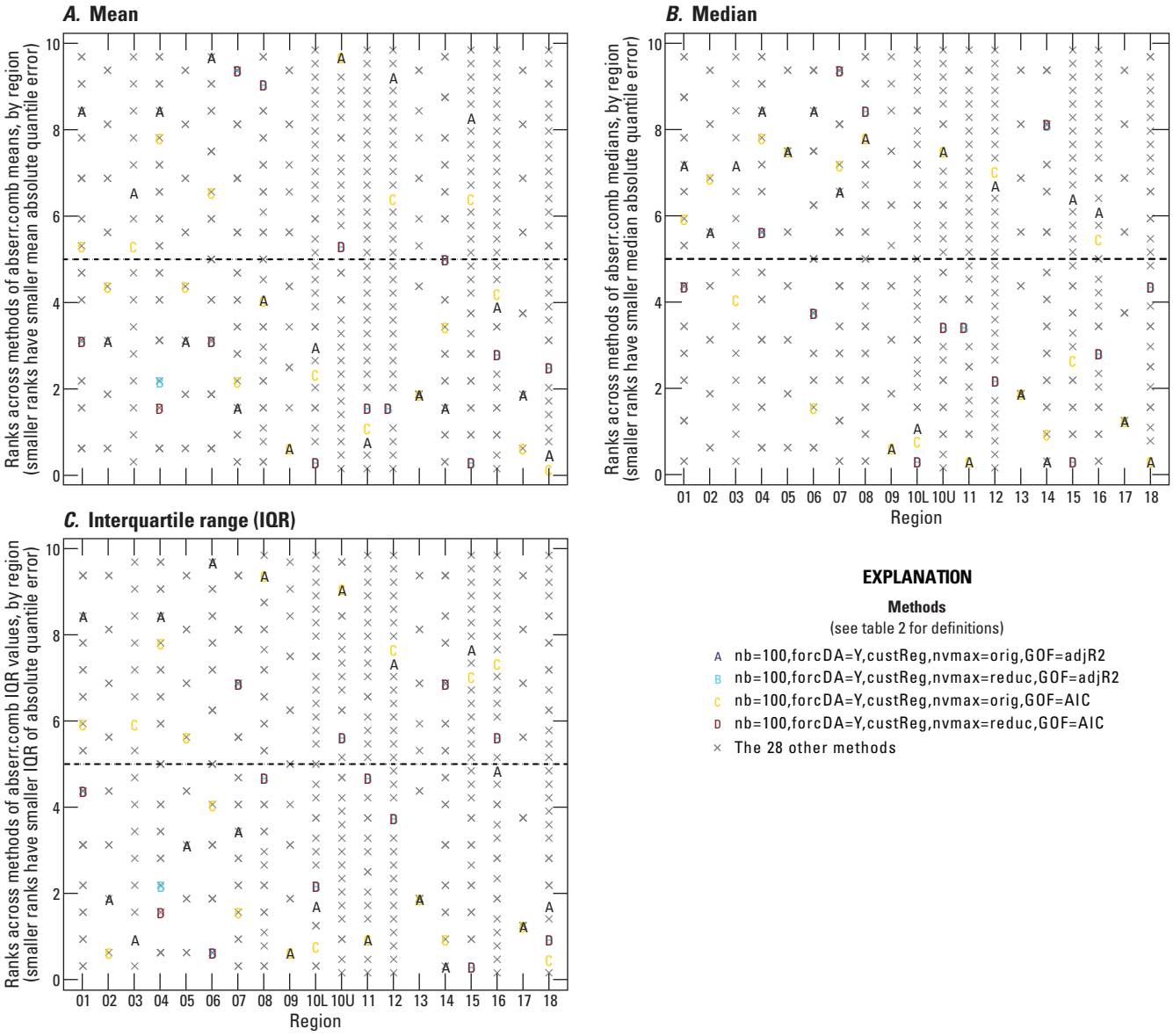


Figure 4. Ranks of absolute values of prediction errors of $\log_{10}$-transformed quantiles across streamgages and quantiles (performance measure "abserr.comb") by method and region, summarized by: *A*, mean; *B*, median; and *C*, interquartile range (IQR). Ranks are on a scale of 0 to 10 with 0 being the best.

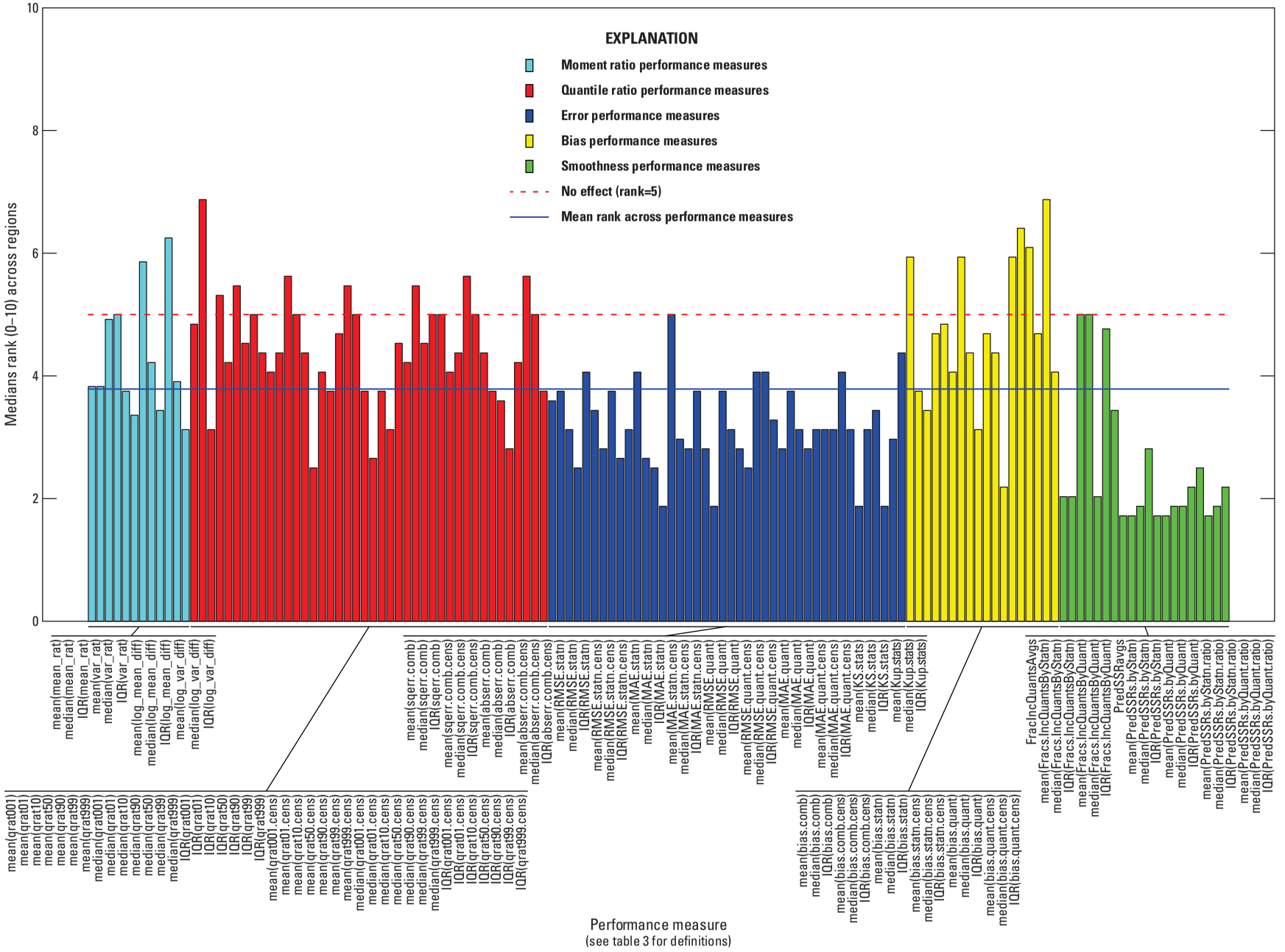


Figure 5. Median rank values across all regions of ranks of summary statistics of all performance measures for the "nb=100&forcDA=Y&custReg&nvmax=reduc" methodological combination (see table 2 for definition). Ranks are on a scale of 0 to 10 with 0 being the best.

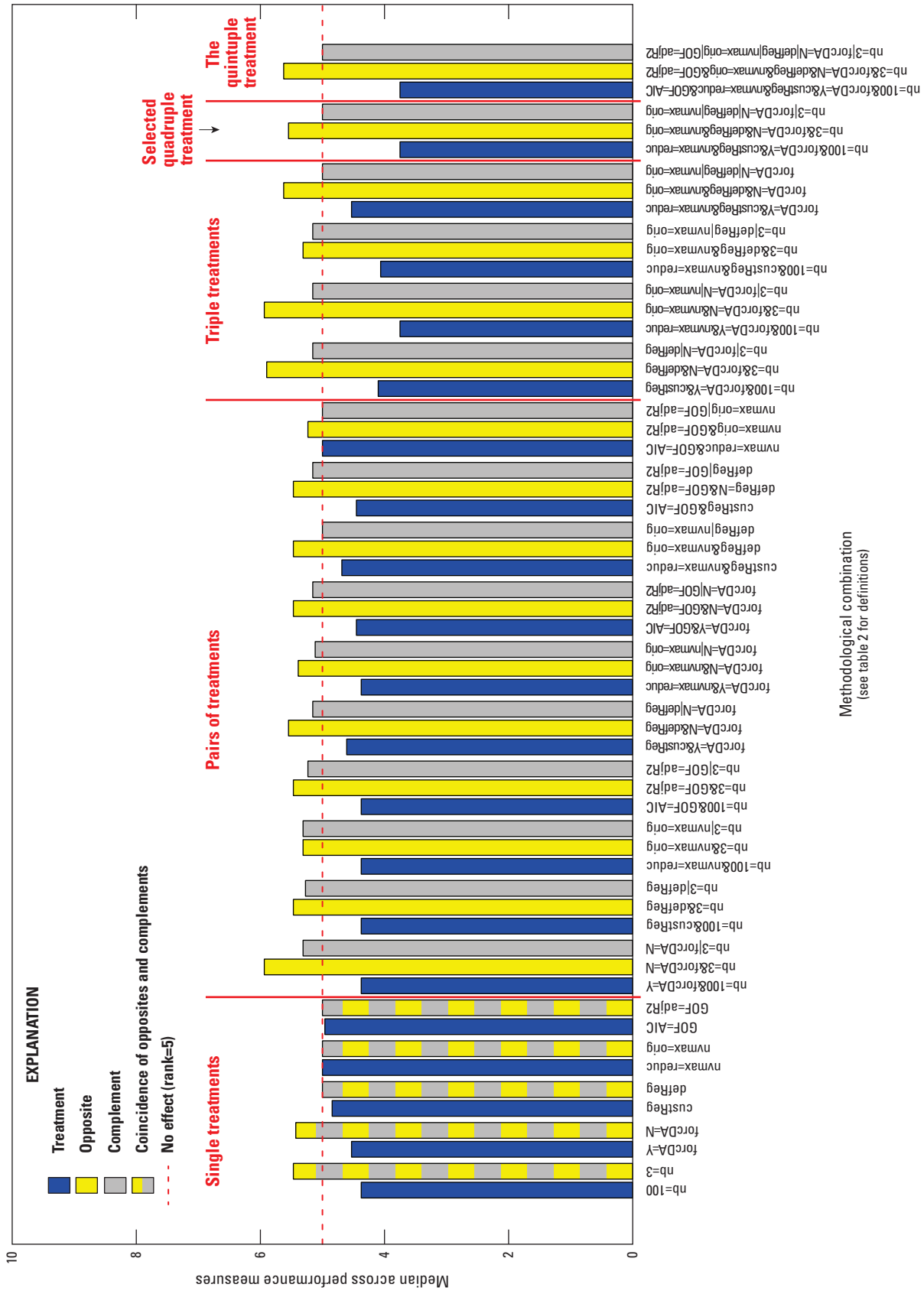


Figure 6. Medians across all summary statistics of all performance measures of median rank values across all regions for selected methodological combinations. A "treatment" in this figure means the specification of one or more methodological options. Each methodological combination designated as being a "treatment" or "opposite" combines results for all five-part methodological combinations with all the indicated properties; those combinations designated as being "complements" include all results for which any one of the indicated properties is present.

Refinement of a Regression-Based Method for Prediction of Flow-Duration Curves

The primary focus of this report is on the effects of the different methodological choices on the accuracy of the FDC-prediction results; however, as discussed, the methodologies are most easily compared by way of ranks, which obscure direct measures of accuracy. Therefore, to provide context for the methodological results and as a general contribution to the understanding of the accuracy of regression-based FDC estimation methods in general, this section begins with a discussion of the overall accuracy of the results.

Accuracy across Regions and Quantiles

The summaries statistics of the absolute errors of the $\log_{10}$-transformed FDC quantile prediction errors for each region and methodological combination over all streamgages and quantiles (PM "abserr.comb") are shown in figure 3. The results when summarized by the median are shown in figure 3*B*. The accuracy of the FDC prediction varies widely by region, with the largest range (on this log plot) arising from different methods for region 09 being only about one-third the range among the regions. The FDCs for region 01 (covering much of New England; fig. 1) are predicted with the least error by far, about 0.06 to 0.07 $\log_{10}$ units (about 15–17 percent, computed as 10^{AE}–1, where AE is the absolute error of $\log_{10}$-transformed values). Region 01 is also the region with the smallest variance among the unit FDCs (table 1, fig. 2). If the results in figure 3*B* are plotted as compared to the unit FDC RMSE, a generally monotonic relation up to an RMSE of about 0.8 is obtained (fig. 7*A*). A roughly similar result is obtained if the median aridity index or the mean fraction of zero discharges for each region (table 1) is used (fig. 7*B*,*C*). These results indicate that the variance among FDCs is a strong determinant of the accuracy of FDC estimation by regression, and that variance between observed FDCs is most likely to arise in arid regions (see also Castellarin and others, 2013).

After region 01 there are several regions covering the East and Midwest (02–08) plus region 17 in the Northwest that are moderately well-predicted, having median $\log_{10}$ absolute quantile errors from 0.1 to 0.2 (about 26–58 percent) (fig. 3*B*) and FDC RMSEs from 0.45 to 0.65 (fig. 7*A*). These regions also make up the least arid regions. The remaining regions have $\log_{10}$ median absolute prediction errors ranging from about 0.25 (about 78 percent, region 14) to 1.0 (a factor of 10, region 15). The results shown in figures 3 and 7 include errors for quantile values that are censored in the original data. The results from the PM "abserr.comb.cens" (not shown), which is the same as "abserr.comb" except that quantiles with values less than 0.005 ft^3/s were set to 0.001 ft^3/s and those between 0.005 ft^3/s and 0.01 ft^3/s were set to 0.01 ft^3/s (table 3), are similar to the "abserr.comb" results for the better-predicted regions, but somewhat better in the other regions, particularly for regions 12 and 13, which have among the highest fractions of zero discharges (table 1).

One view of the overall sensitivity of the results to the methodological options examined is provided in figures 3 and 7. There tends to be a larger intraregional range of outcomes for the more poorly estimated regions, but occasionally a relatively well-estimated region also shows a larger range (region 08 for "abserr.comb" means, region 06 for "abserr.comb" medians). The "abserr.comb" mean values being roughly double the "abserr.comb" median values indicates a high degree of skewness to the results. This skewness indicates that a few streamgages are much more difficult to estimate than the others in the region; therefore, those streamgages should be examined further to determine if they are outliers for some measurable basin characteristics.

The value of understanding how prediction errors vary by quantile is demonstrated in figure 8. Lower flows typically are harder to predict than higher flows (Perry and others, 2004; Esralew and Smith, 2010; Archfield and others, 2010; Over and others, 2014). For example, in the worst-predicted regions (12 and 15), the median error of the low-flow quantiles is about 3 $\log_{10}$ units or a factor of 10^3=1,000, and for the best (region 17), this error is about 0.25 $\log_{10}$ units or a factor of $10^{0.25}$=1.78, whereas for the highest flows, the median absolute error ranges from about 0.1 to 0.5 in $\log_{10}$ units (from a factor of 1.26 to a factor of about 3.2). There is also a minimum in error near nonexceedance probabilities of 0.75 to 0.90 for the better-predicted regions (01–08 and 17), though the appearance of this minimum is somewhat subtle except for region 01. This error minimum indicates that for these more humid regions the highest flow quantiles are somewhat harder to predict than those that are somewhat less extreme (see also Over and others, 2014). This error minimum is similar to that of the standard deviation of the unit FDCs (fig. 2), which also have a minimum near this same range of nonexceedance probabilities. This similarity reinforces the connection between the intraregional variability of unit FDCs and FDC-prediction accuracy seen in figure 7*A*. For the worse-predicted, more arid regions, however, error continues to decrease from low to high flow until about a rather high nonexceedance probability of 0.998, where most flatten out or begin a slight rise. These results document the extent to which FDC-prediction accuracy differs from low to moderately high to high flow and demonstrate the need to focus efforts on better prediction of low-flow quantiles.

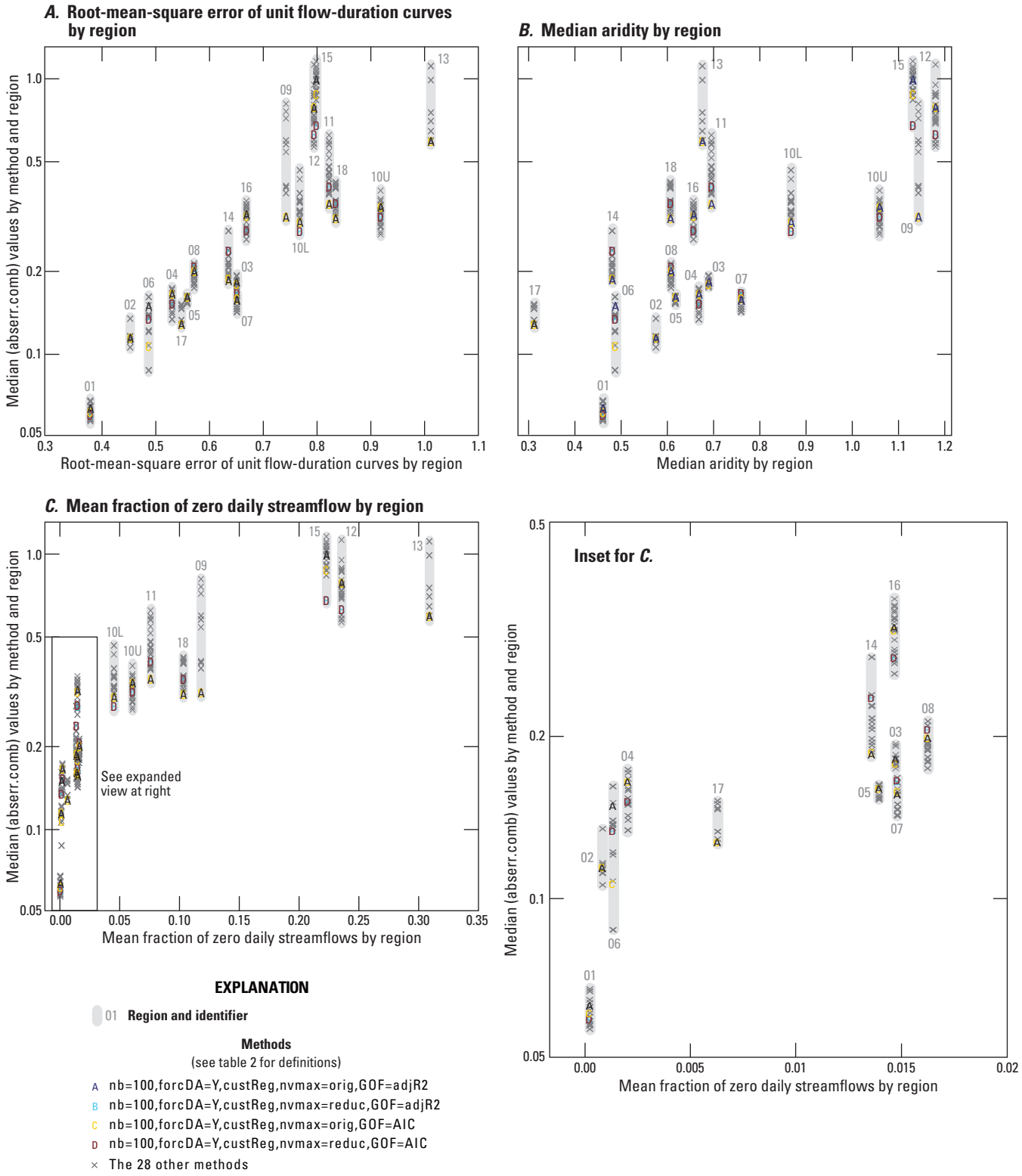


Figure 7. Absolute values of prediction errors of $\log_{10}$-transformed flow-duration curve quantiles across streamgages and quantiles (performance measure "abserr.comb") by region and method, summarized by median, as a function of regional values of *A*, the root-mean-square error of the unit flow-duration curves; *B*, the median aridity, defined as potential evaporation divided by precipitation; and *C*, the mean fraction of zero daily streamflows in the region.

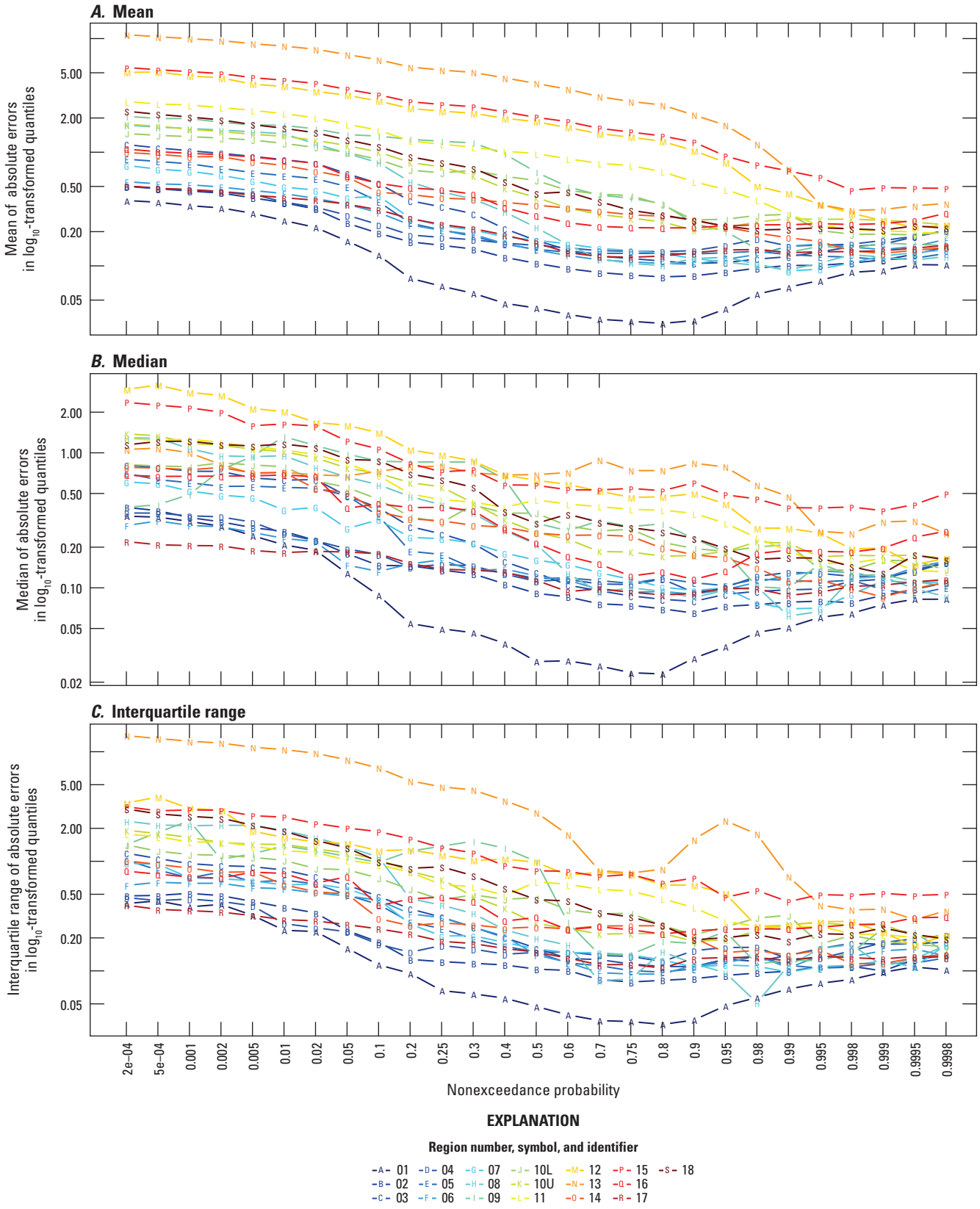


Figure 8. Absolute values of prediction errors of $\log_{10}$-transformed quantiles across streamgages by quantile and region for methodological combination "nb=100&forcDA=T&custReg&nvmax=reduc&GOF=adjR2" (table 2) summarized by *A*, mean, *B*, median, and *C*, interquartile range.

Comparisons of Methodological Combinations

A broad overview of the results on FDC-prediction performance across all regions, PMs, and methodological combinations can be seen in figure 6. This figure shows that, when combining by the median of ranks across all PMs and regions, consistent differences are obtained among different methodological combinations. Because the effects of the different treatments may interact with each other, results from multiple as well as single treatments are considered.

Working from left to right in figure 6, the first set of methodological combinations involve only single treatments and their opposites (opposites are also complements in the single treatment case). The single treatment having the largest effect on the medians of median ranks for this measure of the results is obtained from selecting "nb=100," which means that as many as 100 regression models were retained from the quantile and flow-regime model selection steps, as opposed to three such models being retained. An improvement in results from retaining more models for consideration is not surprising, but additional computation time is required (from about 10 to 30 times as much). The second-largest improvement from a single treatment came from forcing drainage area into the regression models. That the opposite treatments to "nb=100" and "forcDA=Y" (that is, "nb=3" and "forcDA=N," respectively) have substantially poorer medians of ranks confirms the improvements seen for "nb=100" and "forcDA=Y." The use of customized flow regimes ("custReg") provides much more modest improvement, and its opposite treatment, the use of the default regimes ("defReg"), shows a negligible effect, casting doubt on the significance of the effect of using customized regimes as a single treatment. The effects of the use of *nvmax. divisor*=20 in equation 2 (designated as "nvmax=reduc" in the x-axis labels of figures 6 and 9*A*,*B*), which reduces the maximum number of basin characteristics allowed in the FDC quantile regression models *nvmax,* and of the use of AIC rather than adjusted *pseudo-*R^2 for flow-regime regression model selection are both negligible.

The second set of results going from left to right in figure 6 involves pairs of treatments and their opposites and complements. With a couple exceptions, the treatment pair results are similar to each other and similar in magnitude to the effect of single treatments "nb=100" and "forcDA=Y." The least effective treatment pairs are (1) the combination of the least effective of the single treatments (that is, "nvmax=reduc" and "GOF=AIC"), for which the overall effect is negligible (see the right-hand-most treatment pair, "nvmax=reduc&GOF=AIC"); and (2) the combination of "custReg" and "nvmax=reduc." That most of the two-treatment combinations are not substantially better than the better one-treatment combinations, even when the best two are combined ("nb=100" and "forcDA=Y"), indicates the presence of nonlinear interactions between them.

The next set of results to the right in figure 6 is the subset of the combinations of three treatments ("triple treatments") selected from all single treatments except "GOF=AIC." All four of these treatments have a positive effect, and the best three have larger effects than any of the treatment pairs. The treatment "nb=100&forcDA=Y&nvmax=reduc" has the largest effect; "nb=100&forcDA=Y&custReg" and "nb=100&custReg&nvmax=reduc" are next and similar in magnitude. The presence of the "nb=100" treatment in all three of these better combinations is not surprising because it was the best single treatment; the presence of "forcDA=Y" in the best combination is similarly unsurprising because it was the second best single treatment. That the third element of the best three-treatment option is "nvmax=reduc" rather than "custReg," when both had very modest to negligible effects as single treatments, indicates that the interactions of the effects of "nvmax=reduc" with those of "nb=100" and "forcDA=Y" are more helpful than those of "custReg."

The last two sets of results to the right in figure 6 consist of a single combination of four treatments, with its opposites and complements, and the one possible combination of five treatments with its opposites and complements. The first of these, the combination of four treatments "nb=100&forcD A=Y&custReg&nvmax=reduc," was constructed by adding the "custReg" treatment to the best of the triple treatments, but the combination results in a negligible difference from it according to this figure. Similarly, the final combination of treatments to the right, which adds the treatment "GOF=AIC" to the selected quadruple treatment, also has a negligible effect on performance, reinforcing the conclusion that the choice between "GOF=AIC" and "GOF=adjR2" does not make a substantial difference. According to figure 6, then, the triple treatment "nb=100&forcDA=Y&nvmax=reduc" gives overall results that are as good as any other treatment, and because it lacks the redundant single treatments "custReg" and "GOF=AIC" of the quadruple and quintuple treatments having negligibly different results, it is the preferred treatment.

To test the robustness relative to regions that are better or worse predicted than the results presented in figure 6, in particular the conclusion that the triple treatment "nb=100&forcDA=Y&nvmax=reduc" is to be preferred, plots like figure 6 were created by separating the regions into two groups, consisting of the 9 best and the 10 worst-predicted regions (fig. 9*A*,*B*). A general conclusion from comparing these three figures is that the effects of the treatments are larger for the group of worse-predicted regions and smaller for the group of better-predicted regions. One cause of this difference in the magnitude of effects is that in the better-predicted regions (fig. 9*A*), the "forcDA=Y" treatment has no effect because for these regions drainage area is usually selected as a basin characteristic even when not forced into the regression models. Similarly, most of the pairs of treatments have negligible effects for the better-predicted regions other than those involving "nb=100." The combination with the best overall performance for the better-predicted regions is the triple treatment "nb=100&custReg&nvmax=reduc"; the corresponding quadruple treatment with "forcDA=Y" added, "nb=100&forc DA=Y&custReg&nvmax=reduc," seems to be slightly worse,

but because of the negligible effect of including "forcDA=Y" in other combinations, this difference may not be significant.

For the worse-predicted regions (fig. 9*B*), both "nb100" and "forcDA=Y" have substantial effects as individual treatments, and, similar to the results over all regions (fig. 6), the triple treatment "nb=100&forcDA=Y&nvmax=reduc," the quadruple treatment "nb=100&forcDA=Y&custReg&nvmax=reduc," and the quintuple treatment have similar overall performance. For the worse-predicted regions, however, unlike the results for all regions, the quadruple treatment "nb=100&forcDA=Y&custReg&nvmax=reduc" performs somewhat better than the triple treatment "nb=100&forcDA=Y&nvmax=reduc" (or any other triple treatment).

After consideration of the results obtained by separating better- and worse-predicted regions, it is concluded that the quadruple treatment "nb=100&forcDA=Y&custReg&nvmax=reduc" is the best overall combination and recommended for application to the FDC quantile estimation problem. This conclusion is obtained for the following reasons: (1) the quintuple treatment does not improve the results compared to the recommended quadruple treatment for any set of regions, (2) the recommended quadruple treatment is no worse than any other treatment for all regions and for the worse-predicted regions, (3) the improvement of the triple treatment "nb=100&custReg&nvmax=reduc" compared to the recommended quadruple treatment for the better-predicted regions is believed to be negligible. An additional consideration in favor of recommending the "nb=100&forcDA=Y&custReg&nvmax=reduc" treatment is that the rank-based differences for the worse-predicted regions are likely to be larger than those in the better-predicted regions in terms of the original PM values (figs. 3 and 7). According to the numbered reasons given, however, a decision on that consideration is not needed to select "nb=100&forcDA=Y&custReg&nvmax=reduc" as the best overall combination.

To examine the performance of the recommended combination "nb=100&forcDA=Y&custReg&nvmax=reduc" with respect to different PMs, consider figure 5 in further detail. According to figure 5, the selected combination is advantageous for most but not all PMs, as measured by the medians of the ranks across regions of their summary statistics. Among the median PM summary statistics with the best (lowest) ranks are most of the moment ratio PM statistics, the all-quantile error PM statistics (all indicate improvement except one IQR statistic ,"IQR(RMSE.statn)," which has a rank of 5, thus indicating no effect), and the smoothness PM statistics. The relatively poor performance of the "mean(log_mean_diff)" and "mean(log_var_diff)" PM summary statistics when the corresponding medians and IQRs are good indicates that relatively bad results for one or two regions are affecting the mean.

The classes of PMs for which the recommended combination "nb=100&forcDA=Y&custReg&nvmax=reduc" has mediocre performance, according to figure 5, are the quantile ratios and bias PMs. In examinations of figures of the type shown in figures 3 and 4 for these PMs (not shown), the ranks of the summary statistics of these PMs are particularly volatile because for at least one-half of the regions, the summary statistic values are very close together, being close to one for the ratios and to zero for the bias values. As a result, the ranks obtained are less meaningful, and these PMs can be largely discounted.

Summary statistics for the performance measure "abserr.comb" for the recommended methodological combination "nb=100&forcDA=Y&custReg&nvmax=reduc" can be seen in figures 3, 4, and 7, where the "GOF=adjR2" and "GOF=AIC" versions of this combination are indicated for each region by the symbols "B" and "D," respectively (in regions 02, 03, 05, 09, 13, and 17, the cases "nvmax=orig" and "nvmax=reduc" are identical, and therefore, for these regions, the results for the "GOF=adjR2" and "GOF=AIC" versions of methodological combination "nb=100&forcDA=Y&custReg&nvmax=orig," which are labeled as "*A*" and "*C*" in these figures, are also identical and overlie the symbols "B" and "D"). These results show that there is variability among the regions as far as the preferability of the recommended methodological combination. Further investigation of this regional variability might give insight into better ways to select the methods, but "nb=100&forcDA=Y&custReg&nvmax=reduc" gives the best overall results. These results also confirm that the "GOF=adjR2" and "GOF=AIC" versions of this methodological combination are usually quite similar.

Discussion of the Recommended Methodological Combination

The recommended methodological combination ("nb=100&forcDA=Y&custReg&nvmax=reduc") consists of 100 retained regression models retained for further analysis as compared to three ("nb=100"), drainage area forced into the regression equations ("forcDA=Y"), the use of the variance of unit FDC-based custom flow-regime methodology with one to three flow-regime definitions as compared to three fixed flow regimes ("custReg"), and at most one basin characteristic used per 20 streamgages as compared to one characteristic per 10 streamgages ("nvmax=reduc"). Two of the four methodological options constituting the recommended combination, "forcDA=Y" and "nvmax=reduc," explicitly reduce the number of degrees of freedom of the independent variable dataset. In addition, the custom flow-regime methodology ("custReg") also tends to reduce the degrees of freedom relative to the default because many regions were not given three regimes. The presence of these methodological options in the recommended combination suggests that one reason for the better performance observed with this combination is that it addresses the problem of over-fitting, which can arise when too many independent variables are available for model fitting.

A recommended practice in regional regression modeling is to select models that are physically meaningful. Two of the methodological options in the recommended combination

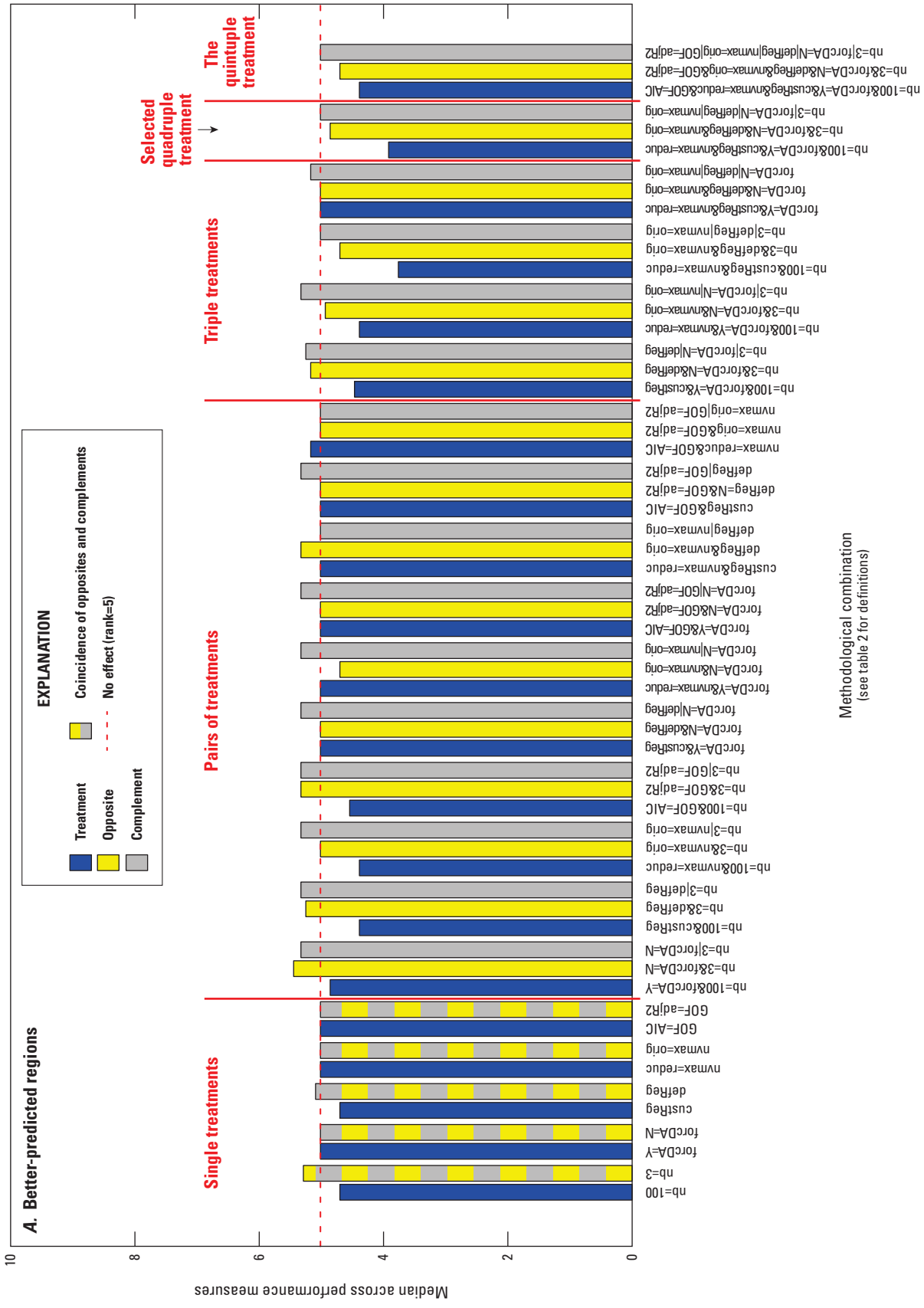


Figure 9. Medians across all summary statistics of all performance measures of median rank values for selected methodological combinations across *A*, better-predicted regions (regions 01, 02, 03, 04, 05, 06, 07, 08, and 17) and *B*, worse-predicted regions (regions 09, 10L, 10U, 11, 12, 13, 14, 15, 16, 18).

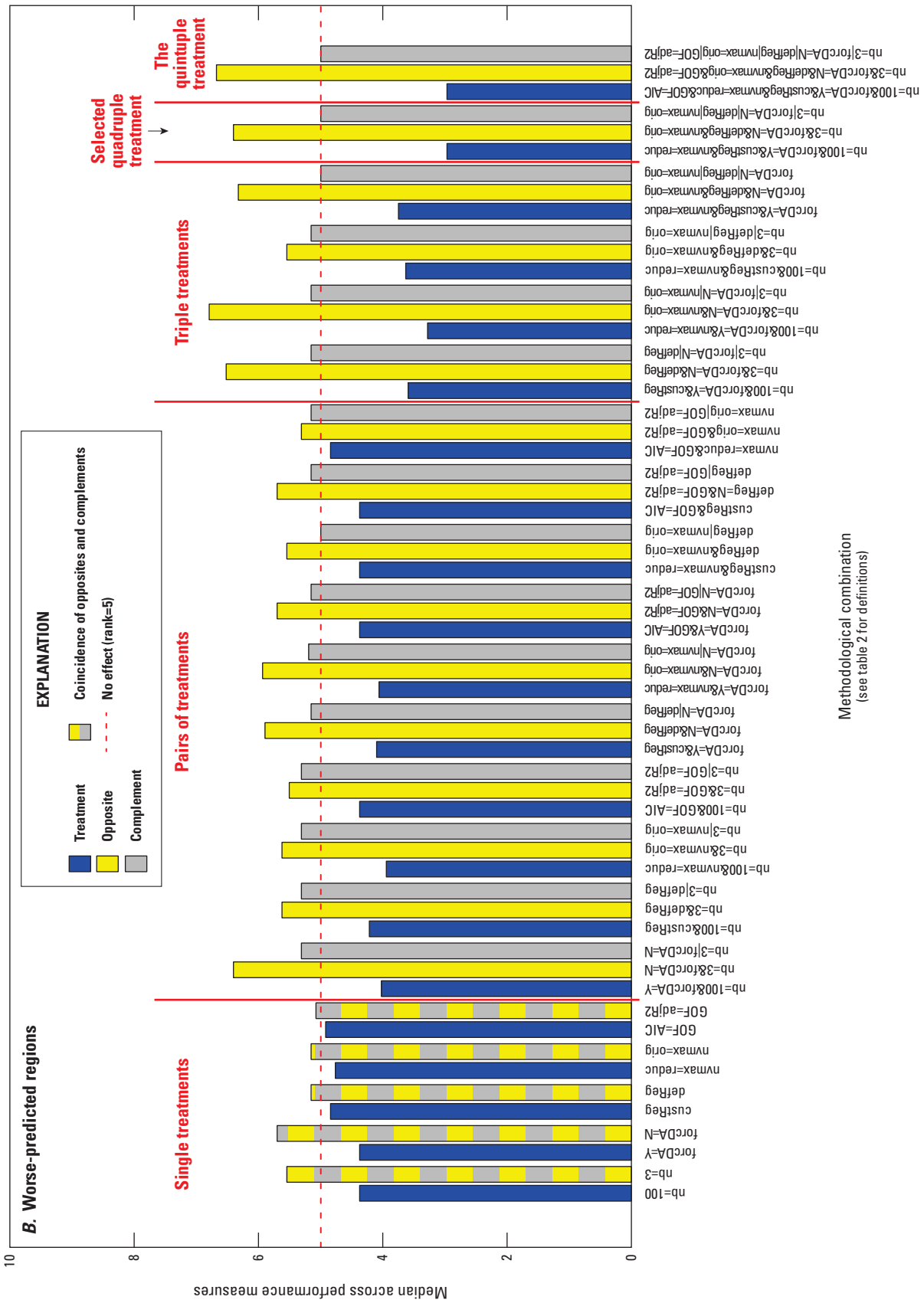


Figure 9. Medians across all summary statistics of all performance measures of median rank values for selected methodological combinations across *A*, better-predicted regions (regions 01, 02, 03, 04, 05, 06, 07, 08, and 17) and *B*, worse-predicted regions (regions 09, 10L, 10U, 11, 12, 13, 14, 15, 16, 18).—Continued

"nb=100&forcDA=Y&custReg&nvmax=reduc," the use of custom flow regimes ("custReg") and the forcing of drainage area into the regression models ("forcDA=Y"), arise from attempts to make the regression model development process more physically meaningful, and thus it is encouraging to see that overall they succeeded. There are other approaches to the problem of correlations between FDC quantiles, such as that of Archfield and others (2010), but the results in this report indicate that defining groups of quantiles to be predicted with the same basin characteristics (that is, streamflow regimes) by examination of the intraregional variance of unit FDCs is also of value.

Drainage area is intuitively the essential basin characteristic for characterizing flow statistics from groups of basins having a wide range of drainage areas, such as those presented in this study (3 or 4 orders of magnitude in most regions; table 1). For example, Gupta and others (1994) defined hydrologically homogeneous regions as those in which peak flow statistics depend only on drainage area, and they determined that drainage-area dependence of peak flow statistics is usually nontrivial (that is, the drainage-area exponent differs meaningfully and significantly from one). For evidence that drainage-area dependence is also nontrivial for FDCs, see Over and others (2014) and Farmer and others (2015). That forcing drainage area into the regression models was determined to be advantageous in this report also supports this intuition; however, in preliminary results for this study, it was also determined that poor results were obtained when a formula for the maximum number of basin characteristics *nvmax* was used that allowed some regions to be modeled with at most one basin characteristic, even when this characteristic was drainage area. These may very well be hydrologically heterogeneous basins by almost any definition, not just that of Gupta and others (1994), but it also seems likely that finding regions that are homogeneous with respect to the full range of flow behavior, as captured by the FDC, is much more challenging than it is for high flows alone.

That there was an advantage in considering more possible models as is seen in the preference for *nbest* and *ntop* set to 100 rather than 3 is unsurprising; however, as mentioned, the tradeoff is computation time, and in this study computation time was not trivial. Because of the LOOCV approach, the computations for each region were repeated as many times as there were basins, and for a single core on a Windows® personal computer, based on preliminary tests, the complete analysis for one methodological combination would have required about 10 days. However, a LOOCV approach is easily parallelizable, and for this study such an approach was implemented in R with the **`doParallel`** package (Revolution Analytics and Steve Weston, 2015).

These results reemphasized the well-known result that low-flow quantiles are substantially harder to estimate than high-flow quantiles, at least as measured by metrics based on log-transformed values. Within the FDC regression-of-quantiles framework, the solution is presumably to find, or possibly develop, better basin characteristics. As noted, in this study, because the GAGES–II dataset does not include any quantitative geologic variables, none were used. It may be that useful quantitative hydrologic indices applicable to large heterogeneous regions such as the CONUS can be constructed from qualitative classifications of surficial geology (for example, Over and others, 2014), from well logs (for example, Martin and others, 2016), or from hydrogeologic simulation models (for example, Gleeson and others, 2011).

There are many other possible directions in which to seek additional improvements in FDC prediction by regression methods. One direction is the question of the maximum number of basin characteristics to use in the regression equations. Finding the better of two versions of a simple formula for setting the maximum number of basin characteristics hardly constitutes a complete solution. For one thing, it may be that the ideal number of basin characteristics varies by quantile, since, for example, the available set of basin characteristics may contain more information useful for predicting higher flows as opposed to lower flows. More generally, the question of how to determine the best number to use is but one aspect of the larger question of variable selection for the FDC regression problem. One general class of statistical approaches that might warrant investigation is shrinkage methods (for example, Faraway, 2005, chap. 9).

Summary

The Water Census Project of the Water Availability and Use Science Program of the U.S. Geological Survey aims to provide consistent nationwide estimates of water flows and storage of the Nation's water resources. Among these are estimates of natural daily streamflow at ungaged locations. Estimates of undisturbed daily flow-duration curves (FDCs) are often used to estimate daily streamflow and in themselves provide a fundamental characterization of water availability; regional regression on basin characteristics is a common method to estimate FDCs at ungaged locations.

In this report, several refinements to the regional regression method of estimating FDCs at ungaged locations are evaluated through a leave-one-out cross-validation procedure in the 19 major river basins of the conterminous United States. The analyses in this report are based on daily streamflow data from water years 1981–2013 from 1,378 relatively undisturbed watersheds. Linear regression using selected basin characteristics at 27 quantiles ranging from 0.02 to 99.98 percent nonexceedance probabilities was applied. The regression computations were primarily by weighted least squares, with left-censored Gaussian regression solved by maximum likelihood in the presence of zero-valued quantiles.

The regional regression method as applied to the FDC estimation problem includes several methodological options that require determination of the better of two or more choices. The options considered in this report include (1) the setting of the maximum number of basin characteristics considered

in the regression models for each region, (2) the method of placing the quantiles into contiguous groups ("flow regimes") having the same basin characteristics used as independent variables, (3) the maximum number of candidate models retained from regressions at the single-quantile level that are retained for testing of the best model at the flow-regime scale, and (4) whether drainage area should be forced into the models. In all, 5 binary options were considered for most regions, resulting in 32 methodological combinations. Leave-one-out cross-validation predictions of FDC quantiles at each streamgage used in the study were used to evaluate compared options. Various performance measures were computed based on the predicted quantiles; these were combined by region and the methods were ranked for each measure.

The results indicate that prediction accuracy depends more on region than on methodology, with FDCs in arid regions and those with a large value of a measure of intra-regional flow-duration curve heterogeneity being harder to predict, particularly with respect to the low-flow quantiles. The results also support the common observation that prediction accuracy is lower for low-flow quantiles than it is for higher flow quantiles and extends it to the entire conterminous United States.

Based on examination of the ranked methods compared across the measures and considering better- and worse-predicted regions, the following treatments produced the more accurate results: (1) using fewer basin characteristics (of the two options considered), (2) utilizing a variance of the unit FDC-based method of determining the flow regimes rather than fixed regimes, (3) retaining more models from the quantile-level regressions regime-wide consideration, and (4) forcing drainage area into the regression models. Forcing drainage area into the regressions had a negligible effect for the better-predicted regions, because drainage area is usually included in those regions without forcing.

The FDCs are a key signature of hydrologic behavior of river basins and a key component of the prediction of the hydrology of ungaged basins. This study has provided a baseline view of the accuracy of FDC prediction by regression on basin characteristics across a large and hydrologically diverse domain, the conterminous United States, and indicated which certain refinements to that general method are the most beneficial to improve accuracy of prediction. Further refinements may be of additional benefit to the accuracy of prediction, but how large that improvement is likely to be, given the finding that variance in accuracy was larger between regions than between methods, is not clear.

Acknowledgments

Some of the R scripts used for this analysis were based on prior work by Elizabeth Cleveland, Michael Olson, and Riten Patel, all of the U.S. Geological Survey.

References Cited

Ahearn, E.A., 2010, Regional regression equations to estimate flow-duration statistics at ungaged stream sites in Connecticut: U.S. Geological Survey Scientific Investigations Report 2010–5052, 45 p.

Archfield, S.A., 2009, Estimation of continuous daily streamflow at ungaged locations in southern New England: Medford, Mass., Tufts University, Ph.D. dissertation, 104 p.

Archfield, S.A., Vogel, R.M., Steeves, P.A., Brandt, S.L., Weiskel, P.K., and Garabedian, S.P., 2010, The Massachusetts Sustainable-Yield Estimator—A decision-support tool to assess water availability at ungaged stream locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41 p., plus CD-ROM.

Bent, G.C., Steeves, P.A., and Waite, A.M., 2014, Equations for estimating selected streamflow statistics in Rhode Island: U.S. Geological Survey Scientific Investigations Report 2014–5010, 65 p., accessed April 5, 2017, at https://doi.org/10.3133/sir20145010.

Blom, G., 1958, Statistical estimates and transformed beta variables: New York, John Wiley, p. 68–75, 143–146.

Blum, A.G., Archfield, S.A., and Vogel, R.M., 2017, On the probability distribution of daily streamflow in the United States: Hydrology and Earth System Sciences, v. 21, no. 6, p. 3093–3103, accessed June 28, 2017, at https://doi.org/10.5194/hess-21-3093-2017.

Castellarin, A., Botter, G., Hughes, D.A., Liu, S., Ouarda, T.B.M.J., Parajka, J., Post, D.A., Sivapalan, M., Spence, C., Viglione, A., and Vogel, R.M., 2013, Prediction of flow duration curves in ungauged basins, *in* Blöschl, G., and others, eds., Runoff prediction in ungauged basins—Synthesis across processes, places and scales: Cambridge, Cambridge University Press, p. 135–162, accessed February 13, 2014, at https://doi.org/10.1017/CBO9781139235761.010.

Castellarin, A., Vogel, R., and Brath, A., 2004, A stochastic index flow model of flow duration curves: Water Resources Research, v. 40, no. 3, accessed December 13, 2004, at https://doi.org/10.1029/2003WR002524.

Cheng, L., Yaeger, M.A., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M., 2012, Exploring the physical controls of regional patterns of flow duration curves—Part 1—Insights from statistical analyses: Hydrology and Earth System Sciences, v. 16, no. 11, p. 4435–4446, accessed December 4, 2013, at https://doi.org/10.5194/hess-16-4435-2012.

Esralew, R.A., and Smith, S.J., 2010, Methods for estimating flow-duration and annual mean-flow statistics for ungaged streams in Oklahoma: U.S. Geological Survey Scientific Investigations Report 2009–5267, 131 p.

Falcone, J.A., 2011, GAGES–II—Geospatial attributes of gages for evaluating streamflow [digital spatial dataset]: U.S. Geological Survey Water Resources NSDI Node web page, accessed May 19, 2015, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml.

Falcone, J.A., Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010, GAGES—A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecology, v. 91, no. 2, p. 621, a database paper in Ecological Archives E091–045–DI, accessed December 16, 2010, at http://esapubs.org/Archive/ecol/E091/045/metadata.htm.

Faraway, J.J., 2005, Linear models with R: Boca Raton, Fla., CRC Press, 229 p.

Farmer, W.H., Archfield, S.A., Over, T.M., Hay, L.E., LaFontaine, J.H., and Kiang, J.E., 2014, A comparison of methods to predict historical daily streamflow time series in the southeastern United States: U.S. Geological Survey Scientific Investigations Report 2014–5231, 34 p., accessed March 17, 2015, at https://pubs.usgs.gov/sir/2014/5231/.

Farmer, W.H., Over, T.M., and Vogel, R.M., 2015, Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States: Water Resources Research, v. 51, no. 3, p. 1775–1796, accessed April 20, 2015, at https://doi.org/10.1002/2014WR015924.

Fennessey, N.M., 1994, A hydro-climatological model of daily streamflow for the northeast United States: Medford, Mass., Tufts University, Ph.D. dissertation [variously paged].

Flynn, R.H., 2003, Development of regression equations to estimate flow durations and low-flow frequency statistics in New Hampshire streams: U.S. Geological Survey Water-Resources Investigations Report 02–4298, 66 p.

Gazoorian, C.L., 2015, Estimation of unaltered daily mean streamflow at ungaged streams of New York, excluding Long Island, water years 1961–2010: U.S. Geological Survey Scientific Investigations Report 2014–5220, 29 p., accessed April 5, 2017, at https://doi.org/10.3133/sir20145220.

Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H.H., Manning, A.H., van Beek, L.P.H., and Jellinek, A.M., 2011, Mapping permeability over the surface of the Earth: Geophysical Research Letters, v. 38, no. 2, accessed January 22, 2011, at https://doi.org/10.1029/2010GL045565.

Gupta, V.K., Mesa, O.J., and Dawdy, D.R., 1994, Multiscaling theory of flood peaks—Regional quantile analysis: Water Resources Research, v. 30, no. 12, p. 3405–3421, accessed February 26, 2007, at https://doi.org/10.1029/94WR01791.

Helsel, D.R., 2012, Statistics for censored environmental data using Minitab® and R (2d ed.): Hoboken, N.J., John Wiley and Sons, 324 p.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods *in* water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chapter A3, 522 p.

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

Kiang, J.E., Stewart, D.W., Archfield, S.A., Osborne, E.B., and Eng, K., 2013, A national streamflow network gap analysis: U.S. Geological Survey Scientific Investigations Report 2013–5013, 79 p. plus one appendix as a separate file, accessed June 18, 2013, at https://pubs.usgs.gov/sir/2013/5013/.

Ley, R., Casper, M.C., Hellebrand, H., and Merz, R., 2011, Catchment classification by runoff behaviour with self-organizing maps (SOM): Hydrology and Earth System Sciences, v. 15, no. 9, p. 2947–2962, accessed March 26, 2018, at https://doi.org/10.5194/hess-15-2947-2011.

Linhart, S.M., Nania, J.F., Sanders, C.L., Jr., and Archfield, S.A., 2012, Computing daily mean streamflow at ungaged locations in Iowa by using the Flow Anywhere and Flow Duration Curve Transfer statistical methods: U.S. Geological Survey Scientific Investigations Report 2012–5232, 50 p., accessed December 10, 2012, at https://pubs.usgs.gov/sir/2012/5232/.

Lumley, T., 2012, leaps—Regression subset selection: R package version 2.9 [using Fortran code by Alan Miller], accessed April 2016, at http://CRAN.R-project.org/package=leaps.

Martin, G.R., Fowler, K.K., and Arihood, L.D., 2016, Estimating selected low-flow frequency statistics and harmonic-mean flows for ungaged, unregulated streams in Indiana (ver. 1.1, October 2016): U.S. Geological Survey Scientific Investigations Report 2016–5102, 45 p., accessed October 6, 2016, at https://pubs.er.usgs.gov/publication/sir20165102.

McKelvey, R.D., and Zavoina, W., 1975, A statistical model for the analysis of ordinal level dependent variables: The Journal of Mathematical Sociology, v. 4, no. 1, p. 103–120, accessed March 26, 2018, at https://doi.org/10.1080/0022250X.1975.9989847.

Mohamoud, Y.M., 2008, Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves: Hydrological Sciences Journal, v. 53, no. 4, p. 706–724, accessed December 1, 2009, at https://doi.org/10.1623/hysj.53.4.706.

Over, T.M., Riley, J.D., Sharpe, J.B., and Arvin, D., 2014, Estimation of regional flow-duration curves for Indiana and Illinois: U.S. Geological Survey Scientific Investigations Report 2014–5177, 24 p. and additional downloads, tables 2–5, 8–13, and 18, accessed November 14, 2014, at https://pubs.er.usgs.gov/publication/sir20145177.

Perry, C.A., Wolock, D.M., and Artman, J.C., 2004, Estimates of flow duration, mean flow, and peak-discharge frequency values for Kansas stream locations: U.S. Geological Survey Scientific Investigations Report 2004–5033, 651 p., accessed April 5, 2017, at https://pubs.usgs.gov/sir/2004/5033/.

Poncelet, C., Andréassian, V., Oudin, L., and Perrin, C., 2017, The Quantile Solidarity approach for the parsimonious regionalization of flow duration curves: Hydrological Sciences Journal, v. 62, no. 9, p. 1364–1380, accessed June 14, 2017, at https://doi.org/10.1080/02626667.2017.1335399.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.T., 1992, Numerical Recipes in Fortran 77—The art of scientific computing (2d ed.): Cambridge, Cambridge University Press, Section 14.3, 963 p.

R Core Team, 2016, R—A language and environment for statistical computing—The R project for statistical computing: R Foundation for Statistical Computing web page, accessed April 2016, at https://www.R-project.org/.

Revolution Analytics and Steve Weston, 2015, doParallel—Foreach Parallel Adaptor for the 'parallel' Package, R package version 1.0.10: accessed April 2016, at http://CRAN.R-project.org/package=doParallel.

Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00–4135, 81 p.

Russell, A.M., Over, T.M., and Farmer, W.H., 2018, Streamflow, flow-duration curves, basin characteristics, and regression models of flow-duration curves for selected streamgages in the conterminous United States: U.S. Geological Survey data release, accessed April 2018, at https://doi.org/10.5066/F70G3JC4.

Stuckey, M.H., 2016, Estimation of daily mean streamflow for ungaged stream locations in the Delaware River Basin, water years 1960–2010: U.S. Geological Survey Scientific Investigations Report 2015–5157, 42 p., accessed April 5, 2017, at https://doi.org/10.3133/sir20155157.

Stuckey, M.H., Koerkle, E.H., and Ulrich, J.E., 2014, Estimation of baseline daily mean streamflows for ungaged locations on Pennsylvania streams, water years 1960–2008 (ver. 1.1, August 2014): U.S. Geological Survey Scientific Investigations Report 2012–5142, 61 p., accessed April 5, 2017, at https://pubs.usgs.gov/sir/2012/5142.

UCLA—Statistical Consulting Group, 2011, FAQ—What are pseudo R-squareds?: accessed March 23, 2018, at https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/.

U.S. Geological Survey (USGS), 2016, USGS water data for the Nation: U.S. Geological Survey National Water Information System web page, accessed April 18, 2016, at http://waterdata.usgs.gov/nwis/

Vogel, R.M., and Fennessey, N.M., 1995, Flow duration curves II—A review of applications in water resources planning: Journal of the American Water Resources Association, v. 31, no. 6, p. 1029–1039, accessed September 5, 2008, at https://doi.org/10.1111/j.1752-1688.1995.tb03419.x.

Westerberg, I.K., Guerrero, J.L., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J.E., and Xu, C.-Y., 2011, Calibration of hydrological models using flow-duration curves: Hydrology and Earth System Sciences, v. 15, no. 7, p. 2205–2227, accessed December 4, 2013, at https://doi.org/10.5194/hess-15-2205-2011.

Yokoo, Y., and Sivapalan, M., 2011, Towards reconstruction of the flow duration curve—Development of a conceptual framework with a physical basis: Hydrology and Earth System Sciences, v. 15, no. 9, p. 2805–2819, accessed October 14, 2011, at https://doi.org/10.5194/hess-15-2805-2011.

Ziegeweid, J.R., Lorenz, D.L., Sanocki, C.A., and Czuba, C.R., 2015, Methods for estimating flow-duration curve and low-flow frequency statistics for ungaged locations on small streams in Minnesota: U.S. Geological Survey Scientific Investigations Report 2015–5170, 23 p., accessed April 5, 2017, at https://doi.org/10.3133/sir20155170.

Appendixes

Appendix tables are available online at https://doi.org/10.3133/sir20185072.

Appendix 1

Table 1.1. Descriptions of basin characteristics used in regression equations for predicting flow-duration curves in this study.

Table 1.2. Basin characteristics considered for use in regression equations for predicting flow-duration curves in each region for this study.

Appendix 2

Table 2.1. Summary statistics, by region and prediction method, of performance measures used to evaluate predicted flow-duration curves in this study.

Table 2.2. Ranks by region of summary statistics of performance measures used to evaluate predicted flow-duration curves among prediction methods in this study.

Appendix 3

Table 3.1. Medians, across selected groups of regions and selected combinations of methodological options, of ranks of the performance measure summary statistics for all regions.

Table 3.2. Medians, across selected groups of regions and selected combinations of methodological options, of ranks of the performance measure summary statistics for regions 01–08 and 17.

Table 3.3. Medians, across selected groups of regions and selected combinations of methodological options, of ranks of the performance measure summary statistics for regions 09–16 and 18.

For more information about this publication, contact
Director, USGS Central Midwest Water Science Center - Illinois Office
405 N. Goodwin Ave.
Urbana, Illinois 61801
(217) 328-8747

For additional information visit https://il.water.usgs.gov

Publishing support provided by the
Madison and Rolla Publishing Service Centers

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20185072