

In cooperation with the Department of the Interior WaterSMART Program

A Comparison of Methods to Predict Historical Daily Streamflow Time Series in the Southeastern United States

Scientific Investigations Report 2014–5231

A Comparison of Methods to Predict Historical Daily Streamflow Time Series in the Southeastern United States

By William H. Farmer, Stacey A. Archfield, Thomas M. Over, Lauren E. Hay, Jacob H. LaFontaine, and Julie E. Kiang

In cooperation with the Department of the Interior WaterSMART Program

Scientific Investigations Report 2014–5231

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

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

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

U.S. Geological Survey, Reston, Virginia: 2014

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

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

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

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

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., <http://dx.doi.org/10.3133/sir2014-5231>.

ISSN 2328-0328 (online)

Contents

Abstract.....	1
Introduction.....	1
Purpose and Scope	2
Study Area, Streamgauge Selection, and Dataset Development.....	2
Methods to Predict Daily Streamflow	4
Selection of Index Gages	4
Methods to Estimate Streamflow Moments at Ungaged Sites.....	7
Methods to Estimate the Quantiles of Flow Duration Curves at Ungaged Sites	8
Statistical, Transfer-Based Methods.....	9
Drainage-Area Ratio	9
Standardization with Mean.....	9
Standardization with Mean and Standard Deviation.....	9
Nonlinear Spatial Interpolation Using Flow Duration Curves	10
Analysis of Flows in Networks of Channels Model, Temporally Disaggregated.....	11
Precipitation Runoff Modeling System	11
Methods of Analysis.....	12
Estimates of Daily, No-Fail Storage-Yield Curves.....	13
Robust Rank-Based Evaluation	14
Results and Discussion.....	14
Overall Goodness of Fit of Predicted Streamflow Records	15
Daily, No-Fail Storage-Yield Curves.....	19
Streamflow Statistics	21
Flow Duration Curves	21
Fundamental Daily Streamflow Statistics.....	23
Robust Rank-Based Evaluation of Streamflow Statistics	25
Full Cross-Metric Comparison	26
Sensitivity to Index-Selection Algorithm	27
Summary and Conclusions.....	29
Limitations and Further Research	30
Acknowledgments	31
References Cited.....	31

Figures

1. Southeast Model Comparison study area with four-digit Hydrologic Unit Codes and Physiographic Provinces.....	3
2. Study area and streamgage locations.....	5
3. Several examples of daily, no-fail storage-yield curves for sites in the study area of the Southeast Model Comparison.....	13
4. The distribution of Nash-Sutcliffe efficiencies of daily streamflow predictions for each method of prediction in ungaged basins considered in the Southeast Model Comparison	15
5. The distribution of the at-site Nash-Sutcliffe efficiencies of the logarithms of daily streamflow predictions for each method of prediction in ungaged basins is considered here	16
6. The distribution of at-site average percent errors of daily streamflow predictions for the seven most accurate methods of prediction in ungaged basins.....	17
7. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the seven overall goodness-of-fit metrics for each method of prediction in ungaged basins	18
8. The distribution of the at-site Nash-Sutcliffe efficiencies of the daily storage-yield curve for the seven most accurate methods of prediction in ungaged basins.....	19
9. The distribution of at-site average percent errors of the estimated daily storage-yield curve for the seven most accurate methods of prediction in ungaged basins.....	20
10. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the seven metrics of the storage-yield curve for each method of prediction in ungaged basins	21
11. The median at-site percent error of five percentiles along the daily flow duration curve for the seven most accurate methods for prediction in ungaged basins	22
12. The median at-site absolute percent error of five percentiles along the daily flow duration curve for the seven most competitive methods for prediction in ungaged basins.....	23
13. The median at-site percent error of the seven fundamental daily streamflow statistics for the seven most accurate methods for prediction in ungaged basins	24
14. The median at-site absolute percent error of the seven fundamental daily streamflow statistics for the seven most accurate methods for prediction in ungaged basins.....	25
15. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the 17 flow statistics for each method of prediction in ungaged basins	26
16. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of all 32 performance metrics for each method of prediction in ungaged basins.....	27
17. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of all 32 methods for each method of prediction in ungaged basins, average across competing methods for the selection of index gages.....	28

Tables

[Tables 1–7 are located in a separate PDF file, available at <http://pubs.usgs.gov/sir/2014/5231/>.]

1. Names, station numbers, fraction of zeros, and period of record for each streamgage used in the Southeast Model Comparison
2. A listing of all of the names and abbreviations of methods for prediction in ungaged basins considered as part of the Southeast Model Comparison
3. Root-mean-square error of the fitted variograms at 182 streamgages used in the Southeast Model Comparison
4. Fitted coefficients and goodness-of-fit statistics for regressions of the 27 percentiles of the distribution of daily streamflow at the 182 streamgages used in the Southeast Model Comparison
5. Mean rank for each performance metric and method of prediction in ungaged basins
6. Standard deviation of the ranks for each performance metric and method of prediction in ungaged basins
7. Mean and standard deviation of the average ranks for each method of prediction in ungaged basins, regardless of the method used for index-gage selection

Appendixes

[Appendixes A, B, and C are located in a separate PDF files, available at <http://pubs.usgs.gov/sir/2014/5231/>.]

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square foot (ft ²)	929.0	square centimeter (cm ²)
square foot (ft ²)	0.09290	square meter (m ²)
square mile (mi ²)	259.0	hectare (ha)
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
million gallons (Mgal)	3,785	cubic meter (m ³)
cubic foot (ft ³)	28.32	cubic decimeter (dm ³)
cubic foot (ft ³)	0.02832	cubic meter (m ³)
cubic yard (yd ³)	0.7646	cubic meter (m ³)
cubic mile (mi ³)	4.168	cubic kilometer (km ³)
acre-foot (acre-ft)	1,233	cubic meter (m ³)
acre-foot (acre-ft)	0.001233	cubic hectometer (hm ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
cubic foot per second per square mile [(ft ³ /s)/mi ²]	0.01093	cubic meter per second per square kilometer [(m ³ /s)/km ²]
cubic foot per day (ft ³ /d)	0.02832	cubic meter per day (m ³ /d)
gallon per minute (gal/min)	0.06309	liter per second (L/s)
gallon per day (gal/d)	0.003785	cubic meter per day (m ³ /d)
gallon per day per square mile [(gal/d)/mi ²]	0.001461	cubic meter per day per square kilometer [(m ³ /d)/km ²]
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m ³ /s)
million gallons per day per square mile [(Mgal/d)/mi ²]	1,461	cubic meter per day per square kilometer [(m ³ /d)/km ²]

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

$$^{\circ}\text{F}=(1.8\times^{\circ}\text{C})+32$$

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

$$^{\circ}\text{C}=(^{\circ}\text{F}-32)/1.8$$

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).

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

A Comparison of Methods to Predict Historical Daily Streamflow Time Series in the Southeastern United States

By William H. Farmer, Stacey A. Archfield, Thomas M. Over, Lauren E. Hay, Jacob H. LaFontaine, and Julie E. Kiang

Abstract

Effective and responsible management of water resources relies on a thorough understanding of the quantity and quality of available water. Streamgages cannot be installed at every location where streamflow information is needed. As part of its National Water Census, the U.S. Geological Survey is planning to provide streamflow predictions for ungaged locations. In order to predict streamflow at a useful spatial and temporal resolution throughout the Nation, efficient methods need to be selected. This report examines several methods used for streamflow prediction in ungaged basins to determine the best methods for regional and national implementation. A pilot area in the southeastern United States was selected to apply 19 different streamflow prediction methods and evaluate each method by a wide set of performance metrics. Through these comparisons, two methods emerged as the most generally accurate streamflow prediction methods: the nearest-neighbor implementations of nonlinear spatial interpolation using flow duration curves (NN-QPPQ) and standardizing logarithms of streamflow by monthly means and standard deviations (NN-SMS12L). It was nearly impossible to distinguish between these two methods in terms of performance. Furthermore, neither of these methods requires significantly more parameterization in order to be applied: NN-SMS12L requires 24 regional regressions—12 for monthly means and 12 for monthly standard deviations. NN-QPPQ, in the application described in this study, required 27 regressions of particular quantiles along the flow duration curve. Despite this finding, the results suggest that an optimal streamflow prediction method depends on the intended application. Some methods are stronger overall, while some methods may be better at predicting particular statistics. The methods of analysis presented here reflect a possible framework for continued analysis and comprehensive multiple comparisons of methods of prediction in ungaged basins (PUB). Additional metrics of comparison can easily be incorporated into this type of analysis. By considering such a multifaceted approach, the top-performing models can easily be identified and considered for further research.

The top-performing models can then provide a basis for future applications and explorations by scientists, engineers, managers, and practitioners to suit their own needs.

Introduction

Effective and responsible management of water resources relies on a thorough understanding of the quantity and quality of available water. In regards to quantity, daily records of streamflow are some of the best sources of the information required for assessing water resources. These records can be used to determine specific characteristics of flow that influence a wide range of projects, from irrigation scheduling to water-supply management and the development of hydropower, to name a few (Sivapalan, 2003; Sivapalan and others, 2003; Hrachowitz and others, 2013). Long-term historical records are also essential to understanding trends in climate and basin hydromorphology (Vogel, 2011). As human development expands across the landscape, daily streamflow records are vital to answering a growing number of increasingly complex questions.

Historically, daily streamflow records have been produced by operating streamgages. Streamgage density and length of continuous record have thus been guided by funding availability and location chosen to address specific, often local, questions. Given the funding requirements and the length of operation needed to produce streamflow records of sufficient length for analysis, it is unreasonable to expect daily streamflow records at the outlet of every basin of interest (for example, HUC-12 outlets). The result is a network of gages in the United States with many spatial and temporal gaps in observed streamflow records (Kiang and others, 2013). Developing methods for filling these spatial and temporal gaps based on mathematical models is required to answer questions regarding the quantity and quality of available water in ungaged basins and is the goal of the Prediction in Ungaged Basins (PUB) initiative (Sivapalan and others, 2003; Hrachowitz and others, 2013).

Recognizing the importance of water resources, the U.S. Department of the Interior (DOI) launched the WaterSMART initiative to describe the quality and quantity of the Nation's water resources (Alley and others, 2013). The U.S. Geological Survey's (USGS) National Water Census, a component of WaterSMART, aims to provide water-accounting tools and to assess the availability of water resources to meet current and projected ecological and human water requirements at regional and national scales (Alley and others, 2013). While the National Water Census considers a number of different aspects of water resources, one specific research focus is to provide spatially and temporally continuous predictions of natural, historical daily streamflow at the subwatershed (HUC-12) outlet scale. (In the southeastern United States, these outlet basins have a mean size of 853 mi².) The daily streamflow product is an essential component of water budget estimates and will provide direct information for human and aquatic needs.

While numerous methods exist to predict specific streamflow statistics in ungaged basins (see Thomas and Benson, 1970, and Ries III, 2007, for a summary), the literature concerning prediction of complete time series is much more limited, often consisting of comparisons of a proposed method against one or two previous methods (Hirsch, 1979; Hughes and Smakhtin, 1996; Shu and Ouarda, 2012; Razavi and Coulibaly, 2013). In order to predict streamflow at a useful spatial and temporal resolution throughout the Nation, ideal methods must be accurate and efficient. In addition, ideal methods, so as to better understand the limits of their usefulness, should provide some means of characterizing the degree of uncertainty in predictions. Prior to national-scale implementation, methodological approaches to continuous streamflow prediction need to be developed and rigorously evaluated.

A pilot study in the southeastern United States, called the Southeast Model-Comparison study (SEMC), was conducted to determine the best prediction methods for regional and national model implementation. The SEMC was focused on conducting a comprehensive analysis and cross comparison of 19 prediction methods for ungaged basins (PUB methods) using a wide-ranging set of performance metrics while also characterizing uncertainty. The intention of the SEMC is to develop a framework of analysis that can incorporate existing and future PUB methods and provide a thorough analysis of the strengths and weaknesses of each. The result of the SEMC is a robust understanding of current PUB methods and their advantages and limitations.

Purpose and Scope

This report presents the results of the SEMC: a comparison of 19 methods for predicting streamflow at a daily time step in the southeastern United States (fig. 1). The report documents the ability of each model to predict observed daily streamflow, daily, no-fail storage-yield curves, and specific flow statistics at 182 locations over a 30-year period (water years 1981–2010; water years are defined as the period of time

between October 1 and September 30 of consecutive calendar years). Additionally, information regarding relative model performance is presented. The accuracy of the models is the principal basis for comparison, though the level of effort and data requirements for the various methods are also discussed. This report presents a method of analysis that is thorough, but not exhaustive, though the conclusions presented remain relative to the location and PUB methods evaluated, rather than unconditional or absolute.

Study Area, Streamgauge Selection, and Dataset Development

The Southeast Model Comparison study area covers approximately 137,000 mi² and includes all or part of 12 4-digit Hydrologic Unit Code (HUC-4) sub-basins in Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (fig. 1). The majority of the study area eventually drains to the Gulf of Mexico, including the Alabama-Coosa-Tallapoosa, Apalachicola-Chattahoochee-Flint, Choctawhatchee-Escambia, Mobile-Tombigbee, Ochlockonee, Pascagoula, Suwannee, and middle and upper Tennessee River Basins, with parts of two additional basins (Altamaha and Savannah Rivers) draining to the Atlantic Ocean. An analysis of the 2011 National Land Cover Database (NLCD; Jin and others, 2013) shows that the study area consists of 49.4 percent forest cover, 17.6 percent agricultural lands, 12.1 percent scrub/shrub and grasslands, 9.0 percent developed lands, 9.0 percent wetlands, 2.5 percent open water, and 0.4 percent barren land. The study area is composed of six physiographic regions: the Appalachian Plateaus, Blue Ridge, Coastal Plain, Interior Low Plateaus, Piedmont, and Valley and Ridge. The Appalachian and Interior Low Plateaus include nearly horizontal layers of Mississippian- and Pennsylvanian-age sedimentary rocks. The Blue Ridge and Piedmont are underlain mostly by Precambrian-age and older Paleozoic-age crystalline rocks. The Valley and Ridge, as its name suggests, consists of a series of northeast-trending linear ridges and valleys underlain by alternating beds of hard and soft Paleozoic sedimentary rocks (Johnson and others, 2002). Lastly, the Coastal Plain is underlain by Mesozoic- and Cenozoic-age sedimentary rocks and unconsolidated sediments (Couch and others, 1995).

The SEMC study area is generally a warm and humid temperate climate. Air temperatures generally decrease with increasing latitude and elevation. Annual average temperatures range from a high of 67.8 °F in the southern part to a low of 50.7 °F in the northern part. Average annual precipitation accumulation in the study area is 54.4 inches (in.), with average values ranging from a low of 44.1 in. in the northeastern-most part of the upper Tennessee River Basin to a high of 69.2 in. at the intersection of Georgia, North Carolina, and South Carolina. Generally, precipitation accumulation is greatest in the southern part of the study area near the Gulf of Mexico

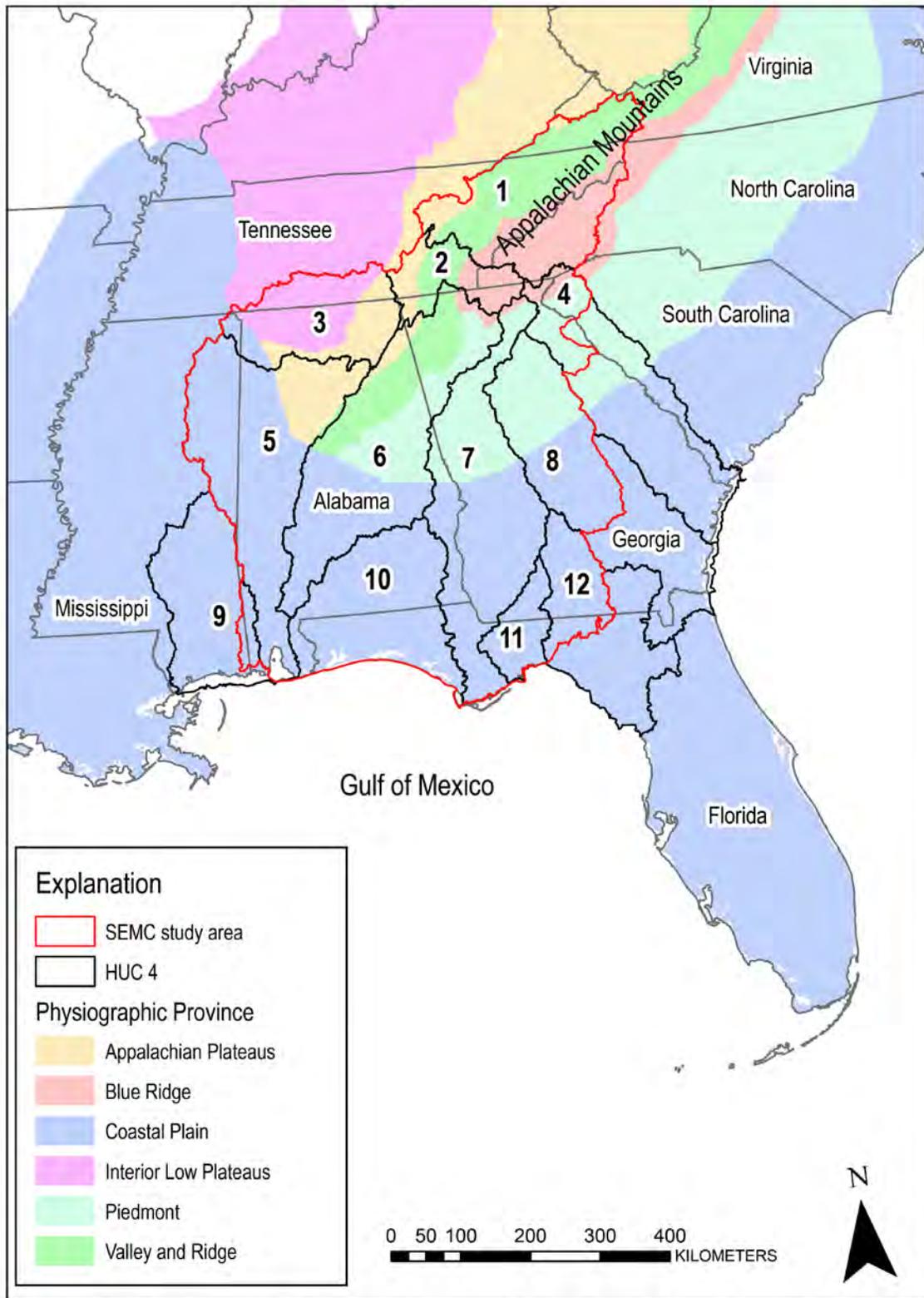


Figure 1. Southeast Model Comparison (SEMC) study area with four-digit Hydrologic Unit Codes (HUC) and Physiographic Provinces. HUC 4 units are as follows: (1) Upper Tennessee, (2) Middle Tennessee-Hiwassee, (3) Middle Tennessee-Elk, (4) Savannah, (5) Mobile-Tombigbee, (6) Alabama-Coosa-Tallapoosa, (7) Apalachicola-Chattahoochee-Flint, (8) Altamaha, (9) Pascagoula, (10) Chocktawhatchee-Escambia, (11) Ochlockonee, and (12) Suwannee.

and in the southern Appalachian Mountains, lowest in central Georgia and the northeastern most part of the study area, and closer to the average in the western half of the study area.

In order to assess the predictive power of the PUB methods, the streamgages considered for calibration and validation were required to exhibit near-natural streamflow records for an extended period of time. As the intention was to produce a 30-year continuous daily record between October 1, 1980, and September 30, 2010, each potential streamgage was required to have at least 10 complete water years within the period of interest. Based largely on the work of Falcone and others (2010), Falcone (2011) and Gotvald and others (2009), 182 streamgages in the southeastern United States satisfied these criteria and were selected for use in this analysis (fig. 2; table 1). The selected records were assessed for missing periods. At seven streamgages (see app. A), short missing periods of 1 to 33 days were filled in consultation with surface-water specialists of each USGS Water Science Center where the data originated; often missing data were available from the local databases. Other methods used for filling included transfer of data from nearby stations and interpolation; see appendix A. After filling, only complete water years of data were used in the analyses.

More than 300 basin characteristics were obtained for each of the catchments used in the study; a description of the characteristics and how they were computed can be found in the GAGES-II documentation (Falcone and others, 2010; Falcone, 2011); these include climate characteristics, drainage characteristics, characteristics describing geology and soils, characteristics describing land use, and topographic characteristics. After some initial exploration of regional regressions following the methods for the prediction of streamflow means, standard deviations, and percentiles outlined below, this list was reduced to over 80 different variables that appeared most frequently in regressions based only on GAGES-II reference basins. These candidate variables were identified using best-subsets regression (regsubsets in R) and the R package leaps (Lumley, 2012). The remaining 80 variables represented the full range of the basin characteristics mentioned above. (The set of basin characteristics considered here is included in app. B.) For streamgages used in this study that are not part of the GAGES-II dataset, the same protocols described in Falcone and others (2010) were employed to determine catchment characteristics.

Methods to Predict Daily Streamflow

Six basic methods were used to predict daily streamflow at ungedaged locations. These methods include four statistical, transfer-based methods, one process-based method, and one quasi-process-based method. The transfer-based and quasi-process-based methods require an index gage for the transfer of timing and magnitude information. Two such indexing algorithms were assessed, selecting either by minimum

Euclidean distance (nearest neighbor) or maximum estimated correlation (Map Correlation; Archfield and Vogel, 2010). The transfer-based methods include drainage area ratio, nonlinear spatial interpolation using flow duration curves (Fennessey, 1994; Hughes and Smakhtin, 1996), standardizing streamflows by a mean streamflow, standardizing streamflows by a mean, and standard deviation of streamflow. Standardizing by mean can be executed with annual or monthly means; standardizing by the mean and standard deviation can be implemented with annual or monthly mean and standard deviations of natural, logarithmically transformed and untransformed streamflow. The Precipitation Runoff Modeling System (PRMS) (Leavesley and others, 1983; Markstrom and others, 2008) is a process-based model that uses climatic, land use, and basin properties to predict time series of streamflow. The quasi-process-based method was a temporal downscaling of the monthly Analysis of Flows in Networks of Channels (AFINCH) model (Holtschlag, 2009). AFINCH uses land use and temporal climatic inputs, but is largely based on statistical relationships rather than physical processes. The temporal downscaling, or disaggregation, is accomplished using an index gage; additional information is provided below. Accounting for the possibilities for index-gage selection and particular use of moments, there were 19 complete methods considered; these are outlined in table 2.

These 19 PUB methods were used to simulate the ungedaged scenario for 182 sites in the southeastern United States. A threefold validation procedure was used whereby the 182 sites were divided into three groups. For each third of the sites, the remaining two-thirds were used to calibrate each PUB method, as needed. The parameters estimated through calibration were then used to predict the ungedaged third. The two-thirds set used for calibration is referred to as the calibration network, while the remaining third is known as the validation set. Iterating which third is considered the validation set produces a single, “ungedaged” estimate for each site, which allowed for an assessment of performance across all 182 sites.

Selection of Index Gages

All of the PUB methods compared in this paper require an index gage from which to transfer information. The statistical, transfer-based methods and the disaggregation of AFINCH require an index gage to determine streamflow timing and relative streamflow magnitudes. PRMS requires an index gage to obtain model parameters at the ungedaged site. The index gages were identified by two different techniques: nearest neighbor or map correlation. Both of these techniques are widely used, though the nearest neighbor is used more often (Archfield and Vogel, 2010; Shu and Ouarda, 2012), largely because of its ease of use and relative success. In either case, the index gage was selected from the calibration network of gages as defined by the threefold validation procedure outlined above.

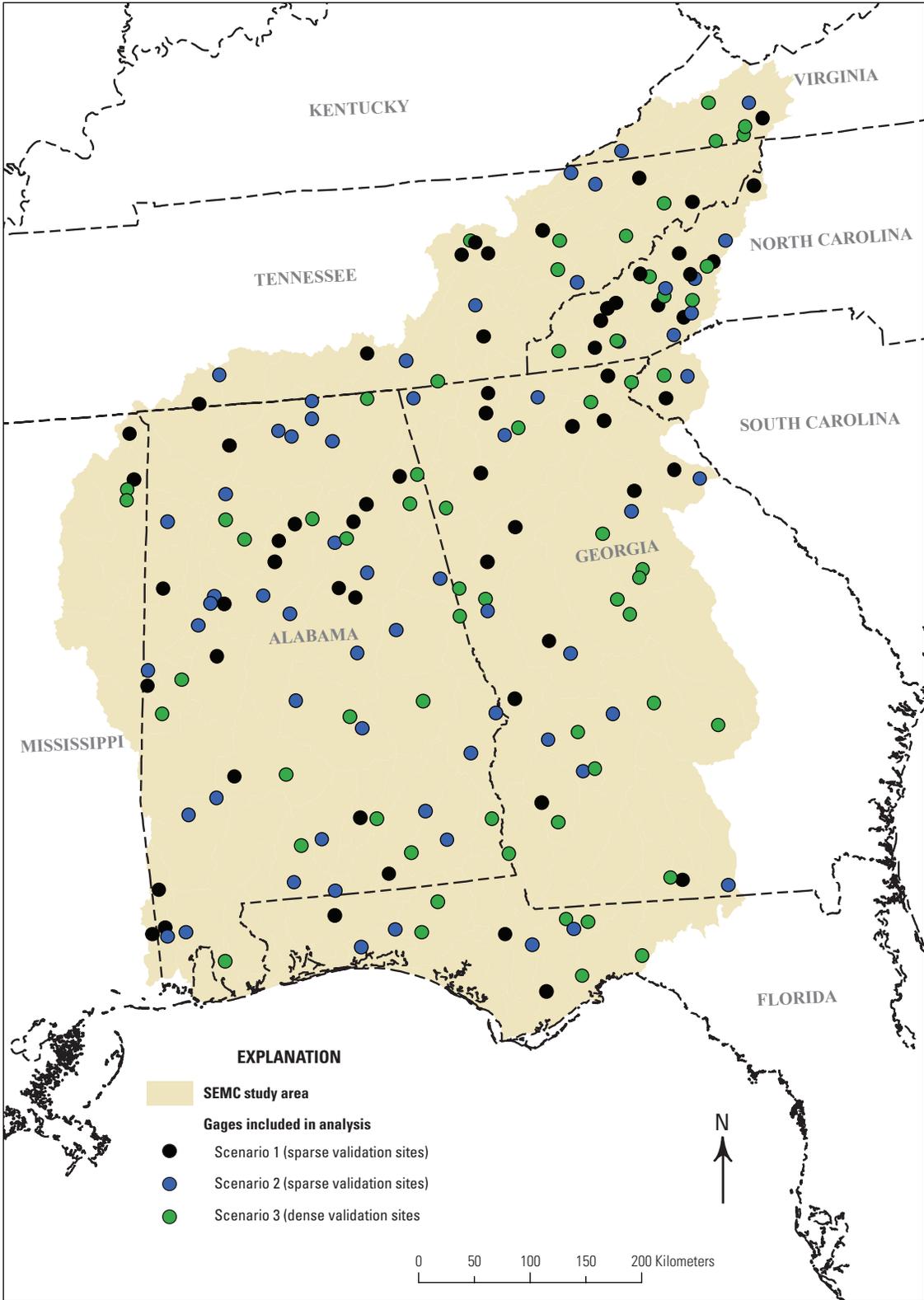


Figure 2. Study area and streamgage locations.

Daily streamflow was predicted at validation sites from October 1, 1980, to September 30, 2010. However, for the PUB methods that require timing information, it was not always possible to find a single index gage in the calibration set that could be used to estimate daily streamflow for the entire period. To remedy this, the potential index gages for a validation site were ranked by the selected criteria (nearest neighbor or map correlation) and index gages were selected until the entire period of interest was represented. For example, the top three index gages for a particular validation site might have had different periods of record: (1) October 1939–September 1996; (2) January 1998–December 2010; and (3) October 1991–September 2005. In predicting the streamflow for this validation site, index gage (1) would have been used from October 1980–September 1996, gage (2) would have been used to predict January 1998–September 2010, and gage (3) would have been used to fill the gap between October 1996 and December 1997.

The nearest-neighbor index-selection technique ranks potential index gages based on spatial proximity. The location of each gage in the calibration network and validation set is defined by the latitude and longitude of the basin outlet. For the statistical, transfer-based methods and AFINCH, the index gages are ranked according to smallest Euclidean distance between the validation site and potential index gage. These methods then use the index gages to build a complete record, as described above. For PRMS, the nearest-neighbor procedure was augmented by restricting the calibration network to only gages where at-site calibration of PRMS exceeded a Nash-Sutcliffe model efficiency of 0.80. Model parameters were transferred to the validation site from the nearest index gage with an acceptably efficient at-site calibration of PRMS.

Map correlation is predicated on the assumption that, when transferring daily time series of streamflow from a gaged to an ungaged location, the correlation in the daily streamflow time series would provide an improved metric of similarity over the use of the nearest index streamgage. Archfield and Vogel (2010) show this to be the case for ungaged locations in the northeastern United States and developed a geostatistically based kriging method termed map correlation to select an index streamgage estimated to be most correlated with the ungaged location. This method has been subsequently applied to the States of Pennsylvania and Iowa (Stuckey and others, 2012; Linhart and others, 2013, respectively) and to the Connecticut River Basin (Archfield and others, 2013).

The map-correlation method (Archfield and Vogel, 2010) is an application of ordinary kriging that provides an estimate of the cross correlation in streamflow time series between any location in a region and an ungaged location and, when applied to a set of gaged catchments, can select the gaged catchment that has the highest estimated correlation in daily streamflow time series with the ungaged catchment. The map-correlation method has been successfully used with the drainage-area ratio method to improve estimates of daily streamflow time series in the northeastern United States (Archfield and Vogel, 2010).

For each gaged catchment in a study region, the map-correlation method is implemented through the following steps:

- Compute the Spearman's rho correlation between concurrent streamflow time series at the gaged catchment and all other gaged study catchments.
- Fit a continuous spatial statistical model, termed a variogram model, to the relation between the observed differences in correlation and distance between each pair of gaged catchments. The resulting variogram model can then be used to map the spatial distribution of correlation between the gaged catchment and any other location in the study area.
- Use the variogram model to estimate the correlation between the gaged catchment and the ungaged location.

When correlations are mapped for more than one gaged catchment, each map can be used to estimate the correlation at the ungaged location. By comparing multiple mapped estimates of correlation at the ungaged catchment, the map-correlation method is able to select the gaged catchment that corresponds to the highest estimated correlation value at the ungaged catchment. Thus, the map-correlation method is able to select the gaged catchment having streamflows estimated to be most correlated with the ungaged location.

Archfield and Vogel (2010) use the Pearson r correlation coefficient to estimate correlation between the logarithms of concurrent streamflow time series at the gaged catchments; however, this measure of correlation has its disadvantages. The Pearson r correlation coefficient assumes a linear relation between the correlated time series (Helsel and Hirsch, 2002). Archfield and Vogel (2010) obtained an approximately linear relation between streamflow time series by taking the logarithms of the streamflows before computing the Pearson r correlation coefficient. This approach is not feasible if there are zero values in the streamflow record and, furthermore, this approach does not fully ensure a linear relation between the streamflow time series. To avoid these concerns, the methods applied here use the rank-based Spearman rho correlation metric (Helsel and Hirsch, 2002) to estimate the correlation between streamflow time series. Just as with Pearson r correlation coefficient values, Spearman rho correlation values range from -1 to 1 , with extremes indicating a perfect negative or positive correlation, respectively.

For each site in the calibration network, variogram models were fit to the Spearman rho correlation values according to the methods explained in Archfield and Vogel (2010) and summarized here. The R statistical program (R Development Core Team, 2005) and the related geoR software package (Ribeiro Jr. and Diggle, 2001) were used to fit the variogram models to the observed correlations between a particular site in the calibration network and any other site within the calibration network. The variogram model was fit using weighted least squares from binned values of the observed correlation

across the calibration sites, as a function of inter-site distances, with the weights determined by the number of points used to estimate each bin (Ribeiro, Jr. and Diggle, 2001). Because of its relatively simple formulation, its visual agreement with the majority of the sample variograms, and the previous success of this model form by Archfield and Vogel (2010), the model fit to the calibration network was a spherical variogram model. The spherical variogram, (see Ribeiro, Jr. and Diggle 2001), has the form

$$\gamma(h) = \begin{cases} 1 - 1.5\frac{h}{a} + 0.5\left(\frac{h}{a}\right)^3, & \text{if } h < 0 \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where

- $\gamma(h)$ is the variogram model (also referred to as the correlation function),
- h is the separation distance, and
- a is the range parameter.

Following from traditional geostatistics techniques for ordinary kriging as presented in Isaaks and Srivastava (1989), when $\gamma(h)$ (eqn. 1) is multiplied by the partial sill, σ^2 , the covariance function, $C(h)$, is obtained (Ribeiro, Jr. and Diggle, 2001):

$$C(h) = \sigma^2 \gamma(h). \quad (2)$$

The parameters a and σ^2 were estimated for each of the variogram models, corresponding to each of the calibration sites, and when used in conjunction with equations 1 and 2, the covariance between two correlation values at any distance apart from one another can be estimated (Ribeiro, Jr. and Diggle, 2001). This application of the map-correlation method also estimated a nugget value, which is a constant value added to equation (1) when the variability in the data is such that $\gamma(h)$ does equal zero when h is zero Isaaks and Srivastava (1989). Fitting a model for each site in the calibration network allows for an approximation of the correlation between a validation site and each calibration site. The potential index gages can then be ranked according to the estimated correlation between streamflow records.

Table 3 presents the root-mean-squared error (RMSE) associated with the fit of each calibration variogram. Apart from the threefold validation, these results are from a leave-one-out validation and give an indication of how well these models are performing in the region. The RMSE is computed for each variogram model by first fitting the model to the entire dataset and then systematically leaving one of the gages out of the fitting procedure. Archfield and Vogel (2010) report much lower RMSE values than reported here; however, because the scope of this report is to evaluate existing methods to estimate daily streamflow, the reasons for this difference were not explored. In the streamflow predictions presented

below, the variogram models were fit following the threefold validation framework.

The map correlation techniques were used to select index gages for the statistical, transfer-based methods and AFINCH, but were not used to identify the parameters for PRMS. The statistical, transfer-based methods ranked potential index gages by the correlation between streamflow time series. AFINCH, for reasons explained below, ranked potential index gages by applying map correlation to the time series of ratios between daily and monthly streamflow.

Methods to Estimate Streamflow Moments at Ungaged Sites

Estimates of the streamflow mean and standard deviation on a monthly and annual basis are required for several of the statistical, transfer-based methods, including the standardizations with means and the standardizations with both means and standard deviations. The means and standard deviations must be calculated using untransformed and natural, logarithmically transformed streamflow data. For these applications, monthly and annual moments of untransformed and natural, logarithmically transformed streamflows were estimated for the ungaged validation sites via regional regression techniques documented and developed by Farmer and Vogel (2013). Depending on the implementations of the PUB methods, a different number of regressions were required. Standardizing with an annual mean requires a single regression, while standardizing with monthly means requires 12 regressions. Similarly, standardizing with an annual mean and an annual standard deviation requires two regressions, while standardizing with analogous monthly moments requires 24 regressions.

Monthly and annual means and standard deviations were estimated using a regression technique similar to the automated approach proposed by Farmer and Vogel (2013). A set of regional regressions for the prediction of these means was developed using only the calibration network and associated basin characteristics. A large set of basin characteristics was passed through a set of filters, described in some detail in appendix A of Farmer and Vogel (2013). In general terms, this process consisted of passing the calibration set of explanatory variables through a stepwise regression to identify the most significant variables. The remaining explanatory variables were passed through a loop of weighted multivariate regressions, removing the most insignificant variables at each step until all variables were significant according to a t test on the regression coefficient. The inverse lengths of known records were used to weight observations. This technique was used to estimate both annual and monthly moments. In the monthly case, these regressions were recursive, with a single lagged variable, allowing the mean from one month to be used as an explanatory variable in the subsequent month and tying the 12 monthly equations into a simultaneous solution of parameters—an adjustment on the iterative solutions of Farmer and Vogel (2013).

Methods to Estimate the Quantiles of Flow Duration Curves at Ungaged Sites

The PUB method of nonlinear spatial interpolation using flow duration curves involves the transfer of streamflow information from the index gage to the validation site via exceedance probabilities and flow duration curves (FDCs). This requires the FDCs at both the index and validation sites be available before implementing the nonlinear spatial interpolation using flow duration curves. Each FDC was represented by discrete percentiles of the streamflow distribution rather than a continuous function. The percentiles used here corresponded to 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. The percentiles of the FDCs at the calibration sites, which possess gaged records, were obtained via empirical estimators. The percentiles of the FDCs at the validation sites were estimated with regional multiple-linear regressions based on the calibration network.

For each calibration site, the value of each of the 27 percentiles of the streamflow distribution was computed by linearly interpolating between the observed daily discharge values and the nonexceedance probability, p_i . These probabilities were computed using the standard formula of the plotting position $p_i = (i - a)/(n + 1 - 2a)$, where p_i is the nonexceedance probability, i is the rank of the daily streamflow value (1 to n , smallest to largest), n is the number of observations, and a is a constant. In this case, setting the constant to $a = 0.375$, the Blom plotting position was used. The Blom plotting position gives unbiased percentiles for the normal distribution (Stedinger and others, 1993), which is usually a reasonable approximation for log-transformed daily streamflow. The Blom formula is also numerically similar to the Cunnane formula with $a = 0.4$, which is approximately quantile unbiased for a range of distributions (Cunnane, 1978). The dense set of 27 percentiles was chosen in order to minimize the need for extrapolation of extremes and gap over which values must be interpolated when applying the nonlinear spatial interpolation using flow duration curves. Because the minimum record length of calibration sites was set to 10 years of daily data, which means there were at least 3,650 observed values, the upper and lower percentiles were set to $p_{\min} = p(i = 1) = 0.0017$ percent and $p_{\max} = p(i = n) = 99.983$ percent.

Seeking to minimize the curvature of the FDCs, the values of the percentiles were logarithmically transformed with the common logarithm. A few of the calibration sites, as noted in table 1, contained zero flow values and produced zero-valued percentiles. The zeroes could not be logarithmically transformed. The smallest nonzero flow in the published records was 0.01 ft³/s, for which $\log_{10} 0.01 = -2$. As a placeholder, the published zeroes were marked as a value of -3 when logarithmically transformed. The logarithmically transformed value -2 was later used as a censoring level in regional regression.

The percentiles of the distribution of daily streamflow at the validation sites were estimated using a regional multiple-linear regression based on the calibration network. This regression provided estimates of the logarithmically transformed

percentile at each site as a function of at-site basin characteristics, following the form below:

$$\log_{10} Q_p(i) = b_p(0) + \sum_j b_p(j) B_p(i, j) + \varepsilon_p(i), \quad (3)$$

where

- $\log_{10} Q_p(i)$ is the base-10 logarithm of the p th quantile of streamflow at the i th station,
- $b_p(0)$ is the estimated intercept of the linear regression model of the p th quantile,
- $b_p(j)$ is the estimated coefficient of the j th ($j = 1, k$) basin characteristic in linear regression model of the p th quantile,
- $B_p(i, j)$ is the value of the j th basin characteristic for the p th quantile of the i th station, and $\varepsilon_p(i)$ is the residual (error) of estimation of $\log_{10} Q_p(i)$, where $\text{Var}[e_p(i)] = \sigma_p^2 \omega(i)$, where σ_p^2 is the ordinary least squares regression residual variance of the p th quantile (to be estimated) and $\omega(i)$ is the residual's variance weight of the i th station, which is assumed to be proportional to its record length, in particular, $\omega(i) = nN(i) / \sum_{i=1}^n N(i)$, where $N(i)$ is the number of years of record of the i th station.

If all the validation sites exhibited nonzero percentiles, the regression model could be solved as a weighted least-squares regression for each quantile. For example, Judge and others (1985, p. 420–421) show how to transform weighted least squares to an ordinary least-squares problem. However, as mentioned, some of the daily streamflow quantiles are published as zeroes, and so the method of estimation must take this into account. In particular, when zero-valued percentiles are present, least-squares methods cannot be used. Instead, the logarithmically transformed percentiles were estimated with “tobit” regression, as the maximum likelihood estimation method is often called in this situation (Greene, 1997, p. 962–967). Tobit regression was implemented using the `survreg()` function of the survival package of R (Therneau, 2013) and a censoring level of $\log_{10} 0.01 = -2$. By using this regression method, the logarithmically transformed values of some streamflow percentiles may be estimated as being less than -2 , corresponding to streamflow values less than 0.01 ft³/s. When no censored values are present, tobit regression reduces to least-squares regression.

For validation sites, the regression models were developed for three flow regimes and composed of one to six basin characteristics. The basin characteristics were comprised of subsets of the GAGES-II variables, with various transformations (for example, logarithmic or square-root transformations) that improved the normality of the regression residuals. The flow regimes included high- (exceedance probabilities less than 5 percent), medium- (exceedance probabilities between 5 and 95 percent), and low-flow events (exceedance probabilities

greater than 95 percent). For a particular candidate model in a particular flow regime, the same basin characteristics were used to predict each percentile, resulting in a unique set of coefficients for each percentile. Candidate models were ranked by combined goodness of fit across the flow regime. Selecting models by flow regime reduced the likelihood of non-monotonicity and noisiness in the estimated FDCs while reducing the number of basic characteristics required for prediction. Although not tested here, alternative strategies for reducing noise in estimated FDCs attempt to account for the cross correlations between quantiles; these include the regression-on-quantiles approach utilized by Archfield and others (2013) and seemingly unrelated regressions (Tasker, 1997). The analysis of PUB methods below applied the threefold validation procedure to percentile estimation, but the regional results of the percentile regression, using all sites as calibration data, are included in table 4.

Statistical, Transfer-Based Methods

Drainage-Area Ratio

The drainage-area ratio (DAR) method has been used for decades to predict daily streamflow at ungaged locations. Asquith and others (2006) provide a summary of the history, development, and application of the DAR method. The DAR method assumes that the streamflow per unit drainage area at the index gage and validation site are equal to one another. In the DAR method, daily streamflows on a given day, t , are predicted as

$$\widehat{Q}_{Y,t} = \frac{A_Y}{A_X} Q_{X,t} \tag{4}$$

and $\widehat{Q}_{Y,t}$ is the predicted streamflow on day t at the validation site, $Q_{X,t}$ is the measured streamflow on day t at the index gage, A_Y is the drainage area of the validation site, and A_X is the drainage area to the index gage.

Standardization with Mean

An alternative to standardizing streamflow by the drainage area is to standardize streamflows by the mean streamflow. This method can be applied with either an annual mean (SM1) or a set of 12 monthly means (SM12). Farmer and Vogel (2013) found that standardizing streamflows by monthly means produced accurate predictions of monthly flows. Especially when applied with monthly means, this method corrects for some of the seasonality observed in the application of the more-traditional drainage-area ratio.

The SM technique maps streamflow information from an index gage, X , to the validation site, Y , as

$$\widehat{Q}_{Y,t} = \frac{\mu_Y}{\mu_X} Q_{X,t}, \tag{5}$$

where

- $\widehat{Q}_{Y,t}$ is the predicted streamflow on day t at the validation site,
 - $Q_{X,t}$ is the measured streamflow on day t at the index gage,
 - μ_Y is the mean streamflow at the validation site, and
 - μ_X is the mean streamflow at the index gage.
- In practice, estimates of the mean at the ungaged site are required prior to the application of SM techniques. As in Farmer and Vogel (2013), regression techniques described above were used to apply these methods.

Standardization with Mean and Standard Deviation

A common method for the extension of streamflow records is the Maintenance of Variance Extension (MOVE) (Hirsch, 1979 and 1982). Hirsch (1979) noted the potential of a similar technique for prediction in ungaged basins, calling it a method of regional statistics. Farmer and Vogel (2013) considered several variations of this method for the prediction of monthly records and these same variations can be applied to daily streamflow records. Standardizing by mean and standard deviation (SMS) can be implemented using untransformed or normal, logarithmically transformed streamflow.

Standardizing untransformed streamflows, the information at a validation site can be related to an index gage as

$$\frac{Q_X - \mu_X}{\sigma_X} = \frac{Q_Y - \mu_Y}{\sigma_Y}, \tag{6}$$

where X and Y indicate two sites and μ and σ are the mean and standard deviation of the flows at the subscripted site. This standardization produces a new standardized variable with zero mean and unit variance, regardless of the probability distribution of the original flows. SMS techniques can also be used to standardize the logarithms of streamflows, in which case the above equation becomes

$$\frac{\ln(Q_X) - \mu_{\ln(Q_X)}}{\sigma_{\ln(Q_X)}} = \frac{\ln(Q_Y) - \mu_{\ln(Q_Y)}}{\sigma_{\ln(Q_Y)}} \tag{7}$$

with the same definitions as above. When relating an index gage to a validation site, either of these equations can be manipulated to solve for the ungaged streamflow. Furthermore, either method can be applied with an annual mean and an annual standard deviation of streamflow or a series of 12 monthly means and standard deviations, resulting in four closely related methods: standardizing streamflow with an annual mean and standard deviation (SMS1R), standardizing the natural logarithms of streamflow with an annual mean and

standard deviation, standardizing streamflows with monthly means and standard deviations (SMS12R), and standardizing the natural logarithms of streamflow with monthly means and standard deviations (SMS12L) (table 2). In practice, the means and standard deviations would need to be estimated; the estimation procedure used here is described above.

SMS techniques have been shown to be extremely powerful in practice. In a study of Virginia watersheds, Hirsch (1979) found the SMS class of methods, referred to as the method of regional statistics, to be “distinctly superior” to the drainage-area ratio technique, among others, for the prediction of monthly series. Farmer and Vogel (2013) found, in a study of unaltered basins across the United States, that SMS techniques outperformed a wide range of techniques when there was little uncertainty in the estimated moments—an idealized case. Of course, as one is required to estimate moments in truly ungaged basins, the degree of superiority is directly linked with the ability to parameterize this PUB method.

Nonlinear Spatial Interpolation Using Flow Duration Curves

The method nonlinear spatial interpolation using flow duration curves assumes that the exceedance probability of streamflow on a given day is identical between two hydrologically similar sites, between an index and validation site. As noted in Archfield and others (2013), with an index gage selected and estimated daily flow duration curve (FDC) at the validation site, a time series of daily streamflow for the simulation period can be constructed by use of the nonlinear spatial interpolation (Fennessey, 1994; Hughes and Smakhtin, 1996; Smakhtin, 1999; Mohamoud, 2008; Archfield and others 2010; Shu and Ouarda, 2012). For example, if the streamflow on October, 1, 1994, was at the 90-percent-exceedance probability at the index gage, then it is assumed that the streamflow on that day at the ungaged location also was at the 90-percent exceedance probability; this probability can then be converted to a streamflow value via the estimated FDC. Therefore, to implement the method of nonlinear spatial interpolation using flow duration curves, an FDC is required at both the index and validation site. The time series of exceedance probabilities is transferred from the index gage to the validation site without transformation.

In this application, the discretization of FDCs necessitated several additional steps for the application of nonlinear spatial interpolation using FDCs. At the index gage, the percentiles of the streamflow distribution are determined by ranking the observations. At the validation site, the streamflow percentiles are estimated with regional regression, as outlined above. The nature of nonlinear spatial interpolation with FDCs requires that streamflows be estimated between the discrete percentiles as well. A Gaussian interpolation technique was used here. The following details describe the additional steps required to implement the nonlinear spatial interpolation using FDCs.

FDCs are typically represented as streamflow values (Q) plotted against exceedance probabilities (p) on semi-log axes. In an attempt to better linearize the curve, especially towards the tails, the FDCs are represented here as the common logarithms of the 27 percentiles (q) plotted against the standard normal (Gaussian) quantiles associated with each exceedance probability (Z_p). The standard normal quantiles are computed as the quantiles of a normal distribution with a mean of zero and standard deviation of one, each with a given exceedance probability p . This approach was designed to address the “hooks” in the extreme tails noted by Archfield and others (2010) when plotting FDCs with log-linear axes.

Estimating the streamflow $Q_{Y,t}$ on day t at the validation site Y begins by observing the logarithm of the streamflow on the same day at the index gage X , where $q_{X,t} = \ln(Q_{X,t})$. More often than not, this value falls between 2 of the 27 observed quantiles on the empirical FDC associated with site X , $q_{X,lower}$ and $q_{X,upper}$. These two quantiles are associated with two normal quantiles, $Z_{p,lower}$ and $Z_{p,upper}$. The normal quantile associated with $q_{X,t}$ and time t , Z_p , can then be obtained by linearly interpolating between these two ordered pairs.

Having obtained Z_p , an estimate of $Q_{Y,t}$ is arrived at by conducting the same process in reverse, with the regression-estimated FDC at the validation site used in place of the index gage’s empirical FDC. On the regression-estimated FDC at the validation site, the same normal quantiles that surround Z_p , namely $Z_{p,lower}$ and $Z_{p,upper}$, are associated with estimated common logarithms of two quantiles, $\hat{q}_{Y,lower}$ and $\hat{q}_{Y,upper}$. The estimate $\widehat{Q}_{Y,t}$ is obtained by conducting a linearly interpolation of Z_t between these two points. The estimate of streamflow at the validation site on day t is then given as $\widehat{Q}_{Y,t} = 10^{\widehat{q}_{Y,t}}$.

Given that the highest and lowest quantiles considered here were at exceedance probabilities of 0.02 percent and 99.98 percent, at any particular site it would be expected that 0.04 percent of the days had streamflow values and probabilities beyond the highest or lowest available quantile. In such a case, it is impossible to conduct an interpolation to transfer information from one site to another. Rather than attempting a more-involved extrapolation technique, the drainage-area ratio was applied in substitution. This procedure of Gaussian interpolation worked well for this study region.

The method of nonlinear spatial interpolation using flow duration curves, known by several different names, is referred to here as QPPQ. Fennessey (1994) first coined this technique the QPPQ TRANSFORM. It has also been published by Smakhtin (1999), Mohamoud (2008), and Archfield and others (2010) under names including “nonlinear spatial interpolation technique” (Hughes and Smakhtin, 1996; Smakhtin, 1999) and “reshuffling procedure” (Mohamoud, 2008). Recently, Shu and Ouarda (2012) demonstrated the advantages of the nonlinear spatial interpolation using flow duration curves and it has become more and more popular.

Analysis of Flows in Networks of Channels Model, Temporally Disaggregated

The Analysis of Flows in Networks of Channels (AFINCH) model estimates monthly streamflow at ungaged locations (Holtzschlag, 2009). AFINCH uses observed temperature, precipitation and streamflow data, as well as land use information to estimate monthly streamflow at ungaged locations. It is a quasi-process-based method because although it is based on empirically derived statistical relationships between streamflow and climate and land use variables, the monthly streamflow estimates are driven by the climate variables as occurs in a rainfall-runoff model. Monthly precipitation and temperature were obtained from the Processing Routines in IDL for Spectroscopic Measurements (PRISM) dataset (Gibson and others, 2002) via the USGS Center for Integrated Data Analytics GeoDataPortal, while land use information was derived from the 1992 version of the National Land Cover Dataset (Vogelmann and others, 2001). AFINCH is designed to provide streamflow estimates based on the National Hydrography Dataset Plus (NHDPlus) framework (U.S. Environmental Protection Agency (EPA) and U.S. Geological Survey, 2010); version 1 of NHDPlus (http://www.horizon-systems.com/NHDPlus/NHDPlusV1_home.php) was used for this analysis. Using a calibration network, AFINCH begins by fitting monthly regressions (12 equations) to predict the runoff (“water yield”, or streamflow per unit area) as a function of basin characteristics. These equations then allow the prediction of water yield along each stream segment in NHDPlus. After predicting the streamflow according to the regression equations, AFINCH then adjusts the predictions to match streamflows at the gaged locations by applying the observed-to-predicted streamflow ratios to the upstream water yields.

This analysis applied the beta version of AFINCH version 2c (accessed on October 31, 2012). AFINCH version 2c allows the user to define regions for analysis by selecting eight-digit hydrologic unit code (HUC) watersheds within a given water resources region (two-digit HUC). Because the region of interest for this study includes parts of regions 03 and 06, and subregions were not defined for the other methods used here, the AFINCH analysis was done as two disparate regions, splitting the 182 sites by two-digit HUC. Therefore, for each of these two regions and each calendar month, regression coefficients relating candidate explanatory variables were determined from the observed monthly streamflow, and those coefficients that were significant were used to predict the streamflow at the ungaged locations in each region.

For this application, the monthly streamflow values predicted by AFINCH were temporally disaggregated to estimate daily values. Monthly streamflows at the validation site were temporally disaggregated by adjusting the simulated monthly by the contemporaneous ratio of daily-to-monthly streamflow from an index gage and a bias correction, that is,

$$Y_d(t) = cR(t)Y_m^S(t), \quad (8)$$

where

$Y_d(t)$	is the disaggregated daily average streamflow on day t at the validation site,
$Y_m^S(t)$	is the monthly average streamflow corresponding to the month of day t ,
$R(t)$	is the daily-to-monthly ratio at the index gage X , and c is a site-dependent constant that corrects for bias.

In this study, a 31-day centered running mean of monthly values was used to reduce temporal edge effects and is represented by the equation $Y_m^S(t) = \sum_{s=t-15}^{s=t+15} Y_m(s) / 31$, where $Y_m(s)$ is the predicted monthly flow associated with day t . The ratio of daily-to-monthly streamflow at an index gage, $R(t)$, was estimated using $X_d(t)/X_m(t) = R(t)$. The constant c is given as $c = \sum_t Y_m(t) / \sum_t (Y_m^S(t)R(t))$. The application and temporal disaggregation of the monthly AFINCH model was executed using the same threefold validation procedure identified previously. Additionally, the index gages, from which the streamflow ratios were obtained, were selected by either the nearest-neighboring index gage or the map-correlated index gage based on the correlation of ratios rather than streamflows.

Precipitation Runoff Modeling System

The Precipitation-Runoff Modeling System (PRMS) is a deterministic hydrologic modeling system (Leavesley and others, 1983; Markstrom and others, 2008). The primary uses of PRMS are to (1) simulate land-surface hydrologic processes, including evapotranspiration, runoff, infiltration, interflow, snowpack, and soil moisture, on the basis of distributed climate information (temperature, precipitation, and solar radiation); and (2) simulate water budgets at the watershed scale with temporal scales ranging from days to centuries. PRMS is useful because it can integrate with models used for natural-resource management or other scientific disciplines and provides a modular design that allows the user to select alternative hydrologic-process algorithms from either the standard module library or user-provided provisional modules.

PRMS models are developed by dividing a watershed into homogeneous areas called hydrologic response units (HRUs) and stream segments derived from the Geospatial Fabric for National Hydrologic Modeling (Viger and Bock, 2014), a dataset based on the spatial units of the NHDPlus v.1 dataset (U.S. Environmental Protection Agency and U.S. Geological Survey, 2010). Saturated zone, unsaturated zone, and surface-runoff flow components were computed for each HRU in response to precipitation, air temperature, and hydrologic processes defined by land-surface and subsurface characteristics (Leavesley and others, 1983) and were routed and accumulated to the watershed outlet using the modeled stream network. Spatial datasets for land cover type, imperviousness, canopy cover, soils information, and shallow permeability maps are used to parameterize the HRUs.

For each site in the calibration network, an automated parameter estimation procedure based on the Shuffled Complex Evolution algorithm (Duan and others, 1992, 1993, 1994) was used to calibrate the PRMS model. Parameters were calibrated in six steps: (1) solar radiation, (2) potential evapotranspiration, (3) water balance, (4) daily flows, (5) high flows, and (6) low flows. For each step, a subset of PRMS parameters was selected that was determined to be most influential in adjusting the simulations to match each calibration step. For steps 1 and 2, the objective functions targeted minimizing the absolute error between simulated and observed long-term mean-monthly solar radiation and potential evapotranspiration data, respectively. For steps 3 through 6, the objective functions targeted minimizing the root-mean-square-normalized error between simulated and observed streamflow at several time steps (annual, monthly, mean monthly, and daily).

As described previously, when the PRMS model was developed, initial parameter values were computed based on various spatial datasets of land cover type, imperviousness, canopy cover, soils information, and shallow permeability maps. During calibration, these initial parameters were allowed to vary across specified ranges as the calibration algorithm searched for an optimal parameter set; as there may be many parameter sets that provide acceptable estimates of streamflow due to the equifinality concept in mechanistic modeling described by Beven and Freer (2001). As the parameters vary across the study area (as opposed to being a constant value for all HRUs), the algorithm adjusts the mean of the distributed values across the HRUs for a particular parameter in the calibration, and then that adjustment to the mean value is applied back to the individual parameter values for each HRU. For ungaged locations, the relative changes made to parameters contained in the calibration watershed are applied to the ungaged watershed parameter values, as opposed to transferring the actual values from the calibration watershed. PRMS was calibrated in a one-to-many fashion using this transfer method. Each gage was used to inform the calibration of all remaining gages used in the study, resulting in a matrix of each gage being used to calibrate every other gage. Then, with a matrix of each gage calibrated by every other gage, the specific streamflow predictions that pertain to the methodology of threefold validation could be extracted from the calibration results to determine the performance of PRMS predictions calibrated by the optimization of flow predictions at adjacent sites.

Methods of Analysis

The performance of the PUB methods was assessed using several goodness-of-fit metrics, observing the reproducibility of a no-fail storage-yield curve, determining the accuracy with which particular streamflow statistics were reproduced and conducting a robust rank-based evaluation across metrics. The overall goodness of fit and reproducibility of the daily, no-fail

storage-yield curve were assessed with seven metrics, including the Nash-Sutcliffe of the untransformed (NSE) and natural, logarithmically transformed (NSEL) streamflow predictions, the root-mean-square error in streamflow predictions (RMSE), the root-mean-square-normalized error in streamflow predictions (RMSNE), the average percent error in streamflow estimates, the Pearson correlation coefficient between observed and predicted streamflow, and the Spearman correlation coefficient between observed and predicted streamflow. In addition, the reproducibility of several streamflow statistics was assessed with the percent error; these statistics included the coefficient of variation of annual streamflow, the coefficient of variation of daily streamflow, the 10th and 50th percentiles of the distribution of 7-day-average annual-minimum flows, the 90th percentile of the distribution of annual-maximum flows, and the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of annual streamflow. This analysis also considered the reproducibility of the seven fundamental daily streamflow statistics, consisting of the mean streamflow, coefficient of variation as an L-moment ratio (L-CV), skewness (L-skew), kurtosis (L-kurtosis) and autoregressive lag-one autocorrelation coefficient of daily streamflow, and the amplitude and the phase of the sinusoidal seasonal signal (Archfield and others, 2013). Details on the metrics noted here, as well as specific notes on the calculation of the no-fail storage-yield curve and the robust rank-based evaluation, are included below. The body of this report focuses on the NSE, NSEL, and average percent error, but full results and descriptions of the other individual goodness-of-fit metrics are included in appendix C.

The Nash-Sutcliffe efficiency was used to assess the accuracy of streamflow predictions and the daily, no-fail storage-yield curve. The NSE is a widely used metric for understanding overall goodness of fit, especially in hydrology (Nash and Sutcliffe, 1970). The NSE is typically estimated for a set of observed, O , and simulated, S , data values as

$$NSE = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \mu_O)^2}, \quad (9)$$

where

n is the number of observations and
 μ_O is the mean of the observations.

Careful inspection reveals that this can be understood as a measure of mean squared error standardized by the variance of the observations, and thus is a function of the RMSE. Upon further analysis, NSE is an amalgamation of bias in the estimate relative to the variability in the observations, correlation between observed and simulated flows, and the similarity of the variances of the observed and simulated records (Gupta and Kling, 2011; Gupta and others, 2009). The NSE ranges from a perfect-fit value of 1 to negative infinity; a value of zero indicates that the mean would provide just as reliable of

an estimate. Because differences between observed and simulated values are squared and an arithmetic mean is implied, NSE can be highly sensitive to outliers and skewed data.

When applied to heavily skewed data, the NSE can be somewhat misleading: Large absolute errors, typically associated with high-flow values, tend to have undue influence on the statistic when the distributions of flows exhibit distinct positive skew. In order to remove some skew from the data and avoid this problem, it is useful to consider the NSE of the natural logarithms of all nonzero flows (NSEL). While a useful transformation, this method is limited by its inability to handle zero values. Nonetheless, it provides an NSE that is affected more evenly by all nonzero flows. Of 182 sites considered, only 7 had observed zero-flow days; these days were removed from the analysis for this particular metric.

The average percent error measures the tendency of each PUB method to overpredict or underpredict nonzero daily streamflow estimates. This is a common measure of goodness of fit, as it gives an indication of the sign and magnitude of the average error. Like the NSEL, the average percent error cannot handle zero-valued flows; the seven sites with zero flows were not considered by this metric. The average percent error is closely related to the percent bias of an estimator, which measures the tendency of each PUB method to overpredict or underpredict streamflow, on average. By definition, the percent bias is merely the percent error in the mean value, which is discussed with the assessment of the fundamental daily streamflow statistics.

Estimates of Daily, No-Fail Storage-Yield Curves

The storage-yield curve (SYC) is a common tool for understanding the relation between basin storage and yield or volume of flow. These curves are often used to aid in the design of reservoir systems and the management of irrigation. The SYC represents the cumulative and sequential behavior of the streamflow record, accounting for the variability, timing, and magnitude of flows and can be thought of as a cumulative signature of the streamflow record.

The SYC is derived from a continuous record of streamflow. There are many techniques for estimating the SYC, depending on the intended reliability of yield and the level of acceptable risk. One common method is the Sequent-Peak Algorithm (Thomas and Burden, 1963). This is a no-fail technique, which means that the algorithm is designed such that enough water will always be available to meet the yield. The result is a curve that specifies the amount of storage required to produce a specific, constant yield. Storage-yield curves are useful in helping water-resources engineers and managers plan abstractions and design of reservoirs.

Several examples of the observed sequent-peak, daily, no-fail SYC are shown in figure 3. In order to compare curves across sites, the storage is represented as time—volume divided by mean streamflow—while the yield is expressed as a fraction of the mean streamflow. These graphs can be read two ways: By specifying a daily, no-fail yield, as a fraction of the mean streamflow, along the horizontal axis, one can determine

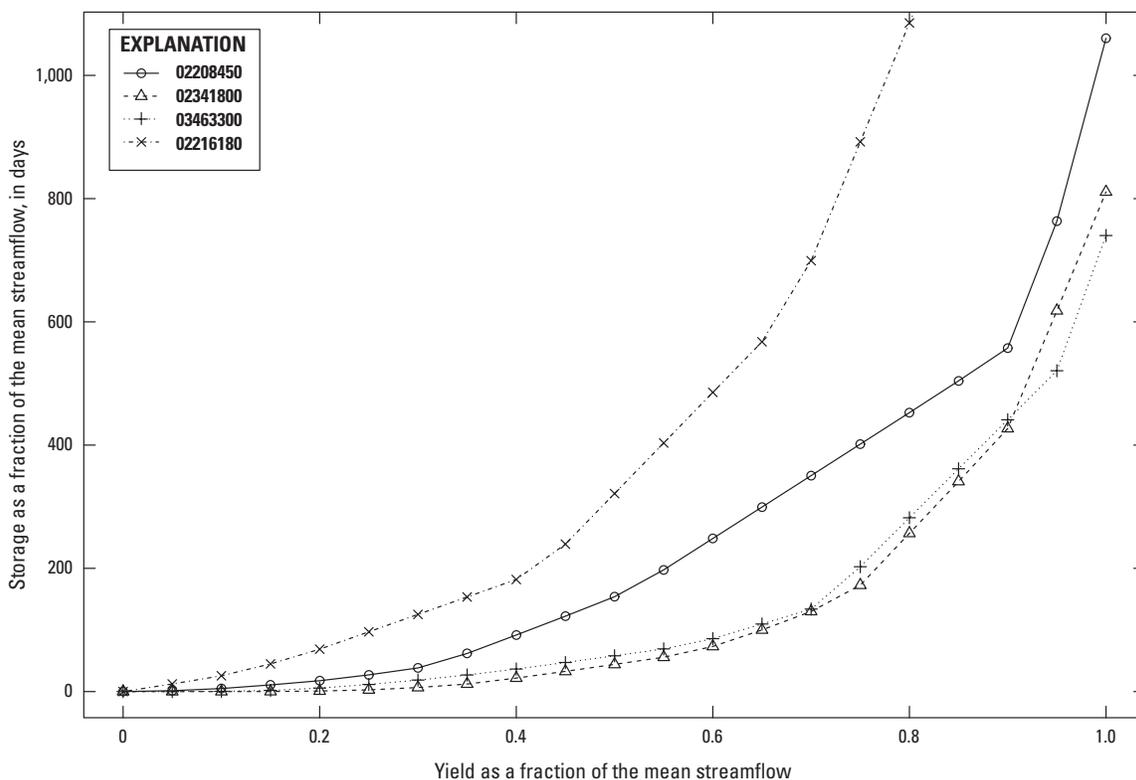


Figure 3. Several examples of daily, no-fail storage-yield curves for sites in the study area of the Southeast Model Comparison. The horizontal axis gives the daily yield, as a fraction of the mean streamflow. The vertical axis gives the required days of storage (volume divided by mean streamflow) required to meet the given yield. Refer to the text for a discussion of the methodology used to develop these curves.

the days of storage required to meet that demand. Or, given a specific storage capacity, one can determine the expected daily, no-fail yield. The daily variability of streamflow dictates how quickly the storage increases with yield.

The overall goodness-of-fit metrics used to assess streamflow, like the NSE and NSEL of streamflow, cannot capture the intangible, cumulative properties represented by the SYC. Therefore, understanding how reliably a PUB method can reproduce an SYC gives an important indication of the ability of that PUB method to reproduce higher order properties of the streamflow record. Here, the SYC was estimated from the observed and predicted records using the daily, no-fail, constant-yield sequent-peak algorithm with several yield fractions. The yield fractions considered ranged from 0 to 1 at a step of 0.05. The reproducibility of these curves was assessed by calculating the NSE, NSEL, root-mean-square error, root-mean-square-normalized error, and the Pearson and Spearman correlations between observed and predicted SYC. The NSE and average percent error of the SYCs are discussed in the body of this report; additional metrics are included in appendix C.

As the SYC is often used to answer design questions concerning the size of reservoirs and safe yields, the average error in the storage-yield curve can be interpreted in terms of the intended design. Consider the case of attempting to determine the size of a reservoir required for a particular yield. A positive error amounts to requiring more storage than is needed for a particular yield fraction. On the other hand, a negative error suggests that one would underestimate the storage needed, causing the effective yield to be less than desired. This is a tradeoff between overdesign and underdesign. Overdesign can be costly because of a waste of materials, while underdesign can be costly in that the design goals are not satisfied. In general, a slight overdesign might be more acceptable than underdesign, but a full cost-benefit analysis would be required for any particular project.

Robust Rank-Based Evaluation

Robust Rank-Based Evaluation (RRBE) is presented herein as a tool for comparing the performance of several PUB methods simultaneously. A largely graphical method, RRBE seeks to find an optimal balance with high average performance and low variability of performance for any particular method. In RRBE, each PUB method is represented as a point defined by the average and standard deviation of a performance metric. (The average and standard deviation are used here, but any measures of central tendency and spread may be acceptable.) When these points are created for each PUB method, the result is a RRBE cloud. The optimal PUB method, which will perform well on average (high NSE or low RMSE) and have minimal variability of performance, will appear near an edge of the cloud. The optimal edge depends on the performance metric, for example, ideal NSEs are near one, while ideal RMSEs are near zero. Methods appearing on the optimal edge

of the cloud are said to exist along the optimal or efficient frontier. In many cases, several methods may exist at or near the frontier, representing a tradeoff between central tendency and spread; one method may offer a better average performance but possess greater variability of performance.

RRBE can be used to assess the performance of several PUB methods across a single performance metric, or it may be used to compare several PUB methods across several performance metrics. When working with different performance metrics, it is often difficult to draw general conclusions: each metric highlights a different aspect of performance and is affected by its own idiosyncrasies, making a cross-metric comparison problematic. In order to draw general conclusions, it must be possible to compare these metrics in commensurate units. Such a comparison would show which methods are performing best across all the performance metrics, on average. This can be achieved by considering the relative ranks of the PUB methods at each site, according to any performance metric, rather than the values of the performance metric. In this manner, the performance is relative and all metrics are on the same relative scale. Thus, the RRBE cloud shows the average relative performance against the variability of relative performance. By averaging the relative performances across metrics, multiple performance metrics can be assessed simultaneously.

Here, RRBE is conducted by selecting each metric in turn, starting with the NSE, for example. For each site, taken in turn, the 19 PUB methods are ranked from 1 to 19. The best performing method (greatest NSE, in this case) is given a rank of 1; the second best is given a rank of 2; and so on. Ties are given the best ranking available. One could apply a RRBE here by graphing each PUB method as a point in Cartesian coordinates, where one axis is the mean of the ranks and the other axis is the standard deviation of the ranks. This would give the efficient frontier according to a single metric, information that can be inferred from the boxplots presented in this report. Instead, the mean and standard deviation of the relative ranks were taken for each PUB method and each performance metric. Multiple performance metrics were assessed by taking the average relative ranking across performance metrics and the average standard deviation across performance metrics, thus creating a cross-metric RRBE.

Results and Discussion

Daily streamflow was predicted for water years 1981 through 2010 for each of the 182 sites using each of the 19 PUB methods and the threefold validation procedure. The goodness of fit of these predictions was assessed at each site by comparing the portion of the observed record with complete water years against the corresponding predictions. The daily predictions for each PUB method were evaluated using three comparisons: (1) the overall goodness of fit of the streamflow records; (2) the accuracy of the reproduced storage-yield curve; and (3) the accuracy of reestimated

streamflow statistics, including the FDC and fundamental daily streamflow statistics. A final analysis combined each of these comparisons to evaluate the overall performance of each PUB method.

Overall Goodness of Fit of Predicted Streamflow Records

Daily streamflow estimates were assessed for goodness of fit using Nash-Sutcliffe efficiencies and the average percent bias to give a coarse measure of how well each method performs. NSE was calculated for untransformed and normal, logarithmically transformed streamflow predictions. Results for root-mean-squared error (RMSE), root-mean-square-normalized error (RMSNE), Pearson correlation, and Spearman correlation are presented in appendix C, but are not part of the discussion below. All seven metrics, however, are used for a cross-metric RRBE. These measures quantify how well daily streamflow is estimated, on average. None of these measures

explicitly capture the sequencing or the serial correlation of errors. These metrics summarize the average error of daily flows and do not quantify the error associated with any particular flow statistic or distribution.

NSE was estimated for each of the 182 sites for each of the 19 PUB methods; the distribution for each PUB method can be seen in figure 4 (outliers were not plotted because their inclusion inflates the scale of the axis so as to obscure the characteristics of the distribution). Based on NSE, the accuracy of PUB methods varies widely. Notably, PUB methods using the nearest-neighbor approach to index selection fared better than those using map correlation. PRMS had the greatest median (0.5615), though NN-AFINCH, NN-QPPQ, and NN-SMS12R were nearly as high. NN-DAR, NN-SMS12L, and NN-SM12 were only slightly inferior. PRMS also had the smallest interquartile range. Annualized methods (SMS1R, SMS1L, SM1) were originally included because they performed well on a monthly time step (Farmer and Vogel, 2013), but they are not competitive here. Measuring by the NSE, PRMS is the best PUB method at 30 percent of sites and

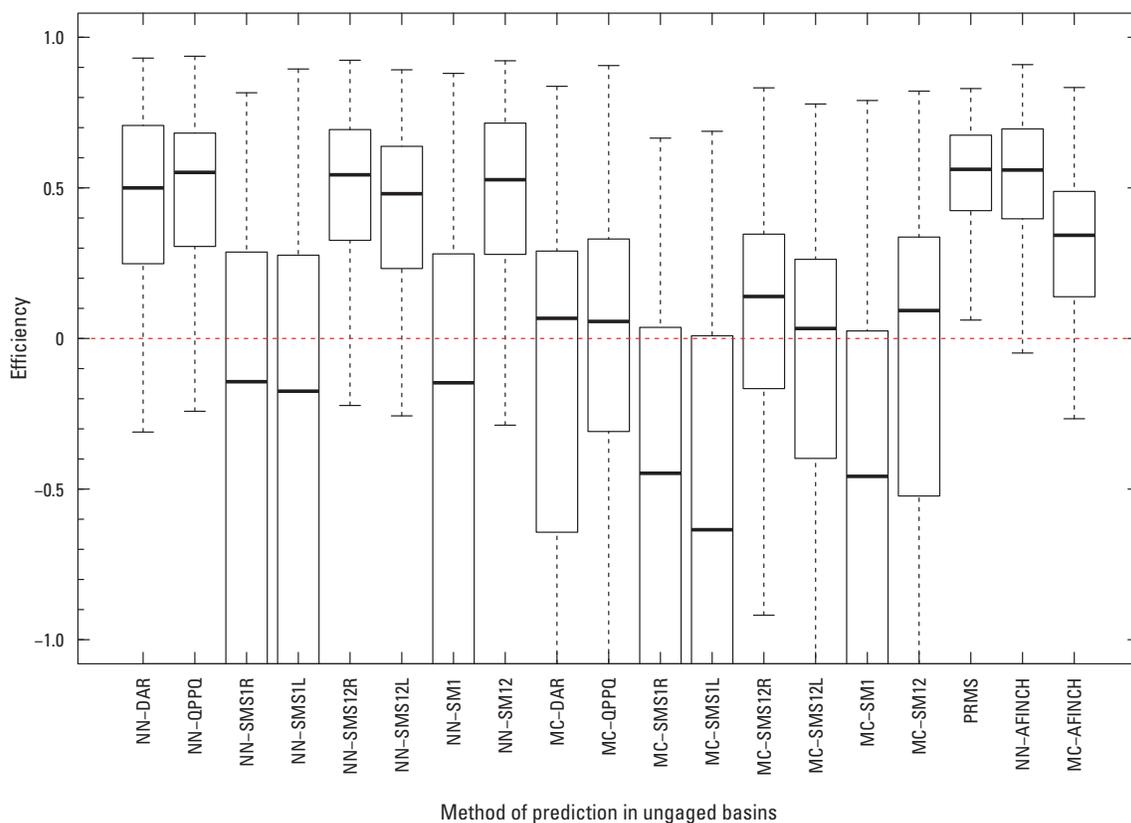


Figure 4. The distribution of Nash-Sutcliffe efficiencies of daily streamflow predictions for each method of prediction in ungaged basins (PUB) considered in the Southeast Model Comparison. The horizontal axis indicates each PUB method. The Nash-Sutcliffe efficiency is along the vertical axis. Nash-Sutcliffe efficiency ranges from one to negative infinity; a value of one indicates a perfect fit, while a value of zero indicates that a mean value would have produced the same level of accuracy. (The dark line indicates the median of the distribution, the box outlines the 25th and 75th percentiles, and the whiskers extend to the data point a distance not more than 1.5 times the interquartile range away from the nearest quartile. Data points beyond this whisker length are defined as outliers and, to increase the visibility of the distribution, have not been drawn.) Method abbreviations are in table 1.

outperforms each other method, taken one by one, at more than 50 percent of sites.

Compared to the values of NSE, taking the NSE of normal, logarithmically transformed streamflows (NSEL) suggests that most of the PUB methods perform better when the range of streamflow values is weighted more evenly (fig. 5; outliers not drawn). However, some of the worse-performing methods by the NSE criterion have even lower NSEL values. Still, the general comparison among methods is quite similar: The nearest-neighbor methods have generally greater NSEs than the map-correlation methods. The same seven methods (NN-DAR, NN-QPPQ, NN-SMS12R, NN-SMS12L, NN-SM12, PRMS, and NN-AFINCH) have the highest distributions of NSEL.

Interestingly, NSEL demonstrates poorer performance of PRMS than was previously indicated by the NSE. This could be a result of the calibration procedure, which minimized RMSE between observed and simulated streamflow in linear space, where high flows tend to dominate the Nash-Sutcliffe

efficiency metric. When the simulated streamflows are analyzed in log space (fig. 5), the low flows have more influence on the Nash-Sutcliffe efficiency. This result indicates that this calibration of PRMS is weighted more towards matching the high flows than the low flows. Perhaps future calibrations that focus on matching logarithmically transformed streamflows would improve the fit across all flow magnitudes. Regardless, this suggests that fully calibrated models may be adversely affected by their calibration criteria.

Balancing the performance of PUB methods in terms of both NSE and NSEL begins to highlight the need for a more comprehensive analysis. Looking at the PUB methods through several different lenses provides an understanding of how a method performs under different circumstances and in different ways. In the case of PRMS, the physical constraints imposed by the model structure may reduce the median performance while restricting the range of outcomes. For NSEL, NN-AFINCH, followed closely by PRMS, demonstrated the smallest interquartile range, but NN-SM12 had the greatest

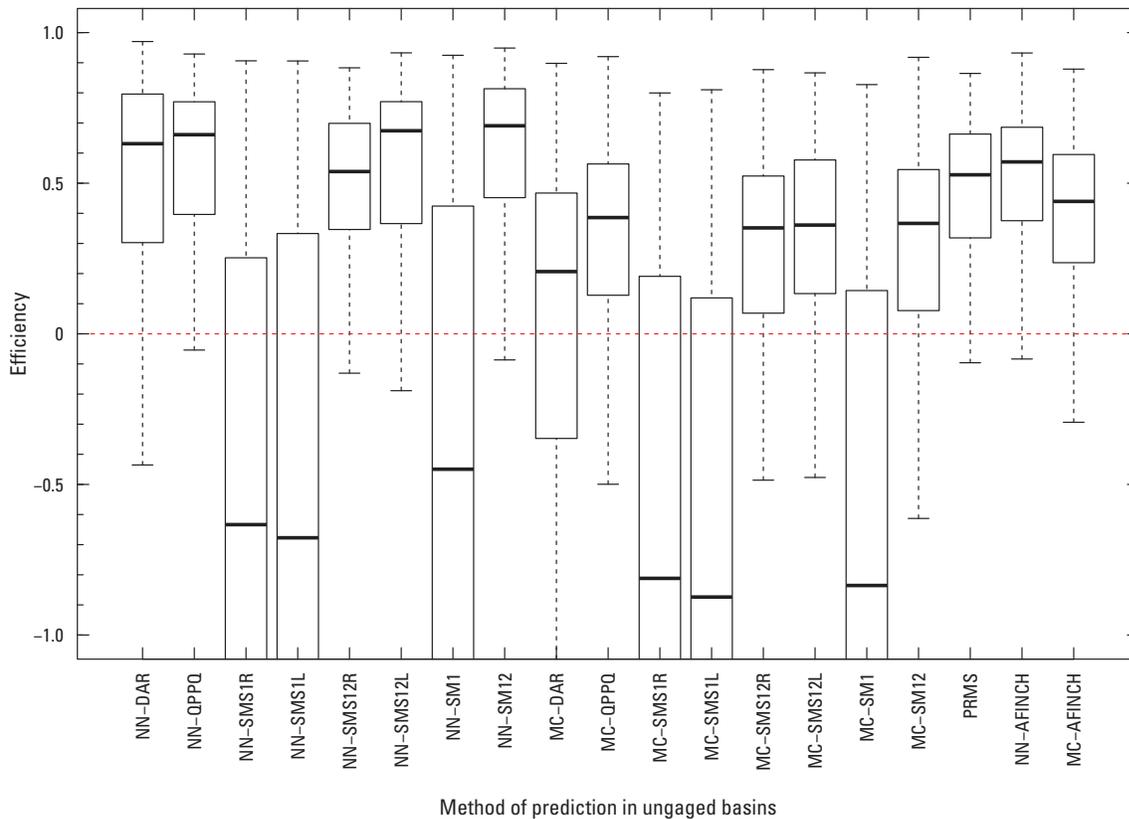


Figure 5. The distribution of the at-site Nash-Sutcliffe efficiencies of the logarithms of daily streamflow predictions for each method of prediction in unengaged basins (PUB) is considered here. The horizontal axis indicates each PUB method. The Nash-Sutcliffe efficiency of the logarithms is along the vertical axis. Nash-Sutcliffe efficiency ranges from one to negative infinity; a value of one indicates a perfect fit, while a value of zero indicates that a mean value would have produced the same level of accuracy. (The dark line indicates the median of the distribution, the box outlines the 25th and 75th percentiles, and the whiskers extend to the data point a distance not more than 1.5 times the interquartile range away from the nearest quartile. Data points beyond this whisker length are defined as outliers and, to increase the visibility of the distribution, have not been drawn.) Method abbreviations are in table 1.

median. Though not drawn here, graphing the median value versus the spread (IQR) of the distribution shows a distinct tradeoff between metric centrality and spread. Improvement in the median NSEL often results in a larger IQR and vice versa. It is noteworthy, therefore, that the NN-DAR continues to be competitive with the more complicated methods, suggesting it may be particularly robust across performance metrics.

The average percent error measures the tendency of each PUB method to overpredict or underpredict nonzero daily streamflow estimates. (Zero-flow days were excluded from the analysis.) This is a common measure of goodness of fit, as it gives an indication of the sign and magnitude of the average error. In all cases, there is a wide variability of the average percent error (fig. 6). From here on out, only the seven PUB methods identified with better performance based on NSE are included, while figures including all PUB methods are included in appendix C. The transfer-based methods (NN-DAR, NN-QPPQ, NN-SMS12R, and NN-SMS12L) show the smallest median error (10–20 percent), though they still present an overprediction. NN-SMS12R offers the smallest range

and median. PRMS has a median error of 50 percent, while NN-AFINCH had a median error of 27 percent. The average percent error is closely related to the percent bias of an estimator, which measures the tendency of each PUB method to overpredict or underpredict streamflow, on average. By definition, the percent bias is merely the percent error in the mean value, which is discussed below in reference to the fundamental daily streamflow statistics. For comparison, the range of percent bias can be seen in appendix C, figure C25.

In addition to the three metrics discussed above, several other metrics were used to characterize the overall goodness of fit of the PUB methods. These include the RMSE (which is closely related to NSE), the RMSNE error, Pearson correlation between observed and simulated streamflows, and Spearman correlation of the same. Full results can be found in appendix C. The general conclusions are identical to those presented here. PRMS is most accurate in terms of Pearson correlation, with regard to both median and range. NN-AFINCH is most accurate in terms of RMSE. The five top transfer-based methods are most accurate in terms of Spearman

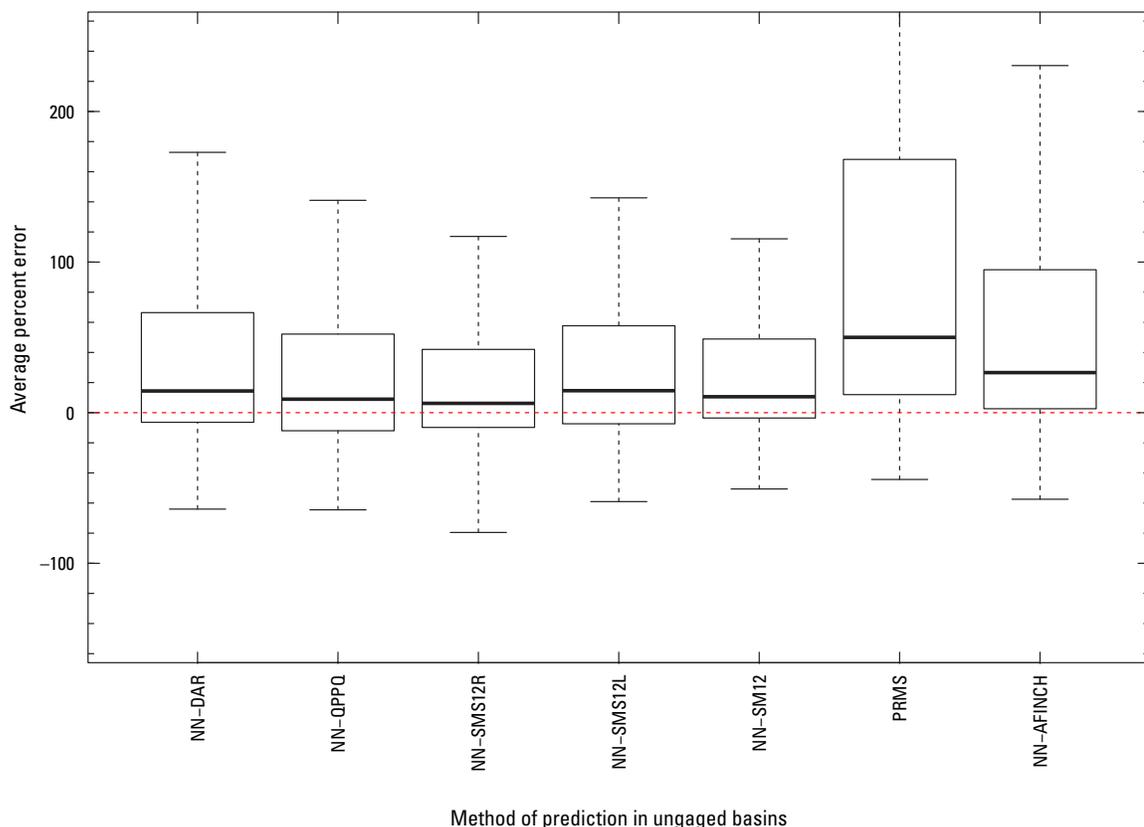


Figure 6. The distribution of at-site average percent errors of daily streamflow predictions for the seven most accurate methods of prediction in ungaged basins (PUB). The horizontal axis indicates each PUB method. The vertical axis shows the average percent bias. Unbiased methods display a median near zero and minimum the variability of at-site bias. (The dark line indicates the median of the distribution, the box outlines the 25th and 75th percentiles, and the whiskers extend to the data point a distance not more than 1.5 times the interquartile range away from the nearest quartile. Data points beyond this whisker length are defined as outliers and, to increase the visibility of the distribution, have not been drawn.) Method abbreviations are in table 1.

correlation and root-mean-square-normalized error. While not discussed in detail, these metrics did play a role in the RRBE discussed below.

In total, seven metrics were used to characterize the overall goodness of fit of each PUB method. The information contained in these metrics was combined by conducting a cross-metric RRBE. For each metric, the PUB methods were ranked at each site. The average and standard deviation of these ranks for each PUB method are provided in tables 5 and 6. The cross-metric RRBE cloud of the overall goodness of fit is composed of the average ranking across the seven metrics and the average standard deviation across metrics (fig. 7). (An example of the efficient edge of the cloud is drawn in fig. 7.) The frontier presented here consists of NN-QPPQ, NN-SM12, and NN-SMS12R. NN-QPPQ and NN-SM12 have more variability, while NN-SMS12R has less variability but a better ranking across these seven metrics. Of the seven best methods, all except NN-DAR and PRMS hover near the frontier of efficiency.

From the metrics used here to characterize overall goodness of fit, it is nearly impossible to select a single PUB method that is best PUB method for this region. NN-SM12

performed well across all metrics, but a strong case can be made for QPPQ as well as SMS12R. Still, some general conclusions can be drawn: Many of the PUB methods that worked well for monthly predictions (Farmer and Vogel, 2013) do not perform well in the context of daily flows. The nearest-neighbor algorithm for selecting an index gage outperforms the map-correlation algorithm, generally.

Regarding the overall goodness of fit of the predicted streamflow time series, the five most accurate methods are NN-DAR, NN-QPPQ, NN-SMS12R, NN-SMS12L, and NN-SM12. Methods that relied on more standardization (SMS12R and L) were not necessarily better than those requiring less (SM12). This is particularly true for SMS12R, which does not account for the skewness of the observations. In addition, two of the more process-based methods, NN-AFINCH and PRMS, also performed well. When considering the RRBE cloud, all of these seven methods, except NN-DAR and PRMS, clustered near the efficient edge. Finally, it is noteworthy that the parsimonious, even simplistic, methods of standardizing by drainage area or mean monthly flow perform so well. This suggests that the additional investment required for some PUB methods provides relatively little improvement

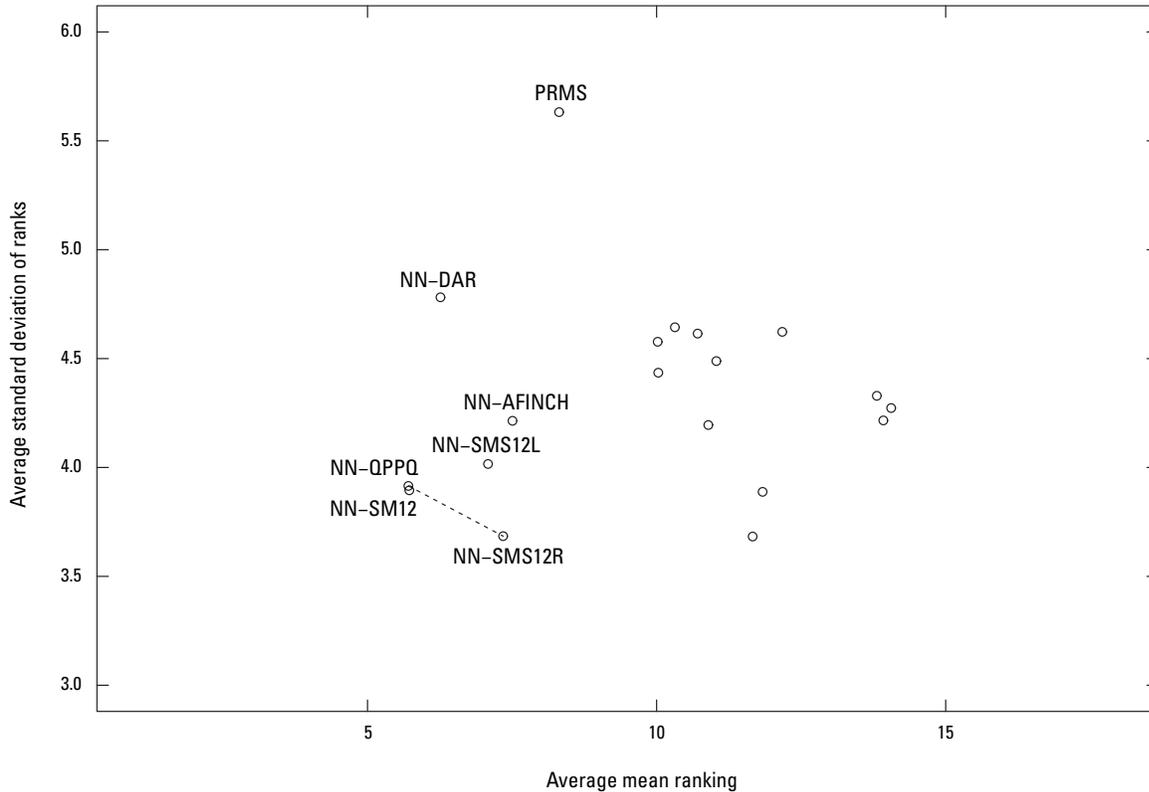


Figure 7. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the seven overall goodness-of-fit metrics for each method of prediction in ungaged basins. The horizontal axis shows the mean average rank, while the vertical axis gives the average standard deviation of the ranks. Optimal methods would display minimal spread and a low mean-average ranking. The dashed line notes the optimal edge of the cloud. The seven most competitive methods have been labeled. Method abbreviations are in table 1.

over DAR. NN-SM12, which requires the only slightly more investment than NN-DAR, provided the largest return on investment.

Daily, No-Fail Storage-Yield Curves

The reproducibility of the SYC indicates how well each PUB method captures the cumulative, sequential behavior of the daily streamflow record. For each PUB method, the accuracy of the estimated daily, no-fail SYC was assessed using the average percent error and the NSE of the estimated storage values. Though not discussed here, the RMSE, RMSNE, NSEL, and Pearson and Spearman correlations between observed and estimated storages are presented in appendix C. All seven metrics are used in the cross-metric RRBE of the SYC.

Because the storage-yield curves considered here are derived from a uniform distribution of yield fractions from

zero to one and SYCs are monotonic, the points along the SYC are unlikely to exhibit strongly skewed behavior. This means that the Nash-Sutcliffe efficiency can be directly applied to the estimated storages rather than the logarithms of the storage. The NSE of the SYC therefore gives a reliable estimate of the goodness of fit of the estimated storage yield curve.

According to the NSE, all of the methods produce the SYC reasonably well, with the majority of values falling above 0.4 (fig. 8). The variability in performance is greater for NN-AFINCH, PRMS, and NN-SMS12R than for the remaining four methods. For NN-DAR, NN-QPPQ, NN-SMS12L, and NN-SM12, more than 75 percent of the SYCs had NSE values greater than 0.8, a common level of “acceptable” performance when applied to the estimation and interpretation of daily and monthly streamflow (Moriasi and others, 2007).

The average percent error of the storage-yield curve indicates how much, on average, the PUB method overpredicts or

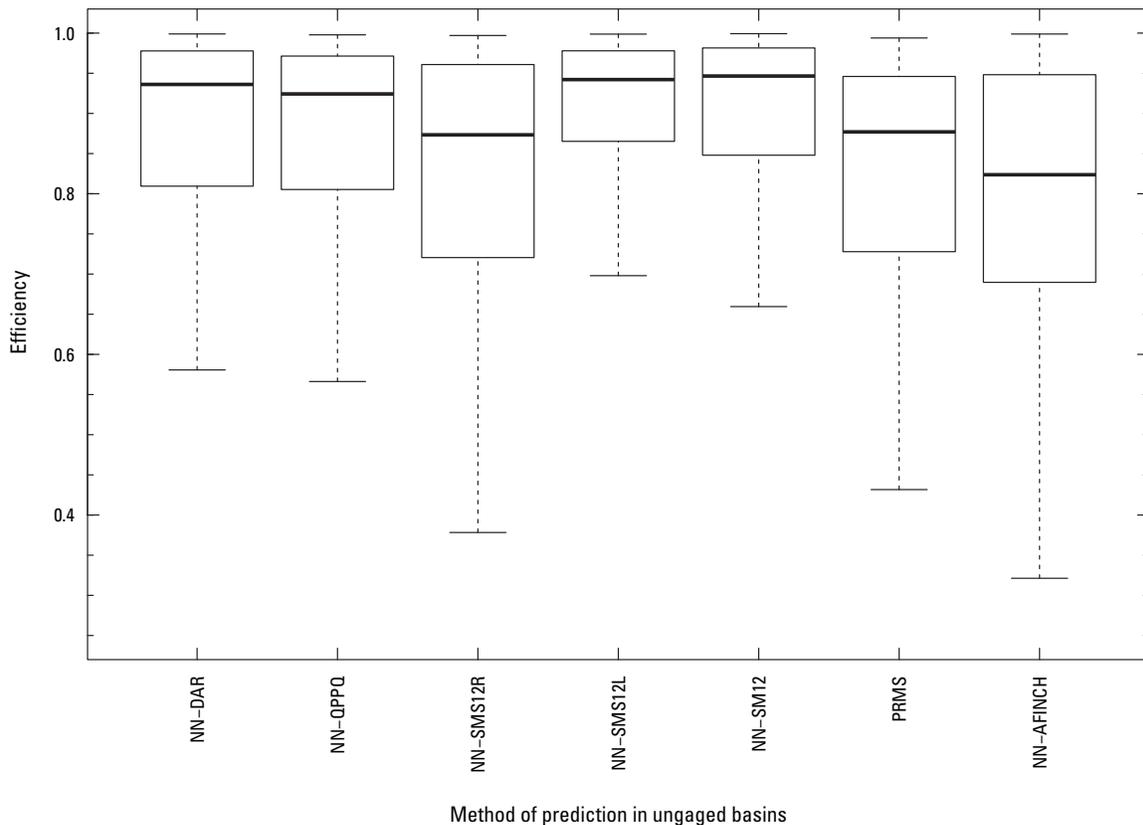


Figure 8. The distribution of the at-site Nash-Sutcliffe efficiencies of the daily storage-yield curve for the seven most accurate methods of prediction in ungaged basins (PUB). See text for a description of the methodology used to predict the storage-yield curve. The horizontal axis indicates each PUB method. The Nash-Sutcliffe efficiency is along the vertical axis. Nash-Sutcliffe ranges from one to negative infinity; a value of one indicates a perfect fit, while a value of zero indicates that a mean value would have produced the same level of accuracy. (The dark line indicates the median of the distribution, the box outlines the 25th and 75th percentiles, and the whiskers extend to the data point a distance not more than 1.5 times the interquartile range away from the nearest quartile. Data points beyond this whisker length are defined as outliers and, to increase the visibility of the distribution, have not been drawn.) Method abbreviations are in table 1.

underestimates the storage required for each yield fraction. All of the methods produce a significant error on average (fig. 9). Several of the transfer-based approaches (NN-DAR, NN-QPPQ, NN-SMS12L and NN-SM12) produce a relatively low level of error in the SYC storages on average, with medians ranging between 5 percent and 20 percent. This is the same group that clustered close to the efficient frontier when considering overall goodness of fit. NN-SMS12R, though, produces a much more variable average error, with a median of nearly 120 percent. The more process-based methods underestimate storage, while reducing the variability of the average error. PRMS underestimates the storage by 36 percent on average, while NN-AFINCH underestimates the required storage by a median of 14 percent. Underestimating the storage required for a reservoir system by more than 15 percent may be more costly than overestimating by 10 percent.

The cross-metric RRBE of the SYC shows that NN-SMS12L and NN-SM12 reproduce the SYC most accurately (fig. 10). Each point in the RRBE cloud shows the average ranking and average standard deviation across the seven

metrics used to assess the reproducibility of the SYC (see tables 5 and 6). NN-SMS12L and MC-SMS12L (unlabeled) define the optimal edge of the RRBE cloud; NN-SM12 is also near the frontier. NN-SMS12R, NN-AFINCH, and PRMS all are relatively far from the frontier. Though not too dissimilar to the result of overall goodness of fit, this analysis shows that the ability to reproduce the SYC varies across methods; some methods that performed poorly with overall goodness of fit are capable of reproducing the SYC accurately.

Assessing the ability of each PUB method to reproduce the storage-yield curve helps to evaluate how well each PUB method reproduces the cumulative behavior of the streamflow record. Most of the PUB methods considered here reproduce the SYC similarly. NN-DAR, NN-QPPQ, NN-SMS12L, and SM12 all overestimate the SYC. PRMS and NN-AFINCH produce significant underestimates of the storage required for a specific yield. NN-SMS12R does not perform nearly as well as the other transfer-based methods. It is again noteworthy that simple methods, like NN-DAR and NN-SM12, perform well when compared against substantially more complex models.

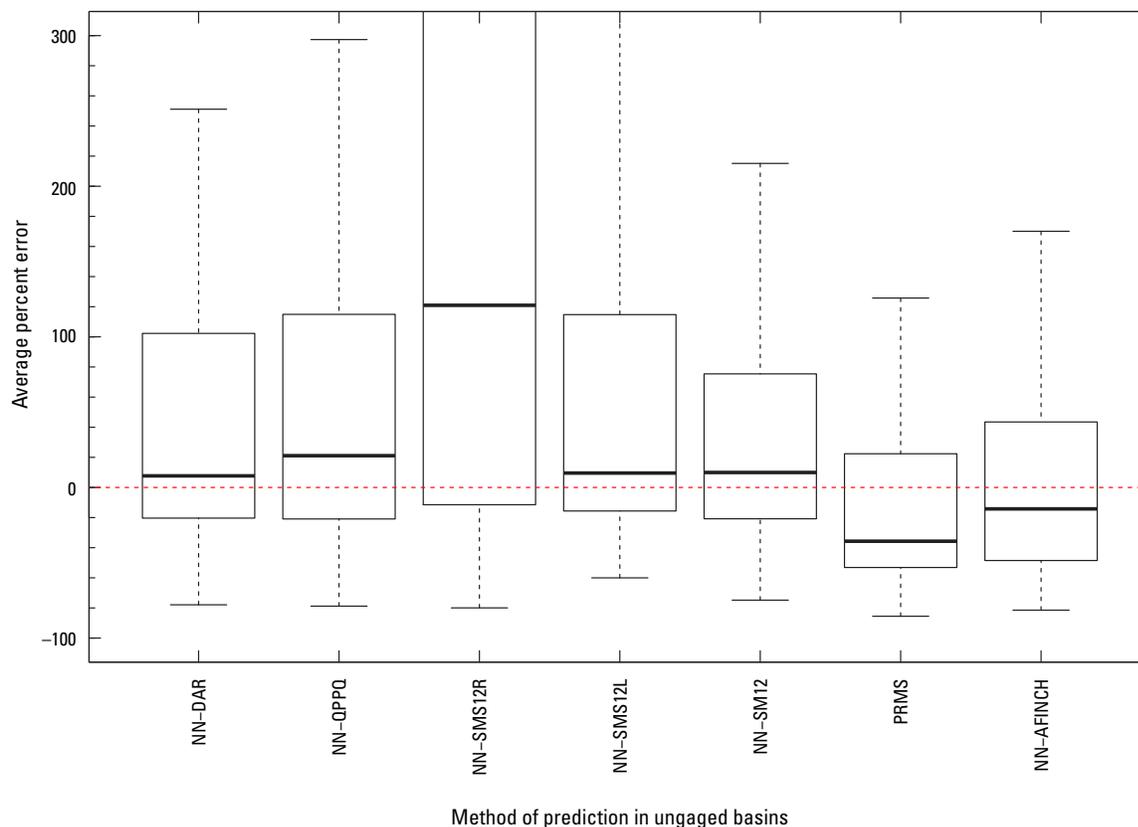


Figure 9. The distribution of at-site average percent errors of the estimated daily storage-yield curve for the seven most accurate methods of prediction in unengaged basins (PUB). See text for a description of the methodology used to predict the storage-yield curve. The horizontal axis indicates each PUB method. The vertical axis shows the average percent bias. Unbiased methods display a median near zero and minimum variability of at-site bias. (The dark line indicates the median of the distribution, the box outlines the 25th and 75th percentiles, and the whiskers extend to the data point a distance not more than 1.5 times the interquartile range away from the nearest quartile. Data points beyond this whisker length are defined as outliers and, to increase the visibility of the distribution, have not been drawn.) Method abbreviations are in table 1.

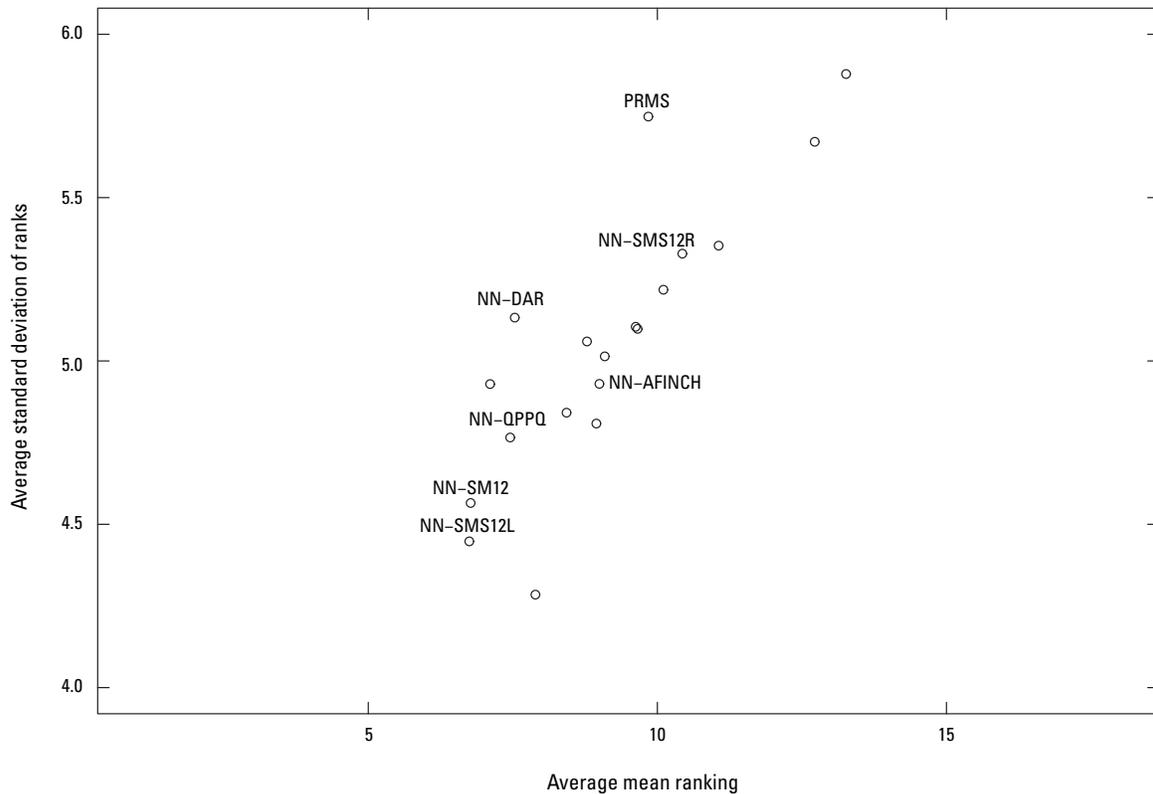


Figure 10. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the seven metrics of the storage-yield curve for each method of prediction in ungaged basins. (See text for a description of the methodology used to predict the storage-yield curve.) The horizontal axis shows the mean-average rank, while the vertical axis gives the average standard deviation of the ranks. Optimal methods would display minimal spread and a low mean-average ranking. The seven most accurate methods have been labeled. Method abbreviations are in table 1.

Even standardizing flows by an annual mean is promising, if one desires only to reproduce the SYC and cares little for producing a reliable estimated record. On average, MC-SM12, NN-SM12, and NN-SMS12L provide the greatest relative performance with the smallest variability.

Streamflow Statistics

Instead of predicting continuous daily streamflow records, a large amount of PUB exploration has focused on the prediction of specific streamflow statistics (Hrachowitz and others, 2013; Ries III, 2007; Thomas and Benson, 1970). Often, a designer or manager is concerned only with a specific streamflow parameter: be it the flow duration curve, an ecological flow statistic, a measure of base flow, or something else altogether. Because of the importance of these statistics, the best PUB methods for daily record reconstruction should also yield reliable estimates of many such statistics. This analysis considers a handful of flow statistics, namely the flow duration curve and the fundamental daily streamflow statistics (Archfield and others, 2013). Additional statistics were also considered in appendix C. The number of statistics that could

be assessed is endless; this analysis considers only a few to augment the overall goodness-of-fit statistics. Future research should expand on these comparisons.

Flow Duration Curves

The daily flow duration curve (FDC) is a representation of the distribution of daily streamflows at a given site. It is identical to the cumulative distribution function, with the flow on the vertical axis corresponding to a particular probability of exceedance along the horizontal axis. An FDC is used to understand the average flows and extreme events in a basin. They can be used to determine design events, which dictate the design parameters of certain water-management structures, including reservoirs, ecological flow standards, and hydro-power installations (Vogel and Fennessey, 1995).

Because of the importance of FDCs, optimal PUB methods should produce reliable approximations of the flow duration curve. Here, the FDC was approximated by reestimating specific percentiles from observed and predicted streamflow records. Five percentiles were considered: 10 percent, 25 percent, 50 percent, 75 percent, and 90 percent. The

observed and estimated percentiles were then compared, giving a general understanding of how well the flow duration curve is reproduced. The median percent error of each quantile for each of the seven top PUB methods is shown in figure 11. The right-hand side of the horizontal axis represents low-flow events.

PRMS estimates the 10-percent-exceedance flow fairly accurately, but the magnitude of the median bias increases towards low flow. This result is consistent with the calibration procedure, which matched simulated and observed streamflows in linear space, putting greater weight on the higher flows. At the 90-percent-exceedance level, PRMS produces a median error more than 50 percent greater than the observed events. This could be because the regional calibration of PRMS is limited by the allowable parameter calibration range. The calibration procedure, for a particular parameter, considers the range of initial values across a large spatial region. For particular regions or sites, this initial value is allowed to vary only within a user-specified window. Due to the constraint of considering a large spatial extent, the calibration algorithm may have not been able to find the true optimal parameter set for any particular gage. Smaller region of calibration may improve performance. Alternatively, model parameters could be refined individually by methods other than autocalibration. Weighting the autocalibration more evenly across all flows

(perhaps by optimizing model fit in log space) may yield a substantial improvement in performance as well.

The statistical, transfer-based methods and the quasi-process-based methods produce a much smaller level of bias than the process-based method. The transfer-based methods are most unbiased, with median biases of ± 5 percent. NN-AFINCH produces a positive error (median: ~ 10 percent) in the lower end of the flow regime. Still, the SMS methods begin to show a slight negative bias at the 90-percent-exceedance level: NN-SMS12R produces a median error of -30 percent; only NN-DAR and NN-QPPQ, among the statistical-transfer methods, did not underestimate the lower flows by more than 5 percent. NN-DAR had a consistently low level of error across all percentiles. With respect to the 90-percent-exceedance event as a measure of low flow, the relatively poor performance of the SM and SMS methods is not surprising as the normalization does not explicitly address extreme events, especially low flows. By similar logic, high flows, which also are not explicitly addressed by the standardization, may nonetheless be better controlled because the mean, which is used for standardization, is significantly affected by the right-hand skew of streamflow data. (Additional metrics of low-flow performance can be found in the appendix C; these include the 10th and 50th percentile of the 7-day-average annual-minimum flow.)

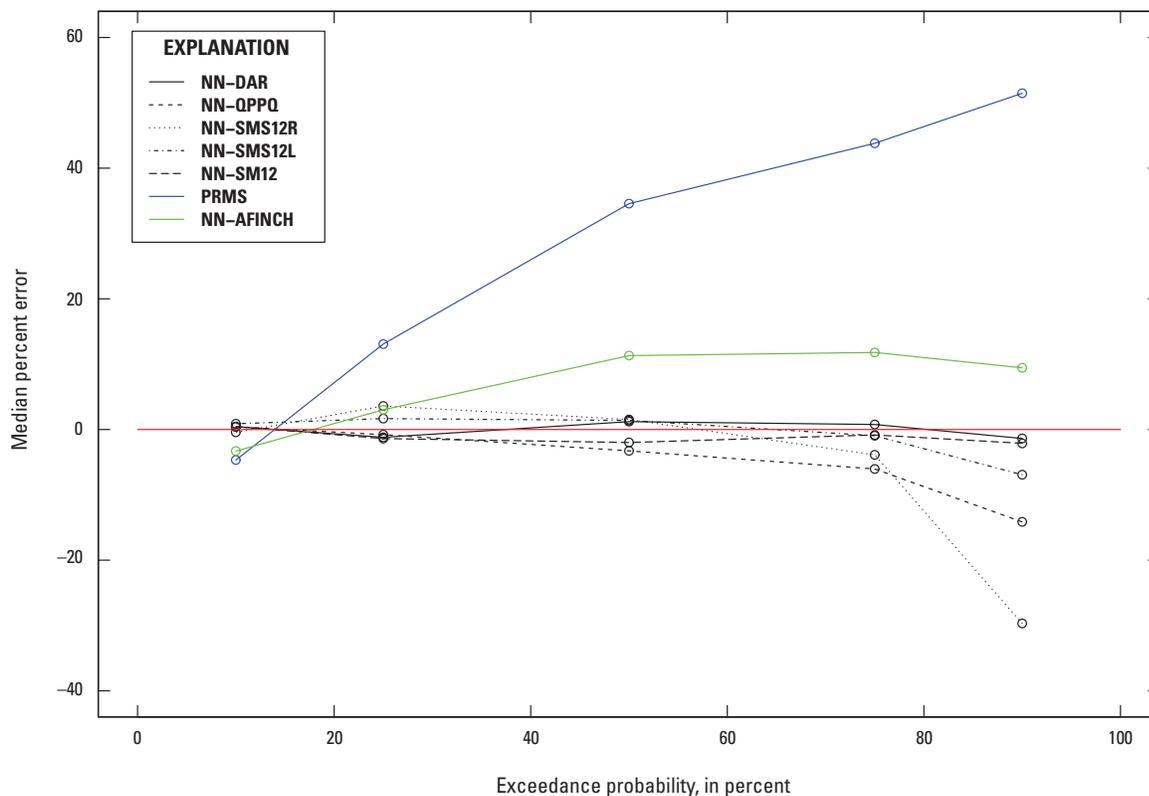


Figure 11. The median at-site percent error of five percentiles along the daily flow duration curve for the seven most accurate methods for prediction in ungaged basins. Transfer-based methods are indicated in black. The percentiles considered are the 10-percent-, 25-percent-, 50-percent-, 75-percent-, and 90-percent-exceedance flows. The horizontal axis represents the exceedance probability, while the vertical axis represents the median percent error. Method abbreviations are in table 1.

In addition to the signed percent error, the median absolute error for each quantile gives a measure of the frequency and magnitude of error, showing the point at which half of the sites exhibit an absolute error less than this amount (fig. 12). All of the PUB methods produce an absolute error of less than 10 percent or 15 percent for the 10-percent-exceedance flow. This level of error increases to 40 percent for the 90-percent-exceedance flow. In this case, PRMS departs widely from the transfer-based methods, but NN-AFINCH does not. Yet again, the SMS12R method also exhibits a relatively poor performance at the 90-percent-exceedance level.

Looking at both the signed error and absolute error, NN-SM12 appears to be nearly unbiased and to produce the lowest median absolute error. NN-AFINCH and the transfer-based methods produce a similar level of absolute bias, but NN-AFINCH displays a positive bias in the middle flows. NN-SMS12R performs similar to all other methods, except at the 90-percent-exceedance level: the median error rises over 50 percent at the low end of the FDC.

Fundamental Daily Streamflow Statistics

In addition to the flow duration curve, recent research demonstrates that the distribution of daily streamflow can be characterized by seven fundamental daily streamflow statistics (Archfield and others, 2013). These statistics sufficiently describe the distribution and behavior of daily streamflow events and can be used to group different basins into homogeneous clusters (Archfield and others, 2013). These seven statistics are the mean, coefficient of variation (L-CV), skewness (L-skew), kurtosis (L-kurtosis) and autoregressive lag-one autocorrelation coefficient of daily streamflow, and the amplitude and the phase of the sinusoidal seasonal signal (Archfield and others, 2013). Because of the importance of these fundamental daily streamflow statistics (FDSS), an optimal PUB method should produce reliable estimates of each of the seven statistics. For each PUB method, the seven statistics were calculated following the methodology of Archfield and others (2013).

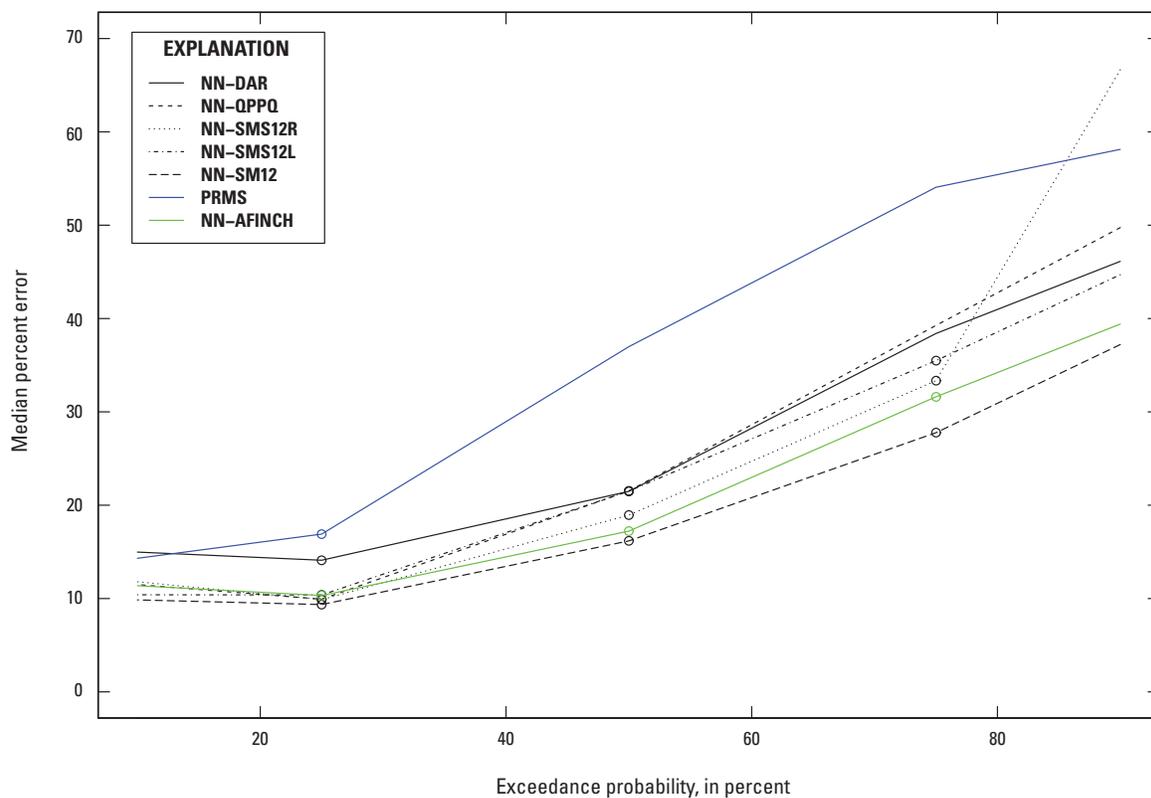


Figure 12. The median at-site absolute percent error of five percentiles along the daily flow duration curve for the seven most competitive methods for prediction in ungaged basins. Transfer-based methods are indicated in black. The percentiles considered are the 10-percent-, 25-percent-, 50-percent-, 75-percent-, and 90-percent-exceedance flows. The horizontal axis represents the exceedance probability, while the vertical axis represents the median absolute percent error. Method abbreviations are in table 1.

Figure 13 shows the median percent error in each of the seven metrics for the top-performing PUB methods. Almost all of the methods are relatively unbiased (median: ± 10 percent) for the mean streamflow and the parameters related to seasonality. In terms of the error in the mean, which can also be thought of as the percent bias of the method, PRMS exhibits a slight positive error (median: 5.2 percent), while NN-AFINCH and NN-QPPQ display an even slighter downward error (median: ~ -2.5 percent). There is more error in the seasonality coefficients: NN-AFINCH overestimates the amplitude of the seasonal variation by a median of 19 percent. The transfer-based methods are relatively unbiased for all FDSS and nearly indistinguishable from each other. PRMS underestimates the L-CV, L-skew, and L-kurtosis with a median of more than 20 percent. While exhibiting less bias than PRMS, NN-AFINCH

significantly underestimates L-CV, L-skew, and L-kurtosis as well. NN-AFINCH is unbiased in the lag-one autocorrelation, but PRMS shows the greatest average error in the lag-one correlation (near -32 percent). The significant underestimation of the L-moments and overestimation of the low-flow quantiles shown by PRMS demonstrates reduced extremes in the predicted record. This may demonstrate a general weakness of calibrated hydrologic simulation models (Thomas, 1982, 1987). The absolute errors in figure 14 demonstrate similar conclusions: the transfer-based methods exhibit significantly less bias than the more process-based methods; NN-AFINCH is more similar to the transfer-based methods than to PRMS. The transfer-based methods, therefore, represent the distribution of streamflow more accurately than the more process-based methods.

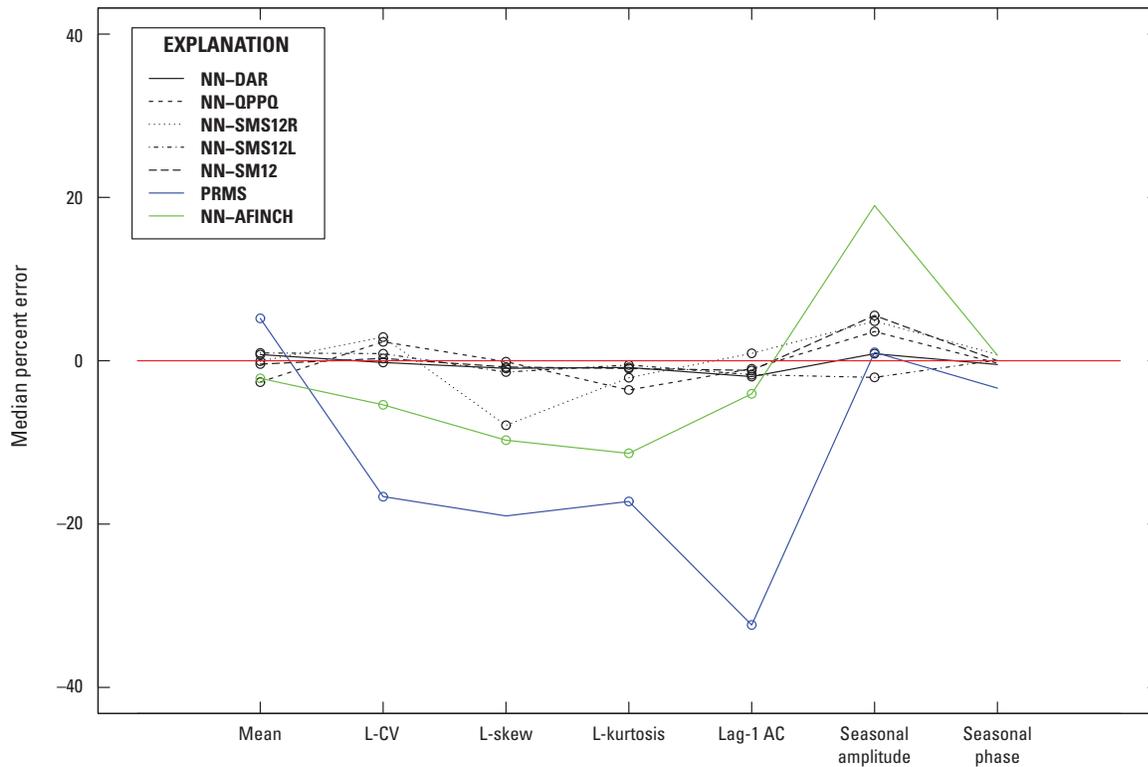


Figure 13. The median at-site percent error of the seven fundamental daily streamflow statistics (FDSS) for the seven most accurate methods for prediction in ungaged basins. Transfer-based methods are indicated in black. In order, the FDSS are the mean, L-CV, L-skew, L-kurtosis, lag-1 autocorrelation, and the amplitude and phase of the sinusoidal seasonal trend. The horizontal axis indicates each statistic, while the vertical axis represents the median percent error. Method abbreviations are in table 1.

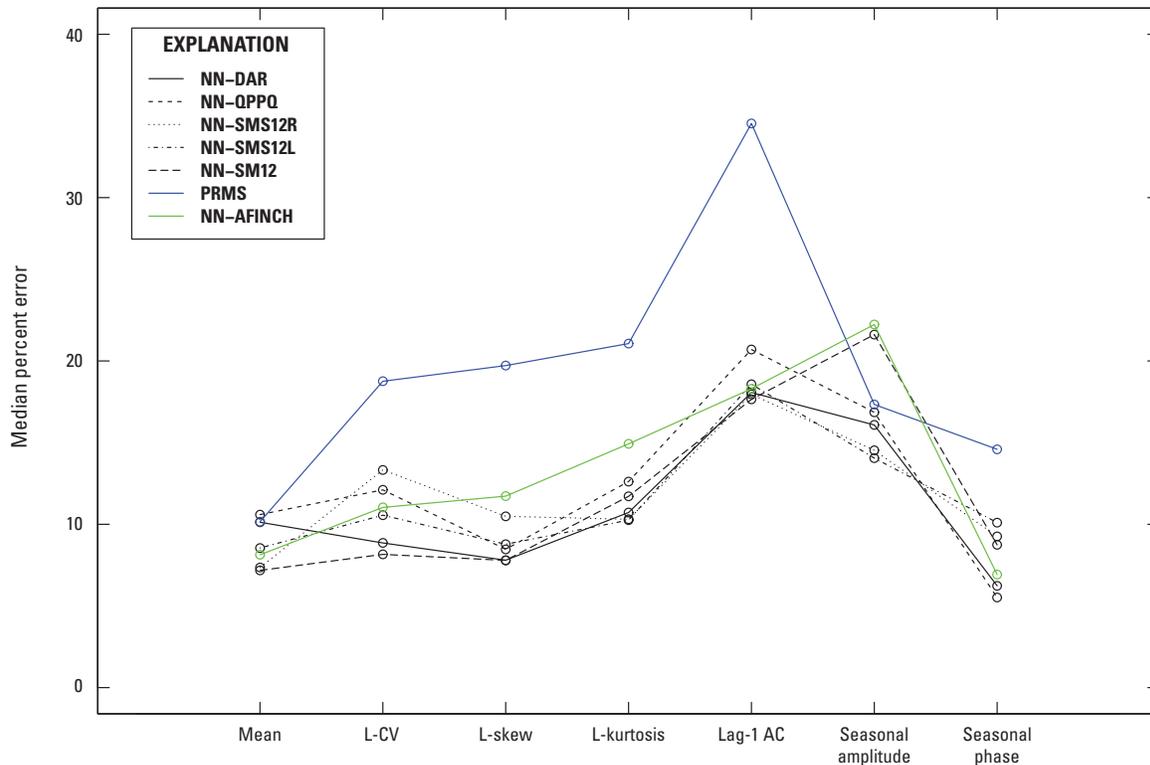


Figure 14. The median at-site absolute percent error of the seven fundamental daily streamflow statistics (FDSS) for the seven most accurate methods for prediction in ungaged basins. Transfer-based methods are indicated in black. In order, the FDSS are the mean, L-CV, L-skew, L-kurtosis, lag-1 autocorrelation, and the amplitude and phase of the sinusoidal seasonal trend. The horizontal axis indicates each statistics, while the vertical axis represents the median absolute percent error. Method abbreviations are in table 1.

Robust Rank-Based Evaluation of Streamflow Statistics

As with the overall goodness of fit and the daily, no-fail storage-yield curve, a robust rank-based evaluation (RRBE) was conducted to determine the method that provided the best and most consistent performance across all of the observed flow statistics. In addition to the metrics of the flow duration curve and the fundamental daily streamflow statistics, the RRBE also considered the coefficient of variation of annual flows, the coefficient of variation of daily flows, the 90th percentile of annual-maximum flow, and the 10th and 50th percentile of the 7-day-average annual-minimum flow. (The latter three statistics are representative of the 10-year annual maximum and the 10-year and 2-year 7-day-average annual-minimum flow, respectively. Whereas these more traditional statistics require the fitting of a log-Pearson distribution, the statistics presented here were calculated with an empirical distribution.) Individual analyses of the additional

flow statistics can be found in appendix C. The PUB methods were ranked in terms of their absolute error in each statistic; the averages and standard deviations of the ranks are included in tables 5 and 6. The average ranking and average standard deviation of the ranks across flow statistics were used to create the RRBE cloud in figure 15.

The RRBE cloud for the tradeoff between the average ranking and average standard deviation of the rankings across the 17 observed flow statistics shows the continued superiority of NN-SM12 and NN-QPPQ (fig. 15). The optimal edge of the cloud is defined by NN-SM12 and NN-QPPQ. The tradeoff between these methods is not steep; the increase in average ranking offered by NN-SM12 comes with only a slight increase in the variability of ranks. NN-SMS12L is also near the frontier, but set back slightly. NN-DAR, NN-AFINCH, and NN-SMS12R are set back still further, forming an inferior cluster. PRMS, as in other tradeoffs, is much further from the frontier. Just behind the frontier, MC-QPPQ and MC-SMS12L are near their NN counterparts.

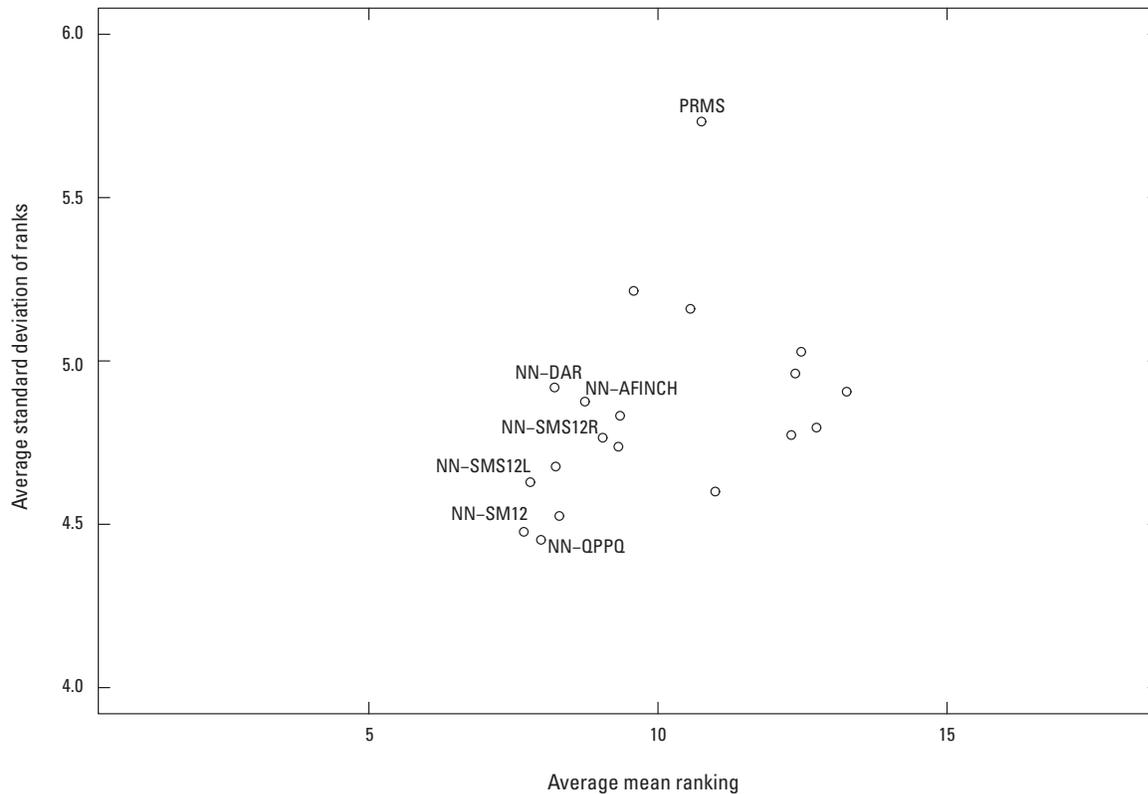


Figure 15. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of the 17 flow statistics for each method of prediction in ungaged basins. (See text for a description of each flow statistic.) The horizontal axis shows the mean average rank, while the vertical axis gives the average standard deviation of the ranks. Optimal methods would display minimal spread and a low mean-average ranking. The seven most competitive methods have been labeled. Method abbreviations are in table 1.

The ability of the PUB methods to reproduce the distribution of streamflow and key streamflow statistics (fig. 15) further strengthens the distinction between transfer- and more process-based PUB methods, like NN-AFINCH and PRMS. The NN-DAR, NN-QPPQ, NN-SMS12L, and NN-SM12 remain extremely strong, in comparison to PRMS and AFINCH. It is difficult to distinguish between these methods based on this series of metrics alone. Additionally, as measured by several different flow statistics, there is a discernible tradeoff in the average performance and the variability of performance. Of the strongest methods, two are very simplistic (NN-DAR and NN-SM12) and suggest the value of parameter parsimony.

Full Cross-Metric Comparison

The 19 different PUB methods considered here have been assessed in terms of accurately predicting the daily streamflow hydrograph, the ability to reproduce daily, no-fail storage-yield curves, and the ability to reproduce flow distribution and other streamflow statistics. Through these individual analyses, four methods are most accurate by all standards. The most

accurate PUB methods are the NN-DAR, NN-QPPQ, NN-SMS12L, and NN-SM12. As these methods were, in many cases, nearly indistinguishable from each other, a strong case must be made if more complexity is to be introduced. If no distinction can be made between these four methods, then the simpler methods (DAR and SM12) may be the most valuable. The next analyses consider the average performance across all metrics and ask how sensitive or robust these methods are to an alternative method of index-gage selection. It is here that the relative deficiency of the two parsimonious methods is made apparent.

In order to combine all of the different metrics into a single measure of performance, an RRBE was conducted across all 32 performance metrics considered here. (A root-mean-square-normalized error across all seven fundamental daily streamflow statistics was included to strengthen the importance of these distributional properties.) The average ranks of each PUB method for each metric are shown in table 5; the standard deviations of the ranks are shown in table 6. The average rank and average standard deviation of the ranks for each PUB method produce the RRBE cloud in figure 16. Some of these metrics, like RMSE and NSE, are strongly correlated. While the inclusion of correlated metrics may mean

that certain aspects of performance are more heavily weighted than others, the correlated metrics quantify extremely important aspects of model performance, like the day-to-day model performance. Along this reasoning, including correlated variables did not adversely affect the analysis of PUB methods presented here. (As discussed below, different sets of performance metrics can be considered to suit the stakeholder need.)

In the full cross-metric RRBE, the three metrics continue to cluster along the optimal edge of the RRBE cloud: NN-QPPQ, NN-SMS12L, and NN-SM12. NN-DAR, while initially competitive, falls further from the frontier when all metrics are considered. NN-SMS12R and NN-AFINCH are also set back from the frontier, but not in the main pack, while PRMS is far from the frontier. By considering all 32 metrics, the statistical advantages of the three frontier methods are made quite clear. As NN-SM12 requires less effort to parameterize, requiring only 12 monthly means, it may have a significant advantage over NN-SMS12L (24 regressions) and NN-QPPQ (27 regressions).

Sensitivity to Index-Selection Algorithm

After considering 32 different metrics of performance, it remains extremely difficult to distinguish between the top three methods: NN-QPPQ, NN-SMS12L, and NN-SM12. As all of these methods were similarly competitive, they might all be reasonably accurate for regional implementation or explored for national implementation. Furthermore, it might be best to use the SM12, as it is the least complex. But, as a final analysis, it is useful to consider how sensitive any of these methods might be to the information content of the index gage. This sensitivity can be assessed by observing how well a basic PUB method, like SM12 or SMS12L for example, performs with two different index gages. As the map-correlation selection methods were shown to be inferior to the nearest-neighbor methods, these two implementations of the basic PUB method provide an example of how a particular basic PUB method might be affected by an index gage with less

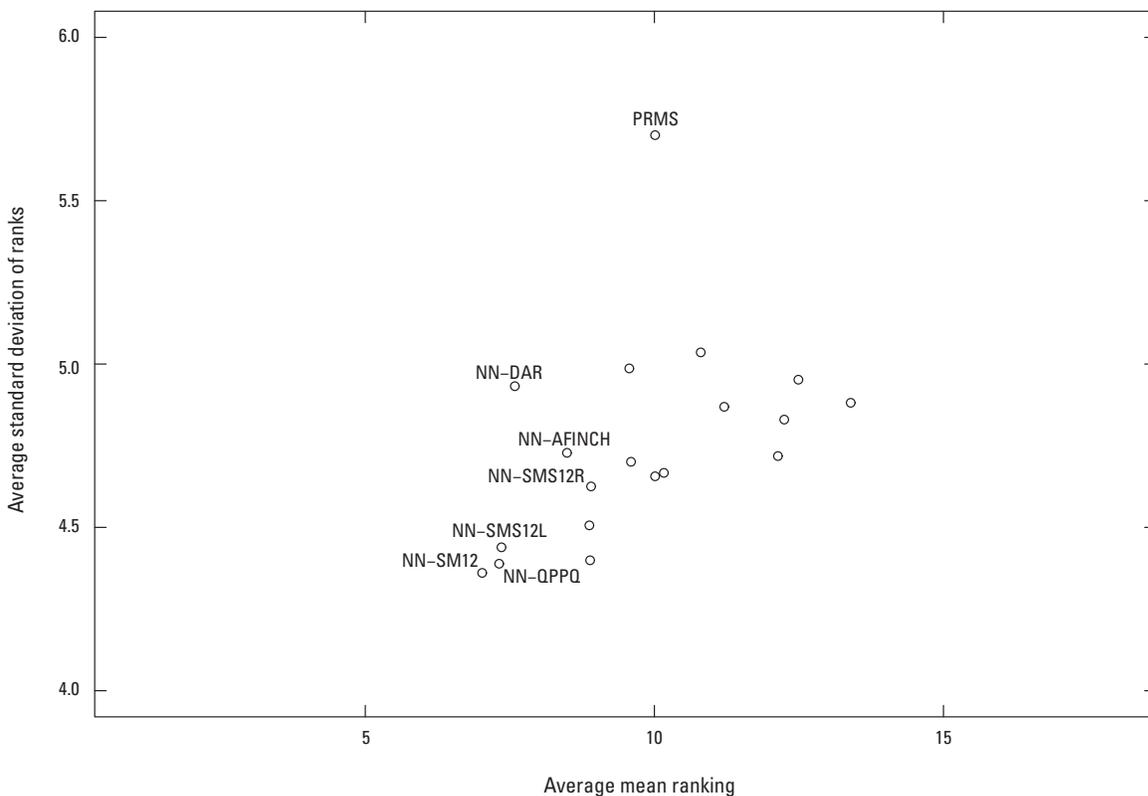


Figure 16. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of all 32 performance metrics for each method of prediction in ungaged basins. (See text for a description of each performance metric.) The horizontal axis shows the mean average rank, while the vertical axis gives the average standard deviation of the ranks. Optimal methods would display minimal spread and a low mean-average ranking. The seven most competitive methods have been labeled. Method abbreviations are in table 1.

information content. Ideally, the stronger basic PUB methods would be less affected by inferior information content.

Two ways of measuring sensitivity of the basic PUB methods to the information available at the index gage were considered here. First, the ranks of table 5 and 6 were averaged by prediction method; for example, the average and standard deviation of all 32 performance-metric rankings for NN-DAR and MC-DAR were used to represent the basic method DAR. A more robust method will be less degraded by this averaging and would remain along the efficient frontier of the cloud in figure 17. The averages and standard deviations for all eight transfer-based methods, along with the process- and quasi-process-based methods, are in table 7. (Note that PRMS remains unchanged because it did not use the NN or MC index-gage selection algorithm in the same manner.) This averaging pushes SM12 off of the frontier, leaving QPPQ and SMS12L along the optimal edge of the cloud. A similar test can be conducted by calculating the Euclidean distance between the nearest-neighbor and map-correlation points of a particular PUB method in the RRBE cloud of figure 16. A robust method will have a small difference between these points. The Euclidean distance between NN and MC

realizations is included table 7. Looking at the top-performing basic PUB methods, QPPQ and SMS12L produce the smallest separation distance (1.57 and 1.53, respectively). Of the two, SMS12L has the smaller differential. Notably, SM12 has a relatively large differential (2.59), suggesting a high sensitivity to index gage information content. Overall, NN-SMS12L and NN-QPPQ are stable and nearly indistinguishable, but NN-SM12 is more sensitive to the information content of the index gage.

This analysis of 32 metrics of performance shows that there are a few PUB methods, of the 19 considered in the SEMC study, that produce reliable and competitive estimates of streamflow. These methods all perform well and can hardly be distinguished from each other. These methods include the nearest-neighbor implementations of drainage area ratio, QPPQ, standardizing the logarithms of streamflow by monthly means, and standard deviations and standardizing streamflow by monthly means. Averaging across all 32 metrics showed these four PUB methods to fall along an efficient frontier of average performance and variability. NN-SM12, NN-SMS12L, and NN-QPPQ were all closest to the efficient frontier. It is notable that the relatively simplistic method,

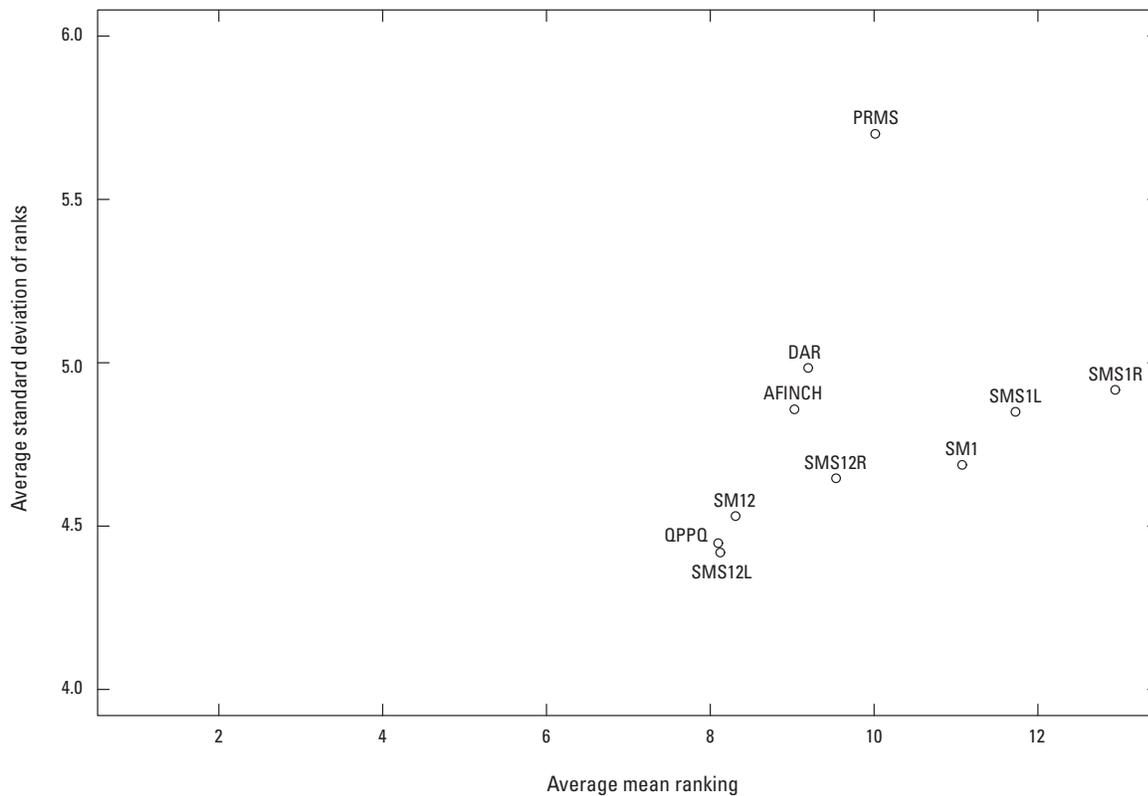


Figure 17. The robust rank-based evaluation cloud showing the tradeoff between the mean ranking and standard deviation of the ranks of all 32 methods for each method of prediction in ungaged basins, average across competing methods for the selection of index gages. (See text for a full description of the methodology.) The horizontal axis shows the mean average rank, while the vertical axis gives the average standard deviation of the ranks. Optimal methods would display minimal spread and a low mean-average ranking. The seven most competitive methods have been labeled. Method abbreviations are in table 1.

NN-SM12, was able to perform so well. This parsimony may be more appealing than any marginal improvement in performance seen by the more complex methods. When considering the robustness of each transfer method to index-gage information content, NN-QPPQ and NN-SMS12L are marginally superior. NN-SM12 was hampered by sensitivity to the index-gage selection algorithm, as the variance of performance between the NN and MC implementations was shown to be much greater. The difference between NN-QPPQ and NN-SMS12L is quite small. Based on this sensitivity analysis, NN-QPPQ appears to possess a slightly greater average performance with only a minimal degradation in variability.

For almost all performance metrics, NN-QPPQ has been shown to be among the top performers. It produces reliable predictions of historical daily streamflow in the southeastern United States. Not surprisingly, this method is already becoming more and more widely used in hydrology and water-resources engineering (see Archfield and others, 2010; Shu and Ouarda, 2012; Smakhtin and Masse, 2000). The literature concerning improvements to QPPQ and improved estimators of the ungaged flow duration curve, which is required for the application of QPPQ, continues to grow (Booker and Snelder, 2012; Castellarin and others, 2007, 2004; Ganora and others, 2009). The popularity of this method in research suggests that further improvements in NN-QPPQ are possible. In some of the cases examined here, additional methods performed similarly. Continued analysis, especially in varying climatic regions, will improve our understanding of PUB performance.

Summary and Conclusions

The Southeast Model-Comparison (SEMC) considered 19 different techniques for estimating continuous daily records of historical streamflow, with the goal of determining the best methods for prediction in ungaged basins (PUB). It was intended that the best methods could then be considered for regional or national implementation as part of the National Water Census project. The analysis presented here covered 32 different metrics of performance and demonstrated a comprehensive and flexible framework for comparing any number of PUB methods. The results showed the nearest-neighbor implementation of nonlinear spatial interpolation with flow duration curves (NN-QPPQ) was the best performing method. This was followed closely by the nearest-neighbor implementation of standardizing logarithms of streamflow by monthly means and standard deviations (NN-SMS12L).

The 19 PUB methods considered here (table 2) included variations of four statistical, transfer-based methods, a process-based model, and a quasi-process-based model. The transfer-based methods considered were the drainage area ratio, QPPQ, standardizing flows by an annual mean and standard deviation, standardizing the logarithms of flow by an annual mean and standard deviation, standardizing flows by monthly means and standard deviations, standardizing the

logarithms of flow by monthly means and standard deviations, standardizing flows by an annual mean, and standardizing flows by monthly means. All of the methods relied on an index gage that was used to transfer information from a gaged to an ungaged site. This index gage was selected by either minimum distance or estimated maximum correlation. Regional regression was used to parameterize methods like the standardization with annual means, which requires estimated ungaged means. Rainfall-runoff models were represented by the Precipitation Runoff Modeling System (PRMS). A transfer-based approach was used to disaggregate the monthly Analysis of Flows in Networks of Channels (AFINCH) model into daily values to examine how a quasi-processed-based monthly model might perform.

The framework presented here considered several different aspects of model performance for each PUB method. First, the overall goodness of fit of each PUB method was assessed using the Nash-Sutcliffe efficiency and several other metrics. This analysis showed that only seven of the PUB methods were reasonably competitive. Additional analysis determined how well each of these seven PUB methods reproduced daily, no-fail storage-yield curves, reproduced other streamflow statistics, and performed across all metrics considered. A robust rank-based evaluation was used to compare the average relative rankings of each PUB metrics. This evaluation showed that the seven best methods were split into two groups; three methods were consistently along or near the efficient frontier of all relative rankings. Finally, these three PUB methods were tested for sensitivity to the index-gage information content. This test showed the superiority of the nearest-neighbor implementations of QPPQ and standardizing logarithms of streamflow by monthly means and standard deviations.

Three important conclusions can be drawn from the analyses conducted here. First, for almost every prediction of ungaged basin (PUB) method considered as part of the Southeast Model Comparison (SEMC), the prediction methods using index gages selected with the nearest-neighbor algorithm outperformed those using the map-correlation algorithm. Second, considering all 32 metrics, the nonlinear spatial interpolation using flow duration curves and the nearest-neighboring index gage (NN-QPPQ) provided better prediction of daily streamflow, daily, no-fail storage-yield curves, flow duration curves, and flow statistics in the southeastern United States than any other PUB method considered in the SEMC. Finally, regardless of the full-metric comparison, PUB method selection should be chosen based on end use. The framework provided here can help to determine the strengths and weaknesses of PUB methods.

The goal of the SEMC study was to identify particular methods for the prediction of historical daily flow records that proved promising for regional or national implementation. This would enable the publication of a temporally and spatially continuous dataset of daily streamflow. Through the numerous analyses, two methods emerged as the most generally competitive PUB methods: the nearest-neighbor implementations of nonlinear spatial interpolation using flow

duration curves (NN-QPPQ) and standardizing logarithms of streamflow by monthly means and standard deviations (NN-SMS12L). It was nearly impossible to distinguish between these two methods in terms of performance. Additionally, neither of these methods requires significantly more parameterization in order to apply: NN-SMS12L requires 24 regional regressions, 12 for monthly means, and 12 for monthly standard deviations. NN-QPPQ, in this case, required 27 regressions of particular quantiles along the flow duration curve. In the absence of a clear, far-and-away “winner,” it is necessary to make a somewhat qualitative judgment between these two methods.

Based on the results presented here, the nonlinear spatial interpolation technique using flow duration curves with the nearest-neighboring index gages (NN-QPPQ) produces the most reliable predictions of continuous records of historical daily streamflow in the Southeast region of the United States. This method, NN-QPPQ, was shown to be a top performer in almost all the metrics considered above. Of course, based on this analysis, other methods, like NN-SMS12L, might be more appropriate for particular applications; additional research should continue to consider new metrics and methods, especially for specific purposes. Modelled time series from the NN-QPPQ method will become available through the USGS National Water Census data portal (<http://cida.usgs.gov/nwc/>) under the heading “Streamflow Stats,” which is currently provided only in a beta state.

The methods of analysis presented here reflect a possible framework for continued analysis and comprehensive multiple comparisons of PUB methods. The approach presented here considers overall goodness of fit along with the ability to reproduce particular streamflow signatures and statistics. Using a tool like RRBE, it was possible to balance average performance against variability of average performance across many metrics. Additional metrics of comparison can easily be incorporated into this type of analysis. By considering such a multifaceted approach, the top-performing models in any region can easily be identified and considered for further research. The top-performing models can then provide a basis for future applications and explorations by scientists, engineers, managers, and practitioners to suit their own needs.

Even with the analysis presented here, the optimal PUB method relies on the intended application. Some methods are stronger overall, while some methods may be better at predicting particular statistics. The optimal PUB method for the fish biologist might be starkly different than the best PUB method for the manager at a large hydropower facility; the optimality is a function of purpose and perspective. The SEMC study sought to find the best method and demonstrated a framework for determining the best method on average. The many facets of this framework allow others with a more targeted application to make a more-informed choice of PUB method. For this work, every effort was made to provide an across-the-board assessment of average performance and not make any weighting of metrics that might limit the application of the identified PUB method.

Limitations and Further Research

The rigorous examination of prediction in ungaged basins (PUB) is an extremely active field of research. The work presented here offers only a glimpse at possible advances in the application of PUB methods. Despite presenting a comprehensive overview of many performance metrics and PUB methods, this report is limited in that new metrics and methods are frequently being developed and introduced. It is therefore necessary that the findings of this report be frequently reassessed and updated. The findings shown here are far from exhaustive and may shift in the light of new PUB methods or in different regions or climates. Additional research will always be needed.

As with all studies, this work is subject to some important limitations. Firstly, this report focuses only on average performance. The intention was not to make any judgment on the relative importance of performance metrics; different stakeholders may value different metrics, and thus this analysis could be tailored to more specific valuations of performance. Second, despite the outperformance of the Precipitation-Runoff Modeling System (PRMS) by some of the statistically based models, PRMS may be more advantageous when it is of interest to understand water availability under future or nonstationary conditions, such as water use, urbanization, or climate change. Recall that PRMS, like many process-based models, can be integrated with other process-based and systemic models and can be edited to incorporate changes in hydrologic processes. The other models tested in this study, with a strict reliance on contemporaneous index records, are only able to provide historical estimates of streamflow time series. Thirdly, regression-based methods should not be applied to basins with characteristics outside of the range of characteristics found at the gaged locations used to develop the regression equations. Finally, as these results have only been applied in the southeastern United States, it is unclear whether the relative model performances would change if the models were applied to another study area with differing climate and physiographic conditions or in areas of sparser streamgage density.

Even with these limitations, there may be some synergy between the process-based and transfer-based approaches. It is possible to envision an approach designed to calibrate the PRMS model in an ungaged mode with the use of estimated streamflow records provided by the statistically based methods. Leveraging the advantages of both techniques, this could lead to a robust methodology to both estimate historical streamflow and forecast future records at ungaged locations. Ongoing research at the USGS and elsewhere has already begun to explore using transfer-based models to calibrate process-based models like the PRMS.

Another avenue of future research might be to consider a more thorough tradeoff analysis, treating the entire PUB puzzle as a multi-objective problem. Here, only the most basic tradeoffs were considered. Additional study could explore the tradeoffs between particular sets of performance metrics. For

example, one might be concerned with the tradeoff between how well the low flows are estimated and how well the high flows are estimated. Additionally, performance metrics could be classified by interest groups or stakeholders: the fish biologist and the hydropower operator, for example. Multi-objective and interdisciplinary techniques are becoming quite popular in the field of hydrology; the PUB problem is an ideal application for interdisciplinary tools.

Acknowledgments

This project was conducted as a part of the USGS National Water Census and also supports the goals of the National Streamflow Information Program.

We wish to thank David Holtschlag and Carol Luukkonen (USGS Michigan Water Science Center) for their help with AFINCH. We also thank our colleagues at the USGS Center for Integrated Data Analytics for processing precipitation and temperature data from the Processing Routines in IDL for Spectroscopic Measurements (PRISM) database and for hosting the data portal. We would also like to thank Skip Vecchia (USGS) for his influential thoughts on methods for the disaggregation from monthly to daily streamflow. We also appreciate the efforts of David Stewart (USGS), who helped obtain basin characteristics and provided invaluable GIS support. In addition, we thank Prof. Richard Vogel of Tufts University for his continued support and insights into this work. We offer our special thanks to Rodney Knight and David Wolock (USGS) who provided exceptional reviews of the original manuscript and vastly improved the final product.

References Cited

- Alley, W.M., Evenson, E.J., Barber, N.L., Bruce, B.W., Dennehy, K.F., Freeman, M.C., Freeman, W.O., Fischer, J.M., Hughes, W.B., Kennen, J.G., Kiang, J.E., Maloney, K.O., Musgrove, M., Ralston, B., Tessler, S., and Verdin, J.P., 2013, Progress toward establishing a national assessment of water availability and use: U.S. Geological Survey Circular 1384, 34 p., at <http://pubs.usgs.gov/circ/1384/>.
- Archfield, S.A., Kennen, J.G., Carlisle, D.M., and Wolock, D.M., 2013, An objective and parsimonious approach for classifying natural flow regimes at a continental scale: *River Research and Applications*, v. 30, no. 9, p. 1166–1183, at <http://dx.doi.org/10.1002/rra.2710>.
- Archfield, S.A., Steeves, P.A., Guthrie, J.D., and Reis III, K.G., 2013, Towards a publicly available, map-based regional software tool to estimate unregulated daily streamflow at ungauged rivers: *Geoscientific Model Development*, v. 6, p. 101–115, at <http://dx.doi.org/10.5194/gmd-6-101-2013>.
- Archfield, S.A., and Vogel, R.M., 2010, Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungauged catchments: *Water Resources Research*, v. 46, no. 10, at <http://dx.doi.org/10.1029/2009WR008481>.
- 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 ungauged stream location in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41 p., plus CD-ROM, at <http://pubs.usgs.gov/sir/2009/5227/>.
- Asquith, W.H., Roussel, M.C., and Vrabel, J., 2006, Statewide analysis of the drainage-area ratio method for 34 streamflow percentile ranges in Texas: U.S. Geological Survey Scientific Investigations Report 2006–5286, 34 p., at <http://pubs.usgs.gov/sir/2006/5286/>.
- Beven, Keith, and Freer, Jim, 2001, Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology: *Journal of Hydrology*, v. 249, nos. 1–4., p. 11–29, at <http://www.sciencedirect.com/science/article/pii/S0022169401004218>.
- Booker, D.J., and Snelder, T.H., 2012, Comparing methods for estimating flow duration curves at ungauged sites: *Journal of Hydrology*, v. 434, p. 78–94, <http://dx.doi.org/10.1016/j.jhydrol.2012.02.031>.
- Castellarin, Attilio, Camorani, Giorgio, and Brath, Armando, 2007, Predicting annual and long-term flow-duration curves in ungauged basins: *Advances in Water Resources*, v. 30, no. 4, p. 937–953, at <http://dx.doi.org/10.1016/j.advwatres.2006.08.006>.
- Castellarin, Attilio, Galeati, Giorgio, Brandimarte, Luigia, Montanari, Alberto, and Brath, Armando, 2004, Regional flow-duration curves; reliability for ungauged basins: *Advances in Water Resources*, v. 27, no. 10, p. 953–965, at <http://dx.doi.org/10.1016/j.advwatres.2004.08.005>.
- Couch, C.A., Hopkins, E.H., and Hardy, S.P., 1995, Influences of environmental settings on aquatic ecosystems in the Apalachicola-Chattahoochee-Flint River Basin: U.S. Geological Survey Water-Resources Investigations Report 95–4278, 58 p., at <http://pubs.usgs.gov/wri/wrir95-4278/>.
- Cunnane, C., 1978, Unbiased plotting positions – a review: *Journal of Hydrology*, v. 37, nos. 3–4, p. 205–222.
- Duan, Q.Y., Gupta, V.K., and Sorooshian, S., 1993, Shuffled complex evolution approach for effective and efficient global minimization: *Journal of Optimization Theory and Applications*, v. 76, no. 3, p. 501–521.

- Duan, Q.Y., Sorooshian, S., and Gupta, V.K., 1992, Effective and efficient global optimization for conceptual rainfall-runoff models: *Water Resources Research*, v. 28, no. 4, p. 1015–1031.
- Duan, Q.Y., Sorooshian, S., and Gupta, V.K., 1994, Optimal use of the SCE-UA Global Optimization Method for calibrating watershed models: *Journal of Hydrology*, v. 158, no. 3–4, p. 265–284.
- Falcone, J.A., 2011, GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow [digital spatial dataset], at http://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, at <http://esapubs.org/Archive/ecol/E091/045/metadata.htm>.
- Farmer, W.H., and Vogel, R.M., 2013, Performance-weighted methods for estimating monthly streamflow at ungauged sites: *Journal of Hydrology*, v. 477, p. 240–250, at <http://dx.doi.org/10.1016/j.jhydrol.2012.11.032>.
- Fennessey, N.M., 1994, A hydro-climatological model of daily stream flow for the northeast United States: Medford, Mass., Tufts University, Ph.D. dissertation, 269 p.
- Ganora, D., Claps, P., Laio, F., and Viglione, A., 2009, An approach to estimate nonparametric flow duration curves in ungauged basins: *Water Resources Research*, v. 45, no. 10, <http://dx.doi.org/10.1029/2008WR007472>.
- Gibson, W.P., Daly, Christopher, Kittel, Tim, Nychka, Doug, Johns, Craig, Rosenbloom, Nan, McNab, Alan, and Taylor, George, 2002, Development of a 103-year high-resolution climate data set for the conterminous United States—Proceedings of the 13th Conference on Applied Climatology, May 13–16, 2002: Portland, Oreg., American Meteorological Society, p. 181–183.
- Gleeson, Tom., Smith, Leslie., Moosdorf, Nils., Hartmann, Jens., Durr, 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, at <http://dx.doi.org/10.1029/2010GL045565>.
- Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and frequency of rural floods in the southeastern United States, 2006; Volume 1, Georgia: U.S. Geological Survey Scientific Investigations Report 2009–5043, 120 p., at <http://pubs.usgs.gov/sir/2009/5043/>.
- Greene, W.H., 1997, *Econometric Analysis* (3d ed.): Upper Saddle River, N.J., Prentice Hall, 1,075 p.
- Gupta, H.V., and Kling, Harald, 2011, On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics: *Water Resources Research*, v. 47, no. 10, at <http://dx.doi.org/10.1029/2011WR010962>.
- Gupta, H.V., Kling, Harald, Yilmaz, K.K., and Martinez, G.F., 2009, Decomposition of the mean squared error and NSE performance criteria; Implications for improving hydrological modelling: *Journal of Hydrology*, v. 377, nos. 1–2, p. 80–91, at <http://dx.doi.org/10.1016/j.jhydrol.2009.08.003>.
- Helsel, D.R., and Hirsch, R.M., 2002, *Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations*, book 4, chap. A3, 522 p., at <http://pubs.usgs.gov/twri/twri4a3/>.
- Hirsch, R.M., 1979, An evaluation of some record reconstruction techniques: *Water Resources Research*, v. 15, no. 6, p. 1781–1790, at <http://dx.doi.org/10.1029/WR015i006p01781>.
- Hirsch, R.M., 1982, A comparison of four streamflow record extension techniques: *Water Resources Research*, v. 18, no. 4, p. 1081–1088, at <http://dx.doi.org/10.1029/wr018i004p01081>.
- Holtschlag, D.J., 2009, Application guide for AFINCH (analysis of flows in networks of channels) described by NHDPlus: U.S. Geological Survey Scientific Investigations Report 2009–5188, 106 p., at <http://pubs.usgs.gov/sir/2009/5188/>.
- Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., and Cudennec, C., 2013, A decade of Predictions in Ungauged Basins (PUB)—a review: *Hydrological Sciences Journal*, v. 58, no. 6, p. 1198–1255, at <http://dx.doi.org/10.1080/02626667.2013.803183>.
- Hughes, D.A., and Smakhtin, V., 1996, Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves: *Hydrologic Sciences Journal*, v. 41, no. 6, p. 851–871.
- Isaaks, E.H., and Srivastava, R.M., 1989, *An introduction to Applied Geostatistics*: New York and Oxford, Oxford University Press, 561 p.
- Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., and Xian, G., 2013, A comprehensive change detection method for updating the National Land Cover Database to circa 2011: *Remote Sensing of Environment*, v. 132, p. 159–175.

- Johnson, G.C., Kidd, R.E., Journey, C.A., Zappia, Humbert, and Atkins, J.B., 2002, Environmental setting and water-quality issues of the Mobile River Basin, Alabama, Georgia, Mississippi, and Tennessee: U.S. Geological Survey Water-Resources Investigations Report 02–4162, 120 p., at <http://pubs.usgs.gov/wri/wri024162/>.
- Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C., 1985, *The theory and practice of Econometrics* (2d ed.): New York, John Wiley and Sons, 1,019 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., at <http://pubs.er.usgs.gov/publication/sir20135013>.
- Leavesley, G.H., Lichty, R.W., Troutman, B.M., and Saindon, L.G., 1983, Precipitation-runoff modeling system; User's manual: U.S. Geological Survey Water-Resources Investigations Report 83–4238, 207 p., at <http://pubs.er.usgs.gov/publication/wri834238>.
- Linhart, S.M., Nania, J.F., Christiansen, D.E., Hutchinson, K.J., Sanders, Jr., C.L., and Archfield, S.A., 2013, Comparison between two statistically based methods, and two physically based models developed to compute daily mean streamflow at ungauged locations in the Cedar River Basin, Iowa: U.S. Geological Survey Scientific Investigations Report 2013–5111, 7 p., at <http://pubs.er.usgs.gov/publication/sir20135111>.
- Lumley, Thomas, 2012, leaps: regression subset selection, R package version 2.9., at <http://CRAN.R-project.org/package=leaps>.
- Markstrom, S.L., Niswonger, R.G., Regan, R.S., Prudic, D.E., and Barlow, P.M., 2008, GSFLOW—Coupled ground-water and surface-water flow model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005): U.S. Geological Survey Techniques and Methods, book 6, chap. D–1, p. 240, at <http://pubs.er.usgs.gov/publication/tm6D1>.
- 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, at <http://dx.doi.org/10.1623/hysj.53.4.706>.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L., 2007, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations: *Transactions of the American Society of Agricultural and Biological Engineers*, v. 50, no. 3, p. 885–900.
- Nash, J.E., and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models part I—A discussion of principles: *Journal of Hydrology*, v. 10, no. 3, p. 282–290, at [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6).
- R Development Core Team, 2005, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing.
- Razavi, T., and Coulibaly, P., 2013, Streamflow prediction in ungauged basins; Review of regionalization methods: *Journal of Hydrologic Engineering*, v. 18, no. 8, p. 958–975, at [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000690](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000690).
- Ribeiro, P., Jr., and Diggle, P., 2001, geoR; A package for geostatistical analysis: *R-news*, v. 1, no. 2.
- Ries III, K.G., 2007, The National Streamflow Statistics Program: A computer program for estimating streamflow statistics for ungauged sites: U.S. Geological Survey Techniques and Methods, book 4, chap. A6, 37 p., at <http://pubs.er.usgs.gov/publication/tm4A6>.
- Shu, Chang, and Ouarda, T.B.M.J., 2012, Improved methods for daily streamflow estimates at ungauged sites: *Water Resources Research*, v. 48, no. 2, at <http://dx.doi.org/10.1029/2011WR011501>.
- Sivapalan, Murugesu, 2003, Prediction in ungauged basins; A grand challenge for theoretical hydrology: *Hydrological Processes*, v.17, no. 15, p. 3163–3170, at <http://dx.doi.org/10.1002/hyp.5155>.
- Sivapalan, Murugesu, Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe, E., 2003, IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences: *Hydrological Sciences Journal*, v. 48, no. 6, p. 857–880, <http://dx.doi.org/10.1623/hysj.48.6.857.51421>.
- Smakhtin, V.Y., 1999, Generation of natural daily flow time-series in regulated rivers using a non-linear spatial interpolation technique: *River Research and Applications*, v. 15, no. 4, p. 311–323, at [http://dx.doi.org/10.1002/\(SICI\)1099-1646\(199907/08\)15:4<311::AID-RRR544>3.0.CO;2-W](http://dx.doi.org/10.1002/(SICI)1099-1646(199907/08)15:4<311::AID-RRR544>3.0.CO;2-W).
- Smakhtin, V.Y., and Masse, B., 2000, Continuous daily hydrograph simulation using duration curves of a precipitation index: *Hydrological Processes*, v. 14, no. 6, p. 1083–1100.
- Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, E., 1993, Frequency analysis of extreme events, in Maidment, D.R., ed., *Handbook of Hydrology*: New York, McGraw-Hill, variously paged.

- Stuckey, M.H., Koerkle, E.H., and Ulrich, J.E., 2012, Estimation of baseline daily mean streamflows for ungaged locations on Pennsylvania Streams, water years 1960–2008: U.S. Geological Survey Scientific Investigations Report 2012–5142, 68 p., at <http://pubs.er.usgs.gov/publication/sir20125142>.
- Tasker, G.D., 1997, Estimating magnitude and frequency of flood peaks at ungaged sites using regression: *Computing Science and Statistics*, v. 29, p. 106–111.
- Therneau, T.M., 2013, A package for survival analysis in S, R package version 2.37-4, at <http://CRAN.R-project.org/package=survival>.
- Thomas, D.M., and Benson, M.A., 1970, Generalization of streamflow characteristics from drainage-basin characteristics: U.S. Geological Survey Water-Supply Paper 1975. [Also available at <http://pubs.er.usgs.gov/publication/wsp1975>.]
- Thomas, H.A., and Burden, R.P., 1963, Operations research in water quality management: Cambridge, Mass., Harvard University.
- Thomas, Jr., W.O., 1982, An evaluation of flood frequency estimates based on rainfall/runoff modeling: *Water Resources Bulletin*, v. 18, no. 2, p. 221–230.
- Thomas, Jr., W.O., 1987, Comparison of flood-frequency estimates based on observed and model-generated peak flows in Singh, V.P. (ed.), *Hydrologic Frequency Modeling*: Dordrecht, Netherlands, D. Reidel Publishing, p. 149–161.
- U.S. Environmental Protection Agency and the U.S. Geological Survey, 2010, NHDPlus Version 1 (NHDPlusV1) User Guide (September 1), accessed December 22, 2011, at ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV1/documentation/NHDPLUSV1_UserGuide.pdf.
- Viger, R.J., and Bock, A., 2014, The geospatial fabric feature set for national hydrologic modeling: U.S. Geological Survey dataset, at <http://dx.doi.org/10.5066/F7542KMD>.
- Vogel, R.M., 2011, Hydromorphology: *Journal of Water Resources Planning and Management*, v. 137, no. 2, p. 147–149, at [http://dx.doi.org/10.1061/\(ASCE\)WR.1943-5452.0000122](http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000122).
- Vogel, R.M., and Fennessey, N.M., 1993, L moment diagrams should replace product moment diagrams: *Water Resources Research*, v. 29, no. 6, p. 1745–1752. [Also available at <http://onlinelibrary.wiley.com/doi/10.1029/93WR00341/abstract>.]
- 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, at <http://dx.doi.org/10.1111/j.1752-1688.1995.tb03419.x>.
- Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, J.N., 2001, Completion of the 1990's National Land-Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources: *Photogrammetric Engineering and Remote Sensing*, v. 67, p. 650–662.

For additional information regarding this publication, contact:

Director, USGS Office of Surface Water
12201 Sunrise Valley Drive
Reston, VA, 20192

Or visit the USGS Office of Surface Water Web site at:
<http://water.usgs.gov/osw/>

Edited and prepared by:
USGS Science Publishing Network
Reston and Raleigh Publishing Service Centers

