USGS
science for a changing world

Prepared in cooperation with the Gulf Coast Ecosystem Restoration Council

Generalized Additive Model Estimation of No-Flow Fractions and L-Moments to Support Flow-Duration Curve Quantile Estimation Using Selected Probability Distributions for Bay and Estuary Restoration in the Gulf States

Scientific Investigations Report 2022–5051

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

Cover. Cheaha Creek, Cheaha Creek Gorge, Talladega National Forest, Clay County, Alabama, August 23, 2020. Photograph by Alan Cressler. Used with permission.

Generalized Additive Model Estimation of No-Flow Fractions and L-Moments to Support Flow-Duration Curve Quantile Estimation Using Selected Probability Distributions for Bay and Estuary Restoration in the Gulf States

By Elena R. Crowley-Ornelas, William H. Asquith, and Scott C. Worland

Prepared in cooperation with the Gulf Coast Ecosystem Restoration Council

Scientific Investigations Report 2022–5051

U.S. Geological Survey, Reston, Virginia: 2023

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

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

Suggested citation:
Crowley-Ornelas, E.R., Asquith, W.H., and Worland, S.C., 2023, Generalized additive model estimation of no-flow fractions and L-moments to support flow-duration curve quantile estimation using selected probability distributions for bay and estuary restoration in the Gulf States: U.S. Geological Survey Scientific Investigations Report 2022–5051, 35 p., https://doi.org/10.3133/sir20225051.

Associated data for this publication:
Asquith, W.H., Knight, R.R., and Crowley-Ornelas, E.R., 2020, RESTORE/fdclmrpplo—Source code for estimation of L-moments and percent no-flow conditions for decadal flow-duration curves and estimation at level-12 hydrologic unit codes: U.S. Geological Survey software release, version 1.0.2, https://doi.org/10.5066/P93CKH92.

Robinson, A.L., Asquith, W.H., Crowley-Ornelas, E., and Knight, R.R., 2021, Estimated quantiles of decadal flow-duration curves using selected probability distributions fit to no-flow fractions and L-moments predicted for streamgages and for pour points of level-12 hydrologic unit codes in the southeastern United States, 1950–2010: U.S. Geological Survey data release, https://doi.org/10.5066/P9MV8BYR.

Robinson, A.L., Asquith, W.H., and Knight, R.R., 2019, Summary of decadal no-flow fractions and decadal L-moments of nonzero streamflow flow-duration curves for National Hydrography Dataset, version 2 catchments in the southeastern United States, 1950–2010: U.S. Geological Survey data release, https://doi.org/10.5066/P9Z4PM55.

Robinson, A.L., Worland, S.C., and Rodgers, K.D., 2020, Estimated daily mean streamflows for HUC12 pour points in the southeastern United States, 1950–2009: U.S. Geological Survey data release, https://doi.org/10.5066/P9P36GXZ.

U.S. Geological Survey, 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, https://doi.org/10.5066/F7P55KJN.

ISSN 2328-0328 (online)

Acknowledgments

This research was funded by the Gulf Coast Ecosystem Restoration Council (RESTORE Council). The work by Scott Worland was done while serving as a Hydrologist with the U.S. Geological Survey.

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
meter (m)	1.094	yard (yd)
	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Flow rate	
cubic meter per second (m^3/s)	70.07	acre-foot per day (acre-ft/d)
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)
cubic meter per second (m^3/s)	22.83	million gallons per day (Mgal/d)

Datum

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

Abbreviations

AEP4	Asymmetric exponential power distribution
cGAM	Censoring-extended generalized additive model
DAR	Drainage-area ratio
dFDC	Decadal flow-duration curve
EPA	U.S. Environmental Protection Agency
FDC	Flow-duration curve
GAM	Generalized additive model
GLO	Generalized logistic distribution
GNO	Generalized normal distribution
KAP	Kappa distribution
LOO	Leave-one-out
MLR	Multilinear regression
NHDPlus	National Hydrography Dataset Plus
NSE	Nash-Sutcliffe model efficiency coefficient
NWIS	National Water Information System
RESTORE Act	Resources and Ecosystems Sustainability, Tourist Opportunities, and Revived Economies of the Gulf Coast States Act
RMSE	Root mean square error
SVM	Support vector machine
USGS	U.S. Geological Survey

Generalized Additive Model Estimation of No-Flow Fractions and L-Moments to Support Flow-Duration Curve Quantile Estimation Using Selected Probability Distributions for Bay and Estuary Restoration in the Gulf States

By Elena R. Crowley-Ornelas, William H. Asquith, and Scott C. Worland

Abstract

Censored and uncensored generalized additive models (GAMs) were developed using streamflow data from 941 U.S. Geological Survey streamflow-gaging stations (streamgages) to predict decadal statistics of daily streamflow for streams draining to the Gulf of Mexico. The modeled decadal statistics comprise no-flow fractions and L-moments of logarithms of nonzero streamflow for six decades (1950–2009). These statistics represent metrics of decadal flow-duration curves (dFDCs) derived from about 10 million daily mean streamflows. The L-moments comprise the mean, coefficient of L-variation, and the third through fifth L-moment ratios. The GAMs were fit to the statistics from 941 streamgages and 2,750 streamgage-decades by using watershed properties such as basin area and slope, decadal precipitation and temperature, and decadal values of flood storage and urban development percentages. The GAMs then estimated decadal statistics for 9,220 prediction locations (stream reaches) coincident with outlets of level-12 hydrologic unit codes. Both entire dataset (whole model) and leave-one-watershed-out model results are reported. No-flow fractions are censored data, and Tobit extensions to GAMs were used to model ephemeral streamflow conditions. Conversely, uncensored GAMs were used for estimation of the L-moments. The GAMs are shown, by coverage probabilities, to construct reliable 95-percent prediction limits. An example shows how no-flow fractions and L-moments may be used to approximate dFDCs by using selected probability distributions (mathematical formulas) including the asymmetric exponential power, generalized normal, and kappa distributions.

Introduction

The U.S. Geological Survey (USGS) and the U.S. Environmental Protection Agency (EPA) are collaborating on a project, in cooperation with the Gulf Coast Ecosystem Restoration Council (2019a), to provide vital information on the timing and delivery of freshwater to streams, bays, estuaries, and wetlands of the Gulf Coast, United States (Gulf Coast Ecosystem Restoration Council, 2019b). One project objective is to quantify alteration and natural statistical characteristics of streamflow regimes, such as minimum and maximum flows or other properties of the distributions of daily streamflow, as defined by flow-duration curves (FDCs). FDCs for this study represent the cumulative percentiles that streamflow is equaled or exceeded within a given time period (decade).

Hydrologic alteration of FDCs have been documented in more than 86 percent of monitored streams nationally including much of the Gulf Coast region (Carlisle and others, 2011). Such alterations are thought to be the primary cause for ecological impairment in many river and stream ecosystems (Carlisle and others, 2010, 2011). Restoration of freshwater inflows can positively affect shellfish, fisheries, habitat, and water quality in streams, rivers, and estuaries (Hildebrand and Gunter, 1953; Copeland, 1966; Alber, 2002; Powell and others, 2002; Carlisle and others, 2010, 2011). Stakeholders and decision makers are increasingly turning their attention to the restoration of streamflows as part of a holistic approach to restoring water quality and habitat and to protecting and replenishing coastal and marine biological resources and the livelihoods that depend on them (USGS, 2020).

An assessment of temporal and spatial trends in streamflow delivery to Gulf Coast estuaries can improve the understanding of potential drivers of change in estuarine health. Estimated streamflows as expressed by estimated L-moments and estimated FDCs at unmonitored (ungaged) stream locations for decadal time scales are critical to trend assessments. Estimated streamflows for the RESTORE Act funded Baseline

Flow project are based in part on combinations of either regional regression (this study) or machine-learning (Worland and others, 2019b) statistical analyses and interpretive syntheses. Both methods use large quantities of daily mean streamflows recorded by more than 900 USGS streamflow-gaging stations (streamgages). These streamgages are distributed throughout the Gulf Coast and proximal regions of the RESTORE project boundary (Crowley-Ornelas and others, 2019c). Worland and others (2019c) focused on estimation of FDC quantiles (ordinates of the FDC) for perennial streams in the study area.

In contrast to the aforementioned RESTORE Act project publications, the current (2022) study is focused on FDC estimation through the estimation of L-moments of the nonzero streamflow with an innovative method for the handling of no-flow fractions. The L-moments can be used to fit probability distributions, and such fits in turn provide FDC quantiles. Streamflow estimation at ungaged locations is based on the statistical coupling between observed FDC statistics and myriad potential predictor variables or watershed properties. The properties considered include physical, physiographic, and land-use categories as well as generalized hydrometeorologic, hydrogeologic, and water-resources development (dams and reservoirs) characteristics. The FDC statistics of primary description in this report are the no-flow (zero-flow) fractions and L-moments (Hosking and Wallis, 1997; Karian and Dudewicz, 2010; Asquith 2011, 2018; Asquith and others, 2017) of the nonzero component of decadal FDCs (dFDC). Finally, the estimation of these statistics is an action consistent with the Gulf Coast Ecosystem Restoration Council's commitment to science-based decision making through the "development of science tools to support freshwater inflow restoration and prioritization of future conservation actions" (Gulf Coast Ecosystem Restoration Council, 2016).

Purpose and Scope

The purposes of this report are (1) to describe results of generalized additive models (GAMS) for FDC quantile estimation through GAM regionalization of L-moments and fitting selected probability distributions to the estimated L-moments and (2) to accompany supporting algorithms and computations documented in and sourcing from the RESTORE/fdclmrpplo software repository (Asquith and others, 2020) with dependency support from the RESTORE/aregis software repository (Asquith and others, 2019). Robinson and others (2019) released decadal predictions of streamflow at more than 900 streamgages and approximately 9,000 level-12 hydrologic unit code (HUC12) pour points (outlets) in the study area from 1950 to 2009. The dFDC quantiles using selected probability distributions for streamgages and HUC12 pour points are available in Robinson and others (2021). The scope of this report parallels the scope of the previously mentioned data release and software repositories.

Study Area

The study area (fig. 1) comprises drainage basins to the Gulf of Mexico from the United States excluding the Florida Everglades, the Mississippi River Basin, and the Rio Grande Basin. The study area includes all or part of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, Missouri, Oklahoma, Tennessee, and Texas. The study area boundary (Crowley-Ornelas and others, 2019c) covers 1,004,220 square kilometers (km^2) and is coincident at the finest scale to HUC12 boundaries of the Watershed Boundary Dataset (USGS, 2018). The map (fig. 1) was developed by using spatial algorithms of Bivand and others (2013) and Pebesma and others (2018) and highlights the locations of 956 USGS streamgages and the Edwards aquifer outcrop in Texas (Texas Commission on Environmental Quality, 2018). Special consideration was made for some streamgages proximal to the Edwards aquifer outcrop, which is the reason for the aquifer outcrop's depiction.

Computational Notes

The RESTORE/fdclmrpplo software storage location (repository) (Asquith and others, 2020) provides source code and extensive documentation that supports this study. The input data are available at https://doi.org/10.5066/P9OD7FAL, https://doi.org/10.5066/P91DOYCJ, and https://doi.org/10.5066/P9KXTDU4 (Crowley-Ornelas and others, 2019a; Crowley-Ornelas and others, 2019b; Crowley-Ornelas and others, 2019d, respectively). The RESTORE/fdclmrpplo repository was used to produce the Robinson and others (2019) data release of the streamgage estimates and HUC12 predictions of no-flow fractions and L-moments of nonzero streamflow that this study describes.

The data on the maps shown in this report are limited to the 2000s. The maps have analogs within the RESTORE/fdclmrpplo repository through comprehensive map-based visualization as algorithm diagnostics for the six decades of streamgage-specific estimates and predictions at HUC12s (Asquith and others, 2020). The RESTORE/fdclmrpplo repository includes other maps of select inputs (files DiagMaps_inputs.pdf and DiagMaps_inputs.tex) and standard errors of fit (files DiagMaps_sefits.pdf and DiagMaps_sefits.tex). The components of the "site map" depiction by software analogous to figure 1 are located in the directory ./scripts/site_map within the RESTORE/fdclmrpplo repository (Asquith and others, 2020).

Flow-Duration Curves and Additional Background

FDCs provide an efficient means of describing the distribution of streamflow at a site and are commonly used as a means of depicting central tendency, variation (interquartile range), extremely low- or high-flow regimes, and depending

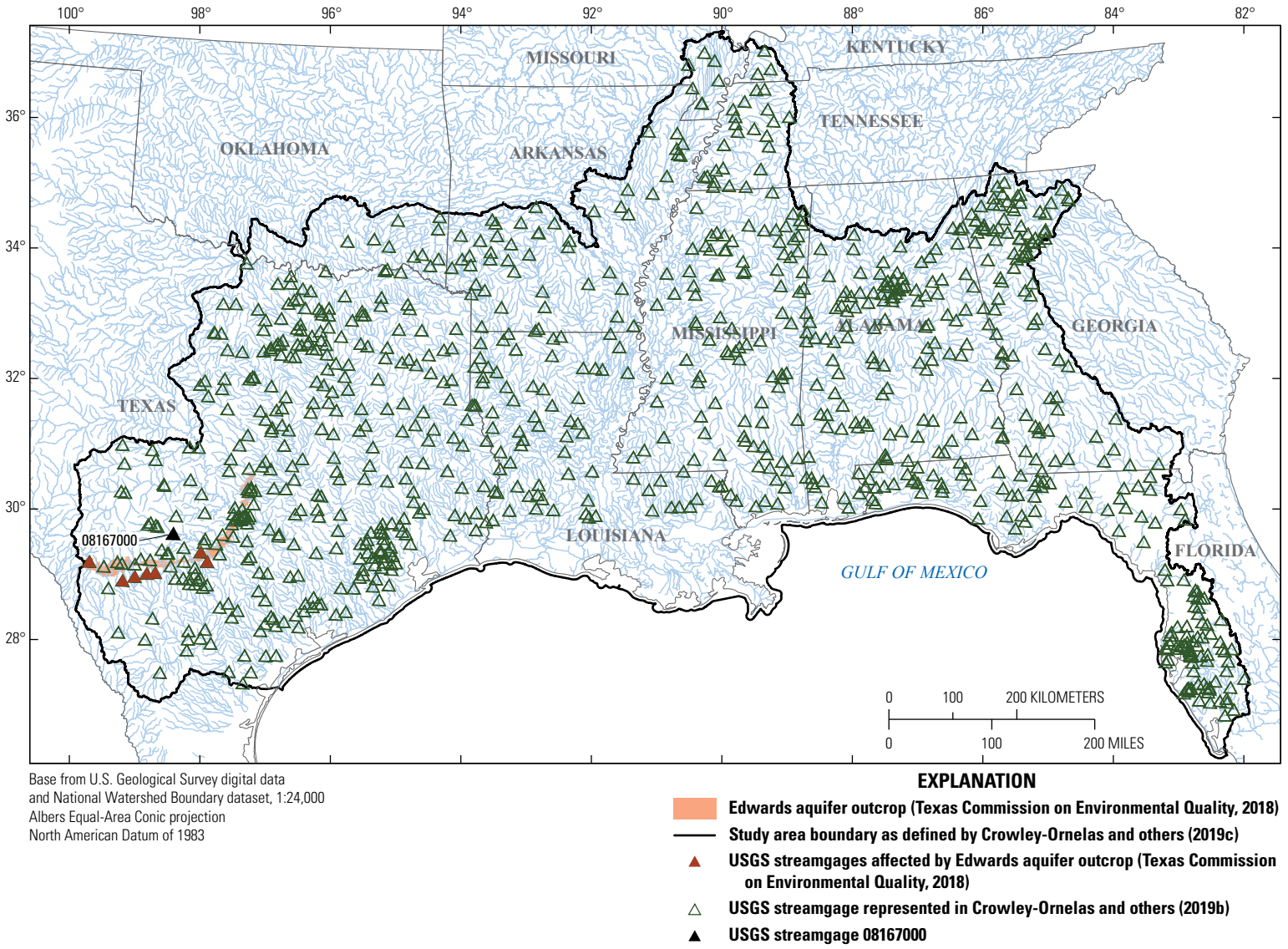


Figure 1. Study area, stream network (U.S. Environmental Protection Agency, 2018a), locations of 956 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) (Crowley-Ornelas and others, 2019b) with at least one complete decade of record during 1950–2009 and highlighting seven streamgages thought to be substantially affected by proximity to the Edwards aquifer outcrop.

on the site, the percentage of no-flow days. An FDC depicts the probability with which a given streamflow for a study period will be either greater than or less than a given value. For example, if a decade of flows for a site has a 30th percentile of 100 cubic meters per second (m^3/s), then 30 percent of the decade the streamflow volume will be less than or equal to 100 m^3/s, while 70 percent of the decade, streamflow will be equal to or greater than 100 m^3/s. Myriad complex and interrelated hydrometeorologic, hydrogeologic, and anthropogenic processes influence streamflow (Searcy, 1959; Vogel and Fennessey, 1994, 1995). At decadal time scales, these processes produce several orders of magnitude in streamflow regime within FDCs (Searcy, 1959; Vogel and Fennessey, 1994, 1995). FDCs do not represent autocorrelation (serial) or seasonal patterns that are endemic in streamflow data.

FDCs have many applications for researchers and resource managers and "are popular tools to estimate the amount of water available in a [watershed]" (Mendicino and Senatore, 2013). Some important applications of FDCs include water allocation, wastewater management, hydropower assessments, sediment transport, and protection of ecosystem health (Vogel and Fennessey, 1995; Castellarin and others, 2004; Blum and others, 2017). A spectrum of statistical approaches has been studied and applied by researchers to FDCs over many decades, and the focus of many researchers has been on the statistical transfer of information from the streamgages to the ungaged locations. FDCs can be used for generating time series of daily streamflows, a topic which has been explored recently for other study areas (Castellarin and others, 2004; Archfield and Vogel, 2010; Farmer and others, 2015a). Worland and others (2019a); Worland and others (2019c); Worland and others (2019d); and Robinson and others (2020) generated time series of daily streamflows for this study area as part of the RESTORE Act Baseline Flow project. The project also has produced statistically derived daily streamflows for two select river basins in the greater Gulf Coast region (Worland and others, 2019b). Method overviews are found in Castellarin and others (2004), Mendicino and Senatore (2013), Farmer and others (2015a), and Poncelet and others (2017). Practical applications require FDC estimation for unmonitored streams (Castiglioni and others, 2009; Over and others, 2014; Pumo and others, 2014; Pugliese and others, 2016; Requena and others, 2018). The requirement arises because streamgages cannot be installed at every location where streamflow information is needed (Farmer and others, 2015a).

For this study, there is an explicit distinction between the percentages of no-flow and the remainder (nonzero streamflow) of dFDCs. No-flow conditions, synonymous with zero-flow conditions and indicative of ephemeral or intermittent streams, are inherently more problematic to regionalize than overall mean annual streamflow because very low (including zero) streamflows tend to be greatly influenced by local hydrologic conditions in places lacking abundant rainfall and by basin area. The occurrence of no-flow days varies greatly for streams across the study area; this variation is attributable in part to the very humid (east) to semi-arid (west) breadth of climate across the study area.

No-Flow Fractions

The no-flow continuum ranges through (1) brief occurrences of zero flow during flow reversals in low-sloped or near-coastal streamgages, (2) continuous presence of water in the stream channel that is not passing downstream through the control cross section at the streamgage, (3) disconnected pools of water, and (4) long stretches of dry channels. Kroll and Stedinger (1999) state "at times the flow in the river may actually be zero, whereas in other instances flows may be nonzero but too small to be recorded by the available measuring instrumentation and consequently recorded as zero." The USGS minimum reported nonzero daily mean streamflow is 0.01 cubic foot per second (ft^3/s) (native units), and days with mean flow less than (<) 0.005 ft^3/s are reported as a zero.

Empirical FDCs can be constructed with data having any number of no-flow observations, sometimes greater than 90 percent, depending on the physical setting of the basin. Logarithmic transformation is common; however, such transformation is problematic when the empirical FDC contains no-flows. No-flow fractions are important to consider because stream locations with large fractions of zero flows were prevalent in many parts of the study area. Not rigorously accounting for no-flows could potentially bias results for basins having a likelihood of no-flow conditions and negatively affect statistical syntheses for water managers. It is important to note that not all regional studies of FDCs account for no-flow conditions. For example, Poncelet and others (2017) purposely restricted analysis of FDCs to permanently flowing rivers in their study of watersheds throughout France. Requena and others (2018) remark that in their approach to FDC estimation it "would be relevant to evaluate the performance of the approach when considering intermittent regimes."

In the study by Mendicino and Senatore (2013), seven models are presented for estimation of the FDCs at unmonitored locations in southern Italy based on regional study of 19 streamgages. The Mendicino and Senatore (2013) study is focused on a cross-validation study of model performance, and they treat the no-flow data as discontinuities in period-of-record FDCs. Also, their statistical models of the no-flow and the flowing FDCs are recombined through total probability, and their five deterministic models are split into two formulations (durations of no-flow and flow). In the current study, the authors use L-moments of the nonzero FDCs to estimate parameters for select probability distributions (mathematical formulas) to serve as curvilinear functions to mimic observed FDC shape. The authors develop a linear regression equation to estimate the "nonzero flow duration fraction" from basin area, change in elevation (top to bottom) in the watershed, and other predictive indices.

The statistical methods described herein address the potential bias when including nonzero flow through censored statistical techniques. Basic background information on censored data and working with such data can be found in Helsel and others (2020) and Helsel (2012). Extensive algorithmic details are found in Henningsen (2018) and Therneau and Lumley (2018) for the R language (R Core Team, 2019). Because of the common use of logarithmic transformation in the hydrologic sciences and the infiniteness of transformed zeros, other authors have considered the problem as censored (Wang and Singh, 1995; Kroll and Stedinger, 1999; Mendicino and Senatore, 2013). Kroll and Stedinger (1999) considered the zeros in the context of logarithmic modeling as one of truncation or so-called type I censoring. Kroll and Stedinger (1999) propose what is known as Tobit regression (Tobin, 1958) to alleviate biases caused by low-value truncation at zero or other minimum streamflow thresholds. Their work could be considered an early study of no-flow censoring and its mitigation in ordinary least squares, regression-like analyses. Kroll and Stedinger recommend the use of Tobit-based methods over ordinary least squares after completing numerical study and simulations. The more flexible GAMs used in this study are partially based on the use of Tobit methods.

For the current (2022) study, no-flows are viewed as a type of censored data for two reasons: (1) such streamflows are arguably censored values because each is technically <0.01 ft^3/s (<2.83*E*−4 m^3/s, in scientific notation) though the USGS National Water Information System (NWIS) database reports zero (USGS, 2019); and (2) conceptually the idea of "zero flow" itself represents a continuum of hydrologic conditions. One statistical approach for censored-data analysis is "survival regression" (Crawley, 2007; Helsel, 2012). GAMs were chosen for this study because they can be made analogous to survival regression. The benefits to using GAMs include regression-like inclusion of parametric and nonparametric smooth terms, and spatially dependent terms. A type of GAM called a censoring-extended GAM (abbreviated as cGAM) is used in this study because it allows for censored regression to account for no-flows.

Data Sources and Statistical Methods

National Hydrography Dataset, Watershed Properties, and Solar Radiation Data Sources

Locations for prediction were identified by using a combination of the National Hydrography Dataset Plus (NHDPlus) (EPA, 2018a) and the Watershed Boundary Dataset (USGS, 2018). Specific locations for estimating FDCs were identified by using location identifiers (COMIDs) of stream reaches coincident with the pour points of HUC12 watersheds, which have a median accumulated drainage area of about 200 km^2. The study area has approximately 10,000 HUC12s, but estimation is restricted to about 9,000 prediction locations. A COMID is a unique identification number of a stream reach, and ideally each HUC12 pour point has a unique COMID. Some COMIDs are associated with the pour points of multiple HUC12s, however, and HUC12-addressing schemes occasionally have been revised (Crowley-Ornelas and others, 2019d). COMID and HUC12 uniqueness is logical for headwater basins, but downstream assemblages and tributary confluences can produce geometries with stream-reach ambiguities. This subtle issue is an innate result of assigning line geometries to dynamic geographic features and is not addressed in this study.

Crowley-Ornelas and others (2019d) compiled a suite of watershed properties, meteorological properties, land use, and other anthropogenic measures for streamgage and prediction locations. Other physical data such as physiographic region, ecological region, and soil classifications are included in the dataset. The mutable basin characteristic data were sourced from Wieczorek and others (2018) and were aggregated by decade for the six decades during 1950–2009. Because of interest in assessing the alteration of FDCs and their estimations at ungaged locations, these decadal properties contribute to the logic of using decades for the streamflow aggregation and empirical FDC construction. Meteorological forcing through land surface solar input and landscape covariates compiled by Crowley-Ornelas and others (2019d) can affect no-flow fractions. For this study, interest resides in the smooth continuous nature of solar radiation to provide a purely spatial metric to complement spatial location (projected coordinates of streamgages and COMIDs). Therefore, the 1998–2009 mean values are assigned to every decade. As a result, although different decades are shown in Crowley-Ornelas and others (2019a), the same site has the identical solar radiation for each decade.

Decadal Daily Mean Streamflow and Streamgages Used for Modeling

Crowley-Ornelas and others (2019b) selected 1,290 USGS streamgages in the study area from the NWIS database (USGS, 2019). The 1,290 streamgages were subsetted based on several categories: (1) river-stage-only sites, (2) sites coded as measuring spring flow, (3) sites representing side weirs, (4) sites with partial-river flow in individual tail races below reservoirs, and (5) sites with tidal influences. Daily streamflows from the beginning of each streamgage record through December 31, 2009, were then downloaded from NWIS (USGS, 2019) and grouped by decade. A maximum of 70 missing values per decade (loosely equivalent to 7 missing days per year, but the requisite computations were not year-by-year based) was considered a whole decade. After filtering for the whole-decade criteria, 956 streamgages and 2,804 streamgage-decades were available for the six decades during 1950–2009.

Additional data screening was made for the 956 streamgages, and attendant disclosures specific for this study are listed in RESTORE/fdclmrpplo (Asquith and others, 2020). The screening resulted in 941 streamgages (fig. 1), which represent 2,750 streamgage-decade records.

Statistical Methods for Regionalization of Decadal No-Flow Fractions and Decadal L-Moments

The 941 streamgages and 2,750 decades of streamflow data were used with the watershed properties and other characteristics to construct the final GAMs presented by the RESTORE/fdclmrpplo software repository (Asquith and others, 2020). The no-flow fraction was studied by using a cGAM (censored), whereas decadal L-moments were studied by using GAMs (uncensored).

Relations between observed no-flow fractions and potential covariates are complex, which is caused in part by nonlinear processes functioning at various scales and interactions with physical factors such as rainfall patterns and other atmospheric influences, land-use and streamflow regulation patterns, and other difficult-to-identify factors. GAMs (Wood, 2008, 2017; Stasinopoulos and others, 2017) were selected to explore these complex relations in lieu of the more conventional (traditional) survival regression (Crawley, 2007; Therneau and Lumley, 2018), which is an extension of multilinear regression (MLR) (Faraway, 2005, 2006).

A GAM uses relations between a response variable and an additive combination of various parametric terms and smooth terms (smooth functions) (Wood, 2017). The incorporation of smooth functions can be an advantage to GAMs over MLR (the underpinning of survival regression) because these functions provide linearly additive terms of nonlinear relations in the data. For the GAM of decadal no-flow fractions reported here, the algorithms of Fang (2017), which provide for data censoring (Helsel, 2012), coupled to those of Wood (2018) were used to estimate the parametric terms and largely default parameters (arguments) of the gam() function of the mgcv package in the R language (R Core Team, 2019). The general form of GAMs for this study is

$$y_i = X_i\theta + f(x_i; \Psi) + \cdots + s(E_i, N_i; \eta) + \varepsilon_i, \qquad (1)$$

where

y_i is the response variable ($base_{10}$ logarithmic transformation) for the ith streamgage-decade record;

X_i is a model vector including an optional intercept for strictly parametric and suitably transformed predictor variables;

θ is a parameter matrix;

f is a smooth function of the predictor variable x_i controlled by smoothing settings Ψ_i;

... represent additional smooth terms as needed;

$s(E_i, N_i; \eta)$ is the smooth on the easting (E_i) and northing (N_i) Albers Equal-Area projected coordinates (Anderson and Mikhail, 1998) of the respective longitude and latitude and smoothing parameters η; and

ε_i are errors.

The $X_i\theta$ term is the familiar multilinear (parametric) regression component of a GAM, and the θ are regression coefficients (conventional slope terms). The gam() function constructs the GAM (Wood, 2018).

Leave-One-Watershed-Out and Coverage Probabilities

Whole-sample modeling has a potential for underestimating prediction uncertainties because model performance is assessed on its fit to the "training data" as opposed to performance on independent data. An elementary approach to assess model performance is out-of-sample or leave-one-out (LOO) testing. This is a form of cross validation, and specifically for this study, leave-one-watershed-out was used. Watersheds (the streamgage without regard to the number of decades available) in this study are treated as the fundamental sampling unit; therefore, watersheds were selected one by one for removal, the same GAM form was fit, and predictions for the streamgage and its decades were recorded.

For flexible modeling techniques, such as GAMs with many smooths, there usually are no simple formulas to compute the expected LOO fit. Practitioners commonly use cross validation to assess model performance on independent data by using rigorous numerical computation in place of theoretical analysis; however, GAMs are in a class of modeling technique similar to regression for which some theoretical analysis is applicable (Faraway, 2005; Helsel and others, 2020). The algorithms of Wood (2018) implement internal validation schemes to mitigate overfitting the data, and if it can be shown that the model accurately computes its prediction uncertainties, then the predicted uncertainty at the prediction locations by the GAMs is reliably documented.

Given intent to provide estimates for the 9,220 prediction locations in the study area, documentation of prediction limits and coverage probabilities of the limits is important. Highly technical disclosures regarding the coverage probabilities are reported herein and 95-percent prediction limits for the streamgage-decade records and the prediction locations are reported in Robinson and others (2019).

The algorithms by Wood (2018) provide standard errors of fit for predictions. These standard errors (output se.fit from the predict.gam function in Wood's nomenclature) can be converted into prediction limits as distinguished from confidence limits. The prediction limits are an expression of uncertainty under circumstances of estimation for new information. Helsel and others (2020) and Faraway (2005) provide the necessary background to distinguish between confidence and prediction limits.

The computed prediction limits of the flowtime from the censored GAM of the no-flow fraction (cGAM-PPLO) and later for other GAMs of the L-moments were based on the 0.05 probability (95-percent limits) of the quantile of the t-distribution ($\pm t_{\alpha/2,d}$) for two-sided probability α and degrees of freedom d by using the sample size (number of streamgage-decade records) minus the summation of the estimated degrees of freedom (edf from the GAM object in Wood's nomenclature) as the d and α = 0.05. The σ in the prediction limit computation is the residual standard error (residual.scale from the GAM object in Wood's nomenclature). An equivalent "leverage" (h) (Helsel and others, 2020) was computed as the square of the ratio (se.fit/residual.scale). The 95-percent lower- and upper-prediction limits for a given estimate ŷ were computed as $\hat{y} \pm t_{\alpha,d} \times \sigma \times \sqrt{1+h}$. The addition of the unity within the square root changes the limit computation from confidence to prediction. The authors of the current (2022) study are not aware of prior clarification in the literature of prediction limit computation from Wood (2018) algorithms. Further details are available in Asquith and others (2020).

Coverage probabilities are the proportions that a computed interval contains the true value of interest. For this study and for whole model and LOO model results, coverage probabilities were computed by using the 95-percent prediction limits previously described. If general assumptions about the computation of a prediction intervals are met, then the coverage probability will equal the definition of the computed prediction limit (95 percent). The coverage probabilities are compared to the estimated 95-prediction limits as part of the discussion of model diagnostics.

Summary of Generalized Additive Model Computations for No-Flow Fractions and L-Moments

Censored Generalized Additive Model and Estimation of Decadal-Flow Days (Flowtimes)

The decadal no-flow fractions were conceptualized as product lifetimes and hence as a survival analysis problem. For this study, the days of nonzero streamflow ("flowtimes") are the response variable in the model and are right-tail censored. The variable y_i, which represents the count of the number of days of nonzero streamflow in a decade, is potentially censored. The regression is then a form of Tobit modeling (Tobin, 1958), and a Tobit model is any of a class of regression models in which the observed range of the dependent variable is censored in some way. General and contextually applicable background on Tobit regression is also found in Kroll and Stedinger (1999), Helsel (2012), and Henningsen (2018). The Tobit type I distribution (family) (Fang, 2017), in particular, was used in lieu of the default Gaussian family for the gam() function (Wood, 2018).

The censored GAM for this study (fig. 2) is abbreviated as cGAM-PPLO and is the PPLO object within RESTORE/fdclmrpplo (Asquith and others, 2020). Whole- and LOO-model estimates by the cGAM-PPLO for the decadal no-flow fractions for the streamgage-decade records are available in a data release by Robinson and others (2019). Also available for the whole and LOO model are estimated lower and upper 95-percent prediction limits for the flowtimes and no-flow fractions. The data release also includes the whole-model residual standard error and individual standard errors of estimate for the whole-model predictions (Robinson and others, 2019).

For development of a given GAM including flowtimes or other statistics, decisions are required including which variables to use, which variables are to be parametric or smoothed, and which smoothing parameter settings to use. The decisions are based not only on performance for no-flow regionalization but also for regionalization of the L-moments of the nonzero streamflow FDC. When justified, structural similarity among all the models described herein is desirable to achieve some qualitative consistency in the ensemble of predictions.

The authors consider it useful to avoid smoothing terms when a parametric term can be used to ease interpretive burden. In an effort to avoid smoothing when possible, the properties measured in percentiles (Crowley-Ornelas and others, 2019d) were arcsine-transformed ($(2 \times \text{asin}(\sqrt{p/100}))$ for a variable p, in percent) prior to being inserted as additive parametric terms in the GAM. This transformation is intended to amplify the end members of 0 (zero) percent on the left and ideally up to 100 percent on the right. The "developed" and "grassland" land-use variables were transformed in this manner; the regression coefficients listed are multipliers on the value of the transformed variables. Experimental evaluation showed some benefit in magnifying the importance or weight in the regression of near-zero and near-100 percent. Based on their common scale of measurement, the developed proportion is inferred as more influential in the model than the grassland proportion (fig. 2). The developed variable represents the percentage of the basin classified as developed, and grassland is the percentage of the basin dominated by graminoid/herbaceous vegetation.

High percentages of development (fraction urbanization, fig. 2) are typically isolated to metropolitan areas, and most of the study area has development fractions that are less than about 1 percent. The positive coefficient (in column labeled "Estimate" in fig. 2) for development is evidence that as development increases, flow becomes more likely—though caution is needed for process interpretations or expectations of individual variables in complex regression analyses. Percentages of grassland are low throughout much of the study area, but grassland percentages increase further south into Florida as well as in Texas west of the 96th meridian. The positive coefficient for grassland (fig. 2) indicates that an increase in grassland is associated with larger no-flow fractions, but the magnitude of the effect is about half that of development.

```
gam(flowtime ~ s(basin_area)          + s(ppt_mean, k=5) +
               s(temp_mean,  k=4)     + s(dni_ann,  k=7) +
               developed + grassland + bedperm + decade - 1 +
               s(x, y))
Family: Censored normal (sigma=0.104) (Tobit family); Link function: identity
Parametric coefficients: (flowtime, number of flowing days per decade)
                     Estimate  Std.Error         z   p-value
developed             0.11698    0.01148    10.186   < 2e-16  (fraction urbanization)
grassland             0.06167    0.02145     2.874   0.00405  (fraction grassland)
bedpermacc_bedperm_1  3.65636    0.01432   255.247   < 2e-16  (not a principal aquifer)
bedpermacc_bedperm_2  3.48706    0.03749    93.025   < 2e-16  (sandstone)
bedpermacc_bedperm_3  3.69766    0.01500   246.559   < 2e-16  (semi-consolidated sand)
bedpermacc_bedperm_5  3.64904    0.01667   218.950   < 2e-16  (sandstone and carbonate
    rocks)
bedpermacc_bedperm_6  3.68444    0.02085   176.698   < 2e-16  (unconsolidated sand and
    gravel)
bedpermacc_bedperm_7  3.68616    0.02365   155.891   < 2e-16  (carbonate rock)
decade1950            0 (zero)        --        --        --
decade1960            0.00098    0.01252     0.078   0.93761
decade1970            0.02043    0.01365     1.496   0.13459
decade1980            0.02882    0.01282     2.247   0.02465
decade1990            0.03505    0.01420     2.468   0.01359
decade2000            0.01292    0.01244     1.039   0.29894
---
Approximate significance of smooth terms:
                edf Ref.df  Chi.sq   p-value
s(basin_area)  1.212  1.398  315.29   < 2e-16      (basin [watershed] drainage area)
s(ppt_mean)    2.937  3.449   29.09  2.84e-06      (mean precipitation, decadal)
s(temp_mean)   2.423  2.762   38.78  2.81e-08      (mean temperature, decadal)
s(dni_ann)     2.318  2.978   10.13    0.0154      (annual average solar radiation, fixed)
s(x,y)        21.720 25.652  227.83   < 2e-16      (Albers equal area coordinates)
---
Additional Abbreviations:
Chi.sq, Chi-squared distribution statistic; edf, estimated degrees of freedom;
gam(...), generalized additive model; k, knot points for smooth function;
Ref.df, reference degrees of freedom; s(...), a smooth function; Std, standard; and
z, Z-score of error distribution
```

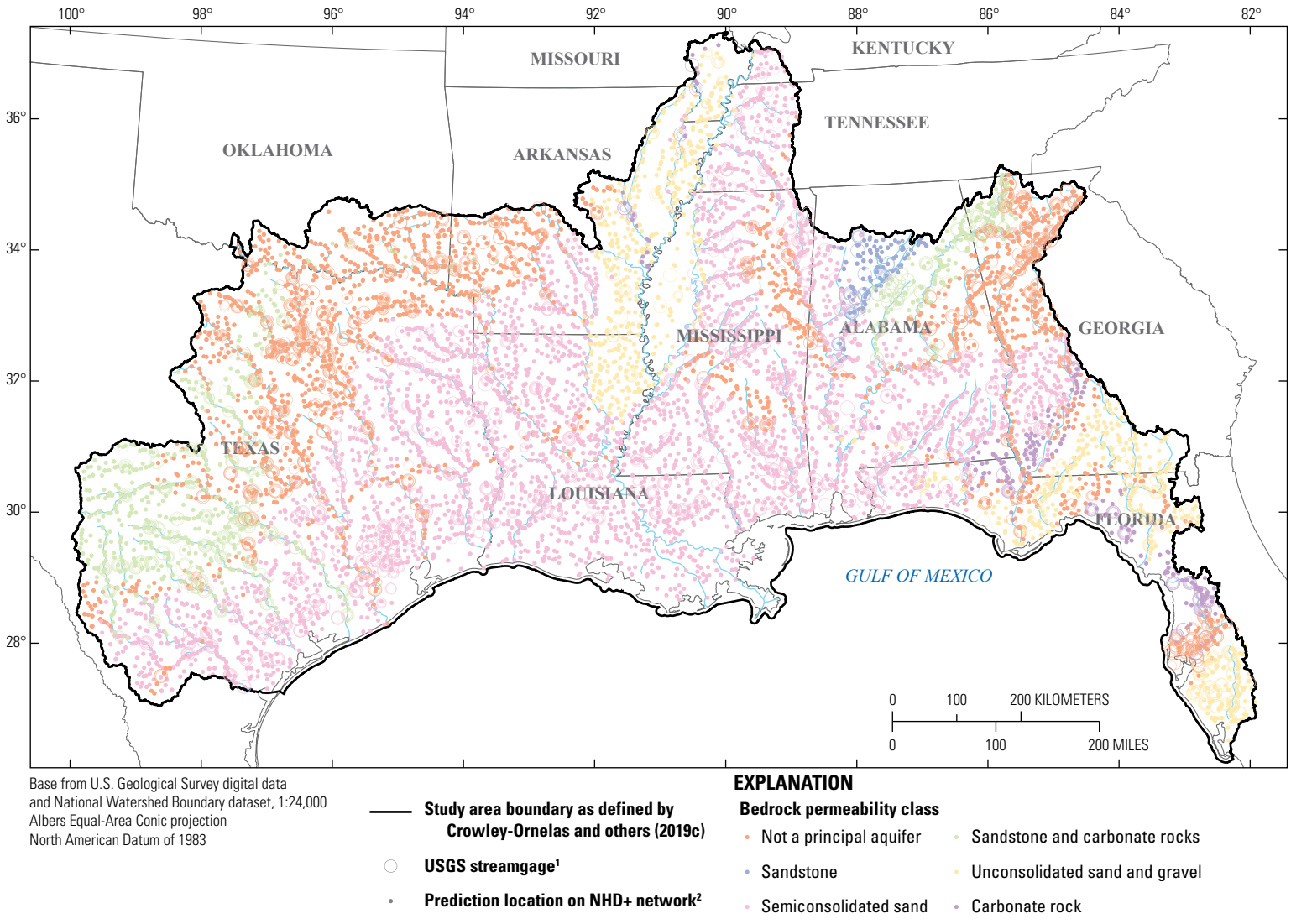
Figure 2. Summary list of output from mgcv::gam(…) (Fang, 2017; Wood, 2018) representing the final predictive model (cGAM-PPLO) of decadal flowtimes that in turn can be converted to decadal no-flow fractions.

The decade coefficients generally have probability values (*p*-values) larger than about 0.01 (fig. 2), which suggests that decade effects in these data are small when in the presence of the other predictors from the model. It must be stated that the precipitation, temperature, developed, and grassland fractions (ppt_mean, temp_mean, developed, and grassland, respectively) are mutable (based on decades). Tests by the authors indicate that the annual solar radiation (dni_ann) does not contribute much to the model compared to the smooth on the projected Albers Equal-Area coordinates of the streamgages (s[x,y]), precipitation, temperature, and grassland fraction. However, annual solar radiation has been retained for use for other GAMs herein.

The bedrock permeability (bedperm) (fig. 3) is a factor variable assigned to the class corresponding to most of the watershed and contains six classes (Crowley-Ornelas and others, 2019d). The bedrock permeability coefficients are additive to the intercept of the model and can be interpreted in part by their magnitudes (fig. 2) and divergence from an overall mean of the six coefficients. The lowest coefficient (about 3.49) is for acc_bedperm_2 (sandstone), and the highest three coefficients (about 3.68) are for acc_bedperm_3, acc_bedperm_6, and acc_bedperm_7 for semiconsolidated sand, unconsolidated sand and gravel, and carbonate rock, respectively. The coefficients (fig. 2) indicate that flowtimes are slightly larger by about 0.20 $\log_{10}$(days) for the latter three classes than for sandstone.

Generalized Additive Models of Decadal L-Moments

An effort was made to select parallel variables for the different GAMs because it is important to generally have structurally similar individual GAMs per L-moment of nonzero streamflow. This helps foster L-moment predictions for each



[1]Symbol size represents one of the six decades: 1950s (smallest symbol) through 2000s (largest symbol). Symbol color represents bedrock permeability class.

[2]Symbol color represents bedrock permeability class.

Abbreviations: NHD+, National Hydrography Dataset Plus, version 2, locations by Crowley-Ornelas and others (2019d)

Figure 3. Bedrock permeability classes at 9,220 prediction locations (stream reaches) and locations of 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during 1950–2009. The classes for the majority of the watershed are acc_bedperm_1, not a principal aquifer; acc_bedperm_2, sandstone; acc_bedperm_3, semiconsolidated sand; acc_bedperm_5, sandstone and carbonate rocks; acc_bedperm_6, unconsolidated sand and gravel; and acc_bedperm_7, carbonate rock.

prediction location to be meaningful as a set for later probability distribution fitting. Of the non-censored L-moments, the mean nonzero streamflow was treated first and built with similarity to the cGAM-PPLO, but the bedrock permeability and grassland variables were no longer significant. The model (fig. 4) is abbreviated as GAM-L1 and is the L1 object within RESTORE/fdclmrpplo (Asquith and others, 2020).

Whole- and LOO-model estimates by the GAM-L1 for the decadal mean nonzero streamflows by streamgage-decade records are available in the data release by Robinson and others (2019). The data release identifies those streamgage-decade records in the model and the observed decadal means. Also listed for the whole and LOO models are estimated lower and upper 95-percent prediction limits, and the whole- and LOO-model Duan retransformation bias corrections (Helsel and others, 2020) are provided. The bias corrections are provided because logarithmic transformation was used for the GAM-L1. Logarithmic transformation was not used for the other L-moments.

The Duan retransformation bias correction (Duan, 1983) requires that model residuals are independent and homoscedastic. The first assumption appears to be reasonable upon review of regression performance, but minor non-homoscedasticity is present in part because the smallest watersheds have inherently more relative variation in streamflows than the largest river systems. The Duan retransformation bias correction (Duan, 1983) is 1.046 (Robinson and others, 2019). The Robinson and others (2019) data release also includes the whole-model residual standard error of the model and standard errors of estimate for the whole-model predictions. Note that the retransformation bias correction has not been applied to the values in Robinson and others (2019).

A summary of the GAM-T2 model is shown in figure 5 and is the T2 object within RESTORE/fdclmrpplo (Asquith and others, 2020). Whole- and LOO-model estimates by the GAM-T2 for the decadal coefficient of L-variation (T2) by streamgage-decade records are available in the data release by Robinson and others (2019). The data release identifies those streamgage-decade records in the model and the observed decadal means. Robinson and others (2019) include whole-model residual standard error and standard errors of estimate for the whole-model predictions.

The addition of flood storage (fig. 6) (U.S. Army Corps of Engineers, 2018) to the T2 statistic (fig. 5) as well as the higher L-moments that follow is deliberate. The flood storage is a highly significant predictor in the steepness and other

```
gam(L1 ~ s(basin_area)   + s(basin_slope, k=5) + s(ppt_mean, k=5) +
         s(dni_ann, k=7) + developed + decade - 1 + s(x, y))
Family: gaussian; Link function: identity
Parametric coefficients: (L1, logarithm of mean nonzero streamflow)
           Estimate  Std.Error  t-value   p-value
developed  0.083542   0.009774    8.547    <2e-16  (fraction urbanization)
decade1950 1.088644   0.007796  139.637    <2e-16
decade1960 1.029676   0.006204  165.959    <2e-16
decade1970 1.013305   0.006412  158.040    <2e-16
decade1980 1.033596   0.006326  163.396    <2e-16
decade1990 1.006226   0.006772  148.581    <2e-16
decade2000 0.999148   0.005886  169.743    <2e-16
---
Approximate significance of smooth terms:
                edf Ref.df        F  p-value
s(basin_area)   8.205  8.826 5752.923  < 2e-16     (basin [watershed] drainage area)
s(basin_slope)  3.075  3.592    5.284 0.000406     (basin slope)
s(ppt_mean)     3.973  3.999  321.284  < 2e-16     (mean precipitation, decadal)
s(dni_ann)      4.898  5.598   13.486 3.02e-14     (annual average solar radiation, fixed)
s(x,y)         26.395 28.523   27.514  < 2e-16     (Albers equal area coordinates)
---
R-sq.(adj) = 0.971   Deviance explained = 99 percent
GCV = 0.01736   Scale estimate = 0.017022   n = 2,750
---
Additional Abbreviations:
edf, estimated degrees of freedom; F, statistic of F-distribution;
gam(...), generalized additive model; GCV, generalized cross validation statistic;
k, knot points for smooth function; n, sample size; Ref.df, reference degrees
of freedom; R-sq.(adj), adjusted R-squared; s(...), a smooth function;
Std., standard; and t-value, statistic of t-distribution
```

Figure 4. Summary list of output from mgcv::gam(…) (Wood, 2018) representing the final predictive model (GAM-L1) of decadal mean (L1) nonzero streamflow.

shapes of the dFDCs but curiously not for the mean nonzero streamflow (fig. 4). The low predictive significance of flood storage for mean nonzero streamflow suggests that flood storage does not intrinsically destroy volumes of water (mean streamflows) but adjusts respective durations of flow.

A summary of the GAM-T3 model is shown in figure 7 and is the T3 object within RESTORE/fdclmrpplo (Asquith and others, 2020). Whole- and LOO-model estimates by the GAM-T3 for the decadal L-skew (T3) by streamgage-decade records are available in the data release by Robinson and others (2019). The data release identifies those streamgage-decade records in the model and the observed decadal values of L-skew. Estimated lower and upper 95-percent prediction limits are also listed for the whole and LOO models. The data release additionally includes the whole-model residual standard error and standard errors of estimate for the whole-model predictions.

A summary of the GAM-T4 model is shown in figure 8 and is the T4 object within RESTORE/fdclmrpplo (Asquith and others, 2020). Whole- and LOO-model estimates by the GAM-T4 for the decadal L-kurtosis (T4) by streamgage-decade records are available in the data release by Robinson and others (2019). The data release identifies streamgage-decade records in the model and the observed decadal values of L-kurtosis. Estimated lower and upper 95-percent prediction limits are also listed for the whole and LOO models. The data release additionally includes the whole-model residual standard error and standard errors of estimate for the whole-model predictions.

A summary of the GAM-T5 model is shown in figure 9 and is the T5 object within RESTORE/fdclmrpplo (Asquith and others, 2020). A precipitation term, though seen in other models, was removed from this model because a high *p*-value for precipitation in preliminary models. A high *p*-value indicates that precipitation is not statistically significant and does not add predictive value to the model. The fifth L-moment ratio is provided as a means to assess comparative distribution fits of 4- and 5-parameter distributions (Asquith, 2014). Whole- and LOO-model estimates by the GAM-T5 for the decadal fifth L-moment ratio (T5) by streamgage-decade records are available in the data release by Robinson and others (2019). The data release identifies streamgage-decade

```
gam(T2 ~ s(basin_area) + s(basin_slope, k=5) + s(ppt_mean, k=5) +
         s(temp_mean, k=4) + s(dni_ann, k=7) + developed +
         s(flood_storage, k=5) + decade - 1 + s(x, y))
Family: gaussian; Link function: identity
Parametric coefficients: (T2, coefficient of L-variation [L-CV] of nonzero streamflow)
            Estimate Std. Error t-value  p-value
developed  -0.027822   0.006697  -4.154 3.36e-05   (fraction urbanization)
decade1950  0.657704   0.005829 112.829  < 2e-16
decade1960  0.652890   0.004496 145.200  < 2e-16
decade1970  0.644659   0.004548 141.757  < 2e-16
decade1980  0.657293   0.004337 151.547  < 2e-16
decade1990  0.668857   0.004663 143.454  < 2e-16
decade2000  0.678731   0.004029 168.450  < 2e-16
---
Approximate significance of smooth terms:
                   edf Ref.df      F  p-value
s(basin_area)    3.635  4.627 44.636  < 2e-16     (basin [watershed] drainage area)
s(basin_slope)   3.749  3.965 19.194 1.38e-15     (basin slope)
s(ppt_mean)      3.805  3.976 11.090 6.90e-09     (mean precipitation, decadal)
s(temp_mean)     2.825  2.975  4.314 0.008948     (mean temperature, decadal)
s(dni_ann)       2.663  3.518  6.268 0.000132     (annual average solar radiation, fixed)
s(flood_storage) 3.015  3.483 27.376  < 2e-16     (flood storage in basin, decadal)
s(x,y)          28.339 28.942 55.705  < 2e-16     (Albers equal area coordinates)
---
R-sq.(adj) = 0.676   Deviance explained = 98.4 percent
GCV = 0.0078113   Scale estimate = 0.007655   n = 2,750
---
Additional Abbreviations:
edf, estimated degrees of freedom; F, statistic of F-distribution;
gam(...), generalized additive model; GCV, generalized cross validation statistic;
k, knot points for smooth function; n, sample size; Ref.df, reference degrees
of freedom; R-sq.(adj), adjusted R-squared; s(...), a smooth function;
Std., standard; and t-value, statistic of t-distribution
```

Figure 5. Summary list of output from mgcv::gam(…) (Wood, 2018) representing the final predictive model (GAM-T2) of decadal coefficient of L-variation (T2) of nonzero streamflows.

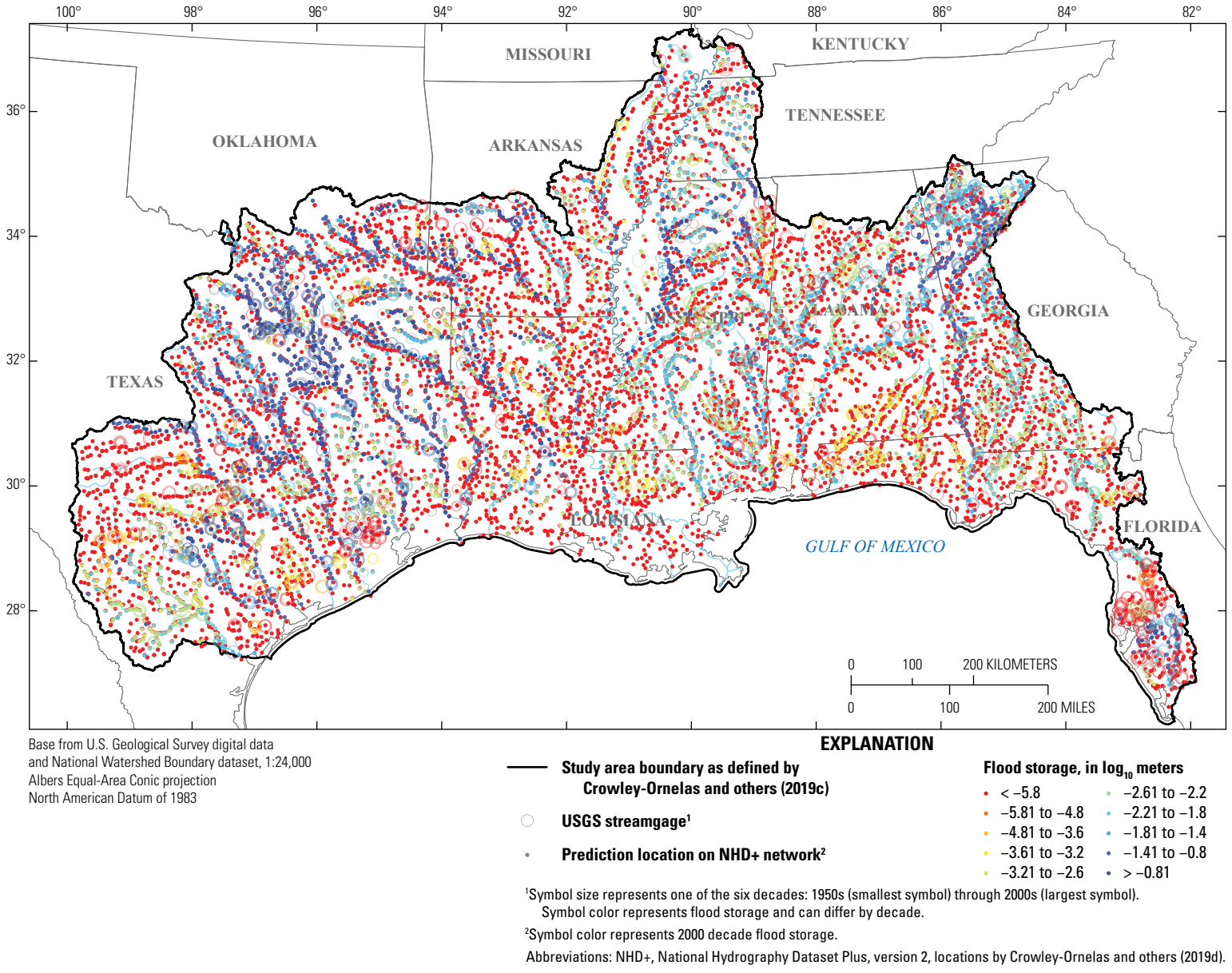


Figure 6. Flood storage values for the 2000s at 9,220 prediction locations (stream reaches) and locations of 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during the 1950–2009.

```
gam(T3 ~ s(basin_area) + s(basin_slope, k=5) + s(ppt_mean, k=5) +
         s(temp_mean,     k=4) + s(dni_ann, k=7) +
         s(flood_storage, k=5) + decade - 1 + s(x, y))
Family: gaussian; Link function: identity
Parametric coefficients: (T3, L-skew of nonzero streamflow)
           Estimate  Std.Error t-value  p-value
decade1950 0.610183   0.005460   111.8   <2e-16
decade1960 0.621047   0.004141   150.0   <2e-16
decade1970 0.596503   0.004115   145.0   <2e-16
decade1980 0.622346   0.004006   155.3   <2e-16
decade1990 0.624968   0.004354   143.5   <2e-16
decade2000 0.635266   0.003640   174.5   <2e-16
---
Approximate significance of smooth terms:
                    edf Ref.df      F  p-value
s(basin_area)     5.232  6.422 57.461  < 2e-16     (basin [watershed] drainage area)
s(basin_slope)    3.595  3.912  8.512 9.93e-07     (basin slope)
s(ppt_mean)       1.455  1.779  7.195 0.000989     (mean precipitation, decadal)
s(temp_mean)      2.755  2.950  5.289 0.001764     (mean temperature, decadal)
s(dni_ann)        4.907  5.616  7.588 2.28e-07     (annual average solar radiation, fixed)
s(flood_storage)  3.631  3.913 70.958  < 2e-16     (flood storage in basin, decadal)
s(x,y)           27.545 28.824 39.045  < 2e-16     (Albers equal area coordinates)
---
R-sq.(adj) = 0.698   Deviance explained = 98.4 percent
GCV = 0.00686   Scale estimate = 0.0067225   n = 2,750
---
Additional Abbreviations:
edf, estimated degrees of freedom; F, statistic of F-distribution;
gam(...), generalized additive model; GCV, generalized cross validation statistic;
k, knot points for smooth function; n, sample size; Ref.df, reference degrees
of freedom; R-sq.(adj), adjusted R-squared; s(...), a smooth function;
Std., standard; and t-value, statistic of t-distribution
```

Figure 7. Summary list of output from mgcv::gam(…) (Wood, 2018) representing the final predictive model (GAM-T3) of decadal-skew (T3) of nonzero streamflows.

records in the model and the observed decadal T5s. Estimated lower and upper 95-percent prediction limits are also listed for the whole and LOO models. The data release additionally includes the whole-model residual standard error and standard errors of estimate for the whole-model predictions.

Model Diagnostics

Selected model diagnostics are listed in table 1 for all six of the final models. The diagnostics include the Nash-Sutcliffe efficiency (NSE) coefficient, the root mean square error (RMSE), as well as coverage probabilities for both the whole and LOO models.

The NSE can range from $-\infty$ to 1. An NSE of 1 indicates a perfect fit between simulated and measured data. An NSE of 0 indicates that the model predictions are only as accurate as the mean of the measured data, and an NSE of less than 0 indicates that the mean of the measured data is a better predictor than the model. RMSEs succinctly measure accuracy and retain the units of the response data on which the models were constructed. RMSE values are never negative, and it is unclear how comparable RMSEs are across the dimensionless models (GAM-T2 through GAM-T5).

The cGAM and GAMs also were evaluated by using assessment of their prediction-limit coverage probabilities. cGAM performance was further evaluated by using (1) assessment restricted to just those predictions of no-flow, (2) assessment of counts of perennial and ephemeral predictions, and (3) independent check using a support vector machine (SVM) prediction method. SVMs are a type of machine learning with completely different foundational mathematics than GAMs (Asquith, 2020). The evaluation of prediction limit and veracity check on no-flow provide a holistic view of model performance with an emphasis on the no-flow fractions.

Coverage probabilities are a diagnostic as to whether the GAM algorithms correctly estimate their respective prediction errors. The coverage probabilities for the cGAM-PPLO model are listed in table 1. The coverage probability was computed as 1 minus the fraction of streamgage-decade records having the estimated flowtimes either less than the computed lower limits or greater than the upper limits. The listed coverage probabilities show that the computed limits are very slightly

```
gam(T4 ~ s(basin_area) + s(basin_slope, k=5) + s(ppt_mean, k=5) +
         s(temp_mean,     k=4) + s(dni_ann, k=7) +
         s(flood_storage, k=5) + decade - 1 + s(x, y))
Family: gaussian; Link function: identity
Parametric coefficients: (T4, L-kurtosis of nonzero streamflow)
           Estimate Std. Error t value  p-value
decade1950 0.436328   0.005835   74.77   <2e-16
decade1960 0.439269   0.004413   99.53   <2e-16
decade1970 0.412231   0.004388   93.94   <2e-16
decade1980 0.444121   0.004282  103.72   <2e-16
decade1990 0.439018   0.004638   94.66   <2e-16
decade2000 0.452668   0.003890  116.36   <2e-16
---
Approximate significance of smooth terms:
                    edf Ref.df      F   p-value
s(basin_area)     4.875  6.038  97.167  < 2e-16     (basin [watershed] drainage area)
s(basin_slope)    3.768  3.967  21.327  < 2e-16     (basin slope)
s(ppt_mean)       1.000  1.000   6.814 0.009094     (mean precipitation, decadal)
s(temp_mean)      2.824  2.969   6.837 0.000113     (mean temperature, decadal)
s(dni_ann)        5.474  5.894  14.900 6.71e-16     (annual average solar radiation, fixed
    )
s(flood_storage)  3.903  3.993 109.782  < 2e-16     (flood storage in basin, decadal)
s(x,y)           26.657 28.618  34.676  < 2e-16     (Albers equal area coordinates)
---
R-sq.(adj) = 0.731   Deviance explained = 96.6 percent
GCV = 0.0078496   Scale estimate = 0.007694   n = 2,750
---
Additional Abbreviations:
edf, estimated degrees of freedom; F, statistic of F-distribution;
gam(...), generalized additive model; GCV, generalized cross validation statistic;
k, knot points for smooth function; n, sample size; Ref.df, reference degrees
of freedom; R-sq.(adj), adjusted R-squared; s(...), a smooth function;
Std., standard; and t-value, statistic of t-distribution
```

Figure 8. Summary list of output from mgcv::gam(…) (Wood, 2018) representing the final predictive model (GAM-T4) of decadal L-kurtosis (T4) of nonzero streamflows.

"permissive" (94 percent recorded for the nominal 95-percent limits), meaning that real coverage is less than the nominal coverage. The mgcv algorithms (Wood, 2017) are thus verified, and the standard errors and prediction limits in Robinson and others (2019) are reliable. Furthermore, the LOO-modeling coverage probabilities have similar magnitudes and verify that the whole-model prediction uncertainties are reliable.

Kroll and Stedinger (1999) measured no-flow regression performance by how well the above-threshold quantile estimates compared to the true values by using a log-space mean square error. Substantial censoring is present in the no-flow fractions because the number of streamgage-decade records having zero as the no-flow fraction is large. Some alternative metrics of model performance are considered. When using the entire sample of streamgage-decade records, the whole- and LOO-model RMSEs in no-flow fraction space are 0.073 and 0.080, respectively. When following the Kroll and Stedinger (1999) methodology to measure away from the threshold (not including streamgage-decade records estimated as perennial), the whole- and LOO-model RMSEs are 0.154 and 0.163, respectively. These two values show that uncertainty in the no-flow fraction when ephemerality is predicted is about ±0.155.

There are other ways to describe cGAM-PPLO performance. Of the 2,750 streamgage-decade records, about 27.2 percent have at least one no-flow day, which can be compared to the whole-model estimate of 18.1 percent and the LOO-model estimate of 18.0 percent. The mean no-flow fraction of 2,750 records is 0.033; the whole-model mean is 0.025, and LOO-model mean is 0.025. These differences imply that cGAM-PPLO slightly underestimates no-flow prevalence among stations and on average. A binary diagnostic is the number of correct decisions (perennial flow observed and predicted or ephemeral flow observed and predicted). The whole model is correct 84.9 percent of the time, and the LOO model is correct 83.9 percent of the time, compared to 72.8 percent of streamgage-decade records with perennial flow.

```
gam(T5 ~ s(basin_area) + s(basin_slope, k=5) +
         s(temp_mean, k=4) +  s(dni_ann, k=7) +
         s(flood_storage, k=5) + decade - 1 + s(x, y))
Family: gaussian; Link function: identity
Parametric coefficients: (T5, fifth L-moment ratio of nonzero streamflow)
           Estimate Std. Error t value  p-value
decade1950 0.333003   0.005399   61.68   <2e-16
decade1960 0.326029   0.004222   77.23   <2e-16
decade1970 0.301374   0.004192   71.90   <2e-16
decade1980 0.333570   0.004169   80.01   <2e-16
decade1990 0.325475   0.004249   76.61   <2e-16
decade2000 0.339912   0.003798   89.50   <2e-16
---
Approximate significance of smooth terms:
                    edf Ref.df      F   p-value
s(basin_area)     5.237  6.408 105.542  < 2e-16     (basin [watershed] drainage area)
s(basin_slope)    3.797  3.974  31.433  < 2e-16     (basin slope)
s(temp_mean)      2.955  2.995   8.415 1.18e-05     (mean temperature, decadal)
s(dni_ann)        5.576  5.927  22.077  < 2e-16     (annual average solar radiation, fixed
    )
s(flood_storage)  3.882  3.990 124.670  < 2e-16     (flood storage in basin, decadal)
s(x,y)           26.390 28.545  98.713  < 2e-16     (Albers equal area coordinates)
---
R-sq.(adj) =  0.732   Deviance explained = 94.6 percent
GCV = 0.0074987   Scale estimate = 0.0073519   n = 2,750
---
Additional Abbreviations:
edf, estimated degrees of freedom; F, statistic of F-distribution;
gam(...), generalized additive model; GCV, generalized cross validation statistic;
k, knot points for smooth function; n, sample size; Ref.df, reference degrees
of freedom; R-sq.(adj), adjusted R-squared; s(...), a smooth function;
Std., standard; and t-value, statistic of t-distribution
```

Figure 9. Summary list of output from mgcv::gam(…) (Wood, 2018) representing the final predictive model (GAM-T5) of decadal fifth L-moment ratio of (T5) nonzero streamflows.

Table 1. Whole- and leave-one-watershed-out model diagnostics and 95-percent confidence limit coverage probabilities.

[NSE, Nash-Sutcliffe model efficiency coefficient; RMSE, root mean square error; LOO, leave-one-watershed-out, cross validation; cGAM, censored generalized additive model; GAM, uncensored generalized additive model; L1, mean of nonzero part of decadal flow-duration curve (dFDCs); T2, dimensionless coefficient of L-variation (second L-moment ratio) of decadal FDC; T3, dimensionless L-skew (third L-moment ratio) of decadal FDC; T4, dimensionless L-kurtosis (fourth L-moment ratio) of decadal FDC; T5, dimensionless fifth L-moment ratio of decadal FDC; PPLO, probability or fraction of decadal no-flow; Note, modeling based on 2,750 streamgage-decade records]

Statistical model	Whole-model NSE	Whole-model RMSE	LOO-model RMSE	Whole-model 95-percent coverage probabilities	LOO-model 95-percent coverage probabilities
GAM-L1[a]	0.975	0.129	0.137	0.945	0.936
GAM-T2	0.661	0.087	0.092	0.958	0.947
GAM-T3	0.693	0.081	0.086	0.958	0.948
GAM-T4	0.730	0.087	0.092	0.947	0.939
GAM-T5	0.732	0.085	0.090	0.946	0.934
cGAM-PPLO[b]	0.303	0.225	0.226	0.944	0.942
Additional diagnostics for the cGAM-PPLO model					
cGAM-PPLO	Fraction of whole-model absolute errors equal to zero = 0.698				
cGAM-PPLO	Fraction of whole-model absolute errors less than 0.02 = 0.783				
cGAM-PPLO	Fraction of whole-model absolute errors less than 0.05 = 0.844				
cGAM-PPLO	Fraction of whole-model absolute errors less than 0.10 = 0.905				
cGAM-PPLO	Fraction of LOO-model absolute errors equal to zero = 0.693				
cGAM-PPLO	Fraction of LOO-model absolute errors less than 0.02 = 0.779				
cGAM-PPLO	Fraction of LOO-model absolute errors less than 0.05 = 0.836				
cGAM-PPLO	Fraction of LOO-model absolute errors less than 0.10 = 0.893				

[a]The NSE and RMSE are computed from $base_{10}$ logarithms of cubic meters per second (GAM-L1 model), and all other L-moment models (GAM-T2, T3, T4, and T5) listed are untransformed.

[b]The NSE and RMSE are computed from $base_{10}$ logarithms of flowtime, where a decade with zero no-flows (a perennial decade) would have a flowtime of about $\log_{10}$(3,653 days) = 2.562. The reported statistics, however, are misleading because of the extreme censored nature of the data. The 95-percent confidence limit coverage probabilities also are computed in the flowtime domain.

Results of Generalized Additive Models for No-Flow Fractions and L-Moments

Estimates of Decadal No-Flow Fractions

The estimates of the decadal no-flow fractions for the 2000s for the prediction locations are shown in figure 10. The no-flow fraction estimates for the prediction locations are available in Robinson and others (2019), including the estimated lower and upper 95-percent prediction limits along with the estimated flowtimes and the respective 95-percent prediction limits. The coverage probabilities (whole- and LOO-model) for the flowtime limits are approximately 95 percent (94 percent computed). After conversion of flowtimes to no-flow fractions and truncation to the interval (0,1), the coverage probabilities of the 95-percent no-flow fraction limits are too wide (about 99.9 percent). As a result, the 95-percent no-flow fraction limits must be considered as only approximate.

A preponderance of perennial streamflow occurs and is predicted in the eastern half of the study area (figure 10). Within the study area in central Florida, however, the no-flow fraction is commonly on the order of about 20 percent with an absence of very high proportions of no flow compared to the southwestern parts of the study area. These observations are consistent with results of studies using neural networks (Worland and others, 2019a; Worland, and others, 2019c). Florida has low no-flow fractions, contrasting to the southwesternmost areas that have more than 50-percent no-flow. In Texas, no-flow fractions generally diminish near the Gulf of Mexico. Large rivers are more commonly predicted as perennial streams than headwater (smaller basin area HUC12) streams.

Many streamgage-decade records for Arkansas, northwest Louisiana, Oklahoma, and Texas show perennial or near-perennial flow. Throughout this region, however, many prediction locations have no-flow fractions greater than about 0.2. This could be a basis for an assertion that streamgages tend to operate on streams that mostly flow in contrast to the stream reaches expected to be at no-flow conditions much of the time. The assertion would represent a type of bias in the

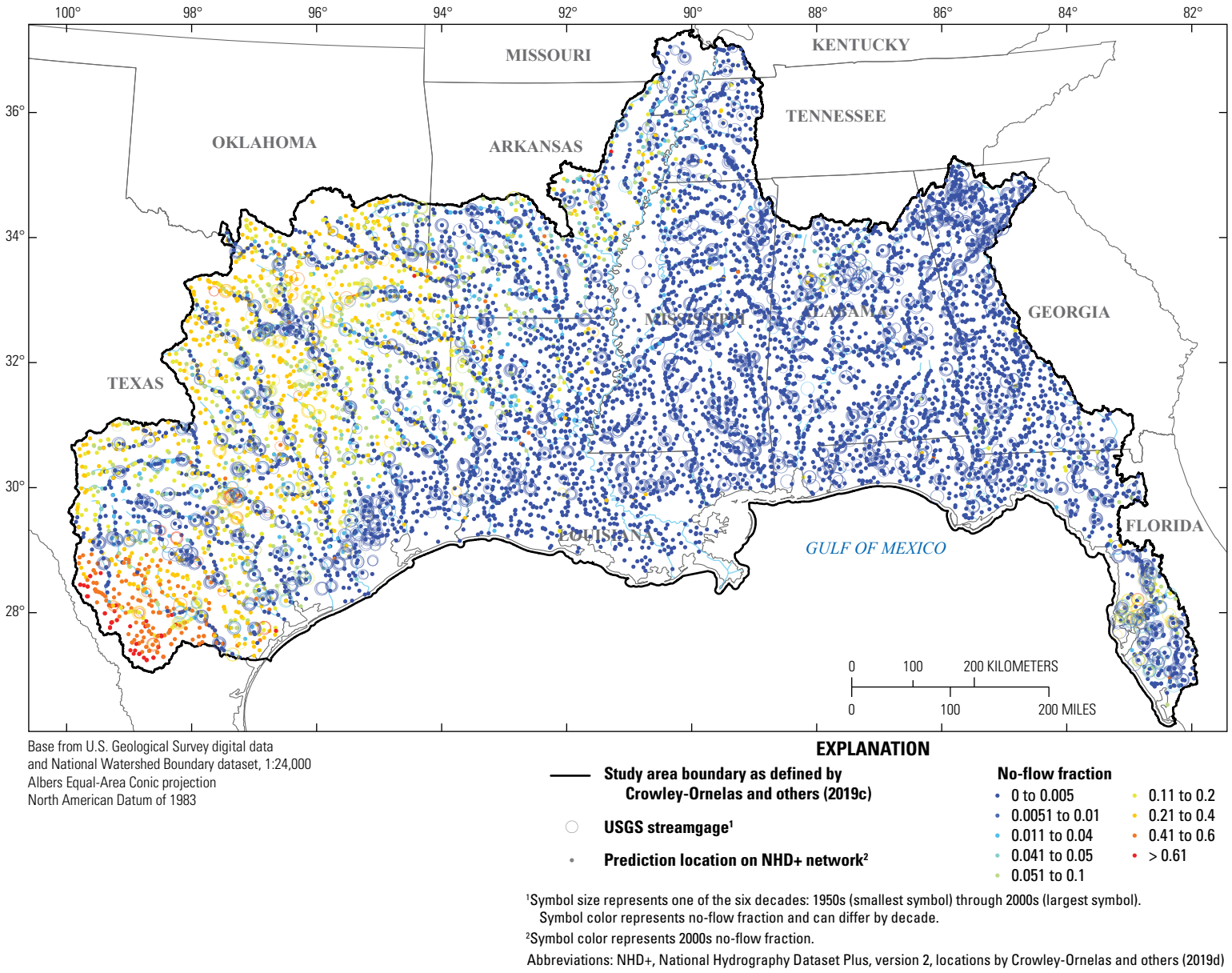


Figure 10. Estimated no-flow fractions (probabilities) for the 2000s at 9,220 prediction locations (stream reaches) and at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during 1950–2009.

operational footprint of the streamgage network in western parts of the study area; rigorous testing, however, could be formidable.

Estimates of Decadal L-Moments at Prediction Locations

The estimates of the decadal mean nonzero (L1) streamflow for the 2000s at the prediction locations are shown in figure 11, along with the corresponding observed values at the streamgages by decade. The estimates for the prediction locations and the estimated lower and upper 95-percent prediction limits and standard errors are in the data release by Robinson and others (2019). The data release also includes the Duan retransformation bias correction (Duan, 1983). The retransformation bias correction is useful because the GAM-L1 was based on logarithmic transformation of the decadal mean nonzero streamflow; note that the retransformation bias correction has not been applied to the otherwise listed means and prediction limits in the data release.

The L1 estimates (fig. 11) are depicted in logarithms; such transformation provides normalization to enhance visualization of data having several orders of magnitude variation. The main stems of major river systems are visually (fig. 11) more identifiable than they were for the no-flow fractions (fig. 10). Smaller watersheds are headwater basins, upland HUC12s have smaller streamflows, and small watersheds (the smaller HUC12s) generally can be identified by the warmer (redder) colors.

The interdecade differences of GAM-L1 results (values in Robinson and others [2019] and decadal graphics in Asquith and others [2020]) are generally small; however, L1 is connected to no-flow fractions because it is the mean nonzero streamflow. Interpreting decadal nonzero streamflow volume trends is therefore difficult. As a conceptual example, consider a major change in flood storage in a basin from one decade to the next. The additional storage could result in reservoir releases somewhat alleviating no-flow conditions and potentially resulting in a decrease in the mean nonzero streamflow. Yet simultaneously, the overall volume of water flowing from the basin between the decades could remain unchanged.

The estimates of the decadal coefficient of L-variation (T2) values for the 2000s are shown in figure 12, and accompanying the estimates are the corresponding observed values at the streamgages by decade. The estimates for the prediction locations are in Robinson and others (2019) along with the estimated lower and upper 95-percent prediction limits. The estimate also includes the residual standard error of the model and standard errors of estimate for the predictions, and decade-over-decade, streamgage-specific changes in the T2 values.

The T2 estimates (fig. 12) are relatively straightforward to interpret because T2 conceptually expresses the relative (dimensionless) variation in the dFDCs. Warmer colors are used to depict hydrologic regimes having lesser streamflow variation. A striking feature of the predictions depicted is the zoning (clustering) of small, moderate, and large T2 values. As a rule, the least varying hydrologic regimes are in watersheds in the eastern third of the study area. Some of the most variable hydrologic regimes in the study area are in Texas. The part of Texas near the 99th meridian and the 30th parallel has less variation within a region consistent with the limestone uplands of central of Texas that is documented (Bomar, 2017) to have unique and more eastern-like hydrologic characteristics than suggested by its longitude.

The estimates of the decadal L-skew (T3) values for the 2000s are shown in figure 13, and accompanying the estimates are the corresponding observed values at the streamgages by decade. The estimates for the prediction locations are available in Robinson and others (2019), including the estimated lower and upper 95-percent prediction limits and standard errors.

The T3 estimates (fig. 13) are more difficult to interpret than the coefficients of L-variation because they express the asymmetry in the dFDCs. Warmer colors continue to depict hydrologic regimes having less skewness. These are watersheds for which the right or flood-flow tail of the daily mean streamflows is not as stretched as elsewhere. The figures show that Texas has relatively larger growth of streamflow in the right tail relative to the eastern parts of the study area. This conclusion holds even though the zero flows, which are common in Texas, have been removed prior to this L-moment computation.

The estimates of the decadal L-kurtosis (T4) for the 2000s are shown in figure 14, and accompanying the estimates are the corresponding observed values at the streamgages by decade. The estimates for the prediction locations are in Robinson and others (2019), including the estimated lower and upper 95-percent prediction limits and standard errors.

The T4 estimates are more difficult to interpret than the T3 values because T4 values express, though not uniquely, a type of peakedness of the dFDCs. Warmer colors continue to depict hydrologic regimes having less T4. Broad regions throughout the study area show a narrow range of T4, with the exceptions being the main stems of major river systems that can be seen by the red to near-red colors in the figure. The primary use of T4 values is thought to be in assessments of fit when using 3-parameter distributions (Hosking and Wallis, 1997; Hosking, 2006; Asquith, 2011; Asquith and others, 2017) to model the FDC.

The estimates of the decadal fifth L-moment ratio (T5) for the 2000s are shown in figure 15, and accompanying the estimates are the corresponding observed values at the streamgages by decade. The estimates for the prediction locations are in Robinson and others (2019), along with the estimated lower and upper 95-percent prediction limits and standard errors.

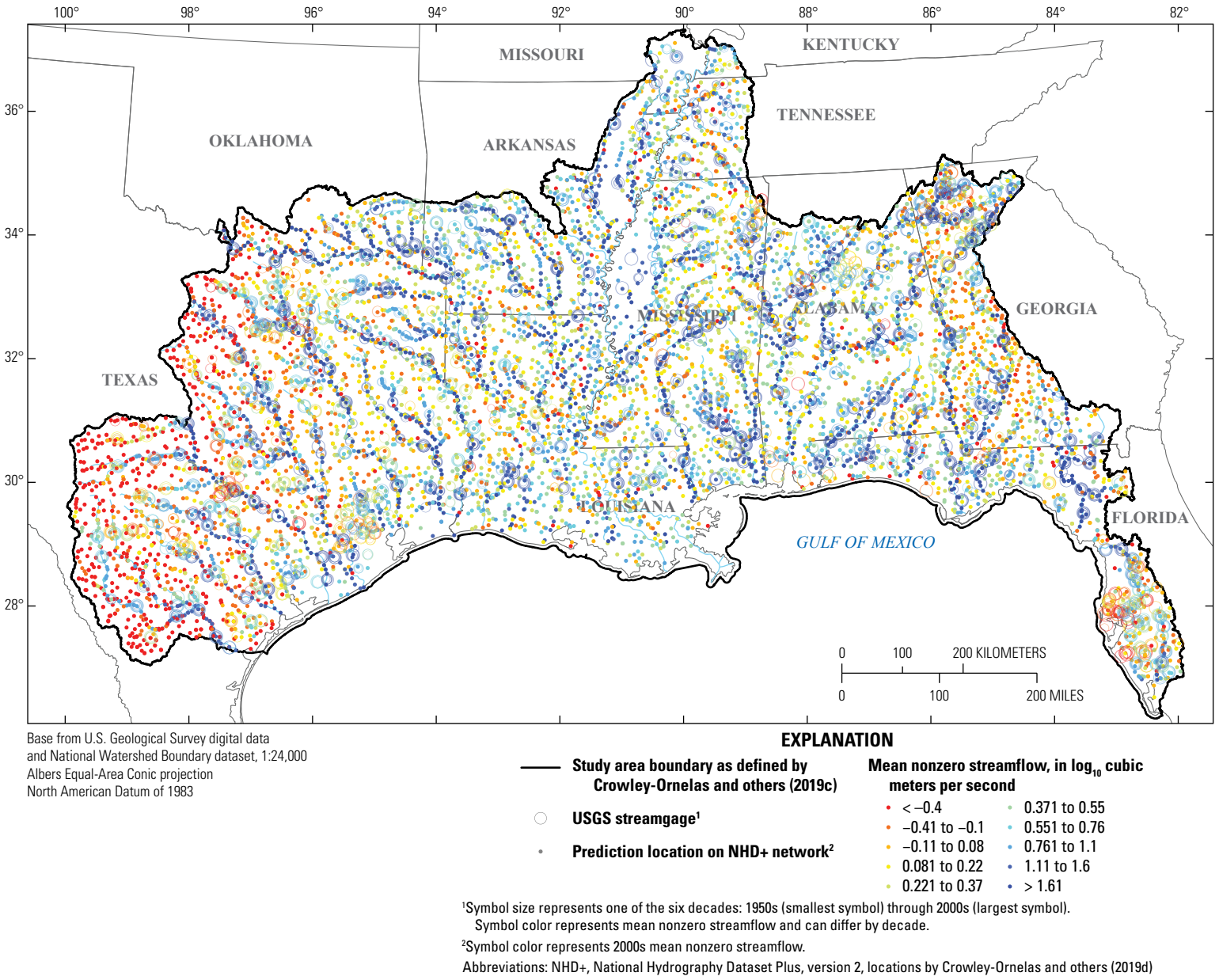


Figure 11. Log-transformed estimated mean nonzero streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during 1950–2009.

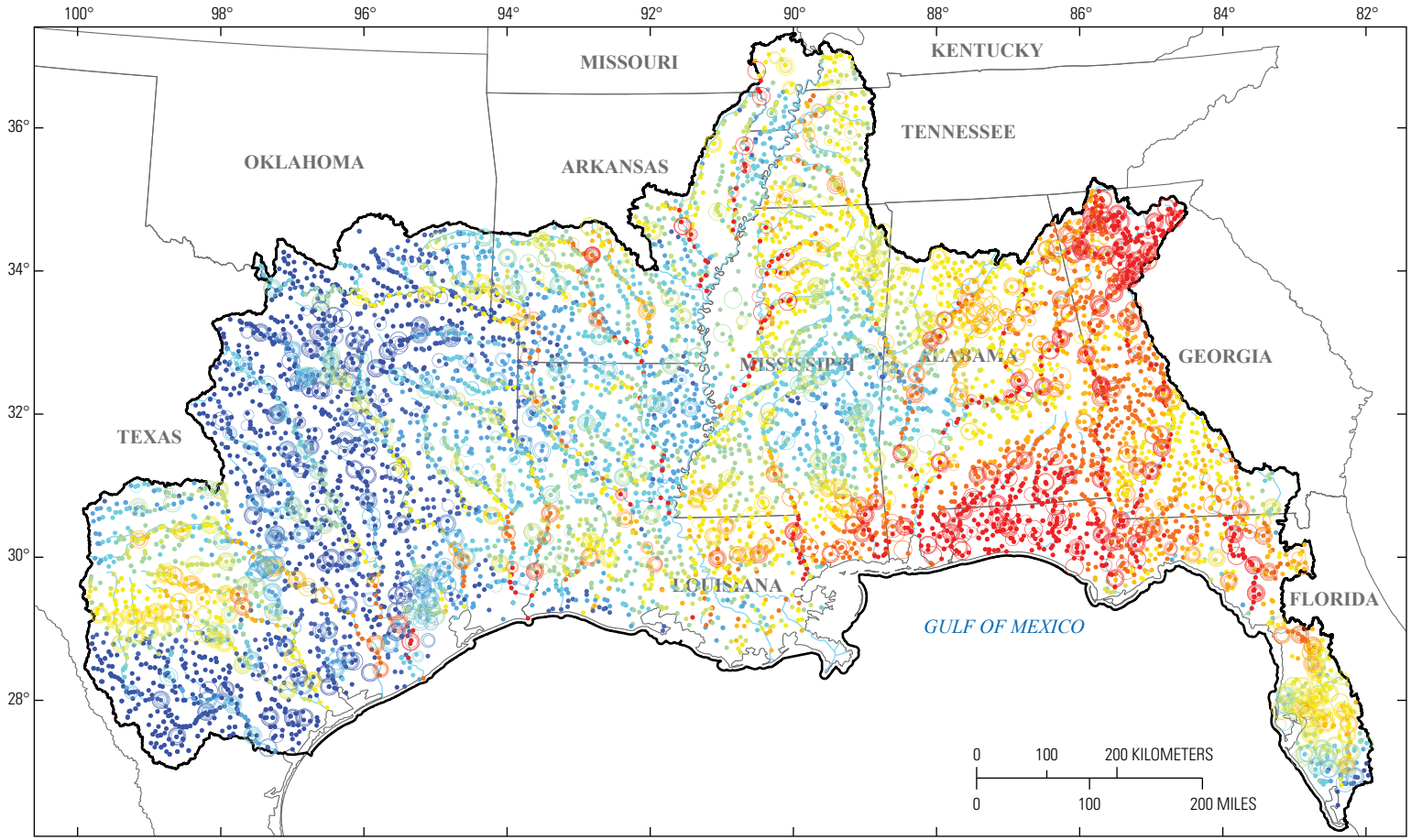


Figure 12. Estimated coefficients of L-variation of nonzero streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey streamflow-gaging stations (streamgages) with at least one complete decade of record during the 1950–2009.

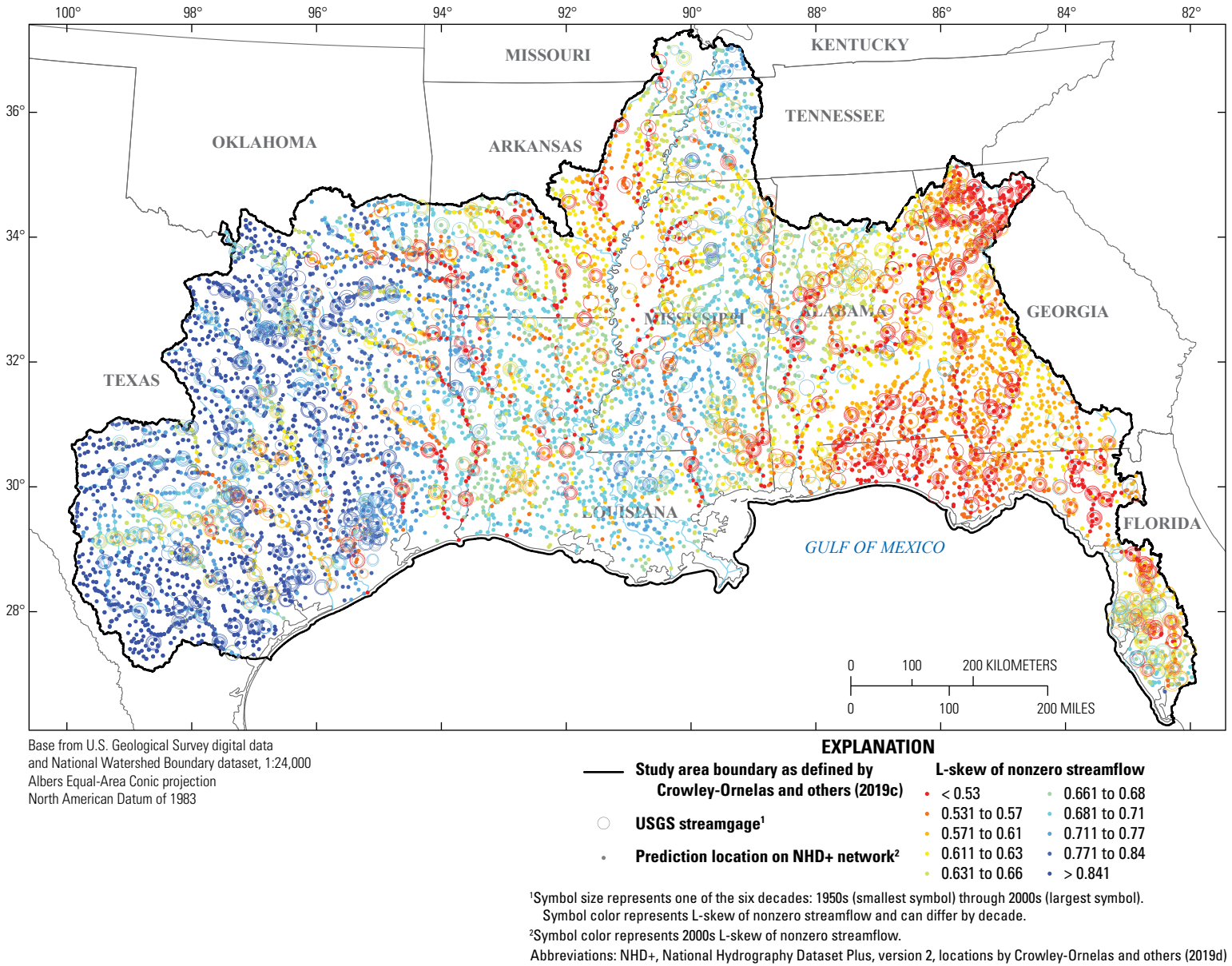


Figure 13. Estimated values of L-skew of nonzero streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during the 1950–2009.

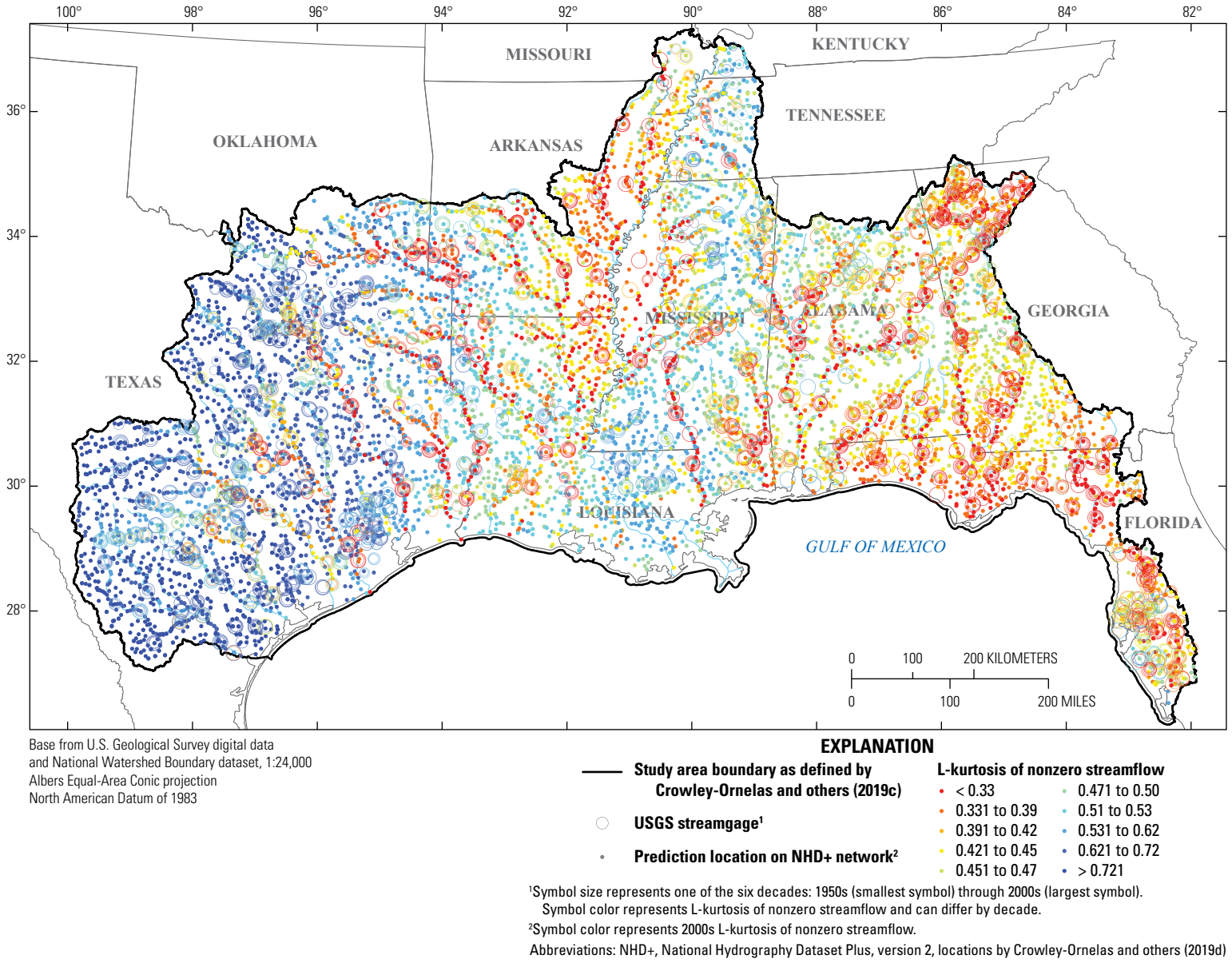


Figure 14. Estimated values of L-kurtosis of nonzero streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during the 1950–2009.

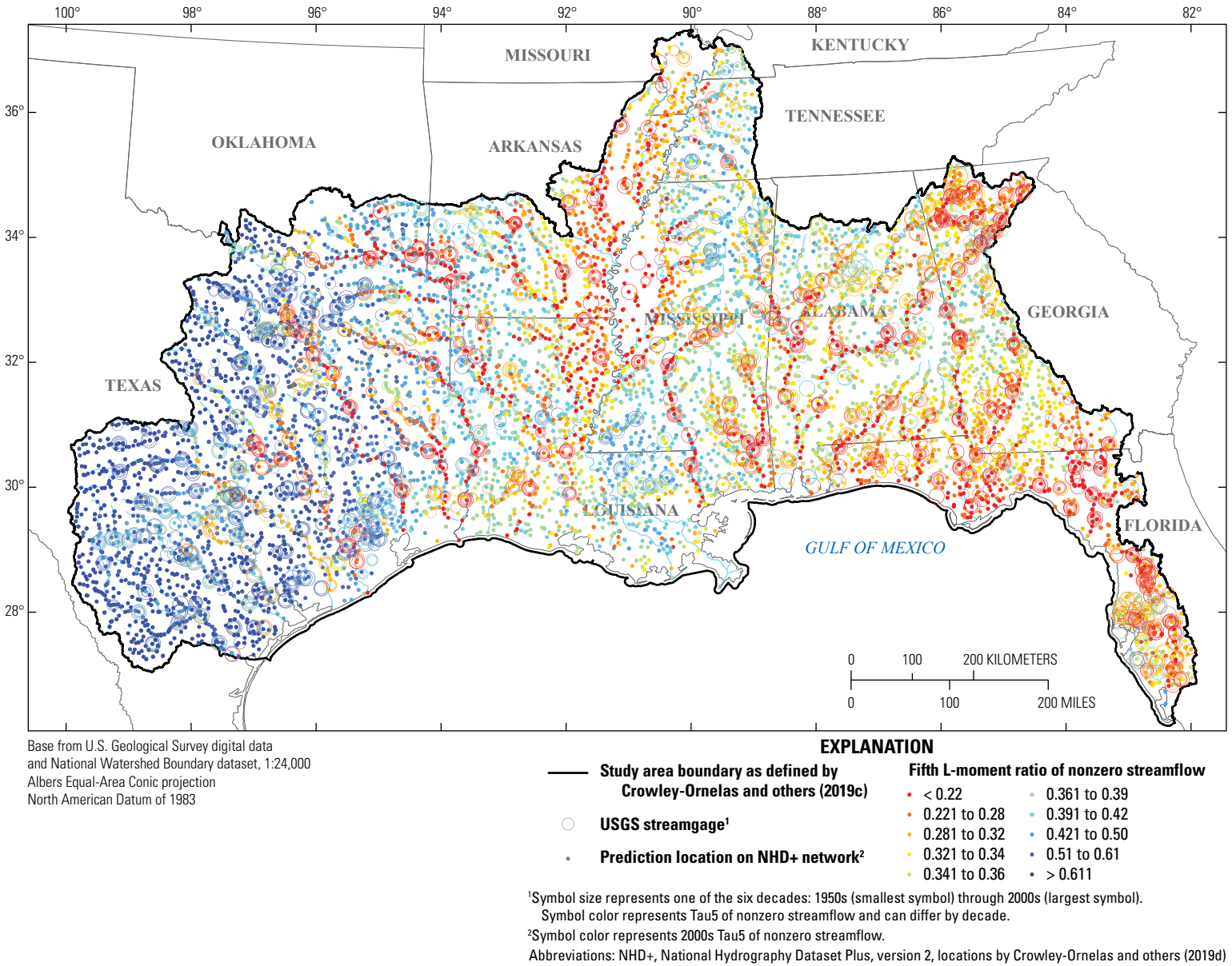


Figure 15. Estimated values of fifth L-moment ratios of nonzero streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during the 1950–2009.

It is difficult to interpret the T5 estimates as geometric measures of the shape of the dFDCs, but T5 measures asymmetry similar to T3. In broad regions throughout the study area, this moment shows a narrow range of T5 values, with the exceptions being the main stems of major river systems that can be seen by the red to near-red colors in the figures. The primary use of the T5 is thought to be in assessments of probability distribution fit (Asquith, 2011). Finally, the T3, T4, and T5 estimates could be useful to identify the somewhat homogenous regions of hydrologic regimes in the study area and to demarcate the major river systems acting as transport corridors.

Decadal Overall Mean Streamflows

The term "overall mean streamflow" is the volumetric yield of a watershed. In other words, overall mean streamflow is the mean that includes the zero flows and nonzero flows. The estimates of the decadal overall mean streamflows for the 2000s are shown in figure 16, and accompanying the estimates are the corresponding observed values at the streamgages by decade.

Here is a simplified example. Suppose for the 1960s that the no-flow fraction is 0.11 from the cGAM-PPLO, that the estimated nonzero mean streamflow is 2.01 m^3/s from the L1-GAM, and that the Duan retransformation bias correction from GAM-L1 is about 1.046, which is the bias_corr in file all_gage_looest_L1.csv (Robinson and others, 2019). The overall 1960s mean streamflow is algebraically a weighted or prorated computation: $1.046 \times (1 - 0.11) \times 2.01 = 1.87$ m^3/s. Predictions from GAM-L1 represent median response, but for the circumstances here, the mean response itself is desired and hence the retransformation bias correction is used.

The 95-percent prediction limits of the overall mean streamflows were computed for whole- and LOO-model results. First, the lower limit was computed from the upper limit for the no-flow fraction, the lower limit for mean nonzero streamflow, and the retransformation bias correction. Second, the upper limit was computed from the lower limit of no-flow fraction, the upper limit for mean nonzero streamflow, and the retransformation bias correction. This computation is structurally a 1 minus no-flow fraction term times mean nonzero streamflow times the bias correction. Third, by using coverage probabilities of the models (cGAM-PPLO and GAM-L1) that are listed in table 1, the coverage probabilities were computed. These coverage probabilities were found to be too large, and a compensation factor determined as 0.23 was needed to contract the limits (inflating the lower limit by 1.23 and deflating the upper limit by 0.77) to yield coverage probabilities of about 0.95 (matching the declared 95-percent prediction limits).

The compression or adjustment of the prediction limits for the overall mean streamflow is deemed acceptable because the stochastic errors in no-flow fractions and mean nonzero streamflows are not perfectly correlated (simultaneous with each other), meaning the largest no-flow fraction does not exclusively occur with the smallest mean nonzero streamflow. As a result, the initial computation in the first and second parts stated previously are intuitively too large. Finally, the estimated overall mean streamflows for the streamgage-decade records and for the prediction locations are available in Robinson and others (2019) and include the estimated lower and upper 95-percent prediction limits.

In order to review overall performance of estimation of mean streamflow using cGAM-PPLO and GAM-L1, it is informative to estimate overall mean streamflow for all streamgage-decade records and to compare those to observed data (fig. 17). The estimated values were computed by using the proration (a weighted mean) between the estimated no-flow fraction and estimated mean nonzero streamflows with the retransformation bias correction. The observed values are decadal overall mean streamflows and thus inclusive of no-flow conditions. To assist in further visualization of model performance, a distinction is made between those streamgage-decade records having no-flow fractions greater than zero (red circles) and those with estimated perennial flow (blue circles), and the equal value line indicates general agreement throughout.

In figure 17, estimation variance appears to contract for the largest streamflows. This is consistent with the fact that the largest river systems in the study area are anticipated to have less overall variation relative to small watersheds. The largest streamflows are associated with perennial streamflow, whereas the smallest streamflows tend to exist under no-flow conditions. This observation shows that the comparatively unbiased estimation of overall mean streamflow coincides with some non-homoscedasticity in the residuals (error variance decreasing as decadal mean streamflow increases).

Data from USGS streamgage 08167000, Guadalupe River at Comfort, Texas, were selected to plot in figure 17 because that streamgage is in a part of the study area with relative propensity for ephemeral streamflow, and this streamgage also is considered in this report for a demonstration of using selected probability distribution to estimate FDC quantiles. The observed and estimated overall mean streamflows are individually identified (fig. 17), and the whole- and LOO-model estimates are shown. According to Crowley-Ornelas and others (2019b), the 1950s is the driest observed decade for this streamgage, with an observed mean of 2.996 m^3/s. The whole- and LOO-model overall mean estimates from Robinson and others (2019) are 3.883 and 3.880 m^3/s, respectively. Conversely, the 2000s are the wettest observed decade, with an observed mean of 8.729 m^3/s and whole- and LOO-model overall mean estimates of 7.149 and 7.124 m^3/s, respectively.

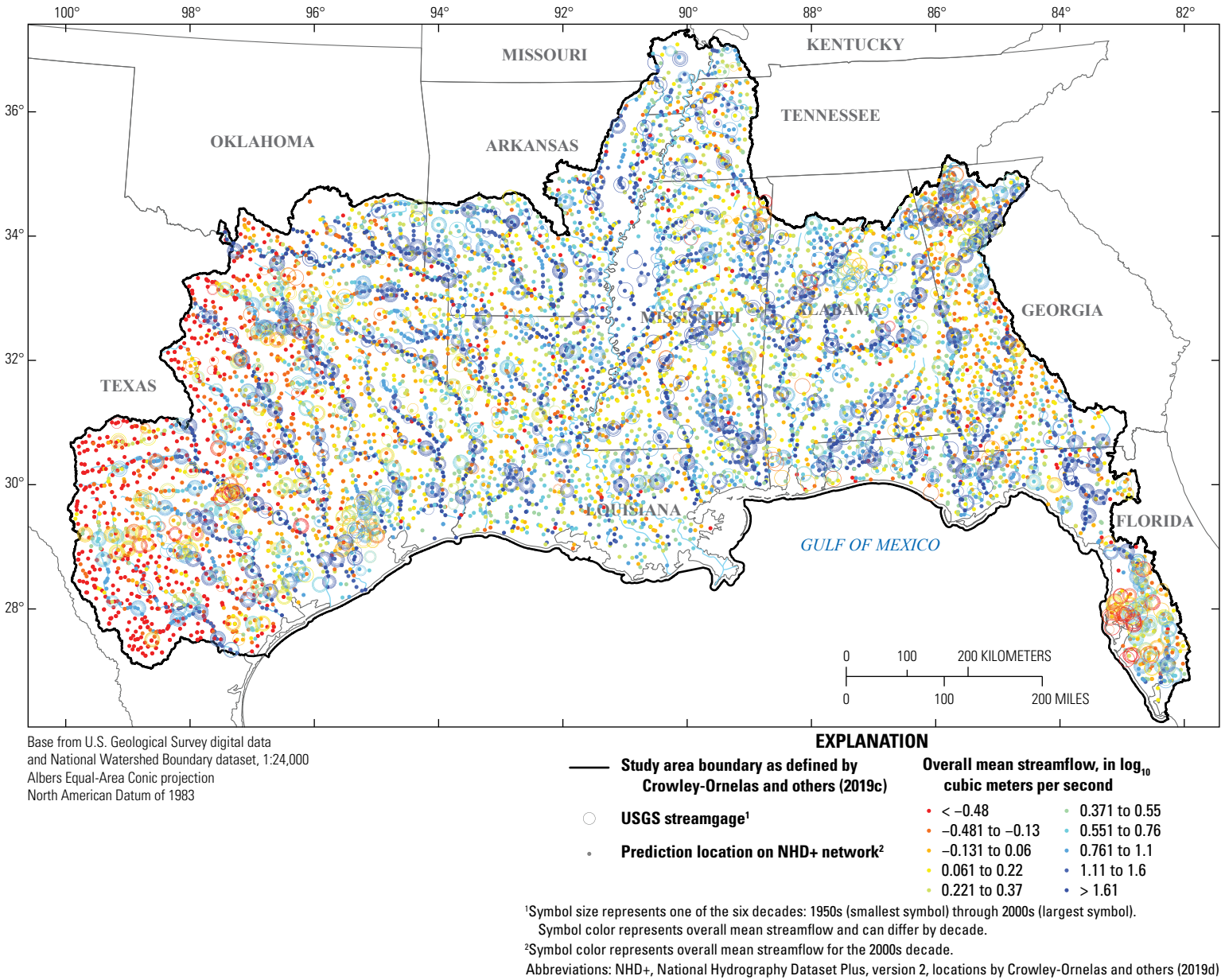


Figure 16. Log-transformed estimates of the overall mean streamflow for the 2000s at 9,220 prediction locations (stream reaches) and observed values at 941 U.S. Geological Survey (USGS) streamflow-gaging stations (streamgages) with at least one complete decade of record during 1950–2009.

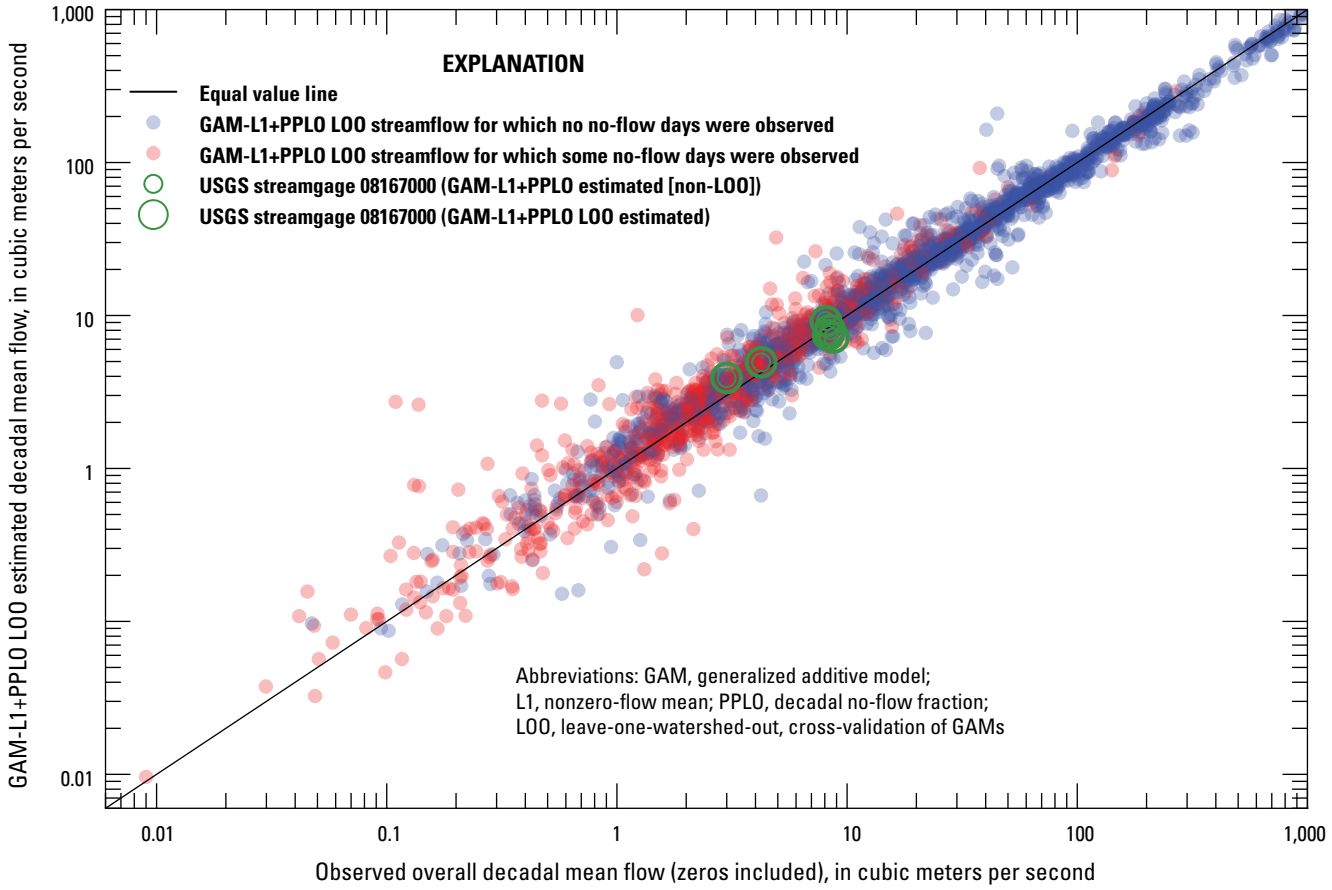


Figure 17. Observed overall decadal mean flow values and leave-one-watershed-out estimates from generalized additive models of decadal mean nonzero streamflow (GAM-L1) with retransformation bias correction and decadal no-flow fractions from the censored generalized additive model (cGAM-PPLO).

Discussion

A censored GAM of no-flow fractions (cGAM-PPLO) was created, and subsequent predictions at HUC12s were deemed reliable as part of the regional study of dFDCs. For the nonzero part of the dFDCs, uncensored GAMs were created to predict mean nonzero streamflow (GAM-L1), its coefficient of L-variation (GAM-T2), and higher L-moment ratios (GAM-T3, GAM-T4, GAM-T5). The watershed properties implemented in the statistical models include basin area (immutable), basin slope (immutable), decadal precipitation and temperature, and decadal percentages of grassland and urban development. Other variables include mean 1998–2009 solar radiation and the projected coordinates (immutable) of the streamgages. General streambed permeability (immutable), one of two categorical variables used in the model, is used only for the cGAM-PPLO.

The other categorical variable (decade) has six classes for the six decades (1950–2009) of streamflow data, and because the coefficients for the decades are considered small (figs. 2, 4–5, and 7–9), semi-qualitative judgment is that decade effects can be considered of minor importance. This provided credence that decadal changes in the watershed properties reasonably provides for decadal changes in streamflow. Finally, flood storage per unit basin area (cumulative maximum storage in reservoirs minus normal reservoir storage divided by basin area) materially affects the relative variation and shapes (skewness and other) but not the mean nonzero streamflow or no-flow fractions. Flood storage influences the timing and duration of particular streamflows but appears to not materially adjust bulk volume of surface water as a generality.

Flow-Duration Curve Quantile Estimation Using Selected Probability Distributions

A demonstration is shown of the no flow and L-moment GAMs used to produce dFDC quantiles using selected probability distributions. The demonstration shows probability distributions fit to the estimated L-moments with left-tail truncation accommodating the estimated no-flow fraction.

A HUC12 pour point near streamgage 08167000 (fig. 1) has been chosen for illustration. Decadal FDC estimation will be made for the ungaged stream reach, but to obtain dFDC estimates calculated from observed values for comparison, the observed results for the streamgage will be transferred to the ungaged location by using the drainage-area ratio (DAR) method (Farmer and others, 2015b). The DAR method is defined as the ratio of streamflows to an exponentiated ratio of areas. The DAR method has been thoroughly studied in Texas by Asquith and others (2006); for Texas, a general estimate of the exponent suitable for daily mean streamflows is 0.9. This exponent contrasts to the exponent of 0.5 (square root) for instantaneous annual peak streamflows in Texas (Asquith and Thompson, 2008).

A HUC12 pour point (ungaged location; COMID 3588922) is located on a minor tributary to the river monitored by streamgage 08167000 (COMID 3589508). The respective basin areas are 2,181.28 km^2 (streamgage) and 188.54 km^2 (ungaged location). The observed dFDCs for the streamgage are transferred to the prediction location by $(188.54/2{,}181.28)^{0.9} = 0.110$. The result of this calculation indicates that the projected flows are estimated to be 11 percent of those observed at the nearby streamgage. For example, the observed (gaged) 1950s overall mean streamflow was 2.996 m^3/s, which will be transferred to the prediction location as 0.3296 m^3/s.

The dFDCs transferred from the streamgage to the prediction location are shown in figures 18–20 for the asymmetric exponential power (AEP4), generalized normal (GNO), and kappa (KAP) distributions, respectively. In the figures, the transferred dFDCs are considered to be observed information.

Three probability distributions were selected, and each is fit to the decadal L1 corrected for retransformation bias, the T2 estimates, and higher L-moment ratios. The probability distributions are shown in figures 18–20, and pertinent references for the distributions and their estimation with L-moments include Hosking (1990), Hosking and Wallis (1997), Hosking (2006), and Asquith and others (2017). GNO (Asquith, 2011, 2018), KAP (Asquith, 2011, 2018), and AEP4 (Asquith, 2014, 2018) distributions were chosen for some theoretical reasons

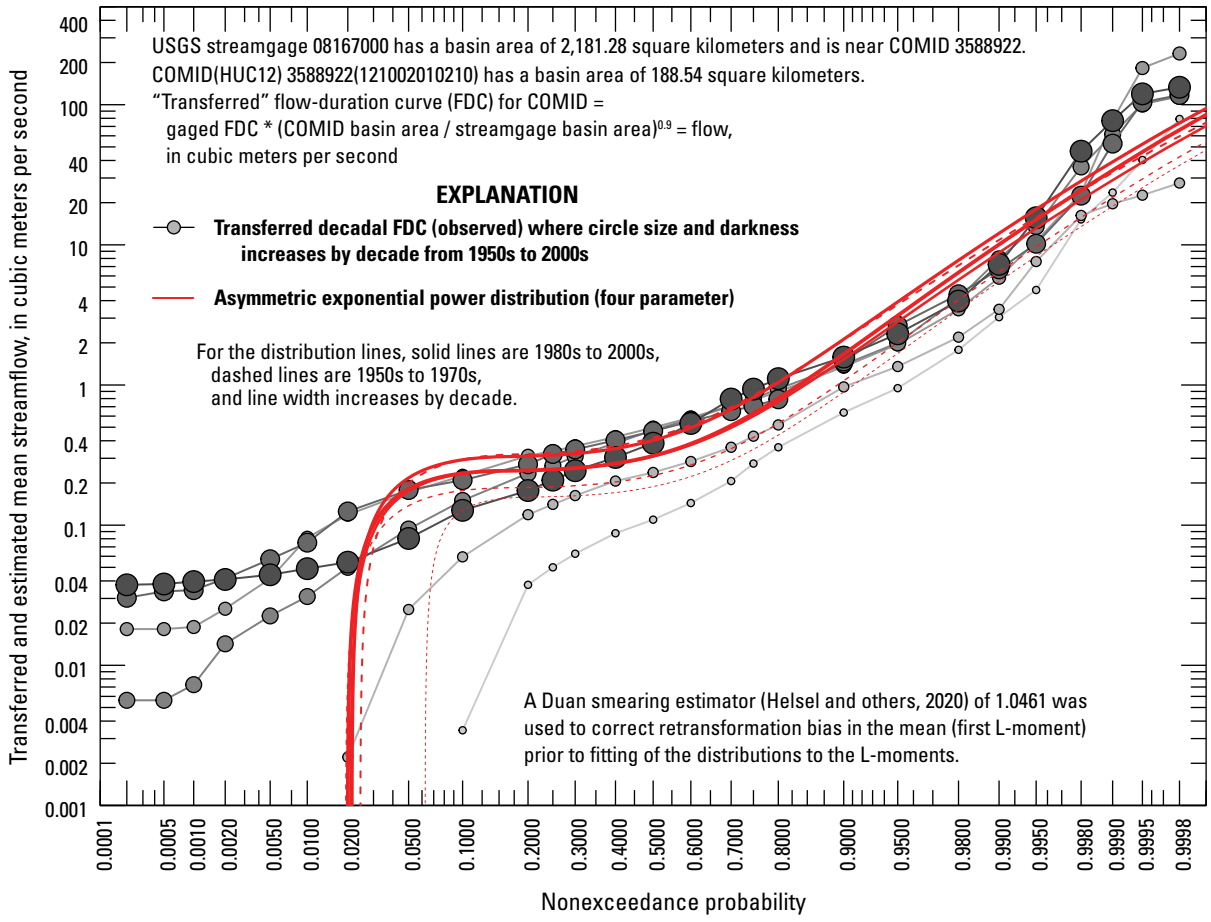


Figure 18. Example computation of flow-duration curve defined by the asymmetric exponential power distribution for six decades during the 1950–2009 for an ungaged prediction location on a tributary of the Guadalupe River using streamflow values for nearby U.S. Geological Survey (USGS) streamflow-gaging station (streamgage) 08167000 on the Guadalupe River in order to show comparison between the locations through scaling by the drainage-area ratio method.

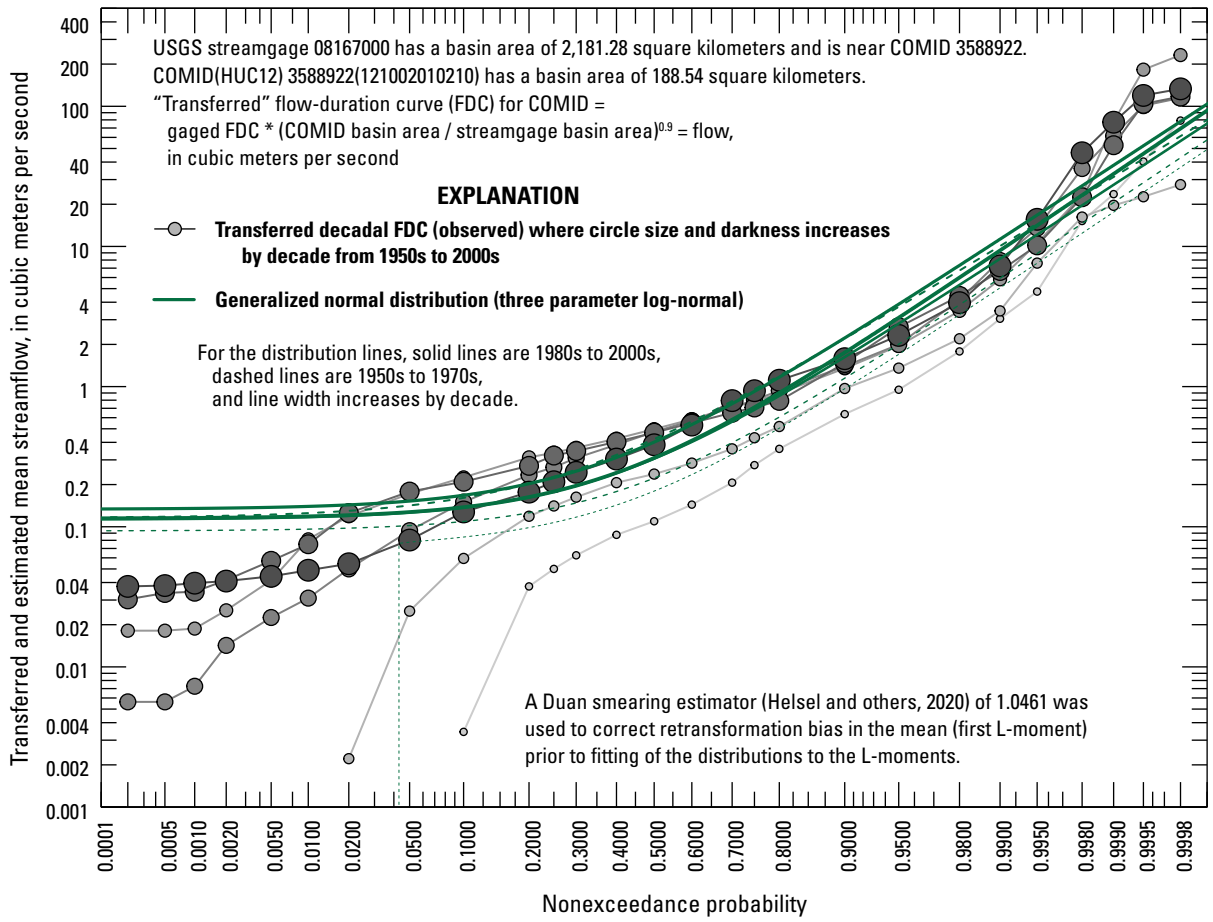


Figure 19. Example computation of flow-duration curve defined by the generalized normal distribution for six decades between during the 1950–2009 for an ungaged prediction location on a tributary of the Guadalupe River using streamflow values for nearby U.S. Geological Survey (USGS) streamflow-gaging station (streamgage) 08167000 on the Guadalupe River in order to show comparison between the locations through scaling by the drainage-area ratio method.

which follow. The GNO (having three parameters) is the same as a 3-parameter log-normal; it is log-normal with a feature of potential curvature in log-probability space. The GNO was fit with the L1, T2, and T3 estimates (fig. 19). The GNO-estimated FDC quantiles for the streamgages (Robinson and others, 2021) and the HUC12s (Robinson and others, 2021) are available.

The KAP is a 4-parameter distribution and is fit with L1 and T2 through the T4 (fig. 20) with a caveat that if the estimated T4 exceeds that of the generalized logistic distribution (GLO), the T4 is snapped to the maximum T4 attainable. The KAP is capable of mimicking other well-known distributions such as the generalized extreme value, GNO, and Pearson type 3. Theoretically, the KAP can be fit to the smallest T4 possible attainable by a random sample (Asquith and others, 2017). For all six decades in this example, the KAP was snapped to the GLO in accordance with recommendations from Hosking and Wallis (1997). This means that the GLO is represented in this example though it has been labeled as KAP. Other prediction locations could show different handling of limiting forms of the KAP. The KAP-estimated FDC quantiles for the streamgages (Robinson and others, 2021) and the HUC12s (Robinson and others, 2021) are available.

The AEP4 is another 4-parameter distribution and, similar to the KAP, is fit with L1 and T2 through the T4 (fig. 18). The AEP4 can be fit to T4 values greater than those available to the KAP and thus above the GLO (Asquith and others, 2017), which makes the AEP4 complementary to the KAP. The AEP4 does have a lower limit of T4, and if needed, the estimated T4 was snapped to the minimum T4 attainable (Asquith, 2018). In this example application, for all six decades, the AEP4 was not required to be snapped to its lowest T4 attainable because the estimated T4 remained well inside the parameter space of this distribution. The AEP4-estimated FDC quantiles for the streamgages (Robinson and others, 2021) and the HUC12s (Robinson and others, 2021) are available.

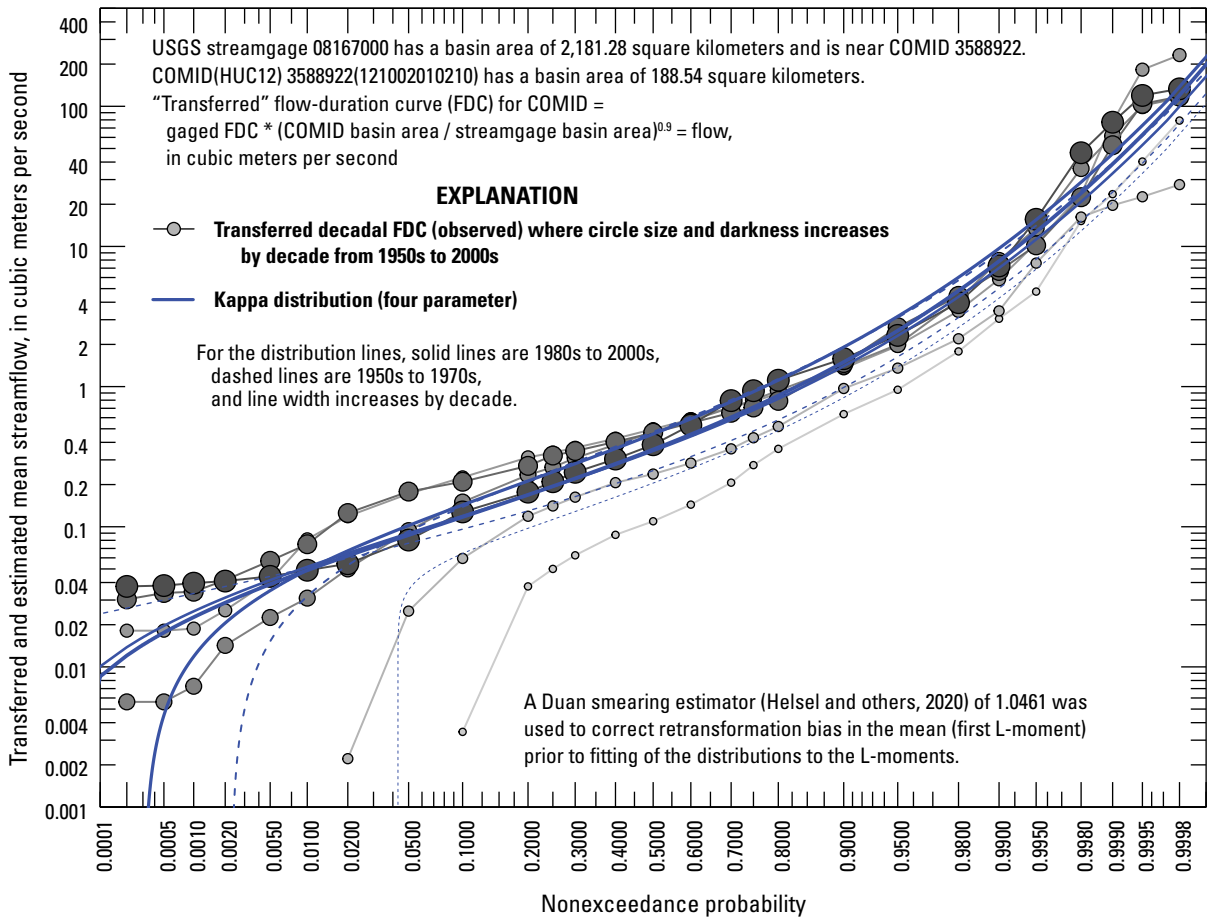


Figure 20. Example computation of flow-duration curve defined by the kappa distribution for six decades during the 1950–2009 for an ungaged prediction location on a tributary of the Guadalupe River using streamflow values for nearby U.S. Geological Survey (USGS) streamflow-gaging station (streamgage) 08167000 on the Guadalupe River in order to show comparison between the locations through scaling by the drainage-area ratio method.

For the fitting process to the L-moments, the retransformation bias correction was used, and truncation of the distribution to the estimated no-flow fraction was made unless the fitted distribution itself had its own limiting value to zero that happened to be at a higher no-flow fraction. The no-flow truncation is seen in the GNO curves in figure 19 for three decades abruptly going vertical at about 0.06 to 0.14 nonexceedance probability. For three decades, the GNO has a lower bound implied by the L-moments fit of about 0.1 m^3/s. The truncation can further be seen by other distribution fits becoming vertical at a streamflow of about 0.001 m^3/s.

The three distributional forms show some differences. Although truly fit to all four L-moments in this example, the AEP4 upper tail (fig. 18) bends more to the right than the GNO (fig. 19) and KAP (fig. 20). The AEP4 shows an ability to curve into the no-flow fractions but simultaneously does not have the lower bounds suggested by three of the GNO fits. The KAP appears to be somewhat in between. The AEP4 is fit to the four L-moments. The KAP and AEP4 have the same fit to the first four L-moments, but the distributions retain their own distribution-specific tail behavior tied to the mathematical definition of the distribution. It is important to stress that the observed data show results from a very simple method (DAR) to transfer flows from a nearby streamgage to the selected prediction location strictly for completing a demonstration of selected results of this study.

The fitted and truncated distributions can be numerically integrated to estimate an overall mean streamflow separately from the direct method of proration between the no-flow fraction L1. Hence, verification checks were made by comparing the overall mean streamflow from no-flow fractions and mean nonzero streamflow with retransformation bias correction (PPLO-ubL1) to numerical integration of the distribution-approximated dFDCs. Decadal estimated overall mean streamflow values of the GNO, KAP, and AEP4 distributions for streamgage 08167000 (not the prediction location used in figures 18–20) are listed in table 2. For this example, the integrated means are numerically congruent with the PPLO-ubL1 estimates with differences on the order of about 0.2 m^3/s. The numerical congruency could be used to assess situations in which the FDC by a given distribution is not to be reported for

Table 2. Decadal overall mean streamflow values from estimates of no-flow fractions and mean nonzero streamflow values from numerical integration of flow-duration curves defined by the three probability distribution approximations of flow-duration curves for U.S. Geological Survey streamflow-gaging station 08167000.

[m^3/s, cubic meters per second; PPLO-ubL1, overall mean streamflow computed by estimates of no-flow fraction (PPLO) from the cGAM-PPLO model (censored generalized additive model [cGAM]) and the unbiased (UB) mean nonzero streamflow (L1) from the GAM-L1 model (uncensored generalized additive model [GAM]) with the retransformation bias correction; GNO, overall mean streamflow by numerical integration of generalized normal distribution approximation of the flow-duration curve (FDC); KAP, overall mean streamflow by numerical integration of kappa distribution approximation of the FDC; AEP4, overall mean streamflow by numerical integration of asymmetric exponential power distribution approximation of the FDC]

Decade	PPLO-ubL1 (m^3/s)	GNO distribution (m^3/s)	KAP distribution (m^3/s)	AEP4 distribution (m^3/s)
1950	3.883	3.991	3.903	4.045
1960	4.927	5.061	4.960	5.122
1970	8.040	8.246	8.127	8.351
1980	7.432	7.633	7.465	7.736
1990	9.160	9.412	9.226	9.534
2000	7.149	7.352	7.176	7.449

a streamgage-decade record. Such checks also could be made if FDCs are estimated for the prediction locations by using the selected probability distribution formulas.

Discussion

General estimates of the dFDC quantiles (Robinson and others, 2021) within about 0.1 to 0.9 nonexceedance probabilities (the central portion of the streamflow distribution) are likely to be reliable. This implicitly shows that the estimated L-moments (Robinson and others, 2019) themselves are reliable. The choices of probability distribution used to represent the FDCs are generally of lesser concern within the central part of the streamflow distribution. Figures 18–20 show that the distribution choice does not produce estimates far from those expected based on FDCs transferred by the DAR method from a nearby streamgage. Truncation according to the estimated no-flow fraction is expected to mitigate fit difficulties in the left tail. However, this may not be sufficient for some scientific interests including biological minimum and other ecological flows (Texas Commission on Environmental Quality, Texas Parks and Wildlife Department, and Texas Water Development Board, 2008). Conversely, quantities focused on the right tail (such as flood flow volumes) are anticipated to be less dependent on accuracy in the no-flow fraction.

Reliable predictions of overall mean streamflows (fig. 17) by the cGAM-PPLO and GAM-L1 models are an important aspect of dFDC regionalization. In addition, the estimates of overall mean streamflow using the techniques described are anticipated to be important benchmarks against which to compare results of other efforts (Worland and others, 2019c) focused on estimation of streamflow statistics within the greater RESTORE project.

Summary

Hydrologic alteration has been documented in the majority of monitored streams in the United States and is thought to be the primary cause of impairment in riverine ecosystems. Study of the timing and quantity of freshwater inflows can be used to more precisely restore the water quality, marine habitats, and biological resources in the Gulf of Mexico. This study focused on providing flow-duration curve (FDC) quantiles using probability distributions fit to L-moments of nonzero streamflow and the no-flow fractions. Generalized additive models (GAMs) were created by using streamflow data from 941 streamgages and watershed properties to estimate streamflow at ungaged locations. Decadal L-moments were studied by using GAMs, and no-flow fractions were studied by using censor-extended GAMs (cGAMs). The L-moments studied using GAMs were L-moment of nonzero streamflow (L1), coefficient of L-variation (T2), L-skew (T3), L-kurtosis (T4), and fifth L-moment ratio (T5). No-flow conditions were considered type I censored data and GAMs that incorporate Tobit regression methods were used to alleviate bias of censored data in the cGAM. Decadal FDC quantiles were estimated by fitting the L-moment estimates by using GAM regionalization with selected probability distributions.

The reliability of the models was extensively tested. Whole-sample modeling can underestimate prediction uncertainties; therefore, to combat this, leave-one-out (LOO) testing was performed on the GAMs and the cGAM. Nash-Sutcliffe efficiency coefficient, root mean square error, and coverage probabilities for the whole and LOO models were used to evaluate the GAMs and cGAM. The Nash-Sutcliffe efficiency coefficient and root mean square error were never negative, indicating that the models performed better than the mean

of the measured data. LOO-modeling coverage probabilities verified whole-model predictions of GAMs, and the cGAM uncertainties are reliable 95-percent prediction limits. The predicted L-moment GAMs and no flow cGAM can be used to reliably produce regionalized decadal FDC (dFDC) quantiles. Therefore, dFDC estimations can be made for ungaged stream reaches in the study area. For example, the level-12 hydrologic unit code pour point (COMID 3588922) near U.S. Geological Survey streamgage 08167000 was used to compare an estimated dFDC (using the GAMs and cGAM) at the ungaged location (COMID 3588922) to an observed dFDC at nearby U.S. Geological Survey streamgage 08167000. Observed dFDC was projected from the streamgage to the ungaged location for the comparison using the drainage-area ratio method. The example highlights the ability to approximate dFDCs at ungaged locations from the no-flow fraction and L-moments using asymmetric exponential power, generalized normal, and kappa probability distributions.

References Cited

Alber, M., 2002, A conceptual model of estuarine freshwater inflow management: Estuaries, v. 25, no. 6B, p. 1246–1261, accessed April 17, 2019, at https://doi.org/10.1007/BF02692222.

Anderson, J.M., and Mikhail, E.M., 1998, Surveying theory and practice: Boston, McGraw-Hill.

Archfield, S.A., and Vogel, R.M., 2010, Map correlation method—Selection of a reference streamgage to estimate daily streamflow at ungaged catchments: Water Resources Research, v. 46, no. 10, 15 p., accessed April 17, 2019, at https://doi.org/10.1029/2009WR008481.

Asquith, W.H., 2011, Distributional analysis with L-moment statistics using the R environment for statistical computing: CreateSpace Independent Publishing Platform. [Also available at https://code.usgs.gov/water/restore/fdclmrpplo/-/blob/c003c07c6ea60ac77ed3215ed14ffbb126cfe55d/literature/ISBN9781463508418interior.pdf.]

Asquith, W.H., 2014, Parameter estimation for the 4-parameter asymmetric exponential power distribution by the method of L-moments using R: Computational Statistics & Data Analysis, v. 71, p. 955–970, accessed November 11, 2019, at https://doi.org/10.1016/j.csda.2012.12.013.

Asquith, W.H., 2018, *lmomco*—L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions: R package version 2.3.2, accessed November 13, 2018, at https://CRAN.R-project.org/package=lmomco.

Asquith, W.H., 2020, The use of support vectors from support vector machines for hydrometeorologic monitoring network analyses: Journal of Hydrology (Amsterdam), v. 583, 10 p., accessed March 1, 2021, at https://doi.org/10.1016/j.jhydrol.2019.124522.

Asquith, W.H., Crowley-Ornelas, E.R., and Knight, R.R., 2019, RESTORE/aregis—Geospatial layers transformed to R Spatial objects for RESTORE project scripting convenience: U.S. Geological Survey software release, version 1.0.1, accessed December 3, 2019, at https://doi.org/10.5066/P9O8GXYK.

Asquith, W.H., Kiang, J.E., and Cohn, T.A., 2017, Application of at-site peak-streamflow frequency analyses for very low annual exceedance probabilities: U.S. Geological Survey Scientific Investigations Report 2017–5038, 93 p., accessed October 11, 2019, at https://doi.org/10.3133/sir20175038.

Asquith, W.H., Knight, R.R., and Crowley-Ornelas, E.R., 2020, RESTORE/fdclmrpplo—Source code for estimation of L-moments and percent no-flow conditions for decadal flow-duration curves and estimation at level-12 hydrologic unit codes: U.S. Geological Survey software release, version 1.0.2, accessed March 30, 2020, at https://doi.org/10.5066/P93CKH92.

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 Investigation Report 2006–5286, 34 p., 1 app., accessed March 30, 2020, at https://doi.org/10.3133/sir20065286.

Asquith, W.H., and Thompson, D.B., 2008, Alternative regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using PRESS minimization: U.S. Geological Survey Scientific Investigations Report 2008–5084, 40 p., accessed July 5, 2018, at https://doi.org/10.3133/sir20085084.

Bivand, R.S., Pebesma, E., and Gómez-Rubio, V., 2013, Applied spatial data analysis with R (2d ed.): New York, Springer. [Also available at https://doi.org/10.1007/978-1-4614-7618-4.]

Blum, A.G., Archfield, S.A., and Vogel, R.M., 2017, On the probability distribution of daily streamflow in the United States: Hydrology and Earth System Sciences, v. 21, no. 6, p. 3093–3103, accessed April 17, 2019, at https://doi.org/10.5194/hess-21-3093-2017.

Bomar, G.W., 2017, Weather in Texas—The essential handbook: Austin, Texas, University of Texas Press.

Carlisle, D.M., Falcone, J., Wolock, D.M., Meador, M.R., and Norris, R.H., 2010, Predicting the natural flow regime—Models for assessing hydrological alteration in streams: River Research and Applications, v. 26, p. 118–136, accessed April 17, 2019, at https://doi.org/10.1002/rra.1247.

Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2011, Alteration of streamflow magnitudes and potential ecological consequences—A multiregional assessment: Frontiers in Ecology and the Environment, v. 9, no. 5, p. 264–270, accessed April 17, 2019, at https://doi.org/10.1890/100053.

Castellarin, A., Vogel, R.M., and Brath, A., 2004, A stochastic index flow model of flow duration curves: Water Resources Research, v. 40, no. 3, p. 1–10, accessed April 17, 2019, at https://doi.org/10.1029/2003WR002524.

Castiglioni, S., Castellarin, A., and Montanari, A., 2009, Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation: Journal of Hydrology (Amsterdam), v. 378, no. 3–4, p. 272–280, accessed April 17, 2019, at https://doi.org/10.1016/j.jhydrol.2009.09.032.

Copeland, B.J., 1966, Effects of decreased river flow on estuarine ecology: Estuarine Ecology, v. 38, pp. 1831–1839, accessed June 22, 2020, at https://www.jstor.org/stable/25035675.

Crawley, M.J., 2007, The R book: New York, Wiley. [Also available at https://doi.org/10.1002/9780470515075.]

Crowley-Ornelas, E.R., Asquith, W.H., Knight, R.R., and Worland, S.C., 2019a, Solar radiation for National Hydrography Dataset, version 2 catchments in the southeastern United States—1950–2010: U.S. Geological Survey data release, accessed August 19, 2019, at https://doi.org/10.5066/P9OD7FAL.

Crowley-Ornelas, E.R., Asquith, W.H., Worland, S.C., and Knight, R.R., 2019b, Summary of streamflow statistics for USGS streamgages in the southeastern United States—1950–2010: U.S. Geological Survey data release, accessed August 19, 2019, at https://doi.org/10.5066/P91DOYCJ.

Crowley-Ornelas, E.R., Knight, R.R., Worland, S.C., and Asquith, W.H., 2019c, Heuristically-determined geospatial boundary of streams and rivers draining to the Gulf of Mexico in the south-central and southeastern United States, July 2018: U.S. Geological Survey data release, accessed August 19, 2019, at https://doi.org/10.5066/P9UO82XL.

Crowley-Ornelas, E.R., Worland, S.C., Wieczorek, M.E., Asquith, W.H., and Knight, R.R., 2019d, Summary of basin characteristics for National Hydrography Dataset, version 2 catchments in the southeastern United States—1950–2010: U.S. Geological Survey data release, accessed August 19, 2019, at https://doi.org/10.5066/P9KXTDU4.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610, accessed April 17, 2019, at https://doi.org/10.2307/2288126.

Fang, Z., 2017, cenGAM—Censored regression with smooth terms: R package version 0.5.3, accessed April 4, 2018, at https://cran.r-project.org/web/packages/cenGAM/index.html.

Faraway, J.J., 2005, Linear models with R: Boca Raton, Florida, Chapman & Hall/CRC.

Faraway, J.J., 2006, Extending the linear model with R—Generalized linear, mixed effects and nonparametric regression models: Boca Raton, Florida, Chapman & Hall/CRC.

Farmer, W.H., Archfield, S.A., Over, T.M., Hay, L.E., LaFontaine, J.H., and Kiang, J.E., 2015a, A comparison of methods to predict historical daily streamflow time series in the southeastern United States: U.S. Geological Survey Scientific Investigations Report 2014–5231, 34 p., accessed April 17, 2019, at https://doi.org/10.3133/sir20145231.

Farmer, W.H., Over, T.M., and Vogel, R.M., 2015b, Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States: Water Resources Research, v. 51, no. 3, p. 1775–1796, accessed April 17, 2019, at https://doi.org/10.1002/2014WR015924.

Gulf Coast Ecosystem Restoration Council, 2016, Comprehensive Plan—Approved Comprehensive Plan update 2016, accessed April 19, 2018, at https://www.restorethegulf.gov/sites/default/files/CO-PL_20161208_CompPlanUpdate_ English.pdf.

Gulf Coast Ecosystem Restoration Council, 2019a, RESTORE, accessed March 30, 2019, at https://www.restorethegulf.gov/sites/default/files/AB-US_20160818_About_the_Council_Fact_Sheet%20August%202016.pdf.

Gulf Coast Ecosystem Restoration Council, 2019b, Resources and ecosystems sustainability, tourist opportunities, and revived economies of the Gulf Coast States Act (RESTORE Act) Initial Funded Priorities List, accessed September 28, 2019, at https://www.restorethegulf.gov/sites/default/files/FPL_forDec9Vote_Errata_04-07-2016.pdf.

Helsel, D.R., 2012, Statistics for censored environmental data using Minitab and R: New York, John Wiley.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chap. A3, 484 p., accessed June 1, 2020, at https://doi.org/10.3133/tm4A3.

Henningsen, A., 2018, censReg—Censored regression (Tobit) models: R package version 0.5-26, accessed April 4, 2018, at https://cran.r-project.org/web/packages/censReg/index.html.

Hildebrand, H.H., and Gunter, G., 1953, Correlation of rainfall with Texas catch of white shrimp, Penaeus setiferus (Linnaeus): Transactions of the American Fisheries Society, v. 82, no. 1, p. 151–155, accessed April 17, 2019, at https://doi.org/10.1577/1548-8659(1952)82[151:CORWTT]2.0.CO;2.

Hosking, J.R.M., 1990, L-moments—Analysis and estimation of distributions using linear combination of order statistics: Journal of the Royal Statistical Society. Series B. Methodological, v. 52, no. 1, p. 105–124, accessed June 22, 2020, at https://doi.org/10.1111/j.2517-6161.1990.tb01775.x.

Hosking, J.R.M., 2006, On the characterization of distributions by their L-moments: Journal of Statistical Planning and Inference, v. 136, no. 1, p. 193–198, accessed April 17, 2019, at https://doi.org/10.1016/j.jspi.2004.06.004.

Hosking, J.R.M., and Wallis, J.R., 1997, Regional frequency analysis—An approach based on L-moments: Cambridge, Cambridge University Press.

Karian, Z.A., and Dudewicz, E.J., 2010, Handbook of fitting statistical distributions with R: Boca Raton, Florida, CRC Press.

Kroll, C.N., and Stedinger, J.R., 1999, Development of regional regression relationships with censored data: Water Resources Research, v. 35, no. 3, p. 775–784, accessed April 17, 2019, at https://doi.org/10.1029/98WR02743.

Mendicino, G., and Senatore, A., 2013, Evaluation of parametric and statistical approaches for the regionalization of flow duration curves in intermittent regimes: Journal of Hydrology (Amsterdam), v. 480, p. 19–32, accessed April 17, 2019, at https://doi.org/10.1016/j.jhydrol.2012.12.017.

Over, T.M., Riley, J.D., Sharpe, J.B., and Arvin, D., 2014, Estimation of regional flow-duration curves for Indiana and Illinois: U.S. Geological Survey Scientific Investigations Report 2014–5177, 24 p., accessed March 30, 2020, at https://doi.org/10.3133/sir20145177.

Pebesma, E., Bivand, R.S., Rowlingson, B., Gómex-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J., O'Brien, J., and O'Rourke, J., 2018, sp—Classes and methods for spatial data: R package version 1.2-7, accessed April 4, 2018, at https://cran.r-project.org/web/packages/sp/index.html.

Poncelet, C., Andréassian, V., Oudin, L., and Perrin, C., 2017, The Quantile Solidarity approach for the parsimonious regionalization of flow duration curves: Hydrological Sciences Journal, v. 62, no. 9, p. 1364–1380, accessed April 17, 2019, at https://doi.org/10.1080/02626667.2017.1335399.

Powell, G.L., Matsumoto, J., and Brock, D.A., 2002, Methods for determining minimum freshwater inflow needs of Texas bays and estuaries: Estuaries, v. 25, no. 6, p. 1262–1274, accessed April 17, 2019, at https://doi.org/10.1007/BF02692223.

Pugliese, A., Farmer, W.H., Castellarin, A., Archfield, S.A., and Vogel, R.M., 2016, Regional flow duration curves—Geostatistical techniques versus multivariate regression: Advances in Water Resources, v. 96, p. 11–22, accessed April 17, 2019, at https://doi.org/10.1016/j.advwatres.2016.06.008.

Pumo, D., Viola, F., La Loggia, G., and Noto, L.V., 2014, Annual flow duration curves assessment in ephemeral small basins: Journal of Hydrology (Amsterdam), v. 519, part A, p. 258–270, accessed April 17, 2019, at https://doi.org/10.1016/j.jhydrol.2014.07.024.

R Core Team, 2019, R—A language and environment for statistical computing, version 3.6.0: Vienna, Austria, R Foundation for Statistical Computing, accessed April 26, 2019, at https://www.r-project.org.

Requena, A.I., Chebana, F., and Ouarda, T.B.M.J., 2018, A functional framework for flow-duration-curve and daily streamflow estimation at ungauged sites: Advances in Water Resources, v. 113, p. 328–340, accessed April 17, 2019, at https://doi.org/10.1016/j.advwatres.2018.01.019.

Robinson, A.L., Asquith, W.H., Crowley-Ornelas, E., and Knight, R.R., 2021, Estimated quantiles of decadal flow-duration curves using selected probability distributions fit to no-flow fractions and L-moments predicted for streamgages and for pour points of level-12 hydrologic unit codes in the southeastern United States, 1950–2010: U.S. Geological Survey data release, https://doi.org/10.5066/P9MV8BYR.

Robinson, A.L., Asquith, W.H., and Knight, R.R., 2019, Summary of decadal no-flow fractions and decadal L-moments of nonzero streamflow flow-duration curves for National Hydrography Dataset, version 2 catchments in the southeastern United States, 1950–2010: U.S. Geological Survey data release, https://doi.org/10.5066/P9Z4PM55.

Robinson, A.L., Worland, S.C., and Rodgers, K.D., 2020, Estimated daily mean streamflows for HUC12 pour points in the southeastern United States, 1950–2009: U.S. Geological Survey data release, https://doi.org/10.5066/P9P36GXZ.

Searcy, J.K., 1959, Flow-duration curves: U.S. Geological Survey Water Supply Paper 1542-A, 33 p., accessed April 17, 2019, at https://doi.org/10.3133/wsp1542A.

Stasinopoulos, M.D., Rigby, R.A., Heller, G.Z., Voudouris, V., and Bastiani, F.D., 2017, Flexible regression and smoothing using GAMLSS in R: Boca Raton, Florida, Chapman & Hall/CRC.

Texas Commission on Environmental Quality, 2018, Download TCEQ GIS data—Edwards aquifer regulatory boundary, accessed January 24, 2018, through https://www.tceq.texas.gov/gis/download-tceq-gis-data at https://www.tceq.texas.gov/assets/public/gis/exports/edwards_tsms.zip and metadata at https://www.tceq.texas.gov/assets/public/gis/metadata/edw_tsms.pdf.

Texas Commission on Environmental Quality, Texas Parks and Wildlife Department, and Texas Water Development Board, 2008, Texas instream flow studies—Technical overview: Austin, Tex., Texas Water Development Board, Report 369.

Therneau, T.M., and Lumley, T., 2018, survival—Survival analysis: R package version 2.41–3, accessed April 4, 2018, at https://cran.r-project.org/web/packages/survival/index.html.

Tobin, J., 1958, Estimation of relationships for limited dependent variables: Econometrica, v. 26, no. 1, p. 24–36. [Also available at https://doi.org/10.2307/1907382.]

U.S. Army Corps of Engineers, 2018, CorpsMap—National Inventory of Dams, accessed July 7, 2020, at https://nid.sec.usace.army.mil/ords/f?p=105:1.

U.S. Environmental Protection Agency [EPA], 2018a, NHDPlus (National Hydrography Dataset Plus), accessed April 11, 2018, at https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus.

U.S. Environmental Protection Agency [EPA], 2018b, Level III and IV Ecoregions of the continental United States, accessed April 24, 2018, at https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states.

U.S. Geological Survey [USGS], 2018, Watershed Boundary Dataset, accessed March 30, 2018, at https://nhd.usgs.gov/wbd.html.

U.S. Geological Survey [USGS], 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed August 31, 2019, at https://doi.org/10.5066/F7P55KJN.

U.S. Geological Survey [USGS], 2020, FPL3b—Proposal—DOI USGS—Develop ecological flow decision-support for Mobile River and Perdido River Basins, accessed September 22, 2020, at https://www.restorethegulf.gov/files/fpl3b-proposal-doi-usgs-develop-ecological-flow-decision-support-mobile-river-and-perdido.

Vogel, R.M., and Fennessey, N.M., 1994, Flow duration curves I—New interpretation and confidence intervals: Journal of Water Resources Planning and Management, v. 120, no. 4, p. 485–504, accessed April 17, 2019, at https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485).

Vogel, R.M., and Fennessey, N.M., 1995, Flow duration curves II—A review of applications in water resources planning: Journal of the American Water Resources Association, v. 31, no. 6, p. 1029–1039, accessed April 17, 2019, at https://doi.org/10.1111/j.1752-1688.1995.tb03419.x.

Wang, C.Y., and Singh, R.P., 1995, Frequency estimation for hydrological samples with zero data: Journal of Water Resources Planning and Management, v. 121, no. 1, p. 98–108, accessed April 17, 2019, at https://doi.org/10.1061/(ASCE)0733-9496(1995)121:1(98).

Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018, Select attributes for NHDPlus version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States: U.S. Geological Survey data release, accessed April 17, 2019, at https://doi.org/10.5066/F7765D7V.

Wood, S.N., 2008, Fast stable direct fitting and smoothness selection for generalized additive models: Journal of the Royal Statistical Society, Series B, Statistical Methodology, v. 70, no. 3, p. 495–518, accessed April 17, 2019, at https://doi.org/10.1111/j.1467-9868.2007.00646.x.

Wood, S.N., 2017, Generalized additive models—An introduction with R (2d ed.): Boca Raton, Florida, Chapman and Hall/CRC.

Wood, S.N., 2018, mgcv—Mixed GAM computation vehicle with automatic smoothness estimation: R package version 1.8–23, accessed April 4, 2018, at https://cran.r-project.org/web/packages/mgcv/index.html.

Worland, S.C., Knight, R.R., and Asquith, W.H., 2019a, Estimated quantiles for the pour points of 9,203 level-12 hydrologic unit codes in the southeastern United States, 1950–2009: U.S. Geological Survey data release, accessed July 5, 2020, at https://doi.org/10.5066/P9YGKZZV.

Worland, S.C., Knight, R.R., and Asquith, W.H., 2019b, Observed and modeled daily streamflow values for 74 U.S. Geological Survey streamgage locations in the Trinity and Mobile-Tombigbee River basins in the southeast United States, 2000–2009: U.S. Geological Survey data release, accessed July 5, 2020, at https://doi.org/10.5066/P92F1ROU.

Worland, S.C., Steinschneider, S., Asquith, W.H., Knight, R.R., and Wieczorek, M.E., 2019c, Prediction and inference of flow-duration curves using multi-output neural networks: Water Resources Research, v. 55, no. 8, p. 6850–6868, accessed July 5, 2020, at https://doi.org/10.1029/2018WR024463.

Worland, S.C., Steinschneider, S., Farmer, W.H., Asquith, W.H., and Knight, R.R., 2019d, Copula theory as a generalized framework for flow-duration curve based streamflow estimates in ungaged and partially gaged catchments: Water Resources Research, v. 55, no. 11, p. 9378–9397, accessed July 5, 2020, at https://doi.org/10.1029/2019WR025138.

Glossary

AEP4 Asymmetric exponential power probability distribution (4-parameter) (Asquith and others, 2017, p. 90).

cGAM-PPLO The censored GAM of the no-flow fraction (PPLO).

COMID An identification number for a stream reach (Wieczorek and others, 2018).

DAR method Drainage-area ratio method (Asquith and others, 2006). Method of spatially transferring streamflow from gaged to ungaged sites (Farmer and others, 2015b) under the assumption that streamflow over the geographic extent is constant and can be simply scaled over an ungaged drainage area.

FDC Flow-duration curve. Can be modified by prepending a time period, such as "period-of-record FDC," "decadal FDC," or "2000s FDC."

GAM-L1 The GAM of the decadal mean nonzero streamflow (L1).

GAM-T2 The GAM of the decadal coefficient of L-variation of nonzero streamflow (T2).

GAM-T3 The GAM of the decadal L-skew of nonzero streamflow (T3).

GAM-T4 The GAM of the decadal L-kurtosis of nonzero streamflow (T4).

GAM-T5 The GAM of the decadal fifth L-moment ratio of nonzero streamflow (T5).

GLO Generalized logistic distribution (3-parameter) which is the upper L-kurtosis (T4) bounds of the kappa distribution (Asquith and others, 2017, p. 85).

GNO Generalized normal probability distribution (3-parameter) and equivalent to 3-parameter log-normal distribution (Asquith and others, 2017, p. 86).

HUC Hydrologic unit code, a unique address of a spatial region of a river system.

KAP Kappa probability distribution (4-parameter) (Asquith and others, 2017, p. 91).

L1 Decadal mean nonzero streamflow.

PPLO No-flow fraction or equivalently, the percentage of decadal no-flow.

PPLO-ubL1 Shorthand for overall mean streamflow (watershed yield) computed by proration of the no-flow fraction (PPLO) and mean nonzero streamflow (L1) using a Duan retransformation bias correction (ub) (Helsel and others, 2020, p. 256).

RESTORE/aregis A minor repository (Asquith, Crowley-Ornelas, and Knight, 2019) of software preprocessing geospatial layers that were in turn used by the RESTORE/fdclmrpplo repository.

RESTORE/fdclmrpplo A large repository of statistical software and documentation (Asquith, Knight, and Crowley-Ornelas, 2020) used in the computations described in this report.

T2 Decadal coefficient of L-variation of nonzero streamflow.

T3 Decadal L-skew of nonzero streamflow.

T4 Decadal L-kurtosis of nonzero streamflow.

T5 Decadal fifth L-moment ratio of nonzero streamflow.

For more information about this publication, contact
Director, Lower Mississippi-Gulf Water Science Center
U.S. Geological Survey
640 Grassmere Park, Suite 100
Nashville, TN 37211
For additional information, visit
https://www.usgs.gov/centers/lmg-water/
Publishing support provided by
Lafayette Publishing Service Center

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20225051