Generalized Additive Regression Models of Discharge and Mean Velocity Associated with Direct-Runoff Conditions in Texas: Utility of the U.S. Geological Survey Discharge Measurement Database

William H. Asquith, Ph.D., P.G.[1]; George R. Herrmann, P.E., P.H., M.ASCE[2]; and Theodore G. Cleveland, Ph.D., P.E.[3]

Abstract: A database containing more than 17,700 discharge values and ancillary hydraulic properties was assembled from summaries of discharge measurement records for 424 U.S. Geological Survey streamflow-gauging stations (stream gauges) in Texas. Each discharge exceeds the 90th-percentile daily mean streamflow as determined by period-of-record, stream-gauge-specific, flow-duration curves. Each discharge therefore is assumed to represent discharge measurement made during direct-runoff conditions. The hydraulic properties of each discharge measurement included concomitant cross-sectional flow area, water-surface top width, and reported mean velocity. Systematic and statewide investigation of these data in pursuit of regional models for the estimation of discharge and mean velocity has not been previously attempted. Generalized additive regression modeling is used to develop readily implemented procedures by end-users for estimation of discharge and mean velocity from select predictor variables at ungauged stream locations. The discharge model uses predictor variables of cross-sectional flow area, top width, stream location, mean annual precipitation, and a generalized terrain and climate index (OmegaEM) derived for a previous flood-frequency regionalization study. The mean velocity model uses predictor variables of discharge, top width, stream location, mean annual precipitation, and OmegaEM. The discharge model has an adjusted R-squared value of about 0.95 and a residual standard error (RSE) of about 0.22 base-10 logarithm (cubic meters per second); the mean velocity model has an adjusted R-squared value of about 0.67 and an RSE of about 0.063 fifth root (meters per second). Example applications and computations using both regression models are provided. **DOI: 10.1061/(ASCE)HE.1943-5584.0000635.** © *2013 American Society of Civil Engineers.*

CE Database subject headings: Streamflow; Discharge measurement; Velocity; Hydraulic properties; Runoff; Regional analysis; Texas; Databases; U.S. Geological Survey.

Author keywords: Streamflow; Stream gauge; Discharge measurement; Mean velocity; Hydraulic properties; Direct runoff; Regional analysis; Regression; Generalized additive models; Texas.

Introduction

The U.S. Geological Survey (USGS) for the operational support of the streamflow-gauging station (stream gauge) network in Texas collected and digitally archived about 140,000 discharge measurements (including zero-flow values) and stream-gauge inspections for more than 600 stream gauges for the approximate period December 1897 to February 2009. These discharge measurements, which are actually individual summaries of extensive field-collected data, reside within the USGS National Water Information System (NWIS) and are readily obtained (USGS 2009b) by stream-gauge number (a unique numerical identifier). The vast majority of the data represent discharges Q measured from current-meter-based (velocity meter) techniques (Turnipseed and Sauer 2010). For most of the discharge measurements, concomitant hydraulic properties are also available; these are cross-sectional flow area A, water-surface top width B, reported mean velocity V, and other details. The basic relation between Q, A, and V is $Q = AV$. The basic relation between hydraulic (mean) depth D and A and B is $D = A/B$.

The National Research Council (1999, p. 29) stated that a "wealth of information on geomorphology could be extracted from the USGS's vast discharge measurement file." This paper demonstrates that the imposing number of records, the flow-condition range, and the large number of stream gauges contained just within the USGS discharge measurement database in Texas facilitate the regionalization of Q and V. The term regionalization in the hydrologic sciences is a framework for statistical analyses that produce procedures for estimation of various properties, such as discharge, at ungauged or unmonitored locations from select characteristics at those locations. The regional models of Q and V reported here demonstrate that indeed a wealth of generalized hydraulic information can be associated with simple metrics of channel morphology and stream location as anticipated by the National Research Council (1999). The National Research Council (2004, pp. 122–123)

[1]U.S. Geological Survey, Lubbock Field Office, Science Building MS-1053, Texas Tech Univ., 2500 Broadway, Lubbock, TX 79409 (corresponding author). E-mail: wasquith@usgs.gov

[2]Research Associate, Dept. of Civil and Environmental Engineering, Texas Tech Univ., CEE Building MS-1023, Lubbock, TX 79409; and Retired, Texas Dept. of Transportation. Austin, TX 78701.

[3]Dept. of Civil and Environmental Engineering, Texas Tech Univ., CEE Building MS-1023, Lubbock, TX 79409.

Note. This manuscript was submitted on February 16, 2012; approved on September 13, 2012; published online on September 16, 2013. Discussion period open until March 1, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 18, No. 10, October 1, 2013. © ASCE, ISSN 1084-0699/2013/10-1331-1348/$25.00.

stated "surprisingly, the USGS and other groups have not published hydraulic geometry relationships [...] for hydroclimatic regions of the United States. A consequence of this is that [situations requiring] hydraulic geometry try [to] use either 'average' hydraulic geometry relationships, which are often the data from Leopold and Maddock (1953) or stream classifications schemes[.]"

Purpose, Scope, and Organization

The purpose of this paper is to document the first systematic and statewide investigation, conducted in cooperation with the Texas Department of Transportation, of principle features of the USGS discharge measurement database in Texas to regionalize (1) the relation between Q and selected predictor variables and (2) the relation between V and selected predictor variables. The objective of the regionalization is to create readily used procedures for engineers and scientists so that they may readily implement the parametric and semiparametric models presented in this paper.

The scope of this paper is limited to discharge measurement records for stream gauges in Texas that are anticipated to represent direct-runoff conditions (see Fig. 1). The regionalization is based on generalized additive models or modeling (GAM) in which selected predictor variables include those associated with fundamental hydraulics and other predictor variables that are readily determined from maps and graphical plots or special *smoothing* functions. These maps and plots are provided herein.

This paper is organized as follows. Previous studies having either conceptual association or those with salient hydraulic analysis are summarized in the "Previous Studies" section. The regional analyses of Q and V are intended to be used in applied circumstances; various applications are discussed in the "Regionalization of Discharge Measurement Databases: Potential Applications" section. In particular, some applications of a Q regional model are discussed in the "Potential Applications of a Regional Model of Discharge" section, and some applications of a V regional model are discussed in the "Potential Applications of a Regional Model of Mean Velocity" section. The "Database of Discharge Measurements" section discusses the data manipulation required to create a unified discharge measurement database in Texas that contains discharge measurements spanning low-flow to high-flow conditions. For this paper, the unified database went through a subsequent paring into anticipated high-magnitude Q to create a database with general association to direct-runoff conditions. The definition of high-magnitude Q and other details are provided in the "Database of Discharge Measurements" section, which, for further scope, see the statistical analyses herein.

Fig. 1. (Color) USGS personnel conducting one of two high-magnitude discharge measurements on January 11, 2007, at USGS streamflow-gauging station 08156800 Shoal Creek at West 12th Street, Austin, Texas; both measurements are represented in the database used for this paper (Photograph by W. H. Asquith and courtesy of USGS)

The regional analysis framework using GAMs is introduced in the "Generalized Additive Models and Regionalization of Discharge and Mean Velocity" section, and a brief introduction to GAM and the basic model forms chosen is provided in the "Generalized Additive Models" section. The preprocessing and preliminary analyses are described in the "Preprocessing and Preliminary Analysis" section, and, in particular, that section describes two non-hydraulic predictor variables selected for regionalization and the topic of selection of suitable variable transformation. The final regional model of Q is presented in the "Generalized Additive Model of Discharge" section, and the final regional model of V is presented in the "Generalized Additive Model of Mean Velocity" section. A discussion on limitations and thoughts for model improvement follows in the "Limitations of QGAM and VGAM and Thoughts for Improvement" section.

The regional models of Q and V herein are intended for use in applied circumstances. Therefore, the "Example Applications" section provides some example applications with extensive example computations to help guide the user. In particular, the "Postevent Discharge Estimation" section provides an example of Q computation, whereas, the "Review of Mean Velocity from a Hydraulic Model" section provides an example of V computation as well as a method to approximate the distribution of a given prediction. Additional discussion of results is provided in the "Discussion" section.

Previous Studies

A conceptual precursor for discharge estimation from channel properties is provided by Riggs (1976), who describes a *simplified* slope-area method for estimation of peak discharge Q_p in natural channels in the Pacific Northwest. The slope-area method (Dalrymple and Benson 1967) can be used to estimate post-direct-runoff peak discharge based on evidence of peak water-surface elevation or extent and corresponding cross-sectional geometric properties. Water-surface elevations are assumed to represent friction slopes S necessary for hydraulic computations in a selected stream reach, and when S are combined with topographic surveys (see Fig. 2) providing multiple cross-sectional areas and other hydraulic properties, an estimate of Q_p results from the slope-area method. Unfortunately, the slope-area method is labor-intensive and expensive. Riggs (1976) sought a quick, reproducible, and inexpensive alternative or compliment to the slope-area method. Arguing that Manning n-values and water-surface slopes are coupled relations, Riggs (1976) proposed that discharge Q can be estimated by

$$Q = c_1 A^{c_2} S^{c_3} \tag{1}$$

Fig. 2. (Color) USGS personnel surveying on December 15, 2004, one of four stream cross sections to support a slope-area computation of peak discharge for a historically important event at USGS streamflow-gauging station 08148500 North Llano River near Junction, Texas (Photograph by W. H. Asquith and courtesy of USGS)

where A = cross-sectional flow area, S = water-surface slope, and c_k = regression coefficients for a particular study area with $k = 1, 2, 3$. Riggs (1976) continued with a further simplification and argued that the contribution of the water-surface slope term can be removed.

Castro and Jackson (2001) investigated statistical relations between various hydraulic elements for 76 USGS stream gauges in the Pacific Northwest. Their primary objectives were to (1) test the validity of the assumption that the 1.5-year (0.67 annual exceedance probability, AEP) discharge represents bankfull conditions; (2) define alternative relations of the T-year bankfull discharge in the study area; and (3) define statistical relations for discharge and channel hydraulics by geographic region. Castro and Jackson (2001, Table 4) list ensembles of regression equations for four geographic regions. Some of these equations have algebraic similarity to the regression models presented in this paper. Those authors developed four regression equations of top width B in the form $B = d_1 Q^{d_2}$ for two regression coefficients d_k. The weighted mean of the exponent d_2 on Q for the four equations is 0.497, which was computed from the tabulated exponents in Castro and Jackson (2001, Table 4). Fitting the Castro and Jackson statistical model to the database described near the end of the "Database of Discharge Measurements" section results in a d_2 value of 0.459, which is similar to the weight-mean exponent computed from Castro and Jackson (2001) of $d_2 = 0.497$ for rivers in the Pacific Northwest. The exponent similarity is interesting because the study areas and the underlying databases are different. Castro and Jackson (2001) used site visits and hydraulic analyses; the analyses in this paper are based exclusively on statistical processing of discharge measurements.

An extensive number of studies have been done related to a regionalization of a range of streamflow statistics in Texas (Slade et al. 1995; Asquith et al. 1996; Devulapalli and Valdes 1996; Asquith and Slade 1997; Raines and Asquith 1997; Asquith 1998; Raines 1998; Lanning-Rush 2000; Rifai et al. 2000; Asquith 2001; Asquith and Thompson 2008; Asquith and Roussel 2009) including a study on the drainage-area ratio method by Asquith et al. (2006). However, these studies generally are focused on the classical problem of estimation of a streamflow statistic (such as the median 7-day low flow, the mean annual streamflow, or the 0.1 annual exceedance probability peak streamflow). Wurbs and Kim (2011) discuss and provide extensive background and citations concerning monthly streamflow estimation as part of Texas water availability modeling to support planning and water rights analysis; the water availability modeling represents a fundamentally different thematic scope than the studies cited at the beginning of this paragraph. In total, all of these studies are fundamentally different from the current (2012) study, which is explicitly focused on regionalization of summaries of USGS discharge measurements and not the estimation of a particular statistic derived from time series of streamflow, such as the mean annual streamflow derived from annual mean streamflow values.

Regionalization of Discharge Measurement Databases: Potential Applications

This section provides description of two interrelated applications of regionalized discharge measurement databases. The applications have distinct circumstances of use that are demonstrated by numerical examples in the "Example Applications" section.

Potential Applications of a Regional Model of Discharge

The regionalization of Q has potential applications for (1) estimation of peak discharge Q_p from readily field-surveyed cross-sectional topography after high-magnitude discharge events, (2) provisional stage-discharge relations from cross-sectional topography, and (3) other applications. After substantial flooding, evidence, such as debris lines on embankments or seed lines on trees, of peak water-surface elevation often remains. A regional model of Q based on A, B, D, and other factors could provide for a relatively straightforward means to estimate Q_p using measured or estimated values of A, B, and D for the event with potentially less labor and expense, albeit with potentially greater uncertainty, compared to a slope-area computation of discharge.

When stream gauges are activated there is a period of time in which the initial development of the stage-discharge relations is needed. Regionalization of Q could facilitate the creation of provisional stage-discharge relations prior to actual measurements being made and subsequently used to initially define the relation between discharge and stage (also referred to as a rating curve or corresponding rating table) at new stream gauges. Typical direct discharge measurements have a potential error of about 5 to 8%—the potential errors associated with regionalized Q from direct measurements likely are much larger than 8%. Indirect measurements of discharge errors are "probably several times larger" than those for direct measurements (Potter and Walker 1981, p. 1505). Stage-discharge relations can be based on both types of discharge measurements, and Potter and Walker (1981) provide extensive discussion of the effects of this fundamental shift in relative error on the peak-streamflow frequency curve.

Other applications of regionalization of Q are foreseen. Estimation of Q (as well as V) for ungauged stream cross sections in Texas has obvious connections to hydraulic modeling but also connection to *instream-flow assessments* for aquatic and riparian habitats. For example, hydraulic values derived from cross-sectional and longitudinal surveys of selected, ungauged stream reaches in the Edwards Plateau, Texas, have been used to predict magnitude and frequency of bed-material entrainment flows for purposes of mitigating maintenance costs associated with gravel bombardment of road crossings (Heitmuller and Asquith 2008). Other hydrologic programs in Texas, notably those to quantify environmental flows [Texas Senate Bill 3 Science Advisory Committee for Environmental Flows (TSAC) 2009], mandate hydraulic assessments of ungauged stream reaches for purposes of aquatic and riparian habitat conservation [Texas Commission on Environmental Quality (TCEQ) and Texas Parks and Wildlife Department (TPWD) and

Texas Water Development Board (TWDB) 2008]. In an expression of the utility of discharge measurements and attendant characteristics for instream-flow assessments, Heitmuller and Greene (2009) rendered historical cross sections and computed hydraulic values at 15 USGS stream gauges in the Brazos and Sabine River basins. The historical cross sections and computed hydraulics are useful for detecting geomorphic and hydraulic conditions associated with instream habitat structure and function. Heitmuller and Greene (2009) as well as Coffman et al. (2011) describe geomorphic associations and properties of select reaches of various Texas riverine systems (Brazos, Sabine, and Trinity) from spatial and temporal perspectives. These two reports provide information as to the complexities of Texas river systems and their responses to flood-control measures, channel modifications, landscape changes, and other activities.

The general assessment or regionalization of hydraulic characteristics for ungauged stream locations provides needed flexibility to support conservation efforts [Texas Senate Bill 3 Science Advisory Committee for Environmental Flows (TSAC) 2009]. Finally, various efforts to model runoff, contaminant loads, and sediment loads commonly are needed for ungauged stream locations (Clark et al. 2000; Morehead et al. 2003; Ockerman and Heitmuller 2010). These types of studies might benefit from regionalization of discharge measurement databases.

Potential Applications of a Regional Model of Mean Velocity

The regionalization of V has potential uses for rapid and reliable review of $\bar{V} = Q/A$ (mean velocity) that can emanate from one-dimensional backwater models. These models often are used to model peak water-surface elevations of high-magnitude discharge, for computations of bridge scour or bank protection, and for other applications. The authors observe that it is common for engineers involved in one-dimensional backwater modeling to have been taught to assemble models based on generalizations of parameter values from textbooks (Jain 2001; Sturm 2010) or literature of the method (ASCE 1996), from computer program documentation, and from experience. However, the aforementioned *experience* often is exclusive to *prior modeling experience*—an example of circular logic. The authors also observe that conventional engineering education, as well as practice, lacks physical (observational) experience with or even exposure to streamflow metrology as exemplified by the discharge measurements supporting operation of the nationally consistent USGS stream-gauge network.

In one-dimensional, open-channel computations, parameters such as Manning n-values are selected from tables (Sturm 2010, Table 4.1), graphs, other published procedures, and ideally from visual site assessments (Barnes 1967). For certain parameters, such as coefficients for expansion or contraction losses, the default values are often used. This practice (understandably) is made because typically there is scant information on which to base alternative values. As a result, modeling efforts by even experienced modelers are assembled and often judged to be valid based entirely on experiences from earlier modeling efforts for hydraulically similar settings.

Unfortunately, unless model calibration is influenced by data from one or more stream gauges, there is seldom any independent information to assess the validity of a given model. Many assessments of, and discussions about, hydraulic model validity necessarily begin and end as expressions of individual professional opinion with often scant quantification to discriminate between valid and invalid models. A regional model of V could provide a fundamental link to physical reality and potentially could provide an authoritative and independent measure of consistency that will allow for enhanced assessment of one-dimensional, open-channel computations and general model reliability. A regional model of V would provide a tool to flag severely inconsistent situations and identify these for further scrutiny.

A regional model of V could also serve as a means for straightforward computation of real-time velocity information to augment real-time discharge data from USGS stream gauges in the context of stream-spill scenarios and attendant emergency response.

Database of Discharge Measurements

This section provides background information to elucidate various nuances concerning observed values of Q and other channel characteristics in Texas. Further, this section discusses various data gaps and information barriers that hinder systematic regionalization of Q and V relations in Texas. These gaps and barriers are not exclusive to the USGS discharge measurement records in Texas; they are likely endemic to other historic or emergent discharge measurements recorded elsewhere by the USGS or other entities.

A unified database of discharge measurements in Texas was prepared by the authors from the USGS National Water Information System (USGS 2009b). The database contains 89,874 discharge records for the approximate period December 1897 to February 2009 for 437 selected Texas stream gauges. The 437 stream gauges were selected as a prerequisite for this paper based on preliminary screening of more than 600 stream gauges and select preprocessing that included factors such as consideration of streamflow data type, record length, number of discharge measurements, and regional setting or location of the stream gauge. In general, a candidate stream gauge needed to be a continuous-record type and represent stream gauges that are considered examples of conventional (traditional) USGS stream-gauging operation and not special projects (perhaps stream gauges operated with theoretical weir stage-discharge relations), partial duration stream gauges (perhaps flood hydrograph, or conversely, low-flow stream gauges), or peak-only stream gauges.

This unified discharge measurement database provides the foundational basis for the analysis reported here and contains the following attributes: discharge, reported mean velocity, cross-sectional flow area, water-surface top width, Froude number, and estimated flow-duration probability of the discharge. Unlike the approach by Castro and Jackson (2001) in their study of regional bankfull relations, no site visits to any of the 437 Texas stream gauges were made for this study. The unified discharge measurement database was assembled through the following steps:

1. Daily mean streamflow values: For the large and reasonably comprehensive list (437) of continuous-record (daily mean values of streamflow) stream gauges in Texas, the daily mean streamflow values were retrieved from USGS (2009a).
2. Streamflow measurements: For the 437 stream gauges, the discharge measurement file for each stream gauge was retrieved from USGS (2009b).
3. Complete records: The measured discharge Q in cubic meters per second (m^3/s), *channel velocity* (referred to herein as *reported mean velocity*) V in meters per second (m/s), *channel area* (referred to herein as *cross-sectional flow area*) A in square meters (m^2/s), and *channel width* (referred to herein as *water-surface top width* or just *top width*) B in meters (m) were extracted, and only those records with $Q > 0$ were retained.
4. Computed mean velocity: Computed mean velocity $\bar{V}$ in m/s was computed by $\bar{V} = Q/A$. The adjective *computed*

(as opposed to *reported* mean velocity in Step 3) in this paper refers to Q divided by A irrespective of the source of Q or A.

5. Velocity consistency: The computed $\bar{V}$ was compared to the reported V, and if the absolute difference was greater than 0.03 m/s (chosen by the authors), then the record (a single discharge measurement) was rejected for inclusion in the unified discharge measurement database and thus not retained for the analysis reported here.
6. Froude number: The Froude number was computed by $\mathsf{F} = V(gA/B)^{-1/2}$, where g is acceleration of gravity. For this paper, F is not used but is retained in the unified discharge measurement database.
7. Flow-duration curve: The entire period of record of daily mean streamflow for each stream gauge referenced in Step 1 was converted to a stream-gauge-specific, flow-duration curve (Vogel and Fennessey 1994).
8. Individual discharge probabilities: The probability of each Q was determined from the respective stream-gauge-specific, flow-duration curve of daily mean streamflow values, using linear interpolation as necessary.

Further discussion of selected details of the eight steps is needed to provide additional context for various decisions or observations that are important to communicate:

- On greater-than-zero discharge: Step 3 excludes reverse flow ($Q < 0$) in tidal and zero-flow conditions ($Q = 0$);.
- On incomplete attributes: To clarify, any discharge measurements (direct or indirect) lacking any core attributes (Q, V, A, and B) or in violation of Step 5 were not retained for the unified discharge measurement database.
- On streamflow probability: Step 7 states that the *entire* record of each stream gauge was used to compute each stream-gauge-specific, flow-duration curve. This explicitly means that no attempt was made to define periods of stationary (unchanging statistical properities) streamflow or, more importantly, statistics of hydraulic relations. For example, no differentiation between prereservoir and postreservoir conditions (if applicable) for a given stream gauge was made. Such stream-gauge-specific investigation is beyond the scope of this paper.
- On stream-gauge location: USGS stream gauges are only very rarely located in settings in which backwater conditions occur, because a unique stage-discharge relation is desired. Also, a given stream gauge is not anticipated to permanently exist at the exact same location along a stream during the course of the stream gauge's operational time frame; however, many stream gauges remain more or less sited at their original locations. Stream-gauge locations are referenced to the nearest town or locality with a postal code, for example, USGS stream gauge 08167000 Guadalupe River near Comfort, Texas. Stream gauges are periodically relocated to nearby locations, but adjustments to identity (number and name) are not made, because of channel migration; channel rectification/restoration; bridge maintenance, decommission, and new construction; property access (landowner changes); and changes in safety policy and practices. Changes in bridge characteristics are likely the most common cause of relocation, because many stream gauges in Texas often are located along Texas Department of Transportation right-of-way.
- On measurement location: A fact, which likely hampers many stream-gauge-specific investigations of geomorphic processes using USGS measurement databases, is that the precise cross-sectional location of an individual discharge measurement is neither reported nor fully documented in USGS discharge measurement summaries used herein. Furthermore, the measurement location is not expected to coincide with the same location either over the years or over a range of discharge conditions. There are many discipline-specific and technically specific reasons discharge measurements might not be made at precisely the same geographic stream location, because of discharge magnitude and year-over-year stream-gauge operation.
- On bankfull conditions and floodplain engagement: The discharge measurements (summaries) available from USGS (2009b) do not provide consistent and, even when available, only limited details identifying whether the measurement summary is applicable for a partially to full channel or whether the floodplain (if it exists in a classical sense) is engaged by the water surface near the measurement location. Because of generally more favorable conditions for measurement, discharge measurements are often performed, whenever possible, in places with flow conditions lacking substantial floodplain inundation. Also, many stream gauges are located near bridges because of the more favorable conditions for truck-mounted-crane, high-magnitude discharge measurement.

A discussion is needed that concerns components of the well-known Manning's equation for computation of simplified open-channel hydraulics in the context of USGS discharge measurement databases. Manning's equation is

$$Q = [(n\text{-value})^{-1}]A(A/WP)^{2/3}S^{1/2} \tag{2}$$

where the equation provides a useful mathematical structure to statistically evaluate Q and V through intrinsic relations between A, B, wetted perimeter WP, and a friction slope S. However, several limitations excluded application of Manning's equation in a statistical context for this paper:

- Friction slope: Friction slope is indisputably an important parameter, because Q and V are proportional to the square root of slope. However, the friction slope is not available from USGS (2009b). Channel slope often is used in place of friction slope in Manning's equation; channel slope is also not available from USGS (2009b). Therefore, for this paper, a metric of channel slope near each stream gauge for statistical consideration is outside the scope but commented on further in the "Limitations of QGAM and VGAM and Thoughts for Improvement" section.
- Manning's n-value: The Manning n-value also is indisputably an important parameter in Eq. (2). Unfortunately, n-values, which are not direct measures of roughness, or other roughness parameters, such as median grain sizes, influencing channel hydraulics are not readily available for any of the stream gauges in general or for individual discharge measurements across time in particular.
- Wetted perimeter: The wetted perimeter WP, which is used to compute the hydraulic radius (the A/WP term in Manning's equation), likely is useful as a direct predictor variable on Q or V or is useful as a predictor variable when expressed as hydraulic radius. The field-measured data for direct measurements of discharge by the USGS contain horizontal stationing and vertical sounding (depth) information. From these raw data, WP for individual measurements could be estimated. Unfortunately, at the present time (2012), the USGS discharge measurement database (USGS 2009b), being summaries of the field observations, lacks either WP values or the raw data to compute them. Hence, WP values are not available for this study.

The unified discharge measurement database of 87,874 records for 437 stream gauges in Texas was subsequently filtered or reduced to contain discharge measurements that could be reasonably associated with direct-runoff conditions. Specifically, discharge measurements exceeding the 90th-percentile daily mean streamflow as determined by the stream-gauge-specific, flow-duration curves were retained for the analysis reported herein. This 90th-percentile discharge measurement database is the database used for statistical analysis in the "Generalized Additive Models and Regionalization of Discharge and Mean Velocity" section. The 90th-percentile database contains 17,753 discharge records for 424 of the original 437 stream gauges. Each of the 424 stream gauges has at least one measurement greater than the 90th-percentile daily mean streamflow for that stream gauge. The 424 USGS station numbers used for regionalization of Q and V reported in this paper are listed in Table 1.

Summary statistics of A, Q, V, F, and B of the 90th-percentile discharge measurement database are listed in Table 2. After filtering for high-magnitude discharge, considerable variation or range remains in A (about 6 orders of magnitude), Q (about 7 orders of magnitude), V (about 2 orders of magnitude), F (about 2 orders of magnitude), and B (about 5 orders of magnitude). These tabulated statistics of their respective distributions could be used for additional data screening and record rejection prior to regionalization. For example, the maximum $B = 14{,}000$ m is almost certainly too large, the minimum $\mathsf{F} = 0.00610$ is almost certainly too small, and the maximum $\mathsf{F} > 1$ (indicative of supercritical flow conditions) is seemingly high for natural channel flow. Additional data screening and record rejection was not made prior to statistical analysis except for the removal of a few extreme outliers as described in the "Preprocessing and Preliminary Analysis" section.

Generalized Additive Models and Regionalization of Discharge and Mean Velocity

Generalized Additive Models

Complex relations between both Q and V and available predictor variables (described in the "Preprocessing and Preliminary Analysis" section) were anticipated. Therefore, in lieu of conventional multilinear regression modeling (Faraway 2005), generalized additive modeling (GAM) (Hastie and Tibshirani 1990; Wood 2006) was chosen. A GAM is a statistical model between a response variable and an additive combination of various parametric terms and smooth terms (functions). The incorporation of smooth functions can be an advantage to GAMs over simpler multilinear regression, because appropriately configured smooth functions accommodate otherwise difficult to *linearly model* components of a prediction-response model. A Gaussian family for the generalized linear model (Faraway 2006) was used to estimate the GAM models reported here using mostly default arguments of the gam function in the R environment (R Development Core Team 2011) from the *mgcv* package by Wood (2009). The model fitting is based on maximum likelihood (not conventional least squares) for parameter fitting (optimization). The basic form of a GAM model:

$$y_i = \mathbf{X}_i\Theta + f_1(x_{1i}, x_{2i}) + f_2(x_{3i}) + \cdots + \epsilon_i \tag{3}$$

where y_i = a suitably transformed response variable for the ith observation, $\mathbf{X}_i$ = a model matrix for strictly parametric and suitably transformed predictor variables, Θ = a parameter matrix, the f_k are

Table 1. Listing of 424 USGS Stream Gauges That Are Represented in the Generalized Additive Models of Discharge (QGAM) and Mean Velocity (VGAM) Reported in This Paper

07227500	07227920	07228000	07233500	07235000	07295500	07297910	07298500	07299540	07299670
07299890	07300000	07301200	07301300	07301410	07307750	07307800	07308200	07308500	07311600
07311630	07311700	07311783	07311790	07311800	07311900	07312100	07312130	07312200	07312500
07312700	07314500	07314900	07315200	07336820	07342465	07342470	07342480	07342500	07343000
07343200	07343500	07344482	07344486	07344500	07346000	07346045	07346050	07346070	07346140
08017200	08017300	08017410	08018500	08018730	08019000	08019200	08019500	08020000	08020900
08022040	08022070	08026000	08028500	08029500	08030500	08031000	08031200	08032000	08033000
08033300	08033500	08033900	08034500	08036500	08037050	08038000	08039100	08040600	08041000
08041500	08041700	08042800	08043950	08044000	08044500	08044800	08045850	08047000	08047050
08047500	08048000	08048543	08048800	08048970	08049500	08049580	08049700	08050100	08050400
08050800	08050840	08051130	08051500	08052700	08053000	08053500	08055000	08055500	08056500
08057000	08057200	08057445	08058900	08059400	08061000	08061540	08061700	08062000	08062500
08062700	08062800	08062900	08063100	08063500	08063800	08064100	08064700	08064800	08065000
08065200	08065350	08065800	08066100	08066170	08066191	08066200	08066250	08066300	08066500
08067500	08067650	08068000	08068090	08068275	08068390	08068400	08068450	08068500	08068720
08068740	08068780	08068800	08069000	08069500	08070200	08070500	08071000	08071280	08072300
08072730	08072760	08073500	08073600	08073700	08074000	08074020	08074150	08074250	08074500
08074800	08075000	08075400	08075500	08075730	08075770	08075900	08076000	08076180	08076500
08077000	08078000	08079575	08079600	08080500	08080700	08082000	08082500	08082700	08083100
08083230	08083420	08083470	08083480	08084000	08084800	08085500	08086050	08086150	08086212
08086290	08088000	08088300	08088450	08088600	08089000	08090800	08091000	08091500	08091750
08092000	08093100	08093250	08093360	08093500	08094800	08095000	08095200	08095300	08095400
08095600	08096500	08098290	08098300	08099100	08099300	08099500	08100000	08100500	08101000
08102500	08103800	08103900	08104100	08104500	08104700	08104900	08105000	08105095	08105100
08105300	08105700	08106310	08106500	08108200	08108700	08109000	08109700	08109800	08110000
08110100	08110200	08110325	08110430	08110500	08110800	08111000	08111500	08111700	08114000
08115000	08116400	08116650	08117500	08117995	08120500	08120700	08121000	08123800	08123850
08124000	08126380	08127000	08128000	08128400	08129300	08130500	08131400	08133250	08133500
08133900	08134000	08134230	08134250	08136000	08136500	08136700	08138000	08141500	08142000
08143600	08144500	08144600	08145000	08146000	08147000	08148500	08150000	08150700	08150800
08151500	08152000	08152900	08153500	08154700	08155200	08155240	08155300	08155400	08156800
08157000	08157500	08158000	08158050	08158600	08158700	08158800	08158810	08158840	08158920
08158922	08158930	08158970	08159000	08159150	08159200	08159500	08160400	08160800	08161000
08162000	08162500	08162600	08164000	08164300	08164350	08164390	08164450	08164500	08164503

Table 2. Summary Statistics of Selected Hydraulic Parameters from 90th-Percentile Discharge Measurement Database in Texas

Statistic	Cross-sectional area, A (m^2)	Discharge, Q (m^3/s)	Mean velocity, V (m/s)	Froude number, F (–)	Water-surface top width, B (m)
Minimum	0.00372	0.000283	0.0152	0.00610	0.0671
1st quartile	10.9	5.94	0.436	0.125	15.9
Median	47.6	30.6	0.658	0.187	33.8
Mean	167	154	0.750	0.231	78.2
3rd quartile	165	125	0.951	0.296	79.6
Maximum	6,940	7,610	4.22	2.37	14,000

smooth functions of the predictor variables x_{ik}, and ϵ_i are error terms taken as independently and identically distributed $N(0,\sigma^2)$ (Gaussian distribution or normal distribution) random variables. The $\mathbf{X}_i\Theta$ term is the familiar multilinear regression component of a GAM.

For this paper, separate GAM analyses of Q and V were conducted. The GAM model of Q is referred to as QGAM, and, similarly, the GAM model of V is referred to as VGAM. As further described and justified in the "Preprocessing and Preliminary Analysis" section, the basic form of the QGAM reported in the "Generalized Additive Model of Discharge" section is

$$\log(Q) = b_1 + a_1 \log(A) + a_2 \log(B) + a_3\Omega + f_5(\text{longitude, latitude}) + f_6(P) \quad (4)$$

and the basic form of the VGAM reported in the "Generalized Additive Model of Mean Velocity" section is

$$V^{1/5} = b_2 + a_4 \log(Q) + a_5 \log(B) + a_6\Omega + f_9(\text{longitude, latitude}) + f_{10}(P) \quad (5)$$

where log = base-10 logarithm; Q = discharge in m^3/s; V = mean velocity in m/s; b_k = intercepts; a_k = regression coefficients; A = cross-sectional flow area in m^2/s; B = top width in m; Ω = the OmegaEM parameter from Asquith and Roussel (2009), and is described in the "Preprocessing and Preliminary Analysis" section; f_k = smooth functions in one or two dimensions as indicated, and the numerical value of the subscript references the applicable figure of this paper; and P = mean annual precipitation in millimeters (mm), and is described in the "Preprocessing and Preliminary Analysis" section. The QGAM and VGAM are presented in the "Generalized Additive Model of Discharge" and "Generalized Additive Model of Mean Velocity" sections, respectively. Last, the predictive potential of watershed drainage area was found to be unsuitable as a predictor variable for the Q and V regionalization of the 90th-percentile discharge measurement database. Select predictor variables are discussed in the "Preprocessing and Preliminary Analysis" section, along with choice of variable transformation.

Preprocessing and Preliminary Analysis

OmegaEM Parameter

Asquith and Roussel (2009) developed regional equations to estimate annual peak-streamflow frequency for undeveloped watersheds in Texas. As part of that analysis, those authors created a generalized residual of the 10-year (0.10 AEP) discharge equation that is referred to as the OmegaEM parameter. This parameter represents a generalized terrain and climate index that expresses peak-streamflow potential not otherwise represented in the watershed characteristics of drainage area, main-channel slope, and P. The OmegaEM parameter is gridded by 1-degree quadrangles (Asquith and Roussel 2009, p. 14) and is reproduced and shown in Fig. 3. Although developed from analysis of undeveloped watersheds, the parameter captures generalized terrain and climate influences on channel conveyance properties affecting discharge magnitude.

The authors hypothesize that OmegaEM should be a useful, but minor, predictor of Q and V, because OmegaEM expresses regional variation in otherwise difficult to quantify variations in high-magnitude discharge. Using the latitude and longitude of each of the 424 stream gauges, the OmegaEM parameter was computed for each stream gauge by bilinear interpolation from the gridded values in Fig. 3.

Mean Annual Precipitation

Climatological conditions in Texas are diverse. Bomar (1994) provides a review of Texas weather and climate and details historically important rainfall and resulting floods, the characteristics of the atmosphere, and general weather statistics for Texas. For the 424 stream gauges, P ranges from about 292 mm for a stream gauge in the extreme western part of Texas to 1,571 mm for a stream gauge in the extreme southeastern part of Texas.

Using the latitude and longitude of each of the 424 stream gauges, mean annual precipitation P in mm was retrieved for each stream gauge from PRISM Climate Group (2010) for the 1971–2000 normals. The PRISM Climate Group (2010) source was chosen for expediency. Given the many sources of uncertainty both in GAM development and implementation by end-users, the authors consider that any general and authoritative source of P for any suitably long period (perhaps 30 years) is sufficient for GAM development or substitution into the QGAM and VGAM that are reported here. This statement concerning the source of P reiterates the position by Asquith and Roussel (2009, p. 3) in a similar context.

The authors hypothesize that P should be a useful, but minor, predictor of Q and V because P exerts considerable influence on vegetation communities both across the greater watershed as well as for the riparian zone near stream channels where such an identifiable riparian might exist. General erosional and attendant geomorphologic settings as represented by stream-channel shapes are also affected by P. Channel shape in turn influences relations between discharge and mean velocity through the hydraulic characteristics of cross-sectional flow area and top width.

The authors also considered other climate normals available from PRISM Climate Group (2010), including mean July high and mean January low temperatures and their difference. These climate indices seem to be no better predictors or contributors to the explanation of Q or V variance than P.

Variable Transformation

The authors hypothesize for the objective of Q regionalization that the hydraulic parameters of A and B should be critically important parameters. A preliminary issue at hand is the choice of transformation in the GAM analysis. Analysis through multilinear regression, Box-Cox power transformations (Box and Cox 1964) [the boxcox function in R from the *MASS* package by Venables and Ripley (2002)], and preliminary GAM analysis showed that logarithmic transformation on Q, A, and B was appropriate.

The authors also hypothesize for the objective of V regionalization that the hydraulic parameters of Q and B should be critically important parameters. The use of A is not appropriate or even possible in the context here because the reported V values are effectively, if not exactly, the ratio of Q to A. A preliminary issue at hand is the choice of transformation in the GAM analysis. Analysis through multilinear regression, Box-Cox power transformations,

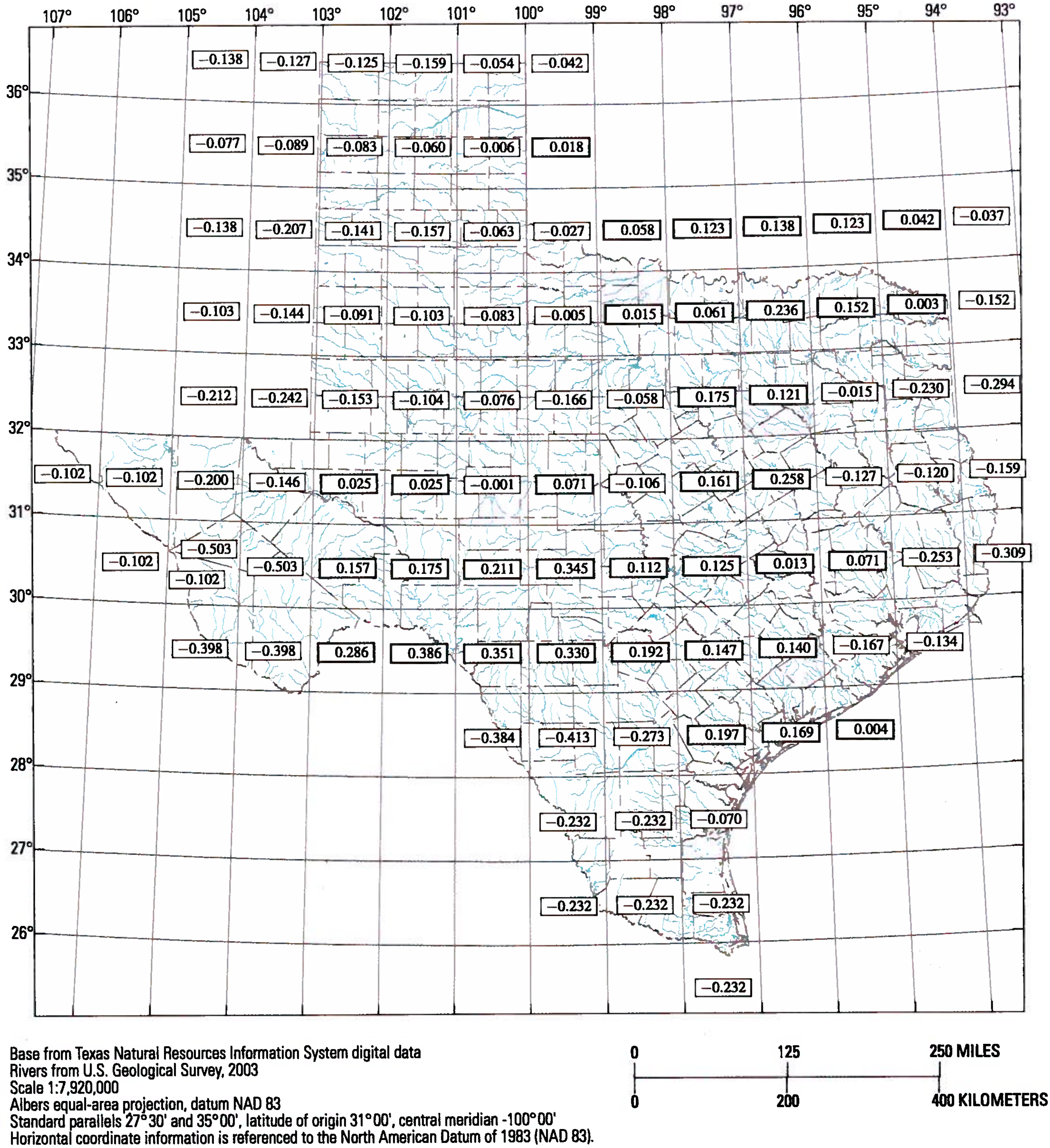

Fig. 3. (Color) OmegaEM parameter of Asquith and Roussel (2009) to be used in generalized additive model (GAM) of discharge (QGAM) shown in Fig. 4 and Eq. (6) and generalized additive model of mean velocity (VGAM) shown in Fig. 8 and Eq. (7); the OmegaEM parameter represents a generalized terrain and climate index expressing relative differences in peak-streamflow potential across Texas [reproduced from Asquith and Roussel (2009)]

and preliminary GAM analysis showed that fifth-root transformation on reported V (or $V^{1/5}$) and logarithmic transformation on Q and B were appropriate.

Preliminary QGAM and VGAM were fit following the algebraic structure of Eqs. (4) and (5) and were used to identify a few extreme outliers. The minimum of the absolute value of the range of the residuals was separately computed for the preliminary QGAM and VGAM. The Q and V records having residuals in absolute value greater than the respective minimums subsequently were removed; summary of these removals (very few) is made in the "Generalized Additive Model of Discharge" and "Generalized Additive Model of Mean Velocity" sections. The effect of outlier

```
DISCHARGE GENERALIZED ADDITIVE MODEL (QGAM), SI UNITS

Select Abbreviations:
log = base-10 logarithm used on Q, A, and B
Q   = discharge in cubic meters per second
A   = cross-section area in square meters
B   = water-surface top width in meters
oem = OmegaEM parameter (Asquith and Roussel, 2009)

Family: gaussian
Link function: identity

Formula:
logQ ~ logA + logB + oem +
       s(LongitudeDegrees, LatitudeDegrees, k = 14) +
       s(MeanAnnualPrecipMillimeters, bs = "cr", k = 5)

Parametric coefficients:
             Estimate Std. Error t-value Pr(>|t|)
(Intercept) -0.289609  0.006156  -47.05   <2e-16
logA         1.269194  0.004927  257.59   <2e-16
logB        -0.224712  0.007641  -29.41   <2e-16
oem          0.286524  0.028057   10.21   <2e-16
---

Approximate significance of smooth terms:
                                    edf Ref.df      F p-value
s(LongitudeDegrees,LatitudeDegrees) 12.87  13.00 187.19  <2e-16
s(MeanAnnualPrecipMillimeters)       4.00   4.00  25.96  <2e-16
---

R-sq.(adj) =  0.949   Deviance explained = 94.9%
GCV score = 0.047158  Scale est. = 0.047103  n = 17727
Residual Standard Error (gaussian family) = 0.217032

RESIDUAL SUMMARY
    Min.  1st Qu.   Median     Mean  3rd Qu.      Max.
-1.04100 -0.12800  0.01848  0.00000  0.14320  1.05000
```

Fig. 4. Summary in R output of generalized additive model of base-10 logarithm of discharge based on statistical relations between the base-10 logarithms of discharge and water-surface top width, OmegaEM parameter by Asquith and Roussel (2009), and separate smooth functions of longitude and latitude $f_5(l,k)$ (Fig. 5) and mean annual precipitation $f_6(P)$ (Fig. 6); GCV is generalized cross-validation

removal was to enhance the centering of the residuals in the final QGAM and VGAM models.

Generalized Additive Model of Discharge

The final QGAM in R output is shown in Fig. 4. For the QGAM, each of the predictor variables is statistically significant. The adjusted R-squared value is about 0.95, and the residual standard error is about $s = 0.22$ base-10 logarithm of m^3/s, which is the square root (Wood 2006, p. 61) of the *Scale est.*, because a Gaussian family was used for this GAM. For the final QGAM model, 26 discharge measurements for 13 stream gauges (USGS station numbers: 07295500, 08018730, 08047500, 08080700, 08110325, 08129300, 08166000, 08185000, 08186500, 08190500, 08197500, 08202700, and 08210400) were removed, but the overall stream-gauge count remained at 424 (see discussion at end of the "Database of Discharge Measurements" section). The QGAM with the coefficients shown in Fig. 4 can be written as

$$\log(Q) = -0.2896 + 1.269\ \log(A) - 0.2247\ \log(B) + 0.2865\Omega + f_5(\text{longitude, latitude}) + f_6(P) \quad (6)$$

where log = base-10 logarithm, Q = discharge in m^3/s, A = cross-sectional flow area in m^2, B = top width in m, Ω = the OmegaEM parameter from Fig. 3, P = mean annual precipitation in mm, and f_5 and f_6 are *smooth functions* of the indicated predictor variables in Figs. 5 and 6, respectively. For Fig. 5, the base map and superimposed smooth lines were created in R using graphic capabilities of packages by Minka (2011) and Wood (2009), and Fig. 6 was created using graphic features by Wood (2009). The red, green, and black lines as ensembles of three for each numerical value shown in Fig. 5 are not all shown for reasons such as grid resolution for the graphic, nonuniform distribution of stream gauges, and general statistical magnitude of the two-dimensional smooth surface.

The $k = 14$ argument (shown in Fig. 4) to the f_5(longitude, latitude) or $f_5(l,k)$ smooth function of location represents the dimension of the isotropic thin plate regression spline (Wood 2006, p. 225). The bs = "cr," $k = 5$ arguments (shown in Fig. 4) to the $f_6(P)$ smooth function represent cubic regression splines (bs = "cr") with the dimension $k = 5$ representing "knots" of the spline (Wood 2006, p. 226). The spline dimensions were chosen through visual evaluation of figures similar to Figs. 5 and 6.

The residuals of the discharge model are shown in Fig. 7, and summary statistics of the residuals are shown in Fig. 4. Because of overplotting, gray transparency was used for Fig. 4 to enhance visual density of the data point distribution. The Akaike Information Criterion is a measure of information content of a regression model. The statistic accounts for a trade off between the number of

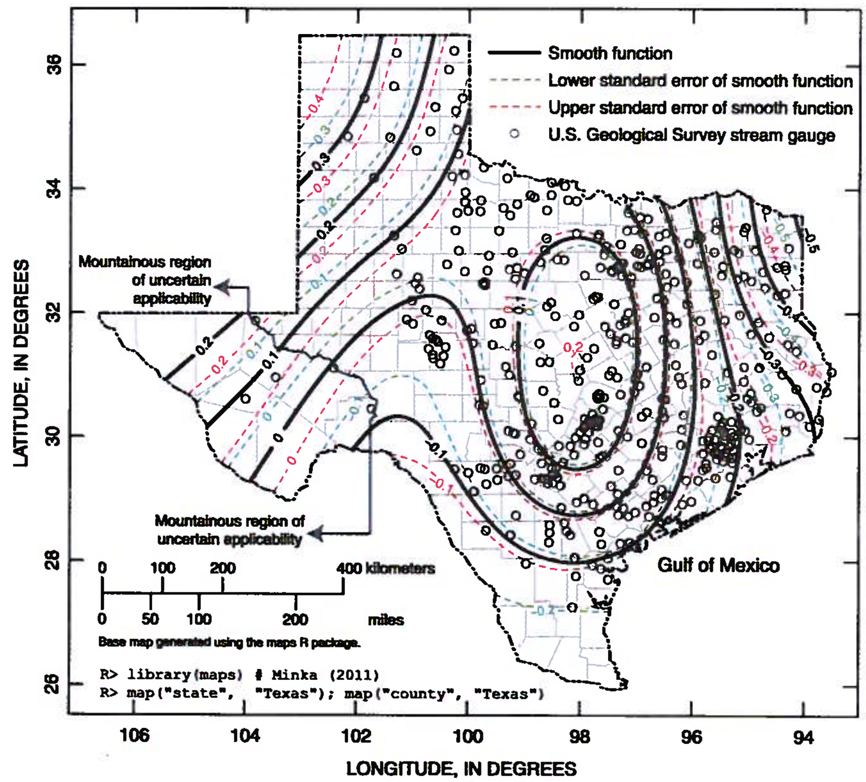

Fig. 5. (Color) Smooth function $f_5(l, k)$ of location in Texas for the discharge model shown in Fig. 4

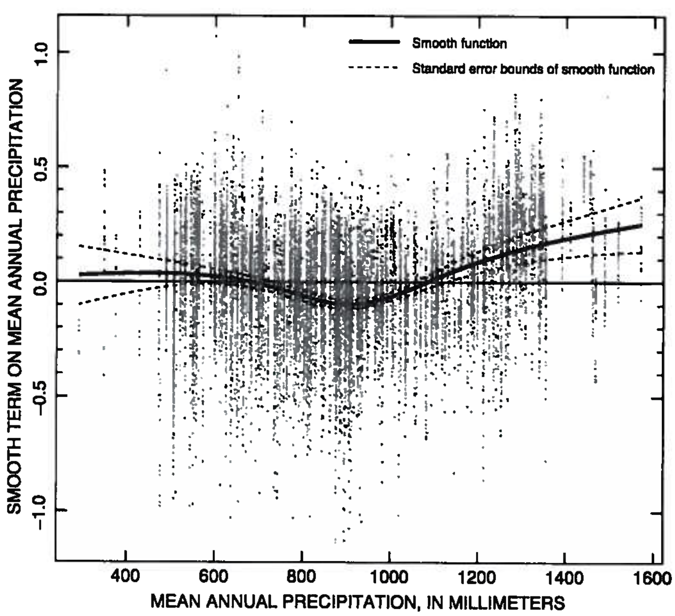

Fig. 6. Smooth function $f_6(P)$ of mean annual precipitation for the discharge model shown in Fig. 4

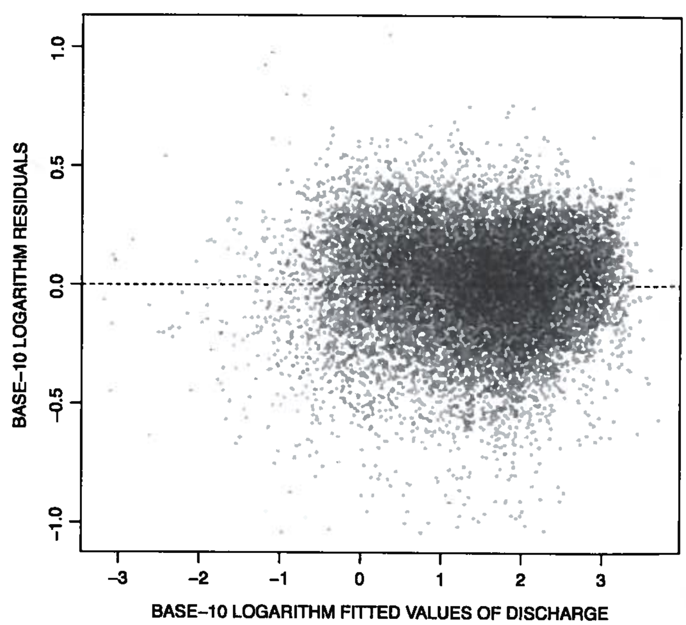

Fig. 7. Residuals for the discharge model shown in Fig. 4

parameters and the fit of the model; small values are sought. The Akaike Information Criterion is −3,830 for the model in Eq. (6) but −281 for the model lacking f_5 and f_6. The percent change in residual standard error from the model lacking f_5 and f_6 to the model in Eq. (6) is −9.6%. A preference for the more complex model involving the smooth functions f_5 and f_6 is made.

Last, loose interpretation of the parametric coefficients can be made that are consistent with well-known hydraulic constraints. The positive coefficient on A shows that Q increases with increasing A; the negative coefficient on B shows that Q decreases with increasing B. The positive coefficient on OmegaEM indicates that Q increases in proportion to OmegaEM. OmegaEM takes on a positive value in the central part of Texas (the region demarked by positive OmegaEM values) and is greatest along the Balcones escarpment in south central Texas. O'Connor and Costa (2003, p. 9) identify this region (Balcones escarpment) of the nation as having "concentrations of large floods." Asquith and

Roussel (2009, p. 23) provide further and relevant discussion. Thus, OmegaEM acts to increase Q in QGAM near the central part of Texas and reduce Q in other parts. The smooth function $f_5(l, k)$ of location also shows a tendency for larger Q in the central part of Texas. The smooth function $f_6(P)$ shows that there is a subtle relation between P and Q that is difficult to interpret given the presence of the two other spatially varying parameters (OmegaEM and f_5).

Generalized Additive Model of Mean Velocity

The final VGAM in R output is shown in Fig. 8. For VGAM, each of the predictor variables is statistically significant. The adjusted R-squared value is about 0.67, and the residual standard error is about $s = 0.063$ fifth root of m/s, which is the square root of the *Scale est.*, because a Gaussian family was used for this GAM. For the final VGAM model reported here, two discharge measurements for two stream gauges (USGS station numbers: 08105000 and 08176500) were removed but the overall stream-gauge count remained at 424 (see discussion at end of the "Database of Discharge Measurements" section). The VGAM with the coefficients shown in Fig. 8 can be written as

$$V^{1/5} = 0.9758 + 0.1588 \log(Q) - 0.1820 \log(B) + 0.0854\Omega + f_9(\text{longitude}, \text{latitude}) + f_{10}(P) \quad (7)$$

where log = base-10 logarithm, V = mean velocity in m/s transformed by the fifth root, Q = discharge in m^3/s, B = top width in m, Ω = the OmegaEM parameter from Fig. 3, P = mean annual precipitation in mm, and f_9 and f_{10} are *smooth functions* of the indicated predictor variables in Figs. 9 and 10, respectively. For Fig. 9, the base map and superimposed smooth lines were created in R using graphic capabilities of packages by Minka (2011) and Wood (2009), and Fig. 10 was created using graphic features by Wood (2009). The red, green, and black lines as ensembles of three for each numerical value shown in Fig. 9 are not all shown for reasons such as grid resolution for the graphic, nonuniform distribution of stream gauges, and general statistical magnitude of the two-dimensional smooth surface.

The $k = 14$ argument (shown in Fig. 8) to the f_9(longitude, latitude) or $f_9(l, k)$ smooth function of location represents the dimension of the isotropic thin plate regression spline (Wood 2006, p. 225). The bs = "cr," $k = 5$ arguments (shown in Fig. 8) to the $f_{10}(P)$ smooth function represent cubic regression splines

```
VELOCITY GENERALIZED ADDITIVE MODEL (VGAM), SI UNITS

Select Abbreviations:
tV  = fifth-root of mean velocity in meters per second
log = base-10 logarithm used on Q and B
Q   = discharge in cubic meters per second
B   = water-surface top width in meters
oem = OmegaEM parameter (Asquith and Roussel, 2009)

Family: gaussian
Link function: identity

Formula:
tV ~ logQ + logB + oem +
     s(LongitudeDegrees, LatitudeDegrees, k = 14) +
     s(MeanAnnualPrecipMillimeters, bs = "cr", k = 5)

Parametric coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  0.9758281  0.0018882  516.80   <2e-16
logQ         0.1588495  0.0009992  158.98   <2e-16
logB        -0.1819640  0.0018281  -99.54   <2e-16
oem          0.0853768  0.0081059   10.53   <2e-16
---

Approximate significance of smooth terms:
                                    edf Ref.df      F  p-value
s(LongitudeDegrees,LatitudeDegrees) 12.72  12.99 203.25  < 2e-16
s(MeanAnnualPrecipMillimeters)       4.00   4.00  11.33 3.52e-09
---

R-sq.(adj) =  0.671   Deviance explained = 67.1%
GCV score = 0.0039773  Scale est. = 0.0039727  n = 17751
Residual Standard Error (gaussian family) = 0.063029

RESIDUAL SUMMARY
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.3762000 -0.0389000 -0.0007972  0.0000000  0.0406300  0.4032000
```

Fig. 8. Summary in R output of generalized additive model of fifth root of mean velocity based on statistical relations between the base-10 logarithms of discharge and water-surface top width, OmegaEM parameter by Asquith and Roussel (2009), and separate smooth functions of longitude and latitude $f_9(l, k)$ (Fig. 9) and mean annual precipitation $f_{10}(P)$ (Fig. 10); GCV is generalized cross-validation

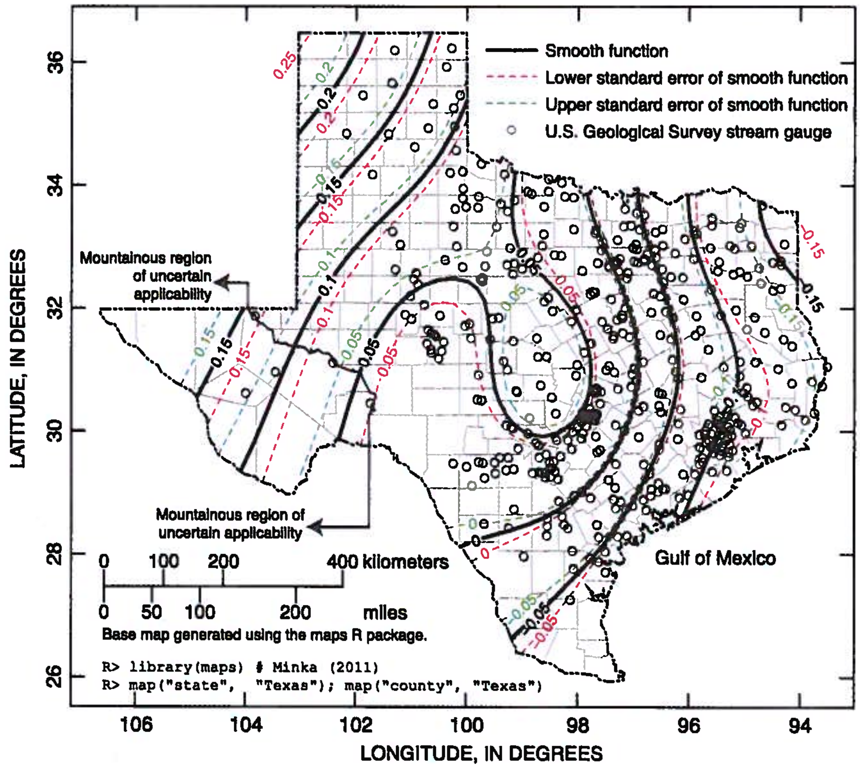

Fig. 9. (Color) Smooth function $f_9(l, k)$ of location in Texas for the mean velocity model shown in Fig. 8

(bs = "cr") with the dimension $k = 5$ representing "knots" of the spline (Wood 2006, p. 226). The spline dimensions were chosen through visual evaluation of figures similar to Figs. 9 and 10.

The residuals of the mean velocity model are shown in Fig. 11, and summary statistics of the residuals are shown in Fig. 8. Because of overplotting, gray transparency was used for Fig. 8 to enhance visual density of the data point distribution. The Akaike Information Criterion is −47,700 for the model in Eq. (7) but −42,100 for the model lacking f_9 and f_{10}. The percent change in residual standard error from the model lacking f_9 and f_{10} to the model in Eq. (7) is −14%. A preference for the more complex model involving the smooth functions f_9 and f_{10} is made.

Again, loose interpretation of the parametric coefficients can be made that is consistent with well-known hydraulic constraints. The positive coefficient on Q shows that V increases with increasing Q; the negative coefficient on B shows that V decreases with increasing B. The positive coefficient on OmegaEM indicates that V increases in proportion to OmegaEM. This finding was anticipated (see discussion in the "Generalized Additive Model of Discharge"

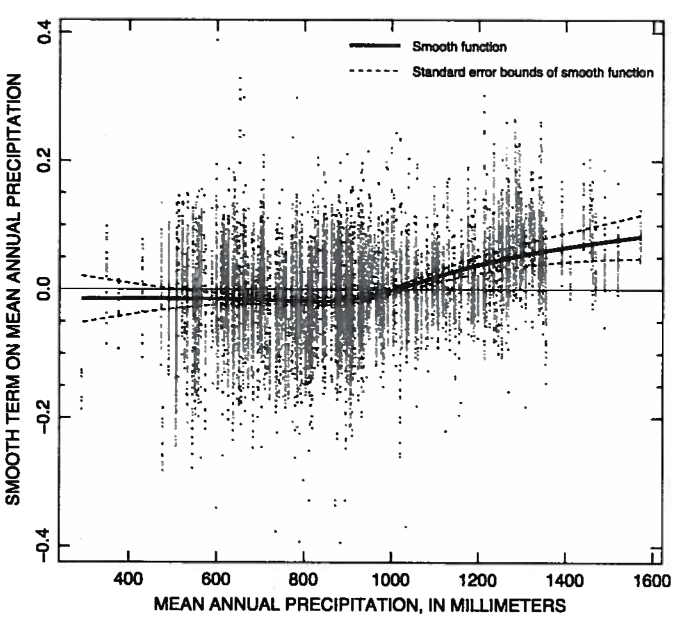

Fig. 10. Smooth function $f_{10}(P)$ of mean annual precipitation for the mean velocity model shown in Fig. 8

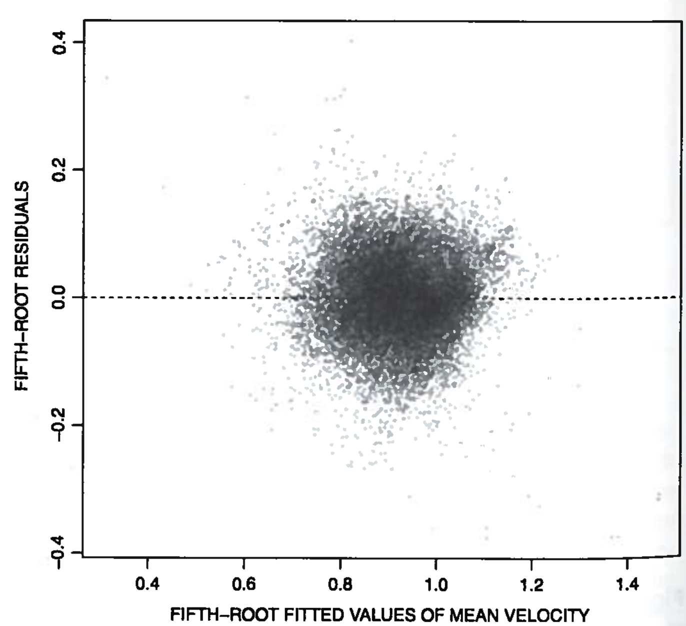

Fig. 11. Residuals for the mean velocity model shown in Fig. 8

section). The smooth function $f_9(l, k)$ of location also shows a tendency for smaller V in the eastern part of Texas. The authors hypothesize that this observation is consistent with greater vegetation density in the riparian zones in the eastern parts of Texas than in the western parts, and vegetation is associated with larger P and other physiographic factors. The smooth function $f_{10}(P)$ shows that there is a subtle relation between P and V that is difficult to interpret given the presence of two other spatially varying parameters (OmegaEM and f_9).

Limitations of QGAM and VGAM and Thoughts for Improvement

According to the National Research Council (2004, p. 123) "a limitation of [the discharge measurement database] is that [stream gauges] are chosen to have particular channel characteristics, such as the existence of a control section that will ensure a unique rating curve." The National Research Council (2004, p. 123) continues, "the channel characteristics of [stream gauge] locations may thus not be representative of randomly selected locations at any point along the entire length of a stream or river." This last statement is particularly relevant for regional analysis of discharge measurement databases in that many high-magnitude discharge measurements are made at bridge crossings; the primary end-user application for VGAM is foreseen to be at or near bridge crossings in Texas. The general applicability or unapplicability of QGAM and VGAM for other cross sections of streams in Texas is difficult to quantitatively assess.

Assuming that the QGAM and VGAM do have acceptable applicability for other cross sections in Texas, additional discussion of applicability in terms of location is needed. The far western part of Texas is a mountainous region (Figs. 5 and 9) with few USGS stream gauges. The applicability of QGAM and VGAM is uncertain, but the models might retain some but difficult to quantify applicability in far western Texas. The number of stream gauges diminishes rapidly towards the southernmost part of Texas; however, because of the low-relief terrain, similarity in soils and vegetation, and orientation of the region with respect to the Gulf of Mexico, the authors suggest that QGAM and VGAM remain applicable. Last, the far north-northwestern parts of Texas also have few stream gauges. By consideration of the physiographic features and the preponderance of branded sand channels in that general region, the authors suggest that QGAM and VGAM might retain some but difficult to quantify applicability.

As discussed in "Database of Discharge Measurements" section, the structure of Manning's equation, and thus the potential influence of S on computation of Q or V, is important, but such *proximal-to-stream-gauge* S data are lacking for this paper. The eventual inclusion of a S (friction, channel, or other slope) as a predictor variable in QGAM and, seemingly more importantly, in VGAM (because of the smaller adjusted R-squared) should further enhance the regionalization of the discharge measurement database used here. Other potentially useful characteristics of the stream network or the channel near the stream gauge include stream order (Strahler 1957; Shreve 1966), drainage density, and sinuosity. The authors hypothesize that the inclusion of additional channel-specific characteristics that are near the stream gauge could serve as measurably important predictor variables for alternative QGAM and VGAM. Presumably, model diagnostics will improve as near-the-stream-gauge characteristics are included in the regionalization.

General enhancement to the GAM diagnostics should be attainable through deliberate and systemic review of the summary statistics of A, Q, V, F, and B (recall such statistics for the entire database listed in Table 2). It might be possible for analysts to select particular variable thresholds. For example, all discharge measurements with $0.1 \leq \mathsf{F} < 1$ or $1 \leq B \leq 2{,}000$ m could be retained and the regional analysis proceed from there.

A suggested approach beyond conventional residual or standardized residual plots would be an evaluation of the inherently coupled relations between Q and V on a per-stream-gauge basis. For example, it is known that the Q and V for most stream gauges show positive association (Q increasing with V and vice versa); however, a not insubstantial number of stream gauges do show negative association between Q and V. Could the generalized association (positive or negative) of Q and V for a given stream gauge be used for further statistical enhancement?

Example Applications

Postevent Discharge Estimation

Two example applications of the QGAM and VGAM are presented in this section. Suppose that a direct-runoff event occurred, and an analyst is interested in estimating the Q_p for a particular stream located at about 31.5°N and −98.5°W. A post-direct-runoff event survey measures the top width of the peak water surface at about 100 m, and the average depth is estimated as 4.5 m. The estimated cross sectional area is thus 450 m^2. The P for the location is about 744 mm (PRISM Climate Group 2010), and the OmegaEM parameter in Fig. 3 for the location is about −0.106.

The smooth function $f_5(l, k)$ of the location for QGAM is judged to be about 0.15 from Fig. 5 using interpretation and interpolation of the smooth function lines (black lines) and the lower and upper standard error lines (green and red lines, respectively) as available. The smooth function $f_6(P)$ of P for QGAM is about −0.02 from Fig. 6. The Q_p can now be readily computed by variable substitution in Eq. (6):

$$\log(Q_p) = -0.2896 + 1.269\,\log(450) - 0.2247\,\log(100) + 0.2865(-0.106) + 0.15 - 0.02 \quad (8)$$

$$\log(Q_p) = 2.728 \quad (9)$$

$$Q_p = 535\ \mathrm{m^3/s} \quad (10)$$

For this estimate of Q_p, the $\bar{V}$ is

$$\bar{V} = \frac{535[\mathrm{m^3/s}]}{450[\mathrm{m^2}]} = 1.19\ \mathrm{m/s} \quad (11)$$

The VGAM provides an alternative estimate of V for a Q of 535 m^3/s. The smooth function $f_9(l, k)$ of the location for VGAM is judged to be about 0.06 from Fig. 9 using interpretation and interpolation of the smooth function lines (black lines) and the lower and upper standard error lines (green and red lines, respectively) as available. The smooth function $f_{10}(P)$ of P for VGAM is about −0.02 from Fig. 10. The V can be readily computed by variable substitution in Eq. (7):

$$V^{1/5} = 0.9758 + 0.1588\,\log(535) - 0.1820\,\log(100) + 0.0854(-0.106) + 0.06 - 0.02 \quad (12)$$

$$V^{1/5} = 1.076 \quad (13)$$

$$V = 1.44\ \mathrm{m/s} \quad (14)$$

Last, the authors observe that the two estimates of V (1.19 m/s versus 1.44 m/s) are seemingly consistent with each other. Consistency between either a computed (from known or design discharge and known cross-sectional area) or modeled $\bar{V}$ and V predicted by VGAM is the subject of the "Review of Mean Velocity from a Hydraulic Model" section.

Review of Mean Velocity from a Hydraulic Model

The previous example application guides a user in computing Q given cross-sectional properties and other characteristics. The focus of the computations was on QGAM. For another example application, the focus is on VGAM. Suppose for the same location that an analyst has a design discharge Q_T of 800 m^3/s for a 0.02 AEP or recurrence interval of $T = 50$ years, and a hydraulic model predicts a B of 100 m and an A of 450 m^2 as used in the previous example for simplicity. The hydraulic model is thus predicting a computed $\bar{V}$ of 1.78 m/s. The VGAM can be used to independently evaluate the $\bar{V}$ from the hydraulic model. The V estimate from VGAM is 1.44 m/s as computed in the previous example.

Wood (2009) provides the predict.gam function (Wood 2006, p. 243), which is designed for use in R. This function computes standard errors of a prediction for a GAM using a Bayesian posterior covariance matrix. However, without a digital presentation of the GAM object from R as well as R running on a host computer, the computations of standard error are tedious and error prone for desktop application by anticipated end-users. A convenient means for end-user implementation to only approximate the distribution of a prediction from VGAM (or QGAM by association) is thus needed.

The prediction percentile for a multilinear regression (Helsel and Hirsch 2002, pp. 295–322) can be computed by

$$y(\Pi/100) = y_o + s \times t_{[(\Pi/100),n-p]}\sqrt{1 + h_o} \qquad (15)$$

where $y(\Pi/100)$ = the predicted response for the Π percentile, y_o = a prediction from the regression model, n = the sample size, p = the number of parameters, $t_{[\Pi/100,n-p]}$ = the quantile distribution function (qdf) of the t-distribution, s = the residual standard error, and h_o = the leverage of the prediction. The sample size for VGAM is large ($n = 17{,}751$, Fig. 8), and the parameter count is small ($p = 7$, Fig. 8); as a result, the qdf of the standard normal distribution $\Phi(F)$ for nonexeedance probability F can be substituted for the t-distribution.

Although the specific leverage or its equivalence of a GAM for ungauged locations is extremely difficult to represent or approximate, the average leverage of a conventional multilinear regression model is p/n. The average leverage for VGAM is effectively zero because the ratio $7/17{,}751$ is small. Therefore, h_o is approximately zero because of the enormous degrees of freedom, and thus $\sqrt{1 + h_o} \approx 1$. The residual standard error is $s = 0.0630$ (Fig. 8). The prediction percentile of the 1.78 m/s velocity can thus be loosely approximated, recalling use of the fifth-root transformation and Eq. (15), by

$$y(\Pi/100) \approx 1.78^{1/5} \approx 1.44^{1/5} + 0.0630\Phi(\Pi/100) \qquad (16)$$

$$\Phi(\Pi/100) \approx 0.739 \qquad (17)$$

$$\Pi/100 \approx \phi(0.739) \approx 0.77 \qquad (18)$$

where $\phi(x)$ = the cumulative distribution function of the standard normal distribution for value x. The results show that $\bar{V}$ of the hydraulic model is at the 77th percentile. The project reviewer would naturally conclude that the $\bar{V}$ of the hydraulic model is consistent with VGAM.

To further demonstrate VGAM application, suppose that an analyst wants to apply for the same location a design Q_T of 2,100 m^3/s. Suppose also that the analyst has run or is reviewing a hypothetical hydraulic model predicting B of 100 m and A of 450 m^2 (as used in previous examples for simplicity). The hydraulic model is thus predicting a computed $\bar{V}$ of 4.67 m/s. The prediction percentile for 4.67 m/s can be estimated, recalling use of the fifth-root transformation, by

$$y(\Pi/100) \approx 4.67^{1/5} \approx 1.44^{1/5} + 0.0630\Phi(\Pi/100) \qquad (19)$$

$$\Phi(\Pi/100) \approx 4.53 \qquad (20)$$

$$\Pi/100 \approx \phi(4.53) > 0.999 \qquad (21)$$

The results show that the hydraulically modeled V is in excess of the 99.9th percentile of VGAM. The analyst running or reviewing the hydraulic model would naturally conclude that the $\bar{V}$ is inconsistent with VGAM and by extension is inconsistent with more than 17,700 measurements of high-magnitude discharge in Texas. The apparent absence of congruence between the two V values could be a sign that enhancements to the reliability of the hydraulic model through changes in model assumptions, parameter values, or select cross-sectional representations might be possible.

The previous computations considered a large hydraulically modeled $\bar{V}$. The problem could also be in the opposite direction. Suppose for the same location that the design Q_T is 210 m^3/s, and again a hydraulic model is predicting a B of 100 m and an A of 450 m^2. The hydraulic model is thus predicting a computed $\bar{V}$ of 0.467 m/s. The prediction percentile for 0.467 m/s can be estimated, recalling use of the fifth-root transformation, by

$$y(\Pi/100) \approx 0.467^{1/5} \approx 1.44^{1/5} + 0.0630\Phi(\Pi/100) \qquad (22)$$

$$\Phi(\Pi/100) \approx -3.44 \qquad (23)$$

$$\Pi/100 \approx \phi(-3.44) < 0.0003 \qquad (24)$$

The results show that the hydraulically modeled $\bar{V}$ is less than the 0.03th percentile. Again, the analyst running or reviewing the hydraulic model would naturally conclude that the $\bar{V}$ is inconsistent with VGAM and by extension is inconsistent with more than 17,700 measurements of high-magnitude discharge in Texas. The apparent absence of congruence between the two V values could be a sign that enhancements to the reliability of the hydraulic model through changes in model assumptions, parameter values, or select cross-sectional representations might be possible.

The procedures shown to compute the distribution of a prediction from VGAM in this section are also applicable by association to the distribution of a prediction from QGAM, although example computations are not shown in this paper.

Discussion

A 90th-percentile or high-magnitude discharge measurement database containing more than 17,700 discharge values and ancillary hydraulic properties was assembled from summaries of discharge measurement records for 424 USGS streamflow-gauging stations (stream gauges) in Texas. These discharge measurements are therefore assumed to represent discharge measurements made during direct-runoff conditions at each stream gauge.

Systematic and statewide investigation of these high-magnitude discharges in pursuit of regional models for the estimation of discharge and mean velocity has not been previously attempted. Generalized additive regression modeling is used to develop readily implemented procedures by end-users for estimation of discharge and mean velocity from select predictor variables at ungauged stream locations. Example applications and computations using both regression models are provided.

The application of generalized additive model regression techniques created apparently useful almost statewide-applicable models of discharge and mean velocity. The diagnostics of the generalized additive models of discharge (QGAM) and mean velocity (VGAM) presented, including adjusted R-squared, residual standard error, the wide ranges in predictor variable values, the large number of stream gauges, and the imposing number of discharge measurements, indicate that reliable estimation of Q and V can be made from the parametric and smooth function components of QGAM and VGAM, respectively. The two smooth functions within QGAM and VGAM show a particular advantage of regionalization using GAM algorithms. Specifically, the smooth function variable fitting to otherwise difficult to incorporate predictor variables measurably enhances the regression model without the explicit need to find optimal transformations for each term with respect to the response variable. The application of generalized additive model regression techniques created apparently useful near-statewide-applicable models of discharge and mean velocity.

The application of GAM for the regionalization of USGS discharge measurement database(s) could be enhanced by inclusion of potentially useful channel, soil, or vegetation properties near stream gauges. Such properties could include proximal channel slope, cohesion classification of bed and bank soils, or channel vegetation classification or density measures. The imposing size of the Texas database suggests that statistical associations with these and other potential predictor variables could be found, and statistical enhancements could be made for alternative QGAM and VGAM analyses.

This study focused on measurements of discharge related to direct-runoff conditions, which is determined by those Q values exceeding the 90th-percentile daily mean streamflow. It currently is unknown what changes or influence (sensitivity) in the basic QGAM or VGAM would manifest with alternative probability thresholding. It might be possible to include a factor variable of low, base, and high flow conditions as a predictor variable in the model-building process to include all discharge measurements (more than 89,900 in Texas) and to create more hydrologic-spectrum encompassing GAMs of Q or V than reported in this paper. Alternatively, a low-flow or drought regionalization of Q and V from perhaps a 10th-percentile discharge measurement database could be more applicable for instream-flow assessments than the 90th-percentile discharge measurement (direct-runoff) database and the reported QGAM and VGAM.

The authors purposely constructed QGAM and VGAM to use B instead of hydraulic depth $D = A/B$. The authors selected B for the VGAM because the response variable reported V was nominally computed as Q/A and, hence, use of either A or the ratio A/B as predictor variables in VGAM leads to conceptual and numerical problems. The B was therefore retained in QGAM for some algebraic consistency with VGAM.

Following the availability of reliable QGAM and VGAM models, some other ideas have come to the authors' attention. The authors suggest that QGAM or other similar statistical models when coupled with a stage (gauge height, h) table of cross-sectional flow area $A(h)$ and a stage table of water-surface top width $B(h)$ could contribute to streamflow monitoring in which peak-stage records or stage-hydrograph recorders are used to support "an alternative data collection paradigm of collecting slightly less accurate [streamflow] information at more geographic sites" (National Research Council 1999, p. 27). Further, QGAM or other statistical models have a natural application for "construction of stream rating curves" for which the National Research Council (1999, p. 28) deems an area where technique improvement is needed.

Further development and refinement of statistical approaches (GAM or otherwise) for regionalization of the extensive and nationwide discharge measurement databases of the USGS could also produce viable and alternative regional models of Q and V measurements. Such models then could support "short-term [monitoring] of flows at street and highway crossings to generate design [discharge] data [which] might be done more appropriately by federal, state, or local highway administrations [than the USGS]" (National Research Council 2004, p. 90). Such monitoring interlocks with the alternative data collection paradigm in the previous paragraph. Last, the incorporation of B in the GAMs might make these models more compatible with sophisticated computer imaging and processing systems used for visually monitoring channel and streamflow conditions. Such systems could provide for objective detection of water-surface extent B rather than water-surface elevation h from image sequences (video) and A estimated in turn from $A(B)$ rating tables (A as a function of B).

As regional models of Q and V become more sophisticated and refined, other applications might be identified. For instance, the Q model could form the basis for assessment of the probability of roadway inundation during high-magnitude discharges at low-chord (low-roadway) elevation stream crossings in rural areas with low traffic volumes that may not warrant efforts towards rigorous hydraulic analysis. Suppose an analyst has estimates of the flood-frequency curves (discharge as a function of AEP) for these stream crossings, such as provided by the equations in Asquith and Roussel (2009) and the $A(h)$ and $B(h)$ tables. A value for h defined by the lowest low-chord elevation of the stream crossing could provide estimates of A and B. These estimates could be used to compute Q'—the discharge for which overtopping of the stream crossing commences—from a model like QGAM. The analyst could then estimate the AEP value of Q' from the flood-frequency curve; if this estimated AEP of overtopping is found to be too small according to some institutional guidance or regulation, then the hydrologic hazard of the stream crossing could be deemed substantial and more rigorous hydraulic analysis conducted.

Acknowledgments

The inspiration for this paper originated from discussions between G. H. Herrmann and W. H. Asquith in December 2008. The nexus was Herrmann's question "What can the USGS discharge measurement database in Texas tell the analyst running or reviewing a hydraulic model about the expected value or distribution of mean velocity for design discharges that emanate from hydraulic models within natural channels or near bridge openings?" Exploratory analyses (circa 2009–2010) by Asquith and Herrmann were summarized in an unpublished white paper, which discussed the potential for regionalization of discharge measurement databases. The authors thank D. B. Thompson (R. O. Anderson) for constructive comments in an early draft of this paper. The authors thank M. C. Roussel (USGS) and R. M. Slade, Jr. (retired USGS Surface-Water Specialist in Texas) for providing thoughtful comments on early drafts of this paper. Last, the authors are grateful for the instructive anonymous peer reviews of this paper.

Notation

The following symbols are used in this paper:

A = cross-sectional flow area in m^2;
$A(B)$ = cross-sectional flow area rating table based on B;
$A(h)$ = cross-sectional flow area rating table based on h;
a_k = regression coefficients for QGAM and VGAM;
B = water-surface top width or just top width in m;
$B(h)$ = water-surface top width rating table based on h;
b_k = regression intercepts for QGAM and VGAM;
c_k = regression coefficients in context of Eq. (1) and Riggs (1976);
d_k = regression coefficients in context of Castro and Jackson (2001);
edf = estimated degrees of freedom (see Figs. 4 and 8);
est. = estimate (see Figs. 4 and 8);
F = Froude number in dimensionless units, which was computed for the database, but not used in this paper;
F = nonexceedance probability;
f_k = general notation for smooth functions of a GAM;
$f_5(l,k)$ = smooth two-dimensional function of the location using decimal longitude and latitude for l and k, respectively, of a stream channel for inclusion in the QGAM shown in Fig. 5;
$f_6(P)$ = smooth one-dimensional function of mean annual precipitation P for inclusion in QGAM shown in Fig. 6;
$f_9(l,k)$ = smooth two-dimensional function of the location using decimal longitude and latitude for l and k, respectively, of a stream channel for inclusion in the VGAM shown in Fig. 9;
$f_{10}(P)$ = smooth one-dimensional function of mean annual precipitation P for inclusion in VGAM shown in Fig. 10;
g = acceleration of gravity in m/s^2;
h = stage or gauge height to be used in rating tables of cross-sectional area $A(h)$ and top width $B(h)$ for proposed application of QGAM;
h_o = so-called leverage for the predictor variables from a multilinear regression model;
k = shorthand notation for latitude;
l = shorthand notation for longitude;
log = base-10 logarithm;
$N(0,\sigma^2)$ = normal distribution with a variance of σ^2;
n = sample size or unique number of measurements represented in QGAM or VGAM;
P = mean annual precipitation for a suitably long climate-averaging period (1971–2000 was used herein);
p = number of unique parameters represented in QGAM or VGAM;
$\Pr(>|t|)$ = probability inequality of absolute value of t-value (see Figs. 4 and 8);
Q = discharge in m^3/s. This symbol, as context dictates, can refer to discharge values in the discharge measurement database, a discharge predicted by QGAM, or just simply discharge as needed;
Q' = discharge at which overtopping of a stream crossing commences;
Q_p = peak discharge in m^3/s in general or that estimated from QGAM based on peak water-surface top width and associated cross-sectional flow area;
Q_T = peak distribution associated with annual exceedance probability (AEP) expressed as the T-year event;
qdf = quantile distribution function;
Ref.df = reference degrees of freedom (see Figs. 4 and 8);
R.sq.(adj) = adjusted R-squared (see Figs. 4 and 8);
S = slope in dimensionless units. This term is context-dependent and can refer to friction slope in open-channel hydraulics, water-surface slope in the downstream direction, or channel slope proximal to the stream location or stream gauge of interest;
s = residual standard error of either QGAM or VGAM as context dictates;
s(vars) = notation for smooth function of vars variables (see Figs. 4 and 8);
T = annual recurrence interval equivalence for an annual exceedance probability;
$t_{[F,df]}$ = quantile distribution function of the t-distribution for nonexceedance probability F and $df = n - p$ degrees of freedom;
V = mean velocity in m/s. This symbol, as context dictates, can refer to mean velocity values in the discharge measurement database, mean velocity predicted by VGAM, or just simply mean velocity, as needed;
$\bar{V}$ = mean velocity in m/s computed from simple division between discharge and cross-sectional flow area;
WP = wetted perimeter in an open channel (identified but not otherwise used in this paper because of lack of availability) and also a percentile in the context of the distribution of a prediction from QGAM or VGAM;
$\mathbf{X}_i$ = model matrix for strictly parametric and suitably transformed predictor variables of a GAM;
x_{ik} = predictor variables of a GAM;
y_i = suitably transformed response variable for the ith observation of a GAM;
y_o = predicted value from a multilinear regression model;
$y(F)$ = quantile function of the predicted value y_o for nonexceedance probability F, which equals percentile divided by 100 or $\Pi/100$;
ϵ_i = error terms of a GAM;
Θ = parameter matrix of a GAM;
Π = percentile ($0 \le \Pi \le 100$);
$\Phi(F)$ = quantile distribution function of the standard normal distribution for nonexeedance probability F;
$\phi(x)$ = cumulative distribution function of the standard normal distribution for value x;
Ω = symbol for the OmegaEM parameter of Asquith and Roussel (2009). In general conversation and text, the term OmegaEM is preferred by those authors;
1st Qu. = first (lower) quartile (see Figs. 4 and 8); and
3rd Qu. = third (upper) quartile (see Figs. 4 and 8).

References

ASCE. (1996). *River hydraulics: Technical engineering and design guides*, ASCE Press, New York.

Asquith, W. H. (1998). "Peak-flow frequency and extreme flood potential for tributaries of the Colorado River downstream of Austin, Texas." *Water-Resour. Invest. Rep. 98-4015*, USGS, Reston, VA.

Asquith, W. H. (2001). "Effects of regulation on L-moments of annual peak streamflow in Texas." *Water-Resour. Invest. Rep. 01-4243*, USGS, Reston, VA.

Asquith, W. H., and Roussel, M. C. (2009). "Regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using an L-moment-based, PRESS-minimized, residual-adjusted approach." *Sci. Invest. Rep. 2009-5087*, USGS, Reston, VA.

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." *Sci. Invest. Rep. 2006-5286*, USGS, Reston, VA.

Asquith, W. H., and Slade, R. M. (1997). "Regional equations for estimation of peak-streamflow frequency for natural basins in Texas." *Water-Resour. Invest. Rep. 96-4307*, USGS, Reston, VA.

Asquith, W. H., Slade, R. M., and Lanning-Rush, J. (1996). "Peak-flow frequency and extreme flood potential for streams in the vicinity of the Highland Lakes, central Texas." *Water-Resour. Invest. Rep. 96-4072*, USGS, Reston, VA.

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." *Sci. Invest. Rep. 2008-5084*, USGS, Reston, VA.

Barnes, H. H. (1967). "Roughness characteristics of natural channels." *Water Supply Pap. 1849*, USGS, Reston, VA.

Bomar, G. W. (1994). *Texas weather*, University of Texas Press, Austin, TX.

Box, G. E. P., and Cox, D. R. (1964). "An analysis of transformations." *J. R. Stat. Soc. Ser. B*, 26(2), 211–252.

Castro, J. M., and Jackson, P. L. (2001). "Bankfull discharge recurrence intervals and regional hydraulic geometry relationships: Patterns in the Pacific Northwest, USA." *J. Am. Water Resour. Assoc.*, 37(5), 1249–1262.

Clark, G. M., Mueller, D. K., and Mast, M. A. (2000). "Nutrient concentrations and yields in undeveloped stream basins of the United States." *J. Am. Water Resour. Assoc.*, 36(4), 849–860.

Coffman, D. K., Malstaff, G., and Heitmuller, F. T. (2011). "Characterization of geomorphic units in the alluvial valleys and channels of Gulf Coastal Plain Rivers in Texas, with examples from the Brazos, Sabine, and Trinity rivers." *Sci. Invest. Rep. 2011-5067*, USGS, Reston, VA.

Dalrymple, T., and Benson, M. A. (1967). "Measurement of peak discharge by the slope-area method." *Techniques of water resources investigations*, Book 3, Chapter A2, USGS, Reston, VA.

Devulapalli, R. S., and Valdes, J. B. (1996). *Volume-duration-frequencies for ungaged catchments in Texas, Vol. 1, Calculation of regional regression equations, Tech. Rep. 173*, Texas Water Resources Institute, Texas A&M Univ., College Station, TX.

Faraway, J. J. (2005). *Linear models with R*, Chapman & Hall/CRC, Boca Raton, FL.

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

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized additive models*, Chapman & Hall/CRC, Boca Raton, FL.

Heitmuller, F. T., and Asquith, W. H. (2008). "Potential for bed-material entrainment in selected streams of the Edwards Plateau, Edwards, Kimble, and Real Counties, Texas, and vicinity." *Sci. Invest. Rep. 2008-5017*, USGS, Reston, VA.

Heitmuller, F. T., and Greene, L. E. (2009). "Historical channel adjustment and estimates of selected hydraulic values in the lower Sabine River and lower Brazos River Basins, Texas and Louisiana." *Sci. Invest. Rep. 2009-5174*, USGS, Reston, VA.

Helsel, D. R., and Hirsch, R. M. (2002). "Statistical methods in water resources." *Techniques of water-resources investigations*, Book 4, Chapter A3, USGS, Reston, VA.

Jain, S. C. (2001). *Open-channel flow*, John Wiley, New York.

Lanning-Rush, J. (2000). "Regional equations for estimating mean annual and mean season runoff for natural basins in Texas." *Water-Resour. Invest. Rep. 00-4064*, USGS, Reston, VA.

Leopold, L. B., and Maddock, T. (1953). "The hydraulic geometry of stream channels and some physiographic implications." *Prof. Pap. 242*, USGS, Reston, VA.

Minka, T. P. (2011). "Maps: Draw geographic maps." *R package version 2.1-6* ⟨http://cran.r-project.org/package=maps⟩ (May 3, 2011).

Morehead, M. D., Syvitski, J. P., Hutton, E. W. H., and Peckham, S. D. (2003). "Modeling the temporal variability in the flux of sediment from ungauged river basins." *Global and planetary change*, 39(1–2), 95–110.

National Research Council. (1999). *Hydrologic hazards science at the U.S. Geological Survey*, National Academies Press, Washington, DC.

National Research Council. (2004). *Assessing the national streamflow information program—Committee on review of the USGS national streamflow information program*, National Academies Press, Washington, DC.

O'Connor, J. E., and Costa, J. E. (2003). "Large floods in the United States: Where they happen and why." *Circular 1245*, USGS, Reston, VA.

Ockerman, D. J., and Heitmuller, F. T. (2010). "Simulation of streamflow and suspended-sediment concentrations and loads in the lower Nueces River watershed, downstream from Lake Corpus Christi to the Nueces Estuary, South Texas, 1958–2008." *Sci. Invest. Rep. 2010-5194*, USGS, Reston, VA.

Potter, K. W., and Walker, J. F. (1981). "A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements." *Water Resour. Res.*, 17(5), 1505–1509.

PRISM Climate Group. (2010). "PRISM Data Explorer." *Northwest Alliance for Computational Science & Engineering (NACSE)*, Oregon State University, Corvallis, Oregon, ⟨http://prismmap.nacse.org/nn/⟩ (Jul. 1, 2010).

Raines, T. H. (1998). "Peak-discharge frequency and potential extreme peak discharge for natural streams in the Brazos River Basin, Texas." *Water-Resour. Invest. Rep. 98-4178*, USGS, Reston, VA.

Raines, T. H., and Asquith, W. H. (1997). "Analysis of minimum 7-day discharges and estimation of minimum 7-day, 2-year discharges for streamflow-gaging stations in the Brazos River Basin, Texas." *Water-Resour. Invest. Rep. 97-4117*, USGS, Reston, VA.

R Development Core Team. (2011). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ⟨http://www.R-project.org/⟩.

Rifai, H. S., Brock, S. M., Ensor, K. B., and Bedient, P. B. (2000). "Determination of low-flow characteristics for Texas streams." *J. Water Resour. Plann. Manag.*, 126(5), 310–319.

Riggs, H. C. (1976). "A simplified slope-area method for estimating flood discharges in natural channels." *J. Res. U.S. Geol. Surv.*, 4(3), 285–291.

Shreve, R. L. (1966). "Statistical law of stream numbers." *J. Geol.*, 74, 17–37.

Slade, R. M., Asquith, W. H., and Tasker, G. D. (1995). "Multiple regression equations to estimate peak-flow frequency for streams in Hays County, Texas." *Water-Resour. Invest. Rep. 95-4019*, USGS, Reston, VA.

Strahler, A. N. (1957). "Quantitative analysis of watershed geomorphology." *Trans. Am. Geophys. Union*, 38(6), 913–920.

Sturm, T. W. (2010). *Open channel hydraulics*, 2nd Ed., McGraw-Hill, New York.

Texas Commission on Environmental Quality (TCEQ), and Texas Parks, and Wildlife Department (TPWD), and Texas Water Development Board (TWDB). (2008). "Texas instream flow studies: Technical overview." *Rep. 369*, Texas Water Development Board, Austin, TX.

Texas Senate Bill 3 Science Advisory Committee for Environmental Flows (TSAC). (2009). "Texas instream flow studies: Technical overview." *Texas Commission on Environmental Quality*, ⟨http://www.tceq.texas.gov/assets/public/permitting/watersupply/water_rights/eflows/geographicscope.pdf⟩ (Aug. 17, 2011).

Turnipseed, D. P., and Sauer, V. B. (2010). "Discharge measurements at gaging stations." *U.S. Geological Survey techniques and methods*, Book 3, Chapter A8, USGS, Reston, VA.

U.S. Geological Survey (USGS). (2009a). "Daily mean streamflow values for Texas." USGS National Water Information System, ⟨http://waterdata.usgs.gov/tx/nwis/dv/?site_no=STATIONIDagency_cd=USGS⟩ (Mar. 1, 2009).

U.S. Geological Survey (USGS). (2009b). "Streamflow measurements for Texas." USGS National Water Information System, ⟨http://waterdata.usgs.gov/tx/nwis/measurements/?site_no=STATIONIDagency_cd=USGS⟩ (Mar. 1, 2009).

Venables, W. N., and Ripley, B. D. (2002). *Modern applied statistics with S*, 4th Ed., Springer, New York.

Vogel, R. M., and Fennessey, N. M. (1994). "Flow-duration curves I: New interpretation and confidence intervals." *J. Water Resour. Plann. Manag.*, 120(4), 485–504.

Wood, S. N. (2006). *Generalized additive models—An introduction with R*, Chapman & Hall/CRC, Boca Raton, FL.

Wood, S. N. (2009). "mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL." *R package version 1.7-6*, ⟨http://cran.r-project.org/package=mgcv⟩ (Mar. 1, 2009).

Wurbs, R. A., and Kim, T. J. (2011). "River flows for alternative conditions of water resources development." *J. Hydrol. Eng.*, 16(2), 17–37.