Comparison of Daily Streamflow Estimation Methods in the Thompsons Creek Watershed

Texas Water Resources Institute TR-535
July 2021

Texas Water Resources Institute
make every drop count

Michael Schramm, Anna Gitter, Jackie Rambo,
Stephanie DeVilleneuve, Ed Rhodes, Lucas Gregory
Texas Water Resources Institute

Comparison of Daily Streamflow Estimation Methods in the Thompsons Creek Watershed

Prepared by:
Michael Schramm, Anna Gitter, Jackie Rambo,
Stephanie DeVilleneuve, Ed Rhodes, Lucas Gregory

Texas Water Resources Institute
578 John Kimbrough Blvd., 2118 TAMU
College Station, Texas 77843-2118

Texas Water Resources Institute TR–535
July 2021

The U.S. Environmental Protection Agency and Texas Commission on Environmental Quality provided financial support for this study.

Cover photo: Thompsons Creek photo by Anna Gitter, TWRI.

Contents

Table of Figures

Table of Tables

Abbreviations

cfs	cubic feet per second
DAR	Drainage Area Ratio
E. coli	*Escherichia coli*
EDF	estimated degrees of freedom
FDC	flow duration curve
GAM	generalized additive model
GHCN	Global Historical Climatological Network
HSPF	Hydrological Simulation Program – FORTRAN
IHACRES	Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Streamflow data
KGE	King-Glupta Efficiency
NHD	National Hydrography Dataset
NOAA	National Oceanic and Atmospheric Administration
nRMSE	Normalized Root Mean Square Error
NSE	Nash-Sutcliffe Efficiency
SWAT	Soil and Water Assessment Tool
SWQM	Surface Water Quality Monitoring
TCEQ	Texas Commission on Environmental Quality
TMDL	Total Maximum Daily Load
USGS	United States Geological Survey
WWTF	Wastewater Treatment Facility

Summary

To facilitate development of a total maximum daily load in the Thompsons Creek watershed, we assessed methods for estimating mean daily streamflow in absence of a suitable long-term streamflow gage within the watershed. Continuous water levels were recorded at three sites in the watershed and combined with periodic 15-minute streamflow measurements to develop streamflow rating curves. The rating curves were used to calculate mean daily streamflow from measured depth data from May 2020 through March 2021. Three methods — drainage area ratio, linear regression, and generalized additive models — were used to estimate daily streamflows, and performance of each method was assessed against the rating curve calculated streamflows. Due to the lack of suitable streamflow gages, the drainage area ratio was determined to perform poorly in the Thompsons Creek watershed. Linear regression and generalized additive models provide acceptable performance for predicting daily streamflows and flow exceedance values. One station with a high proportion of no-flow days required the use of a two-step hurdle model that predicts the likelihood of flow and the volume of flow on a given day. Based on the results, ease of use, and model interpretability, we suggest using linear regression models at two sites and a generalized additive model at the third site to develop estimated daily streamflows for future total maximum daily loads.

Introduction

Report Purpose

Three individual tributaries are found in the Thompsons Creek watershed: Cottonwood Branch (1242B), Still Creek (1242C), and Thompsons Creek (1242D). In total, this watershed spans nearly 33,297 acres in Brazos County and is adjacent to the cities of Bryan (population 85,445) and College Station (population 116,218; Gitter et al. 2020). The Thompsons Creek watershed is located in east-central Texas and is characterized as having a subtropical humid climate. Currently, portions of all water bodies are listed as impaired for elevated concentrations of the fecal indicator bacteria *Escherichia coli* (*E. coli*). Further information regarding the characteristics of the watershed can be found in *Watershed Characterization of the Thompsons Creek Watershed* (Gitter et al. 2020).

The waterbody impairment necessitates the development of a watershed-based plan to address potential causes of the impairment. Most watershed-based plans in Texas rely on streamflow-pollutant relationships (typically load duration curves) to describe the relationship between potential pollutant sources and instream pollutant concentrations. Like most small watersheds in the state, daily streamflow data is not available in the Thompsons Creek watershed. The primary purpose of this study is to develop a record of daily streamflows for the Texas Commission on Environmental Quality (TCEQ) Surface Water Quality Monitoring (SWQM) stations in the Thompsons Creek watershed to facilitate development of total maximum daily loads (TMDLs) and watershed-based plans.

Background

Quantifying streamflow is critical for characterizing water quality and estimating pollutant loads in a watershed. Streamflow, or discharge, is defined as the volume of water that moves over a designated point over a fixed period and can be expressed as cubic feet per second (cfs). Various organisms depend on specific streamflows, and therefore understanding the fluctuations and patterns of streamflow within a watershed can be helpful for ensuring adequate ecosystem health. Additionally, the quantification of streamflow is essential for understanding the stream's capacity to entrain sediment, as well as for calculating the upper limit of pollutants that can enter a water body without drastically impacting water quality or exceeding water quality standards.

United States Geological Survey (USGS) stations provide reliable and validated measures of routine streamflow data. However, USGS stations can easily cost $25 thousand for equipment and installation and require at least $15 thousand per year for operation and maintenance (Lasater et al. 2019). Due to the costs, resources, and time needed for continuous and routine monitoring, continuous streamflow data are not always available, especially in smaller, less prioritized streams. Many watershed-based plans in Texas utilize the drainage area ratio (DAR) approach to estimate the naturalized flow in a water body if there is an absence of flow records within the impaired watershed. The DAR approach is a widely used technique that utilizes streamflow records available from a nearby watershed with similar climate and land cover characteristics, including watershed size, land use, urban area, and soil type (Asquith et al. 2006).

For the purposes of estimating streamflow in the Thompsons Creek watershed, the DAR method is anticipated to be an inappropriate approach due to the lack of neighboring watersheds that have similar land cover types and hydrograph patterns. Due to the costs associated with continuous streamflow monitoring and the limitations of DAR, this study aims to examine the potential applicability of other methods for estimating streamflow within watersheds lacking historical streamflow data.

Generally, there are two categories of methods for predicting streamflow in an ungaged basin. First, *information transfer methods* rely on the mathematical or statistical transfer of information from a gaged watershed to an ungaged watershed. The second approach is the statistical or mathematical modeling of streamflow response to forcing variables such as precipitation that will be referred to as *rainfall-runoff methods*. These two methods are discussed in greater detail in this report. While a third approach known as *regionalization* does exist, it is not further discussed.

Information Transfer Methods

Statistical transfer procedures simply transfer flow duration curves (FDC) or daily streamflow values from a gaged watershed to the ungaged watershed using assumed relationships between area and runoff. The most common statistical transfer method is the DAR method. With the DAR, daily streamflows are transferred from one basin to the other by multiplying the area ratio to daily streamflows:

$$Q_y = Q_x(A_y/A_x)^{\phi} \qquad (1)$$

where Q_y is streamflow at ungaged basin y, Q_x is streamflow at gaged basin x, and A_y/A_x is the ratio of the area between the two basins. Parameter ϕ is typically equal to one. However, Asquith, Roussel, and Vrabel (2006) provide empirically estimated values of ϕ for use in the DAR when applied in Texas. A major benefit of this method is that it requires no additional data outside of drainage sizes and daily streamflow at the gaged site and performs well under appropriate conditions.

If some streamflow data is available at the ungaged site, we can extend the DAR to an empirical linear regression for streamflow estimation using one or more gaged watersheds (State of Ohio Environmental Protection Agency 2009):

$$Q_y = \beta_0 + \beta_n Q_{xn} + \varepsilon \qquad (2)$$

where Q_y is the predicted mean daily streamflow (typically log transformed) at the ungaged site y, β_0 is the intercept, Q_{xn} is the mean daily streamflow at gaged watershed n, β_n is the regression coefficient, and ε is the residual error term assumed normally distributed around mean zero. A linear regression of this form still acts as a streamflow transfer method like the DAR approach but allows for an easy incorporation of additional model terms such as lagged streamflows, which might improve predictive performance. Additionally, having some streamflow data at the ungaged site not only allows for the ability to incorporate an empirical approach but also validates of the accuracy of the DAR approach.

Rainfall-Runoff Methods

If nearby gaged watersheds are unavailable or are not reflective of the streamflow responses in the ungaged watershed, locally available weather information can be used to empirically estimate streamflow response. Several empirically based rainfall-runoff routing models are available that account for soil and land-use conditions to predict streamflow response (SIMHYD, Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Streamflow data [IHACRES], and Sacramento rainfall-runoff models are examples). More complex mechanistic models that simulate hydrologic and water quality responses to land use and precipitation are also available (Soil and Water Assessment Tool [SWAT] and Hydrological Simulation Program – FORTRAN [HSPF] are two examples). The mechanistic models have a steep requirement for data and ability of the technician developing the model. The routing models and mechanistic models are outside the scope of work for this particular project. Here we focus on using a semi-parametric regression-based approach to predict streamflow using locally available weather data. Empirical regression-based approaches require a period of measured streamflow and some predictor variables to estimate the runoff response from. Typically, daily rainfall and temperature data are employed to fit a regression model to measured streamflow response.

If the relationship between predictor variables and the response are expected to be nonlinear, polynomial terms can be included. However, generalized additive models (GAMs) allow relatively easy fitting of these nonlinear terms to the data. With GAMs, the response variable depends on the sum of smoothing functions applied to each predictor variable:

$$Q_i = \beta_0 + f(x_1) + f(x_n) + \varepsilon \qquad (3)$$

where Q is predicted discharge on day i, β_0 is the intercept, the sum of $f(x_1)...f(x_n)$ is equivalent to the sum of the smoothing functions (f) applied to the linear predictors ($x_1...x_n$), and ε is the residual error term assumed normally distributed around mean zero. In the case of GAMs fit using the mgcv package in R, f is a smoothing function fit to the data using generalized cross validation or restricted maximum likelihood (Wood 2008; R Core Team 2021). Due to the smoothing functions and link function, GAMs are extremely flexible for fitting regression models to data of different distributions and responses. However, compared to linear regression, the inclusion of the smoothing functions limits interpretability because traditional regression coefficients are not part of the model structure. Therefore, the effect of each individual smoothing function on the mean of the response variable is shown graphically. A handful of studies have successfully applied GAMs to predict regional FDCs and monthly streamflows at ungaged sites (Shortridge et al. 2016; Ouali et al. 2017). While the use of GAMs for daily streamflow prediction is relatively novel, GAMs are prevalent in the academic literature, including applications in water quality and extreme flow event modeling (Richards et al. 2010; Hagemann et al. 2016; Beck and Murphy 2017; Murphy et al. 2019).

Measured Streamflows

Development of empirically based streamflow estimation methods (linear regression and GAMs) and evaluation of the performance of all the methods described above requires a period of measured streamflow. The proposed method to develop a period of streamflow record includes the use of two main instruments: SonTek-IQ Plus and an Onset HOBO water level logger. The SonTek-IQ Plus is an advanced flow meter, designed for measuring continuous flow in open channels. This instrument functions by recording the horizontal and vertical distribution of velocities in the channel. A theoretical flow is calculated by the unit using channel geometry and the channel velocity profile.

The resulting continuous streamflow data is combined with stream stage recorded by the HOBO water level logger to develop a stage discharge curve. The HOBO water level logger is a high-resolution pressure sensor that defines the stream's water level. The HOBO water level logger records stage data every 15 minutes, and a HOBO barometric pressure transducer, which should be installed within 16 kilometers of the water level logger, accounts for fluctuations in atmospheric pressure. The stage discharge curve establishes a relationship between stream depth and flow. This is useful for determining flows when the flow meter is not deployed (the IQ Plus is a bottom mount flow meter and not suitable for permanent deployment in a natural stream due to sedimentation, fouling, etc.).

Under steady-state flows, the rating curve is a power function relating stream height and discharge (Venetis 1970):

$$Q = K(H-H_0)^z \qquad (4)$$

where Q is steady-state discharge, H is stream height (stage), and H_0 is the stage at zero discharge. K and z are rating curve constants. Following convention, Q and H are log-transformed prior to parameter estimation.

Unsteady flows occur when the rising and falling stage of the stream hydrograph results in different discharges at identical stream heights. The resulting hysteresis-affected rating curve will present as a loop rather than a line. The modified Jones formula described by Petersen-Øverleir (2006) and Zakwan (2018) is used:

$$Q = K(h-a)^n \times (1+x(\partial h/\partial t))^{(1/2)} \qquad (5)$$

where Q is discharge, and h is stream height. The partial first order derivative, $(\partial h/\partial t)$, is approximated as J using finite differences:

$$J(h_t) = (h_t+1 - h_t-1)/\,\Delta t \qquad (6)$$

Where h_t is the stream height at time t, and Δt is the time interval. Simplified, this is the slope or the instantaneous rate of change for the function between stream height and time that is estimated using measured stream height values. K, a, n, and x in equation (5) are rating curve constants.

To summarize, this project has two parts. First, a period of streamflow record is generated using rating curves developed from continuously measured stream heights and periodically measured concurrent streamflows. Second, three different methods of estimating streamflows (DAR, linear regression, and GAM) at these sites are validated using the generated streamflow record.

Methods

Site Descriptions

SWQM stations 16396 and 16397 are both located on Thompsons Creek (1242D), which flows 18 miles from the confluence of the Brazos River upstream to the confluence of

Thompsons Branch, north of FM 1687 (Gitter et al. 2020). SWQM station 16396 is located on AU 1242D_01, while SWQM station 16397 is located upstream on AU 1242D_02. SWQM station 16882 lies on AU 1242C_02 of Still Creek (1242C), which is a 9-mile perennial stream segment that flows from the confluence with Thompsons Creek upstream to the headwaters in Brazos County near US 190. Individual site locations are shown in the map below (Figure 1).

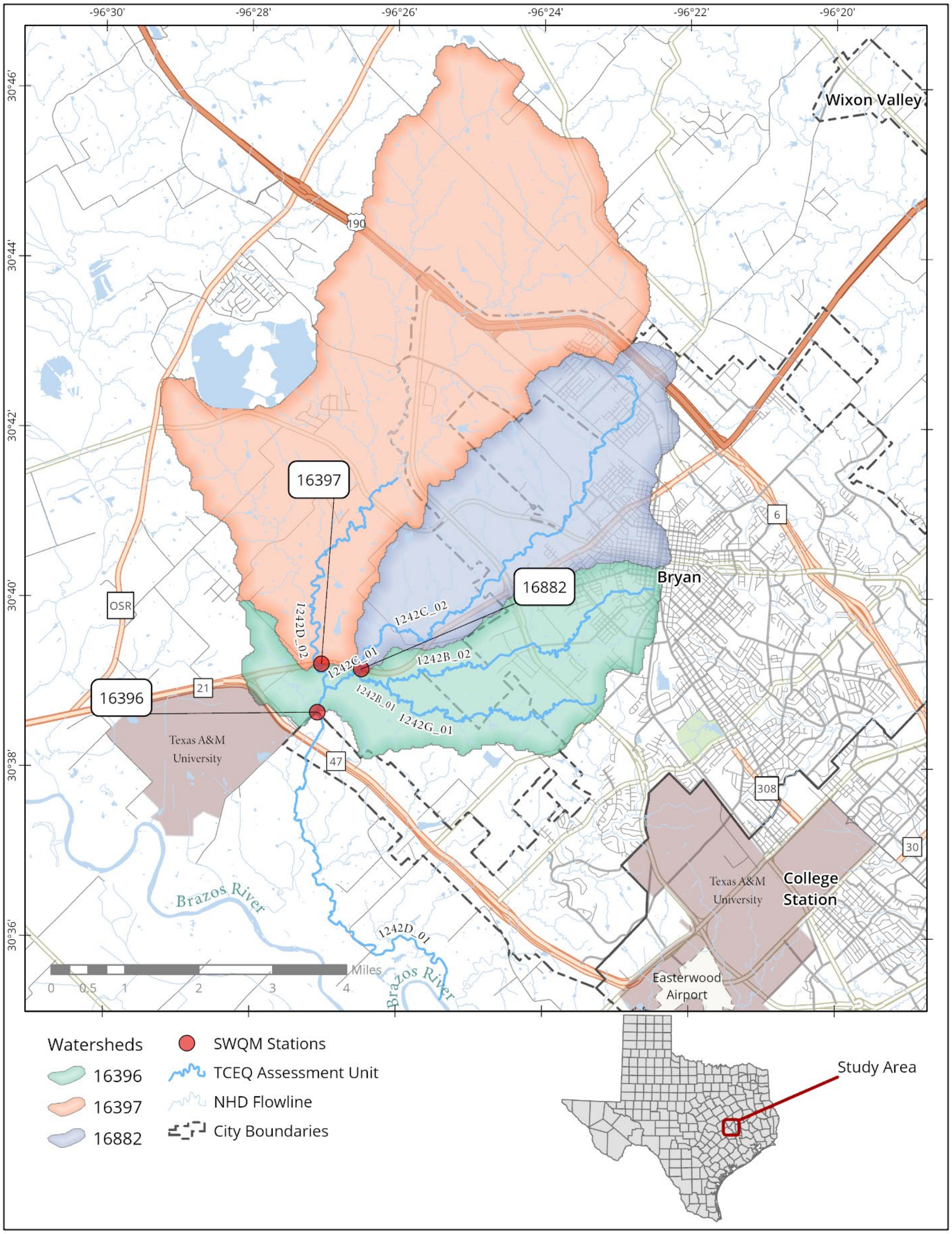

Figure 1. Map of the Thompsons Creek watershed. SWQM = Surface Water Quality Monitoring, TCEQ = Texas Commission on Environmental Quality, NHD = National Hydrography Dataset.

Site Installation and Data Collection

Each SWQM station (Figure 1) was instrumented with a HOBO U20L water level logger and a SonTek-IQ Plus acoustic doppler flow meter. The HOBO U20L is a submersible pressure transducer that calculates water depth using pressure readings it logs, corrected with paired atmospheric pressure readings. The SonTek-IQ Plus flow meter records the velocity of particles in the water column and calculates flow volume using measured velocities, water depths, and programmed stream profiles processed with SonTek's proprietary algorithms. Photos of sites where the SonTek-IQ Plus flow meter was installed are portrayed in Figure 2.

Figure 2. Left: Thompsons Creek at surface water quality monitoring (SWQM) station 16396 during a high flow event. Right: Still Creek at SWQM station 16882. The SonTek IQ Plus unit can be seen in the middle of the streambed.

The instream HOBO U20L units were mounted in the stream channel at each station using a T-post driven into the streambed approximately 20 feet away from the SonTek-IQ Plus to prevent potential interference. The logger was mounted to the downstream side of the post to protect against debris and was housed in a protective mesh canister that allowed the water to freely flow through. One HOBO U20L was attached to a tree in the floodplain near SWQM station 16882 to collect local atmospheric barometric pressure. These atmospheric pressure readings allow for site-specific corrections to the underwater pressure readings, ensuring that water depth calculations are more precise. Water depth calculations were performed using the proprietary HOBOWARE software package. The HOBO U20L has a typical error of ± 0.1% full scale, equivalent to less than 0.1 feet of water or a maximum error of ± 0.2% full scale, equivalent to less than 0.2 feet of water (Onset 2018). Units were programmed to collect data every 15 minutes beginning upon deployment. Data were downloaded from these units approximately quarterly. The stream height at 15-minute intervals and corresponding summary data are shown in Figure 3 and Table 1, respectively.

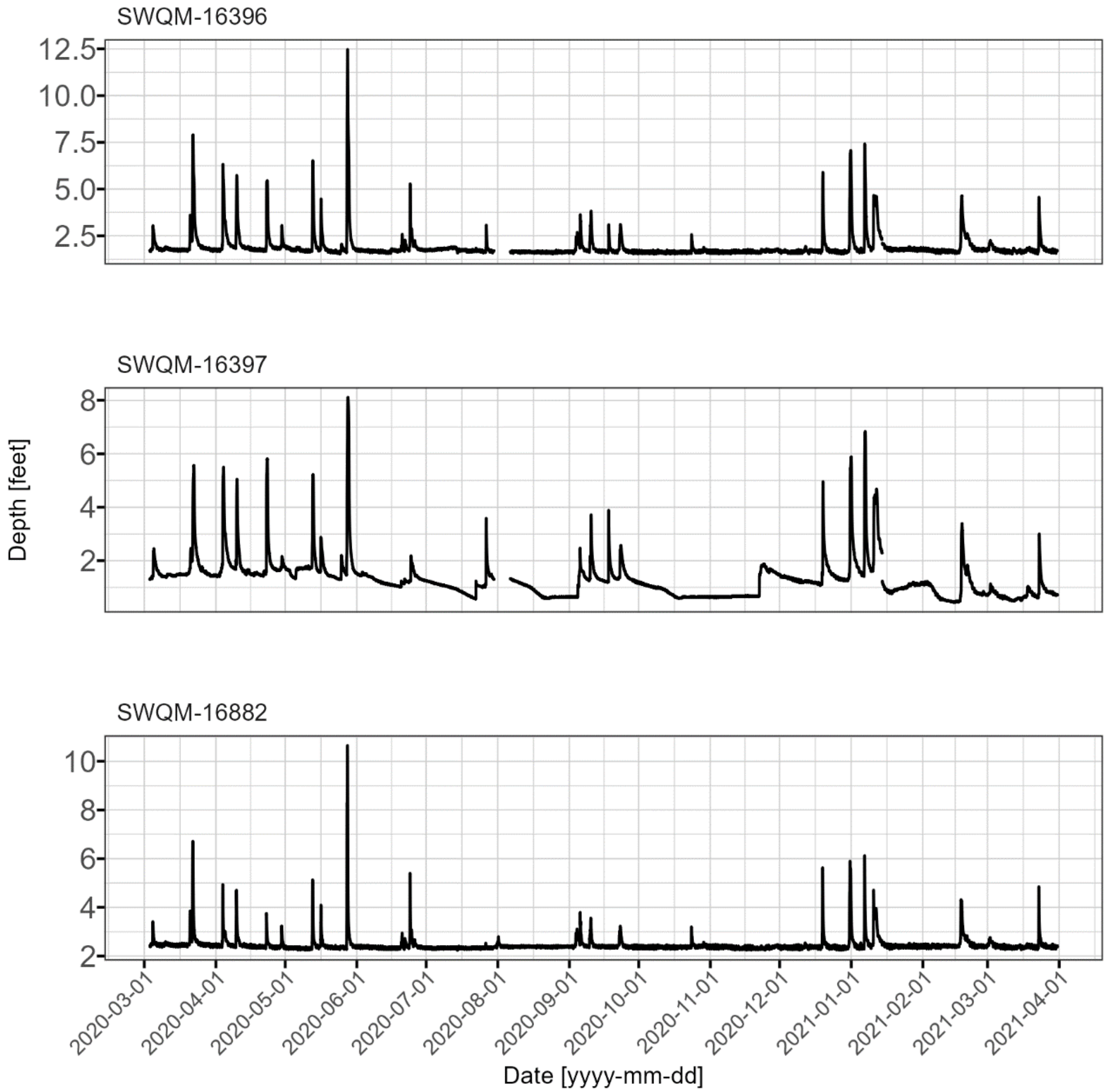

Figure 3. 15-minute stream heights measured at each surface water quality monitoring (SWQM) station from March 3, 2020 through March 31, 2021.

Table 1. Mean and standard deviation of stream height measurements.

Surface water quality monitoring (SWQM) station	n	Mean height (feet)	Standard deviation (feet)
SWQM-16396	37,030	1.89	0.68
SWQM-16397	37,032	1.30	0.75
SWQM-16882	37,712	2.46	0.38

SonTek-IQ Plus units were installed instream in the most suitable location accessible. Per the product manual, units should be installed in a straight section of each stream, avoiding any natural bends or abrupt changes in elevation that create turbulent, non-laminar flow. This is always a challenge in a natural stream environment, especially under limited access situations. Our best attempts were made to install these units in areas of uniform flow and low turbulence. Each unit was mounted to an 80-pound custom fabricated concrete block that was buried into the streambed such that the IQ Plus was slightly above the streambed.

Concrete blocks were staked down using 2 foot long metal rods to minimize potential for instrument loss. IQ Plus units were attached to stainless steel 1-inch risers that were bolted to the concrete blocks to allow sediment to pass under the unit. Power and communication cables for each unit were sleeved using a ¾-inch liquid-tight conduit buried into the streambed and bank to protect system components and minimize potential down time. A 12V deep cycle battery housed in a lock box mounted well above the stream channel powered each unit. Once units were installed, a cross-sectional survey was conducted at each site to create a stream profile (Figure 4). Elevation measurements of the streambed and banks relative to the top of the IQ Plus unit were conducted at 1-foot increments away from the units in each direction. At points in the cross-section where major changes occurred, measurements between the predefined 1-foot increments were made to refine the measured profile. Manufacturer specifications indicate unit accuracy specifications for water level (greater of 0.1% of measured depth or 0.01 foot), pressure (0.1% of full scale), and velocity (±1% of measured velocity). Units were programmed to collect data every 15 minutes using a 2-minute average velocity reading beginning upon deployment. Data were downloaded from these units approximately quarterly and before and after select storm events. Two units were available and rotated between sites to maximize data collected. However, technical issues with the units limited us to a single unit for a portion of the summer and towards the end of sample period.

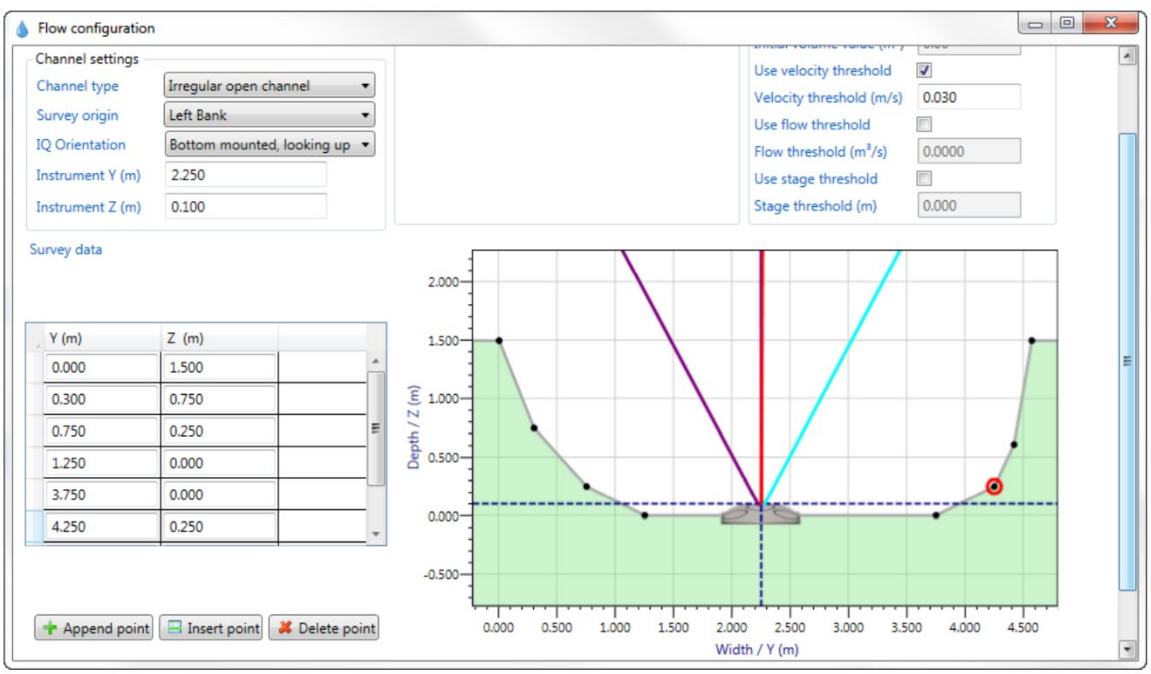

Figure 4. An example of a cross-section developed at each site where equipment was installed.

During data exploration, it was evident there was excess noise in the data at low flows for each station. During periods of stagnant or near-stagnant conditions, the doppler flow meters recorded highly variable stream velocities and reported unrealistic flows. Due to the excess data noise, periods of extreme low or stagnant flow were cleaned from the data record. Future deployments will need to consider under what conditions long-term deployments are appropriate. For small streams such as these, periodic storm flow deployments are likely the most appropriate deployment. Figure 5 shows the cleaned 15-minute data record generated by the acoustic doppler units.

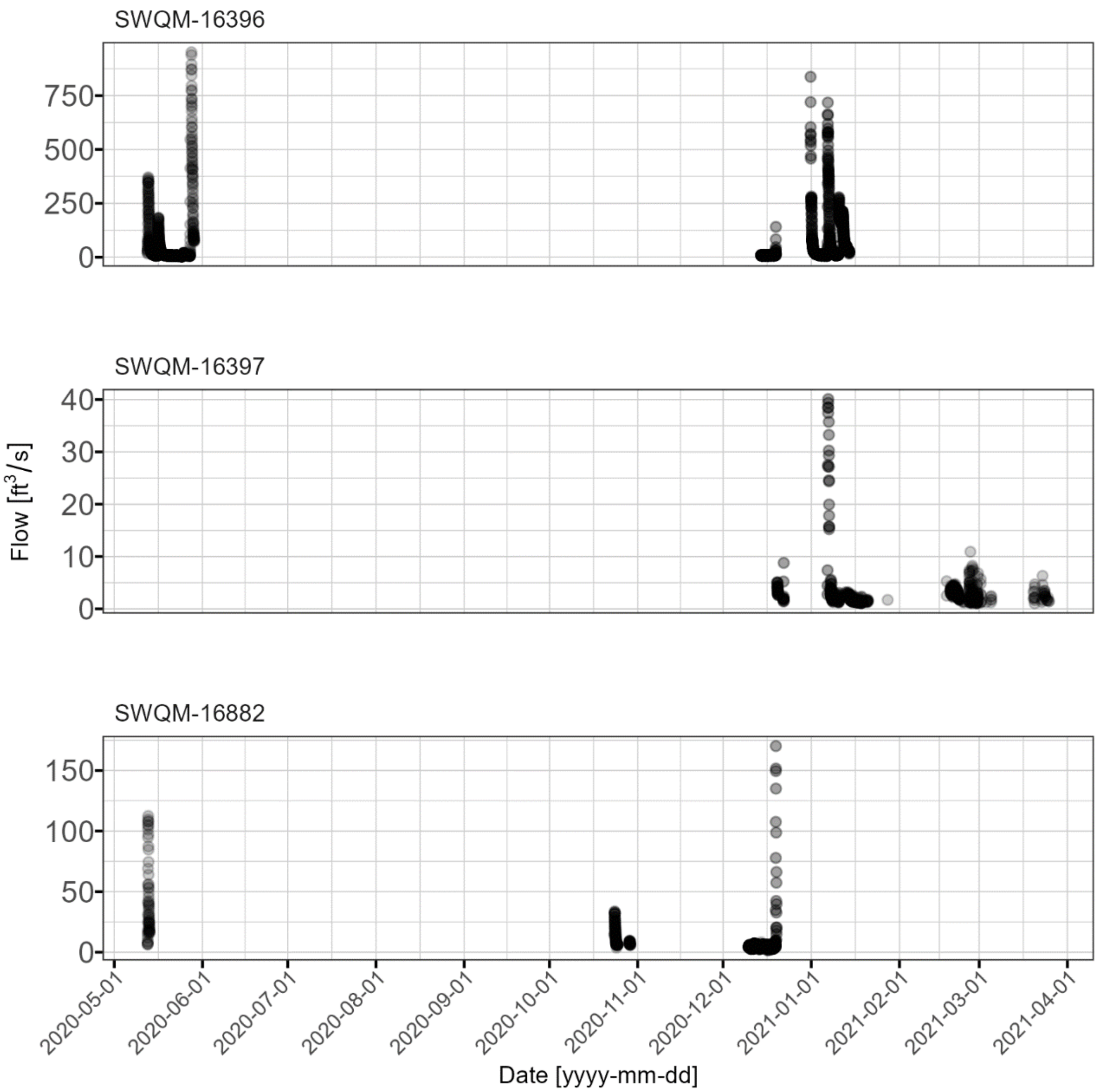

Figure 5. 15-minute streamflow measured at each surface water quality monitoring (SWQM) station from March 3, 2020 through March 31, 2021. ft^3/s = cubic feet per second.

Rating Curve Development

Rating curves were developed using measured streamflow data from the SonTek IQ and the HOBO U20L. The rating curve constants are solved using nonlinear least squares regression to minimize the sum of square error. Nonlinear optimization methods search through parameter combinations to minimize the sum of square error. Petersen-Øverleir (2006) applied the Nelder-Mead algorithm to solve the Jones formula. Zakwan (2018) presented spreadsheet-based nonlinear optimization methods using generalized reduced gradient and genetic algorithm. Most methods require careful planning for parameter starting values that are somewhat near the global minimum value or risk identifying alternative local minimum values.

Instream and stream bank conditions change through the year due to plant growth, plant dieback, sedimentation, erosion, and other processes. These changing conditions can necessitate the development of multiple rating curves. Exploratory data analysis was used to identify periods of change and the potential for hysteresis-affected rating curves. Once rating curve periods and appropriate formulas were determined, rating curve parameters in equations (4) and (5) were estimated using nonlinear least squares regression in the R statistical software (R Core Team 2021). To reduce the likelihood of convergence on local minimum, the nls.multstart package in R provides functionality to iterate nonlinear least squares optimization over many different starting values (Padfield and Matheson 2020). Individual rating curves were used to estimate streamflows using the measured stream heights. Nash-Sutcliffe Efficiency (NSE) and normalized Root Mean Square Error (nRMSE) were used to evaluate goodness-of-fit between measured and estimated streamflow. As a final step, the rating curve estimated 15-minute streamflows were summarized to main daily streamflow.

Mean Daily Streamflow Prediction

Prior to applying DAR or regression methods, a daily naturalized streamflow was estimated at each site. This was done to minimize the influence of artificial discharges on the streamflow record. We identified two wastewater treatment facilities located upstream of the lowest discharge point. The Still Creek wastewater treatment facility (Permit Number WQ0010426002) is permitted to discharge 4.0 MGD to Still Creek, and discharge flows past SWQM sites 16882 and 16396. The Sanderson Farm Inc. facility (Permit Number WQ0003821000) is permitted to discharge 1.678 MGD to a tributary of Cottonwood Branch, and the discharge flows past SWQM site 16396. Mean daily wastewater facility discharges were downloaded from the U.S. Environmental Protection Agency Environmental Compliance and History Online database using the echor R package and subtracted from the measured mean daily streamflows to better represent naturalized mean daily flows (Schramm 2018). Mean daily discharges are summarized and reported on monthly intervals (Figure 6). Mean daily discharges for each day would provide better estimates of naturalized streamflow but were not available in the watershed.

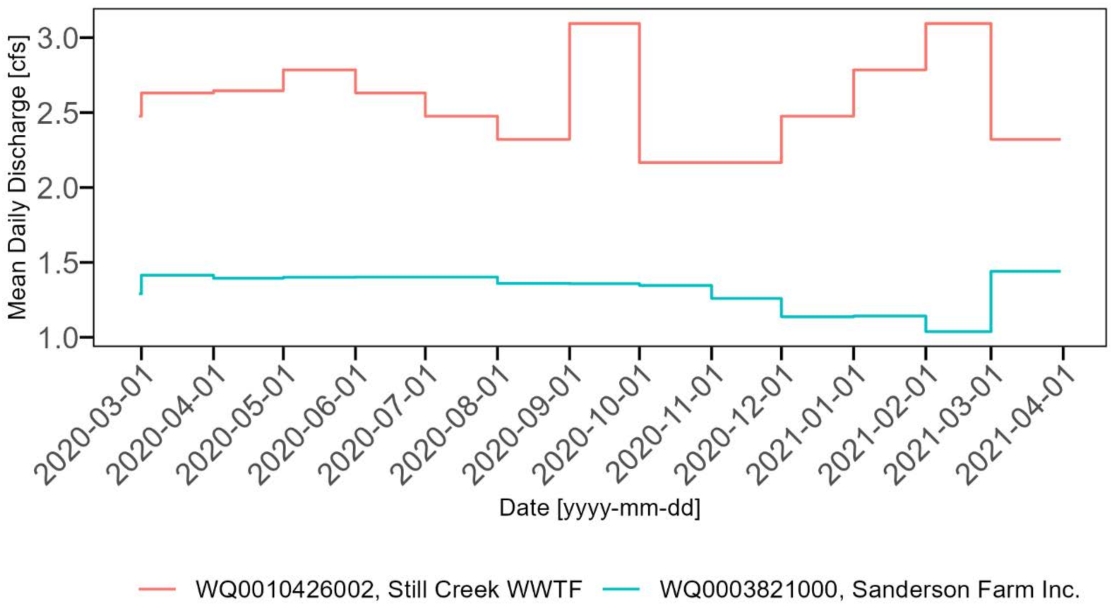

Figure 6. Timeseries plot of mean daily wastewater discharges. cfs = cubic feet per second, WWTF = Wastewater Treatment Facility.

DAR

Naturalized mean daily streamflow was estimated using the DAR approach in equation (1). Values of ϕ recommended in Asquith, Roussel, and Vrabel (2006) were used. Three different USGS stream gages (08065800, 08109800, and 08110100) were evaluated for performance (Table 2, Figure 7, Figure 8). Mean daily streamflows from each gage were downloaded from the USGS National Water Information System using the dataRetrieval package in R (De Cicco et al. 2018).

Table 2. Drainage areas and site locations used for the drainage area ratio (DAR) procedure.

Site	Description	Drainage area (miles^2)
SWQM-16396	Thompsons Creek at Silver Hill Road	42.33
SWQM-16397	Thompsons Creek at Highway 21	24.21
SWQM-16882	Still Creek at Highway 21	10.03
USGS-08065800	Bedias Creek near Madisonville	321.00
USGS-08109800	East Yegua Creek near Dime Box	244.00
USGS-08110100	Davidson Creek near Lyons	195.00

Note – SWQM = Surface Water Quality Monitoring, USGS = United States Geological Survey.

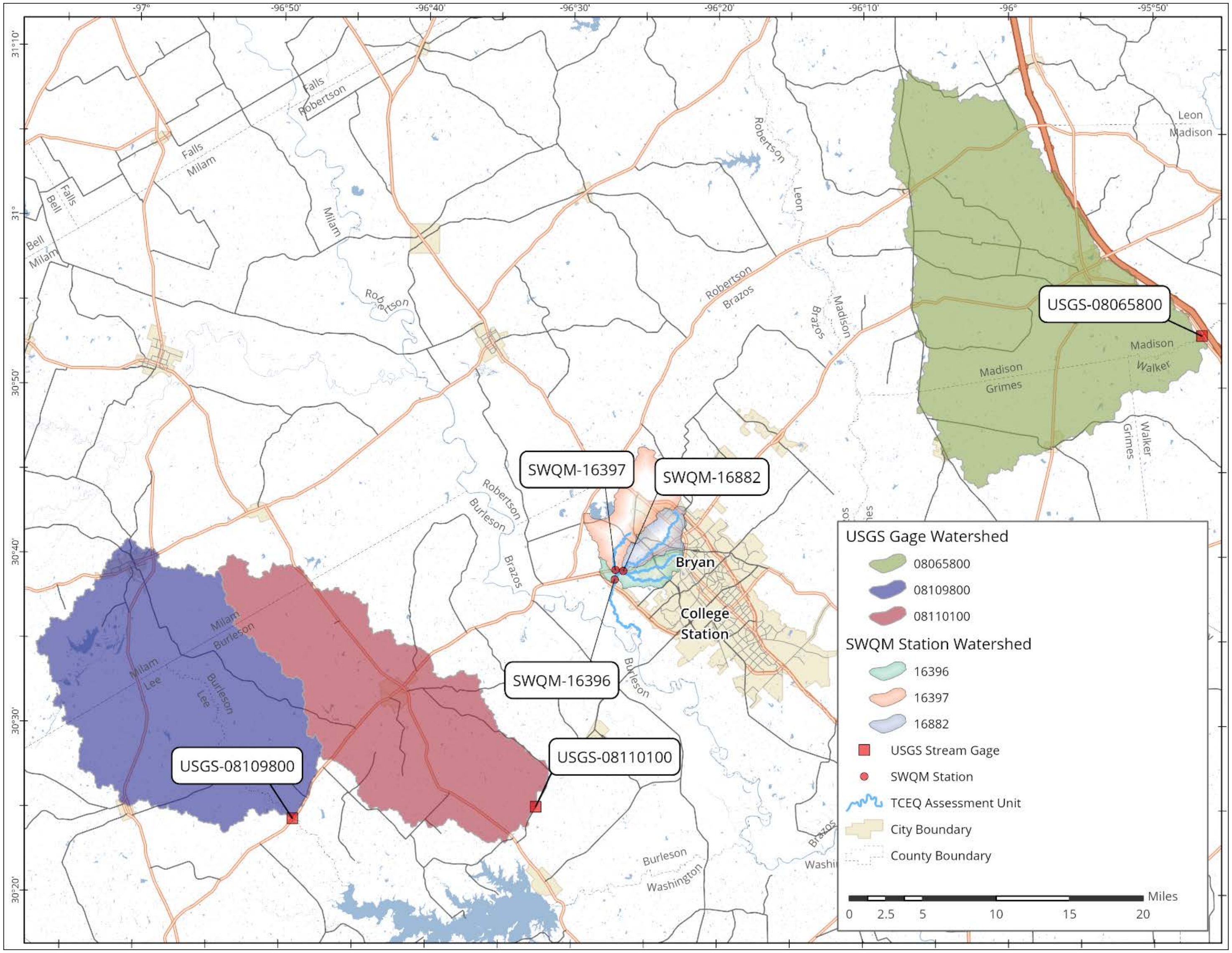

Figure 7. Source U.S. Geological Survey (USGS) gages and watersheds used by drainage area ratio approach. SWQM = Surface Water Quality Monitoring, USGS = United States Geological Survey, TCEQ = Texas Commission on Environmental Quality.

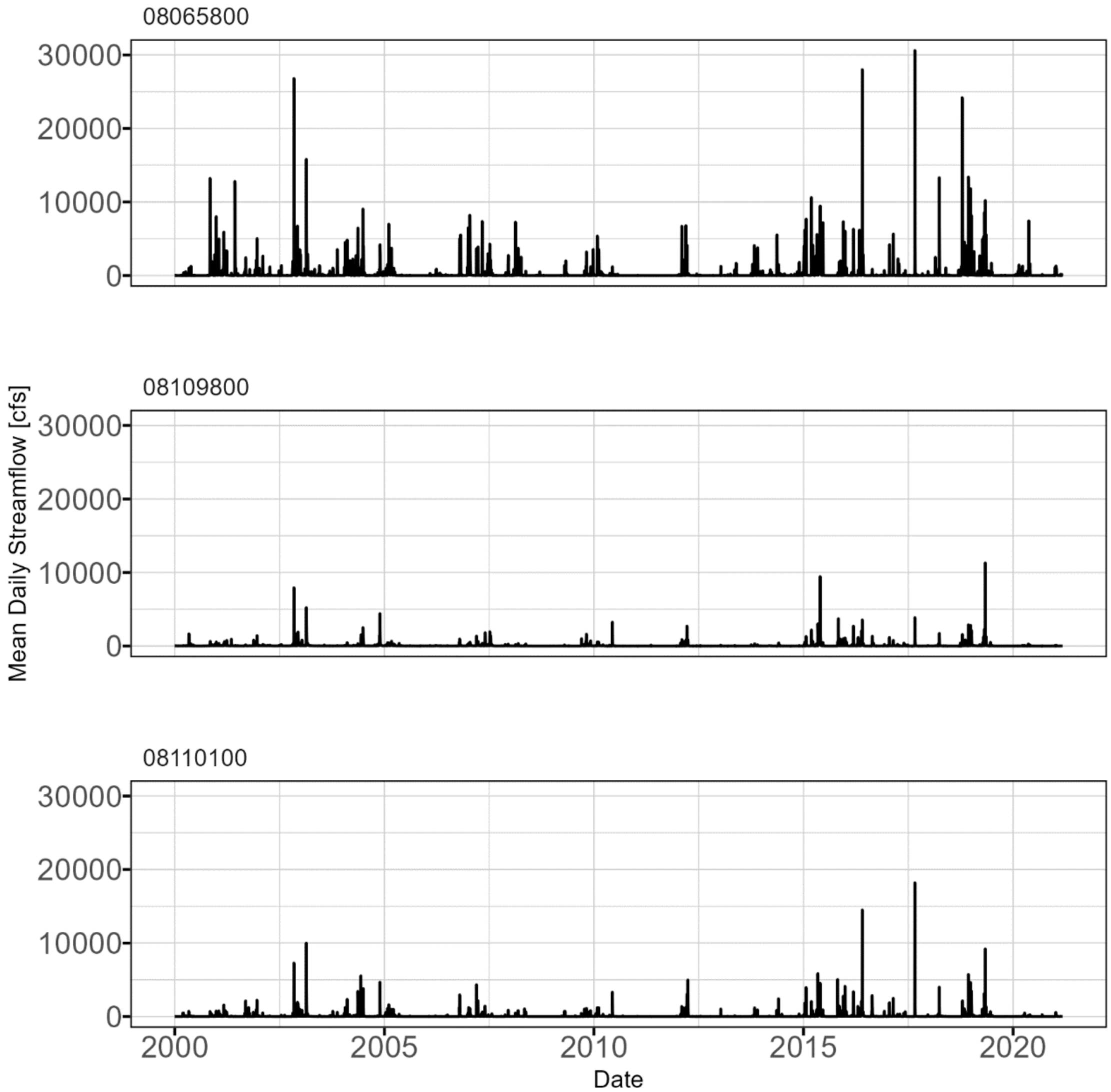

Figure 8. Timeseries hydrographs of mean daily streamflows at U.S. Geological Survey gages used in the drainage area ratio approach. cfs = cubic feet per second.

Linear Regression

Linear regression was used to estimate log transformed flows at each SWQM station using mean daily flows and 1-day lagged mean daily flows from each USGS stream gage as a predictor variable:

$$\log(Q_{yi}) = \beta_0 + \beta_1 Q_{08065800,i} + \beta_2 Q_{08109800,i} + \beta_3 Q_{08110100,i} + \beta_4 Q_{08065800,i-1} + \beta_5 Q_{08109800,i-1} + \beta_6 Q_{08110100,i-1} + \varepsilon \quad (7)$$

where Q_{yi} is streamflow at ungaged site y on day i. Each regression term is mean daily streamflow at the indicated gage on day i or day i-1. All terms were log transformed prior to fitting the linear regression.

GAMs

GAMs modelled the streamflow response to local climatological predictor variables. One National Oceanic and Atmospheric Administration (NOAA) Global Historical Climatological Network (GHCN) location provided total daily precipitation and maximum daily temperature data for the project area (Figure 9, Figure 10). GHCN daily summaries were downloaded for GHCND:USW00003904 (Easterwood Airport) using the NOAA application programming interface services and the rnoaa package in R (Chamberlain et al. 2019). Easterwood Airport is between 1 and 7.6 miles (straight-line distance) from each SWQM station and 3.2 miles from the closest watershed boundary. Total daily precipitation spiked at 0 days because there were predominantly 0 days of rainfall.

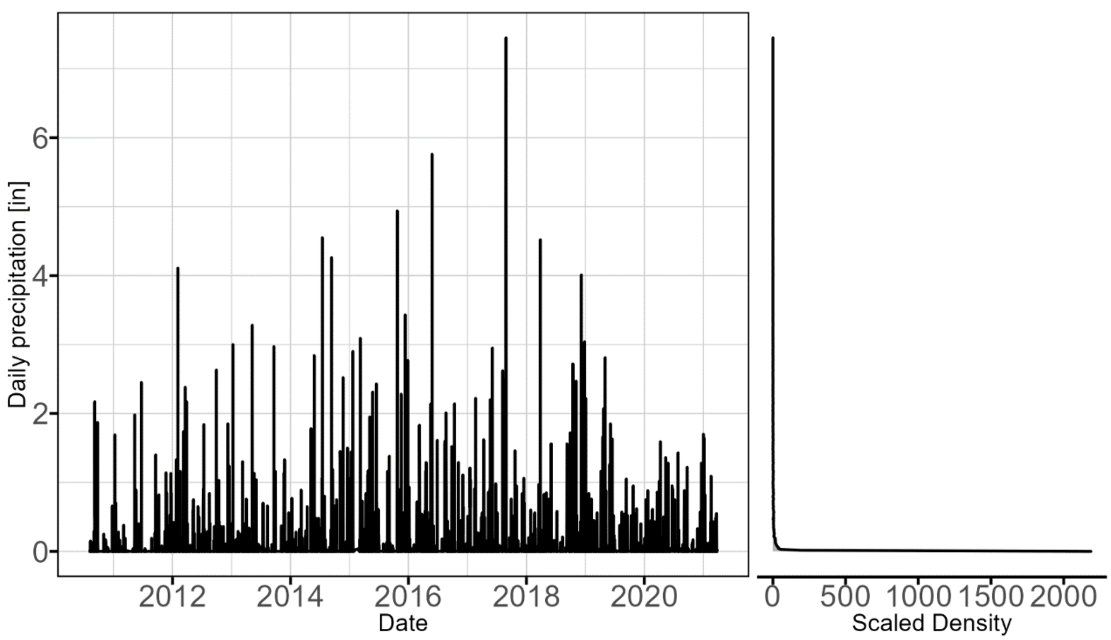

Figure 9. Timeseries and density plots of total daily precipitation in inches (in.) at Easterwood Airport.

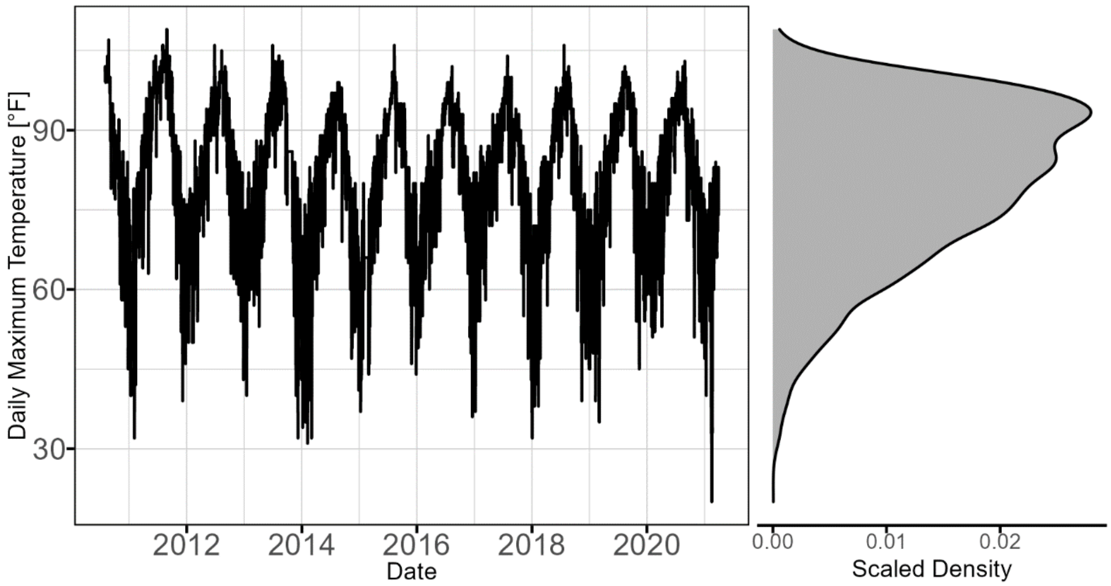

Figure 10. Timeseries and density plots of daily maximum temperature at Easterwood Airport.

The GAM was formulated using the following locally available climatological predictor variables:

$$Q = f(P) + f(T) + f(P_{lag}) + f(P_{sum,3}) + f(T_{mean,5}) + f(H) + f(M) \quad (8)$$

where P is log transformed total daily precipitation, T is squared maximum daily temperature, and both are assumed to be the main forcing variables influencing streamflow. P_{lag} is the log transformed 1-day lagged precipitation. $P_{sum,3}$ is the log transformed rolling 3-day sum rainfall and is included as an indicator of wetness in the watershed. $T_{mean,5}$ is the rolling 5-day mean (squared) of the maximum daily temperature and is included as an indicator of seasonal temperature condition that exhibits less variance than T and may be an improved indicator of potential evapotranspiration conditions. H and M represent relative humidity and month of the year, respectively. Note that predictor variables were transformed prior to model fitting to reduce skewness and improve model residuals. The smoothing function, f, is a cubic regression or thin plate spline function fit to the observed data using restricted maximum likelihood, which estimates the optimal smoothing parameters. The GAM error structure was fit using a Gamma distribution and log link function.

SWQM station 16882 presented an additional challenge due to a high frequency of no-flow days once the streamflow was naturalized. To accommodate zero flow values a two-step hurdle model was set up (Liu et al. 2019). In the first model, a logistic regression GAM was developed to predict the probability of any streamflow on a given day. On days with less than 50% likelihood of flow, the streamflow is assumed zero. On days with greater than 50% likelihood of flow, another GAM is used to predict the amount of flow on that day. For the logistic regression, the model covariates remain the same, except the GAM uses a binomial distribution and logit link. The number of covariates in the second GAM had to be reduced due to fewer available model degrees of freedom. Variables included in the final model were based on information criterion parameters (Zuur 2009).

Performance evaluation

The performance of DARs was assessed using NSE and King-Glupta Efficiency (KGE) goodness-of-fit metrics between rating-curve-estimated naturalized streamflows and DAR-estimated streamflows for all available streamflow data. Initial performance of linear regression and GAMs were evaluated using models fit to the full dataset using NSE and KGE metrics.

For methods that appeared to provide suitable performance, Monte-Carlo cross validation was also conducted to evaluate how well the methods will perform on “out of sample” data. In hydrological modeling, some portion of the dataset is often held out of model-fitting procedures (calibration) and used to validate how well the model performs when used on data not included in the model fitting procedures. Physically based models often exhibit similar goodness of fits between in-sample and out-of-sample datasets. A concern with statistical models is that predictions do not extrapolate well outside of the data used to fit the model. In this case, with limited (approximately 1-year) streamflow data, it is preferable to maximize the data used to fit the model. Monte-Carlo cross validation involves repeatedly holding out a randomized small portion of the dataset, fitting the model to the remaining data, and then validating against the held-out data (Haddad et al. 2013). This is repeated *n* times on different subsets of data. Then an average or distribution of the goodness-of-fit metrics provides insight into how well the method will perform on data outside of the dataset used to fit the model. Monte-Carlo cross validation was performed on suitable models by randomly holding out 30% of the data on 100 different model iterations. NSE, KGE, nRMSE, and r^2 metrics are reported on the cross-validation results.

Results

Rating Curves

SWQM Station 16396

SWQM station 16396 showed unsteady flows, and the Jones formula rating curve was used. The developed rating curve parameters resulted in very good fit between streamflow measurements and rating curve values (Table 3, Figure 11). The measured streamflow values between December 2020 and March 2021 exhibited a handful of potential measurement errors with abnormally low streamflow measurements at a given depth. There only appears to be two of these outliers, and they are not expected to impact the accuracy of the developed rating curves.

SWQM Station 16397

Steady-state flow conditions were observed at SWQM station 16397. The resulting rating curves were developed using a power function (Table 4, Figure 12). Rating curve accuracy is questionable at low flows for this site. For the first rating curve, the rating curve fits very well at higher flow values. There is high variance in the measured streamflows at low stream height for both time periods. The second period only has a variance in stream depth of slightly more than 1 foot and approximately 5 cfs. This has a strong negative impact in normalized goodness-of-fit metrics because the same absolute deviance at higher streamflows would be more acceptable. It is also notable that the least squares function settled on an unrealistic negative value for H_0. Possible improvements in this rating curve might be observed by setting the value of H_0 to some empirically observed field value.

SWQM Station 16882

Unsteady flows were observed at SWQM station 16882. The rating curves resulted in very good performance for the first two periods (Table 5, Figure 13). The third period had acceptable performance but high variance in measured streamflow values at low depths and on the decreasing stage of the hydrograph resulted in decreased goodness of fit.

Table 3. Rating curve parameters and goodness-of-fit metrics for surface water quality monitoring (SWQM) station 16396.

SWQM station	Period	K	a	n	x	NSE	nRMSE
SWQM-16396	2020-03-03 : 2020-05-17	4.807710	0.3664121	0.4816073	-0.1807818	0.985	2.5
SWQM-16396	2020-05-18 : 2020-12-13	4.691948	0.3325470	0.4981528	-0.1135623	0.988	1.5
SWQM-16396	2020-12-14 : 2021-03-31	4.391591	0.1785159	0.6552531	0.0808611	0.975	1.8

Note – NSE = Nash-Sutcliffe efficiency, nRMSE = normalized root mean square error.

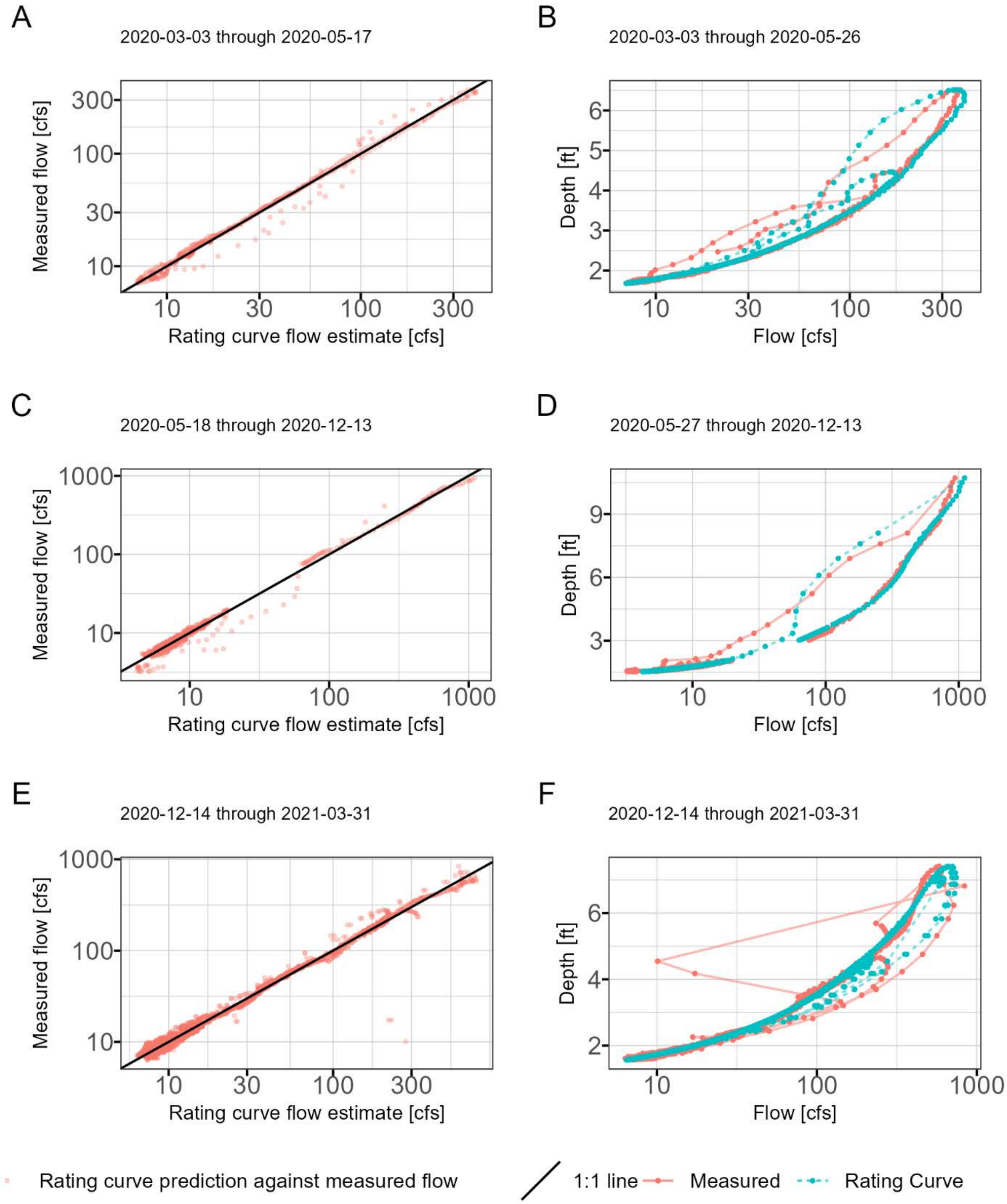

Figure 11. Rating curve flow estimates plotted against measured flows (subplots A, C, and E) and rating curves (subplots B, D, F) at surface water quality monitoring (SWQM) station 16396. cfs = cubic feet per second, ft = feet.

Table 4. Rating curve parameters and goodness-of-fit metrics for surface water quality monitoring (SWQM) station 16397.

SWQM station	Period	K	H_0	Z	NSE	nRMSE
SWQM-16397	2020-03-03 : 2020-01-20	0.0000000	-6.483048	9.207152	0.945	4.9
SWQM-16397	2020-01-21 : 2021-03-31	0.5618336	-1.166750	1.496724	0.329	12.7

Note - NSE = Nash-Sutcliffe efficiency, nRMSE = normalized root mean square error.

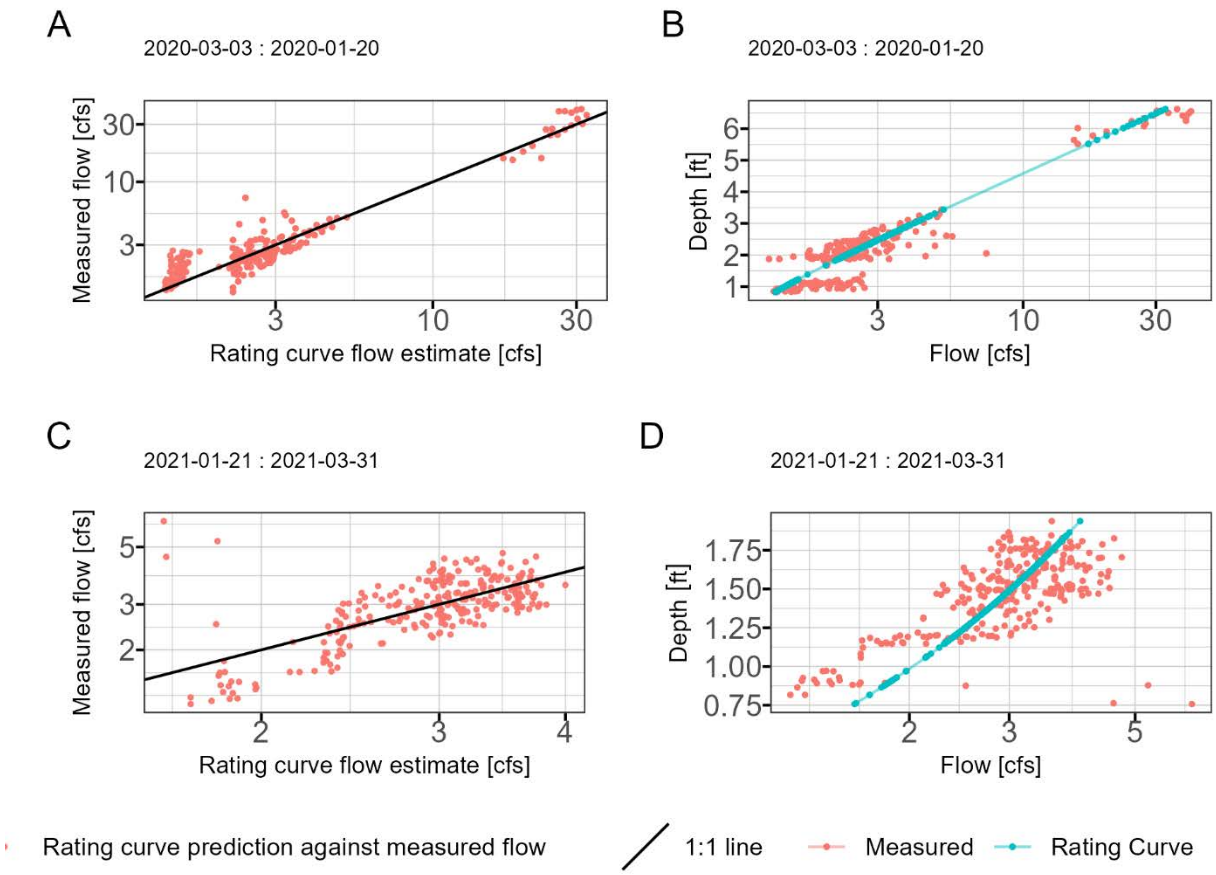

Figure 12. Rating curve flow estimates plotted against measured flows (subplots A and C) and rating curves (subplots B and D) at surface water quality monitoring (SWQM) station 16397. cfs = cubic feet per second, ft = feet.

Table 5. Rating curve parameters and goodness-of-fit metrics for surface water quality monitoring (SWQM) station 16882.

SWQM station	Period	K	a	n	x	NSE	nRMSE
SWQM-16882	2020-03-03 : 2020-10-22	4.520613	0.9598066	0.1515910	-0.5764236	0.967	5.5
SWQM-16882	2020-10-23 : 2020-12-09	4.429873	0.8771891	0.2634828	-3.6730616	0.926	8.0
SWQM-16882	2020-12-10 : 2021-03-31	5.057880	0.8393452	0.7575511	0.4721933	0.882	10.9

Note - NSE = Nash-Sutcliffe efficiency, nRMSE = normalized root mean square error.

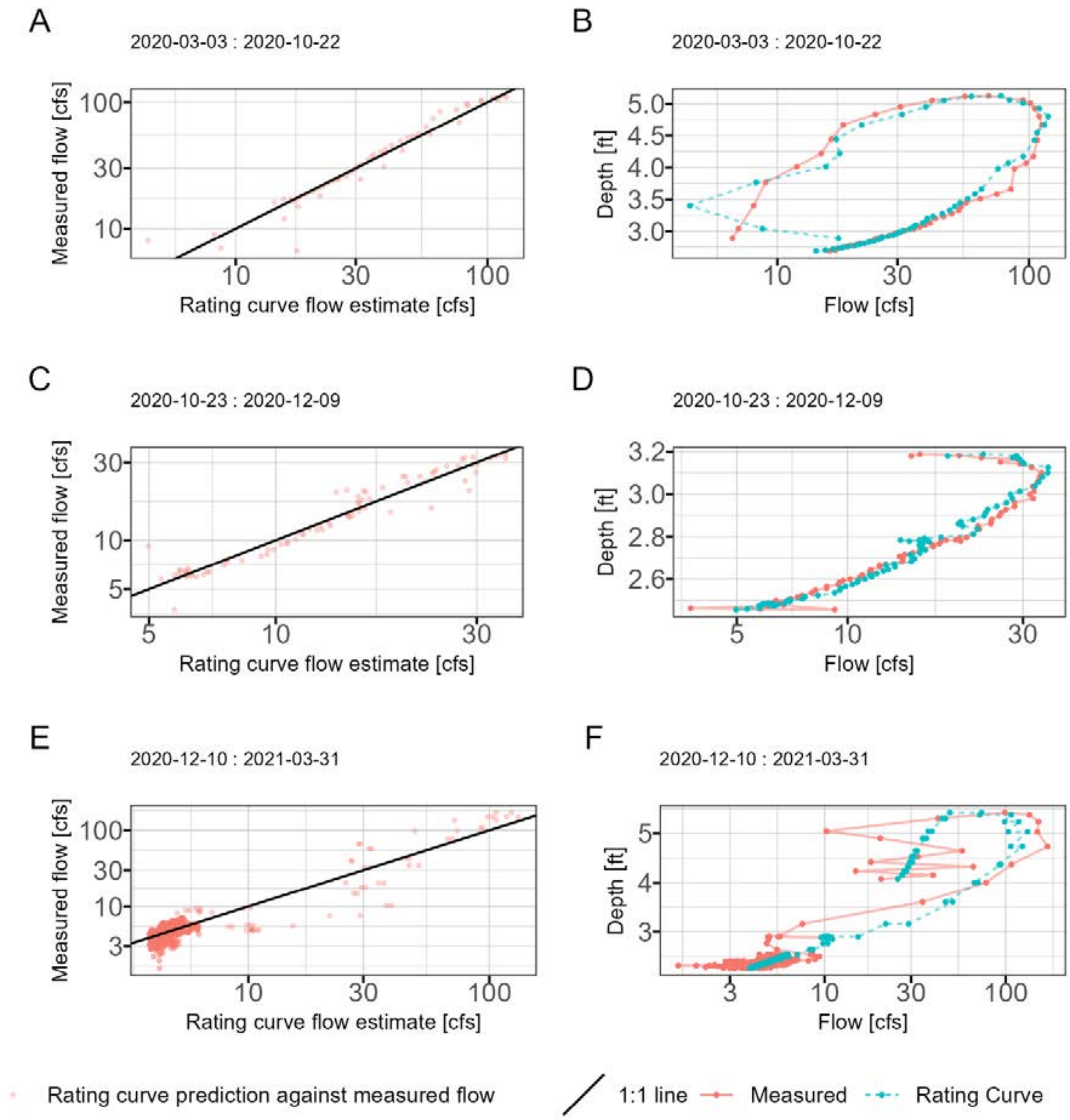

Figure 13. Rating curve flow estimates plotted against measured flows (subplots A, C, and E) and rating curves (subplots B, D, F) at surface water quality monitoring (SWQM) station 16882. cfs = cubic feet per second, ft = feet.

Mean Daily Streamflow

Figure 14 shows the rating curve estimated 15-minute streamflow values at each site with measured streamflow values overlaid. Improved temporal coverage of measured streamflows is desired. Equipment issues prevented the collection of more 15-minute streamflow values. Figure 15 shows streamflow aggregated to mean daily streamflow values with available measured mean daily streamflow values overlaid. Based on this graph, the rating curves perform well for estimating mean daily values, although there are very limited data available to assess against.

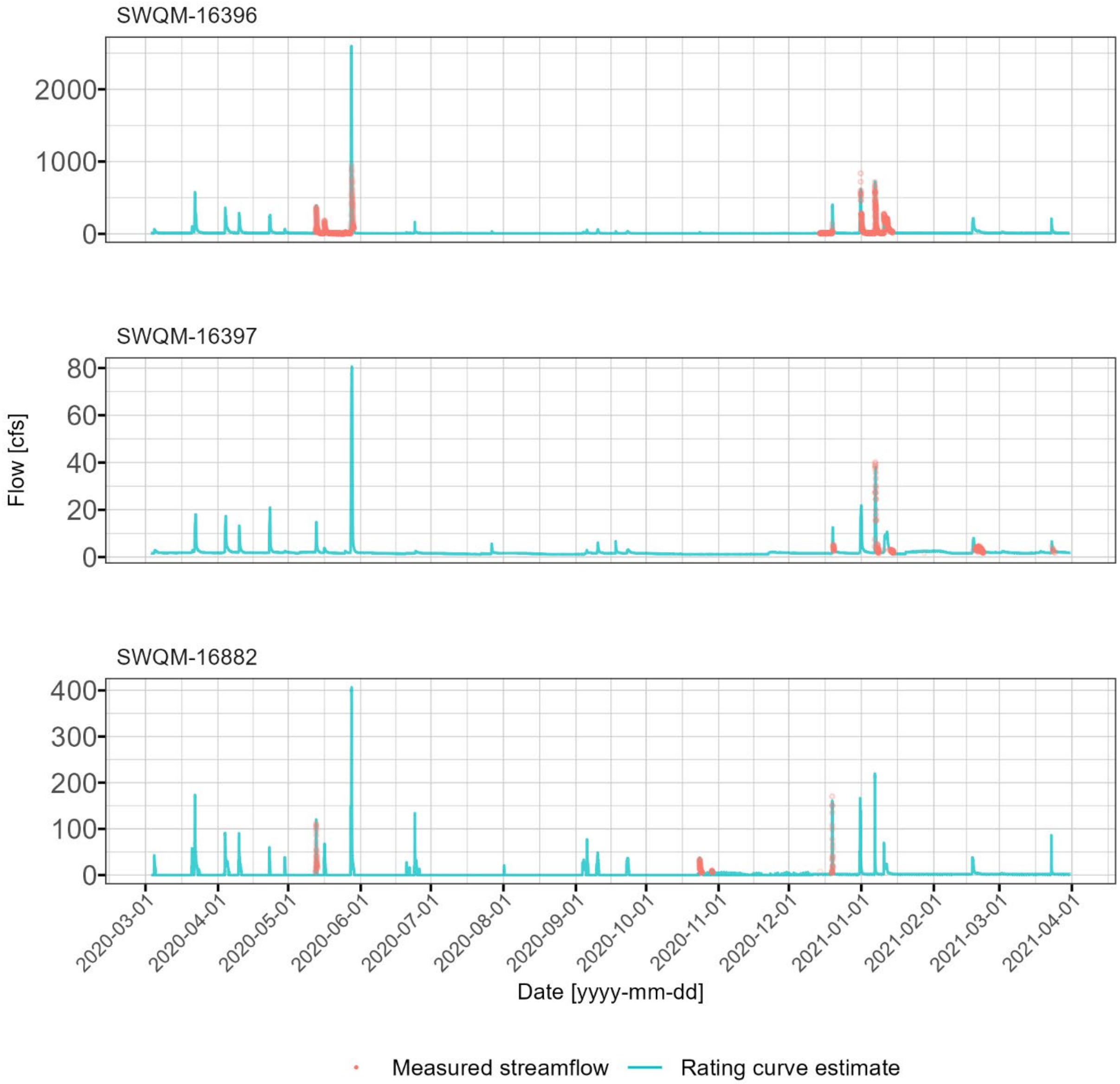

Figure 14. 15-minute streamflow measurements and rating curve estimates. cfs = cubic feet per second.

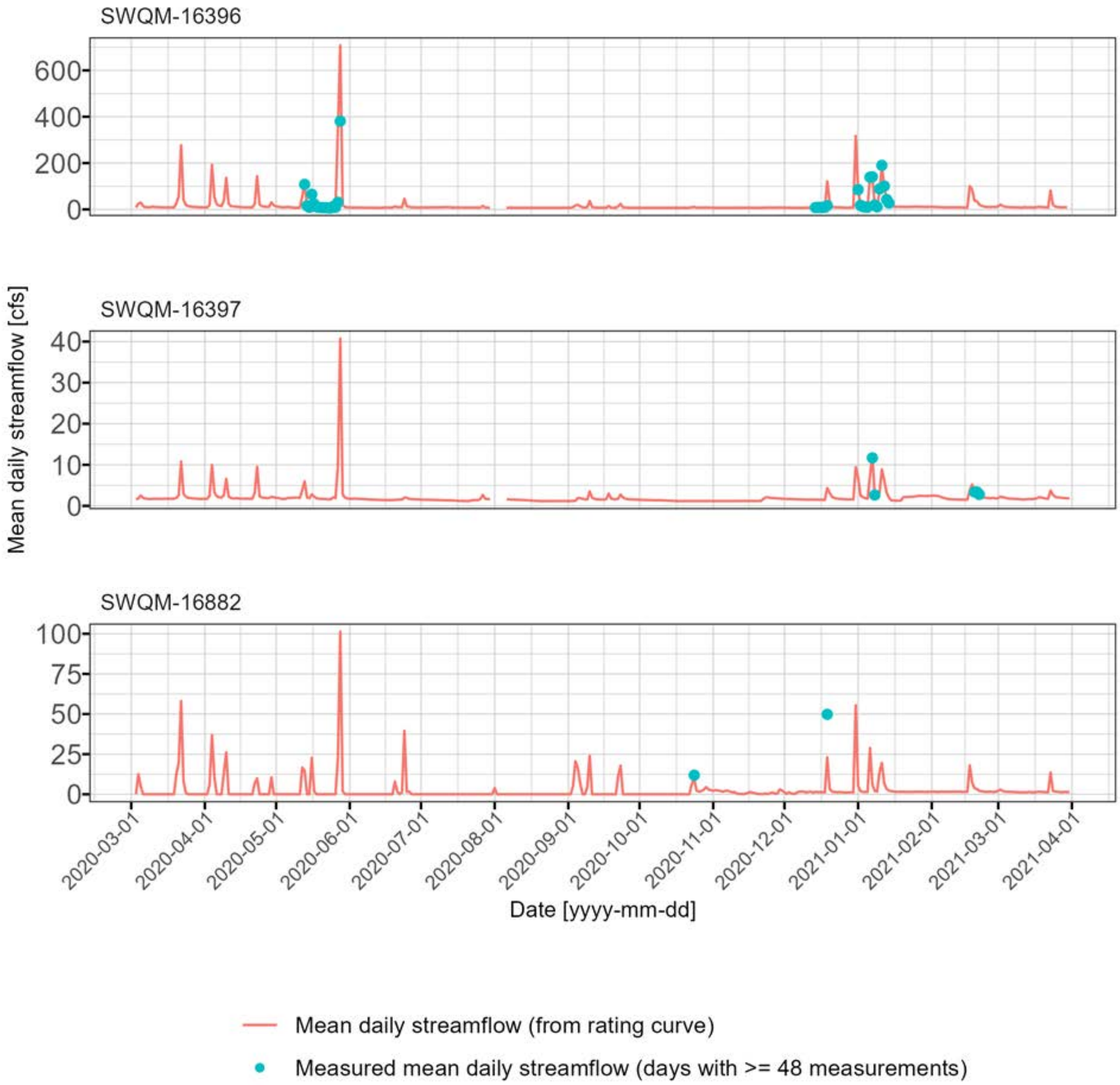

Figure 15. Mean daily streamflow measurements (days with greater than 48 15-minute measurements) and rating curve derived mean daily streamflow. cfs = cubic feet per second.

Streamflow Prediction

Mean daily discharges were subtracted from the streamflow record to better reflect naturalized flows. Figure 16 shows the naturalized streamflow hydrograph at each SWQM station. It is apparent that SWQM station 16882 is driven mostly by wastewater discharges with flashy responses to precipitation events.

All stations show a large peak on May 27, 2020 that coincides with an apparent moderate precipitation event. Considering the size of the precipitation event, the streamflow response would be expected to be proportional to the other streamflow peaks at the same station. While this could be a measurement error, we anticipate the streamflow measurement is correct. The Easterwood Airport weather station is approximately 3.2 miles south and east of the watershed. Based on news reports and anecdotal experience from the authors there was an extreme and localized storm with rain and substantial hail in portions of the watershed that suggest localized precipitation amounts greater than the approximately 1 inch reported at Easterwood Airport (Fiedler 2020 May 28).

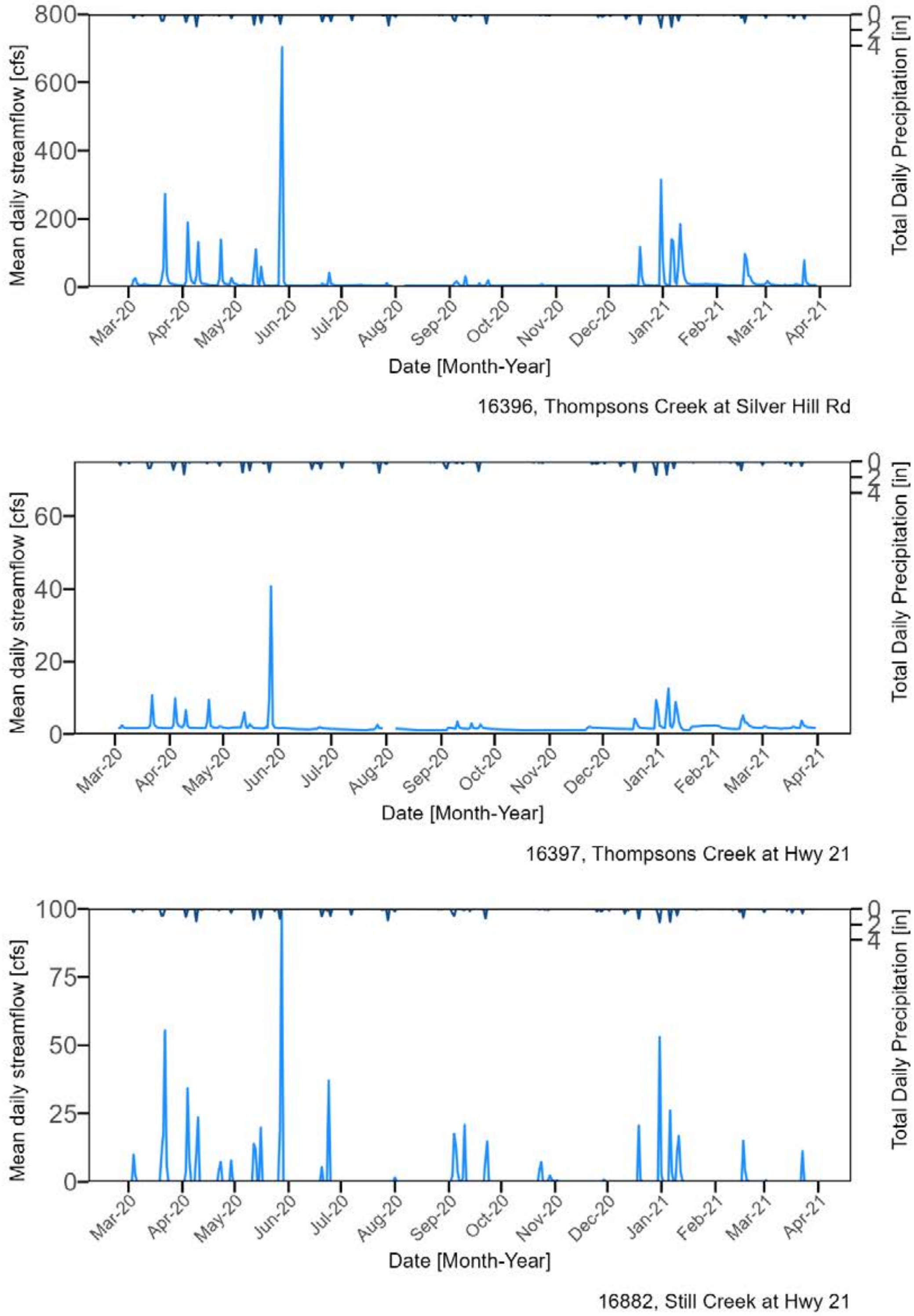

Figure 16. Hydrograph of naturalized mean daily streamflow at each surface water quality monitoring (SWQM) station with precipitation reported at Easterwood Airport depicted along the top axis. cfs = cubic feet per second, in = inches.

DAR

DAR from each USGS source gage generally underpredicts streamflows during low to moderate flow events with a handful of overpredictions when applied to SWQM stations 16396 and 16397 (Figure 17, Figure 18). Hydrographs indicate that the DARs miss some of the peak streamflow events over the year, indicating misalignment of precipitation-runoff events at the USGS gaged watersheds and the SWQM stations. DAR effectiveness at SWQM station 16882 suffers due to the prevalence of no-flow days (Figure 19). These plots indicate that DAR is unlikely to produce accurate daily flows or FDCs due to inaccuracy in peak flow timing and flow volume.

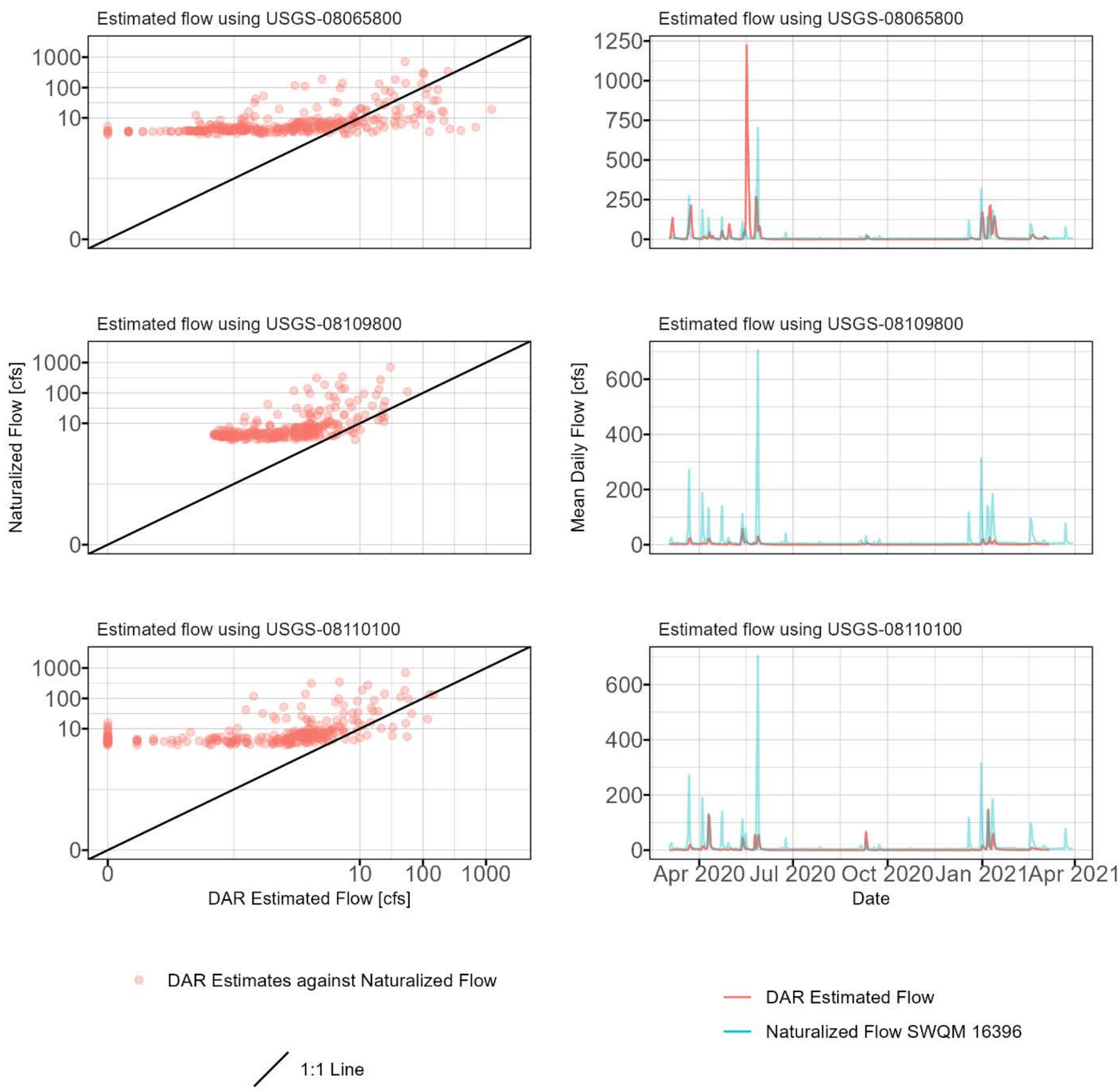

Figure 17. Drainage area ratio (DAR) predicted flows plotted against naturalized flows at surface water quality monitoring (SWQM) station 16396. USGS = United States Geological Survey, cfs = cubic feet per second.

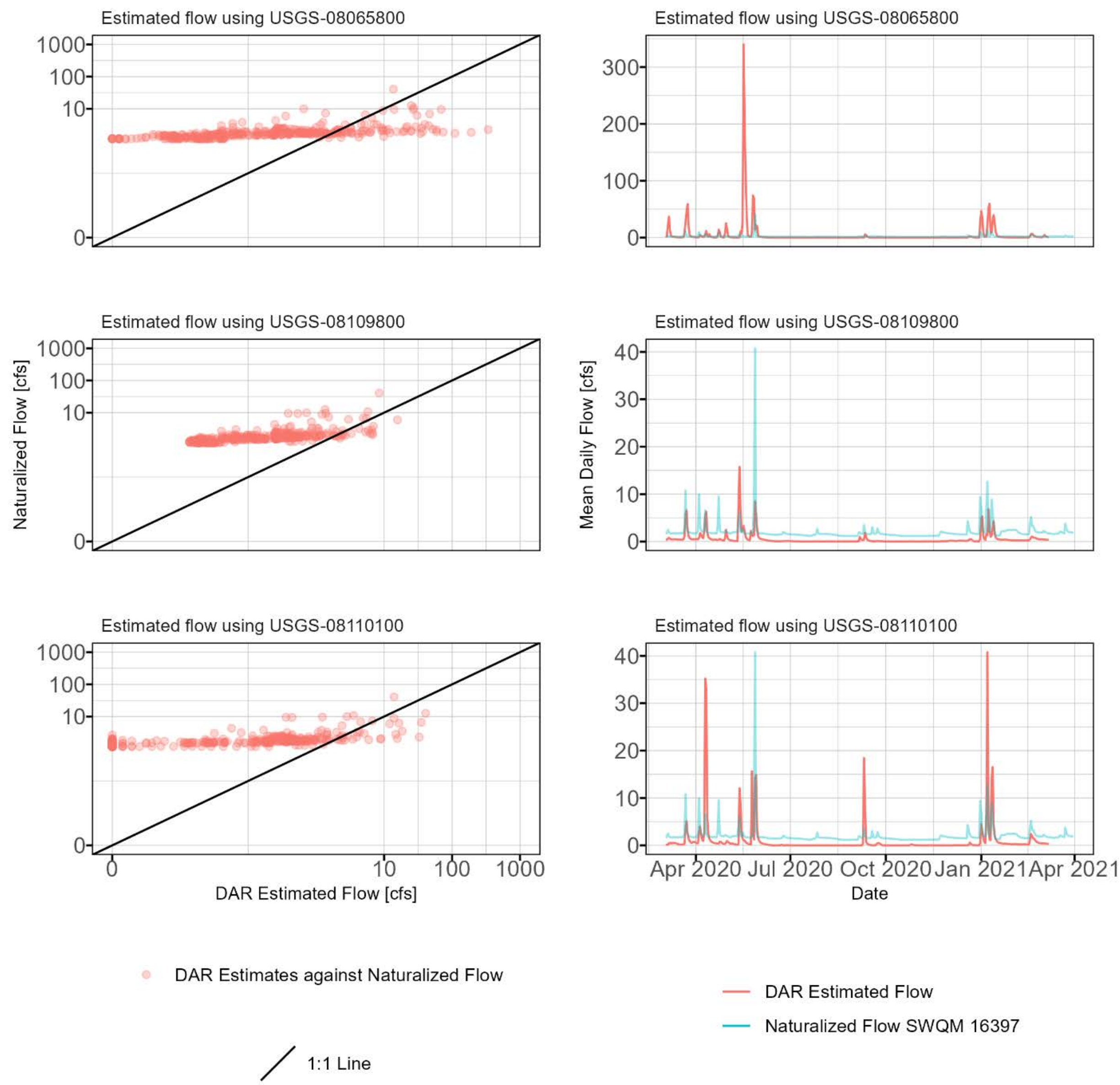

Figure 18. Drainage area ratio (DAR) predicted flows plotted against naturalized flows at surface water quality monitoring (SWQM) station 16397. USGS = United States Geological Survey, cfs = cubic feet per second.

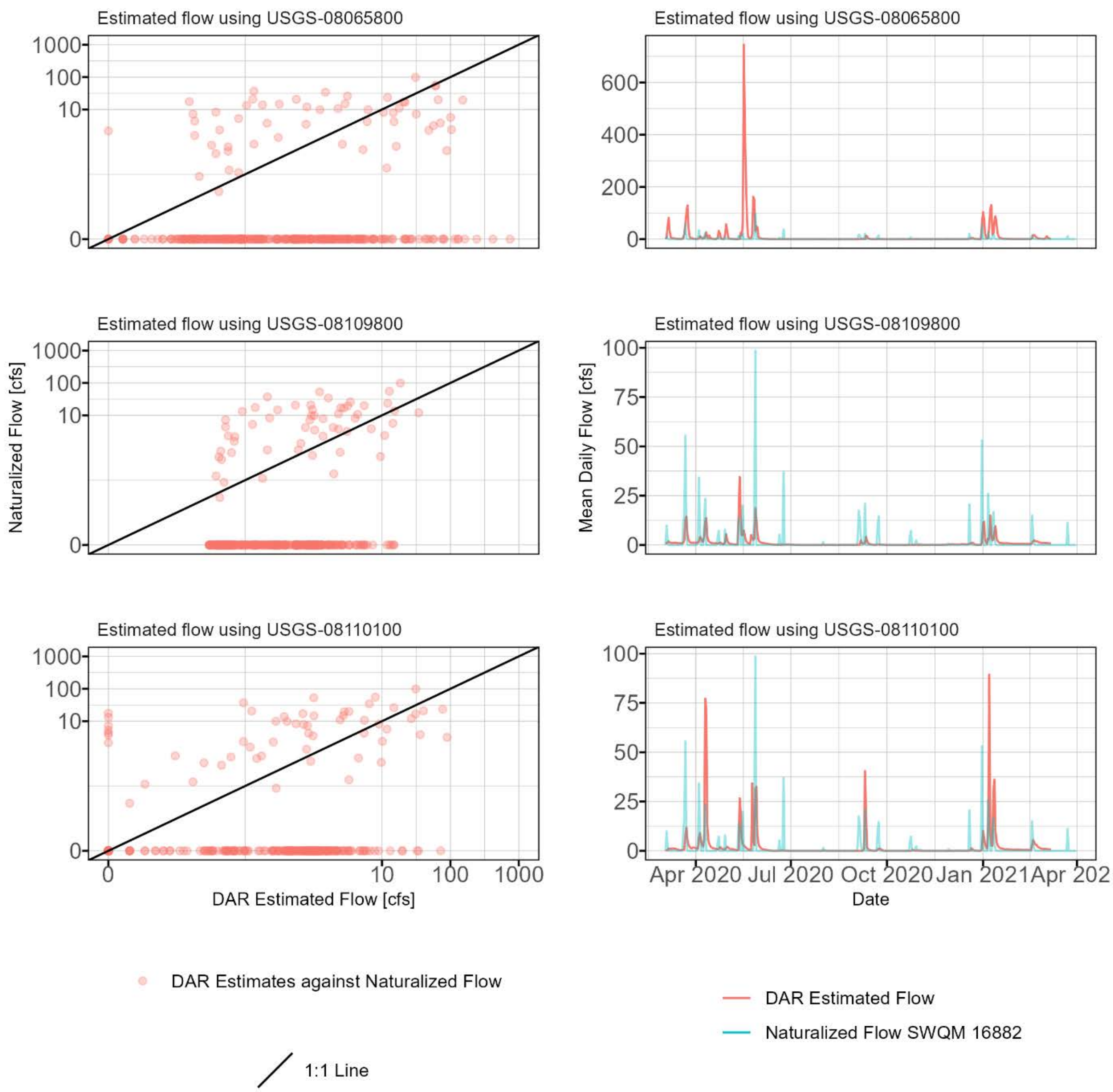

Figure 19. Drainage area ratio (DAR) predicted flows plotted against naturalized flows at surface water quality monitoring (SWQM) station 16882. USGS = United States Geological Survey, cfs = cubic feet per second.

Linear Regression

R squared values (0.43 to 0.64) indicate each of the linear regressions may perform well for predicting flows at each SWQM location (Table 6, Table 7, Table 8). Visual inspection of the measurement against prediction plots indicates improved performance compared to DAR (Figure 20). Linear regression does appear to systematically underpredict high flow events but with improved timing and distribution of error predictions at low flows.

Table 6. Linear regression coefficients at surface water quality monitoring (SWQM) station 16396.

	Estimate	Standard error	t value	p-value
(Intercept)	1.577	0.050	31.657	0.0000***
$Q_{08065800}$	0.289	0.044	6.632	0.0000***
$Q_{08109800}$	0.453	0.076	5.960	0.0000***
$Q_{08110100}$	0.534	0.054	9.841	0.0000***
$Q_{08065800,\ i-1}$	-0.128	0.041	-3.149	0.0018**
$Q_{08109800,\ i-1}$	-0.529	0.081	-6.527	0.0000***
$Q_{08110100,\ i-1}$	-0.348	0.053	-6.543	0.0000***

*Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 0.5082 on 356 degrees of freedom
Multiple R-squared: 0.6446, Adjusted R-squared: 0.6386
F-statistic: 107.6 on 356 and 6 DF, p-value: 0.0000

Table 7. Linear regression coefficients at surface water quality monitoring (SWQM) station 16397.

	Estimate	Standard error	t value	p-value
(Intercept)	0.862	0.019	44.355	0.0000***
$Q_{08065800}$	0.057	0.017	3.381	0.0008***
$Q_{08109800}$	0.162	0.030	5.481	0.0000***
$Q_{08110100}$	0.230	0.021	10.858	0.0000***
$Q_{08065800,\ i-1}$	-0.005	0.016	-0.298	0.7661
$Q_{08109800,\ i-1}$	-0.189	0.032	-5.967	0.0000***
$Q_{08110100,\ i-1}$	-0.149	0.021	-7.153	0.0000***

*Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 0.1982 on 356 degrees of freedom
Multiple R-squared: 0.6146, Adjusted R-squared: 0.6081
F-statistic: 94.62 on 356 and 6 DF, p-value: 0.0000

Table 8. Linear regression coefficients at surface water quality monitoring (SWQM) station 16882.

	Estimate	Standard error	t value	p-value
(Intercept)	0.094	0.059	1.588	0.1131
$Q_{08065800}$	0.193	0.053	3.640	0.0003***
$Q_{08109800}$	0.549	0.092	5.946	0.0000***
$Q_{08110100}$	0.507	0.066	7.693	0.0000***
$Q_{08065800,\ i-1}$	-0.145	0.049	-2.934	0.0036**
$Q_{08109800,\ i-1}$	-0.558	0.099	-5.660	0.0000***
$Q_{08110100,\ i-1}$	-0.421	0.065	-6.508	0.0000***

*Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 0.6185 on 362 degrees of freedom
Multiple R-squared: 0.4439, Adjusted R-squared: 0.4347
F-statistic: 48.17 on 362 and 6 DF, p-value: 0.0000

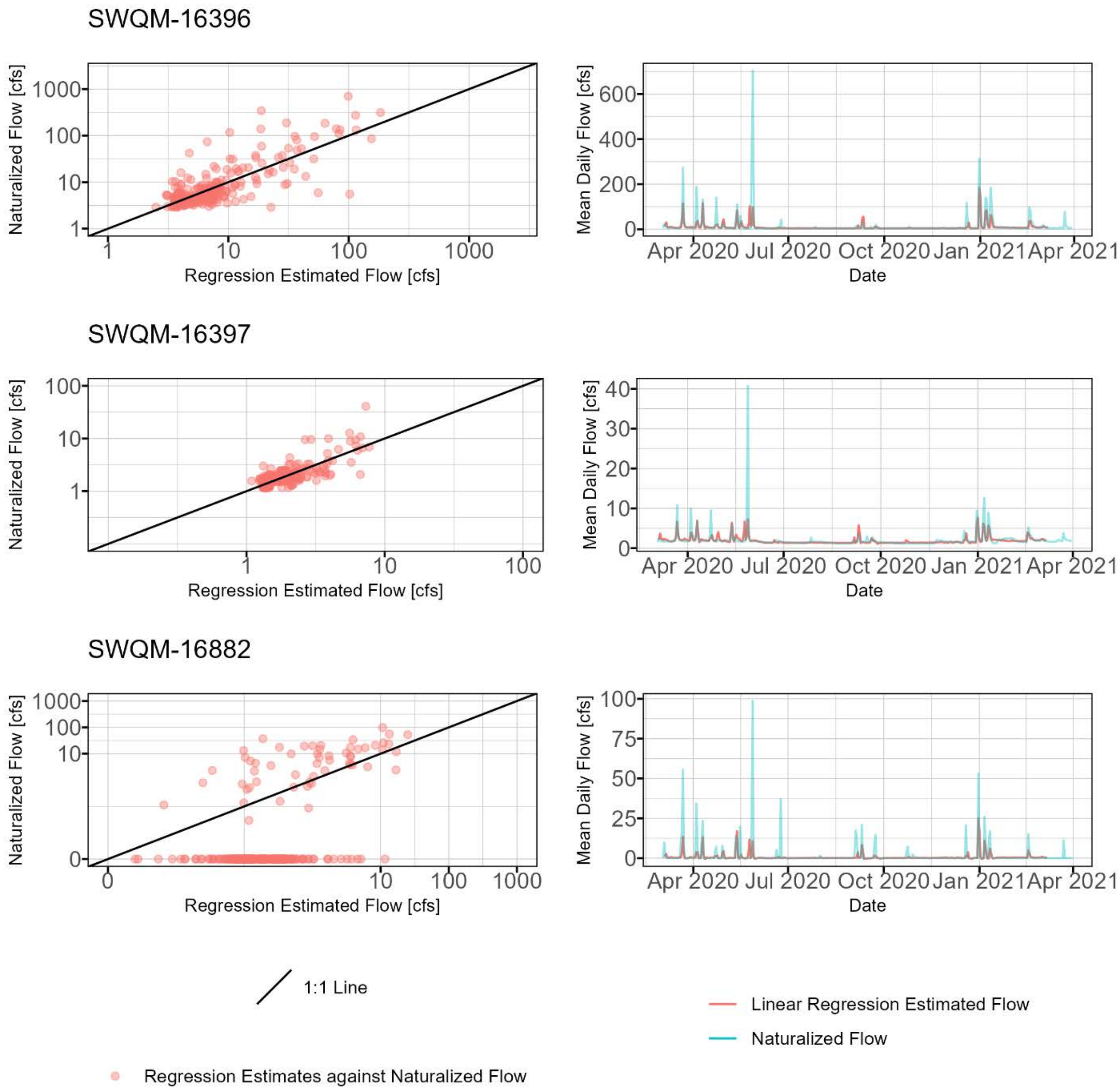

Figure 20. Linear regression predicted flows plotted against naturalized flows. SWQM = Surface Water Quality Monitoring.

Generalized Additive Models

Table 9 shows the smooth functions used in the GAM for SWQM station 16396. The first component is the linear intercept term; the second component describes the non-parametric smoothing functions. Following convention, smoothed terms are shown as *s(covariate)*. The relative variable influence can be interpreted from the estimated degrees of freedom (EDF). EDF of one or lower indicates a generally linear function between the covariate and the dependent variable. The higher the EDF, the less smooth the function. EDF of zero indicates the covariate function was penalized to zero and has no influence on the model. The p-values are calculated from the test statistic (t-value or F-value) and for smooth terms, describe the likelihood that the smooth function is equal to zero. The p-value calculated for smooth function should be considered approximate and does not describe the significance of the covariate term or its contribution to the model. The model also describes how much of the deviance from the mean is described by the model. Extremely high deviance explained might indicate an overfit model that is not generalizable, while a low deviance explained indicates a model that does not fit the data well. For SWQM station 16396, the GAM explains approximately 0.801 of the deviance in log daily streamflow.

The non-parametric model terms do not return a single model coefficient because they are composed of different effect sizes at different values of the model term. A visual representation of the effect of each model term is provided in Figure 21. These plots visually depict how much a given value of the individual model term affects the mean streamflow when all other variables are held equal (this is sometimes described as a partial effect plot).

The GAM at SWQM station 16396 indicates an expected positive effect of increasing (log-transformed) P and $P_{sum,3}$ (3-day cumulative precipitation) on streamflow. $T_{mean,3}$ shows an expected negative impact on streamflows. The expected impact of month of year is largest in the spring. The effects of T and H on streamflow are zero. When using GAMs to explain the impact of covariates on the dependent variable, more effort would be spent ensuring the model terms have sensible impacts. In this case, our primary concern is the accuracy of the model predictions, so limited effort is spent on the interpretability of the independent variable model terms. However, general visual inspection of the terms indicates plausible covariate effects.

The SWQM station 16397 GAM explains approximately 0.714 of streamflow deviance (Table 10). As expected, precipitation variables positively effect streamflow (Figure 22). T had no effect, but $T_{mean,5}$ had an expected negative effect on streamflow.

A two-part model was utilized for SWQM station 16882. The first model predicts the probability of zero streamflow and used the same covariates in the GAM models for SWQM stations 16396 and 16397. The second model was a GAM fit to days that there was positive naturalized streamflow. There were only 58 days of positive streamflow, which limited the available degrees of freedom to fit the model. Due to the limited degrees of freedom, the streamflow GAM at SWQM station 16882 was simplified by removing H as a predictor variable. Approximately 54% of the deviance in presence/absence of streamflow was described by the logistic regression GAM and almost 60% of the deviance in the amount of streamflow was described by the second GAM (Table 11, Table 12). Figure 23 displays the partial effect of each model term for predicting flow (value = 1) or no flow (value = 0).

Neither partial effect plots at SWQM station 16882 indicate unexpected variable influence (Figure 23, Figure 24).

Table 9. Generalized area model (GAM) terms for surface water quality monitoring (SWQM) station 16396.

Component	Term	Estimate	Standard error	t-value	p-value
A. Parametric coefficients	(Intercept)	2.034	0.038	53.804	***
Component	Term	EDF	Reference degrees of freedom	F-value	p-value
B. Smooth terms	s(P)	2.638	9.000	13.269	***
	s(T)	0.000	9.000	0.000	
	s(P_{lag})	0.001	9.000	0.000	
	s($P_{sum,3}$)	5.372	9.000	22.621	***
	s($T_{mean,5}$)	4.383	9.000	3.662	***
	s(H)	0.000	9.000	0.000	
	s(M)	5.959	8.000	6.926	***

Deviance explained 0.801; N: 387
Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1
EDF = estimated degrees of freedom

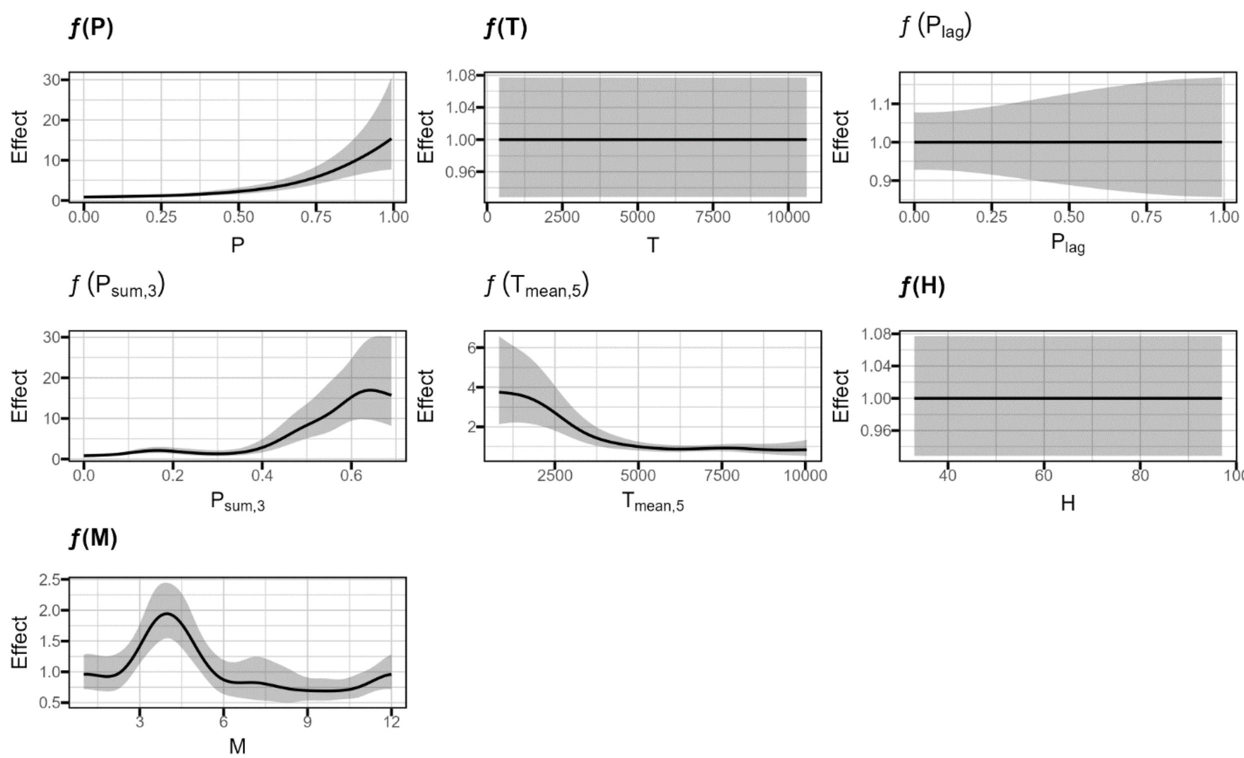

Figure 21. Generalized additive model (GAM) partial effects plot at surface water quality monitoring (SWQM) station 16396.

Table 10. Generalized additive model (GAM) terms for surface water quality monitoring (SWQM) station 16397.

Component	Term	Estimate	Standard error	t-value	p-value
A. Parametric coefficients	(Intercept)	0.618	0.018	34.198	***
Component	Term	EDF	Reference degrees of freedom	F-value	p-value
B. Smooth terms	s(P)	3.314	9.000	4.929	***
	s(T)	0.000	9.000	0.000	
	$s(P_{lag})$	4.609	9.000	22.979	***
	$s(P_{sum,3})$	0.000	9.000	0.000	
	$s(T_{mean,5})$	0.894	9.000	0.617	*
	s(H)	0.000	9.000	0.000	
	s(M)	6.266	8.000	7.702	***

Deviance explained 0.714; N: 388
Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < ''
< 1
EDF = estimated degrees of freedom

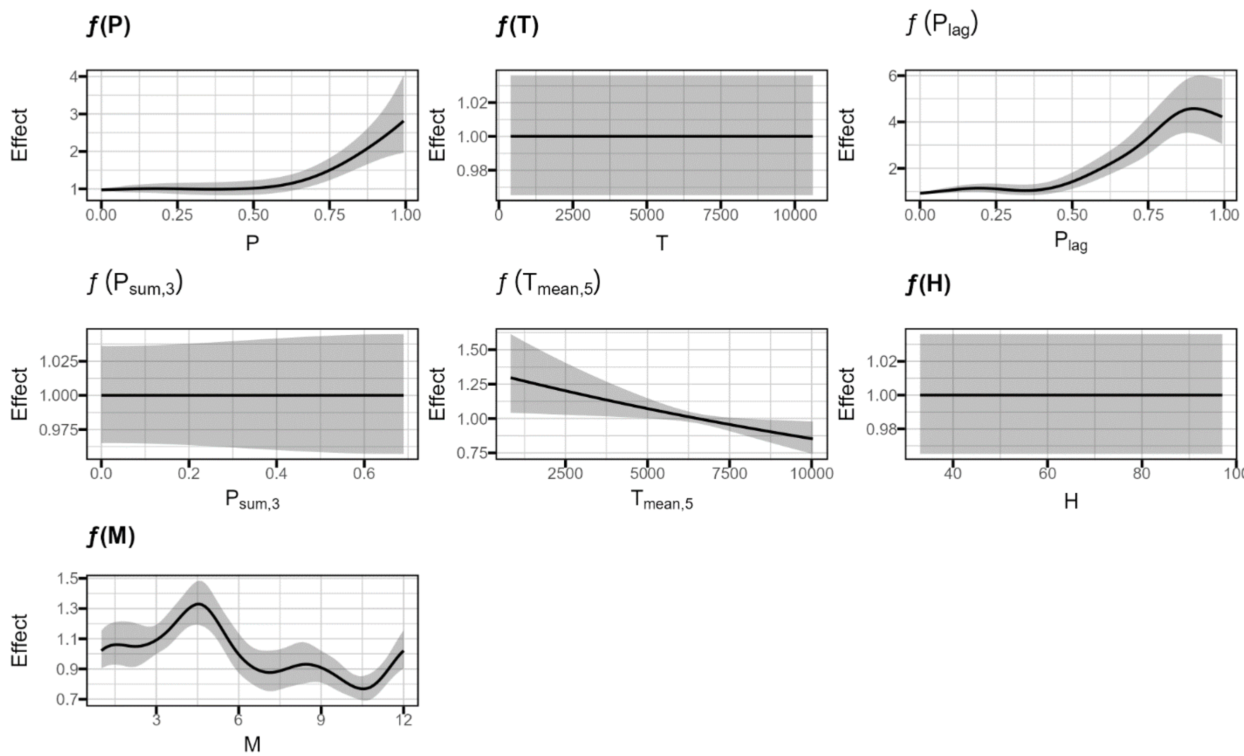

Figure 22. Generalized additive model (GAM) partial effects plot at surface water quality monitoring (SWQM) station 16397.

Table 11. Logistic regression generalized additive model (GAM) terms for surface water quality monitoring (SWQM) station 16882

Component	Term	Estimate	Standard error	t-value	p-value
A. Parametric coefficients	(Intercept)	-2.945	0.297	-9.925	***
Component	Term	EDF	Reference degrees of freedom	F-value	p-value
B. Smooth terms	s(P)	0.979	9.000	38.972	***
	s(T)	1.270	9.000	2.530	
	s(P_{lag})	0.000	9.000	0.000	
	s($P_{sum,3}$)	0.979	9.000	39.434	***
	s($T_{mean,5}$)	2.376	9.000	16.594	***
	s(H)	0.000	9.000	0.000	
	s(M)	5.470	8.000	32.363	***

Deviance explained 0.537; N: 394
Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1
EDF = estimated degrees of freedom

Table 12. Generalized additive model (GAM) terms for surface water quality monitoring (SWQM) station 16882.

Component	Term	Estimate	Standard error	t-value	p-value
A. Parametric coefficients	(Intercept)	1.842	0.119	15.463	***
Component	Term	EDF	Reference degrees of freedom	F-value	p-value
B. Smooth terms	s(P)	0.964	9.000	3.051	***
	s(T)	0.203	9.000	0.030	
	s(P_{lag})	0.002	9.000	0.000	*
	s($P_{sum,3}$)	2.079	9.000	3.863	***
	s($T_{mean,5}$)	3.927	9.000	1.501	**
	s(M)	2.219	8.000	2.287	***

Deviance explained 0.597; N: 58
Significance codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < ''
< 1
EDF = estimated degrees of freedom

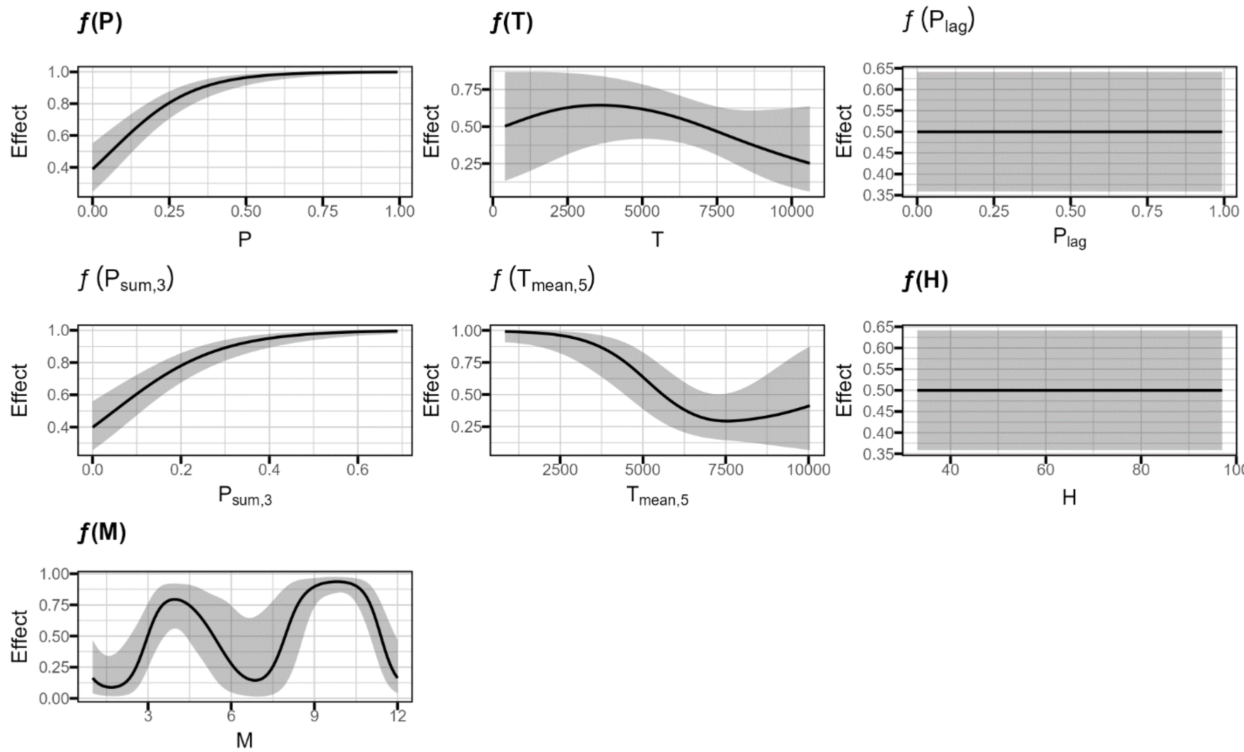

Figure 23. Logistic regression partial effects plot at SWQM station 16882.

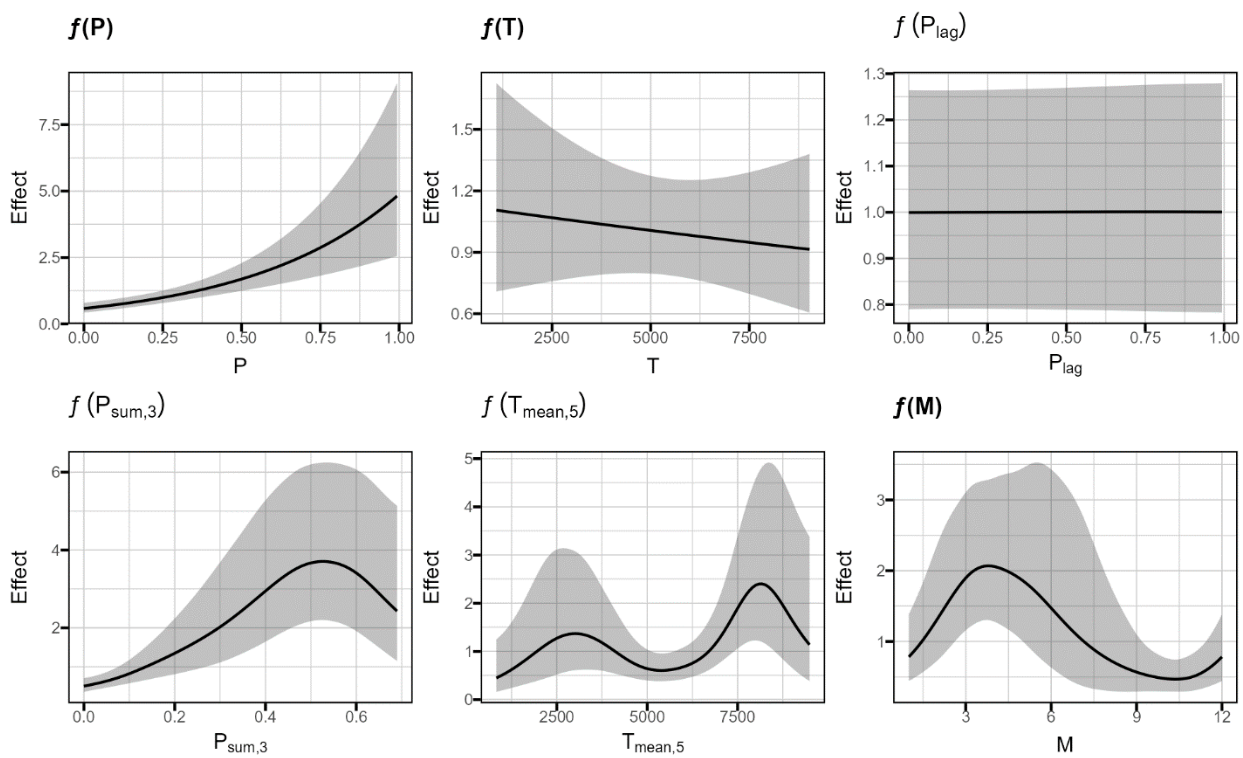

Figure 24. Generalized additive model (GAM) partial effects plot at surface water quality monitoring (SWQM) station 16882.

Visual inspection of GAM model fits against measured values indicates well-distributed model errors for all sites (Figure 25). There is slightly high variance in model prediction errors at higher flow volumes. The plots of measured and predicted flow over time indicate good alignment of peak flows. There appears to be less systematic underprediction of peak flows compared to linear regression models. The GAM at SWQM station 16882 does indicate prediction of streamflow on a handful of days that no flow was measured and prediction of no flow on days that flow was measured. However, the performance is generally good at predicting presence or absence of streamflow.

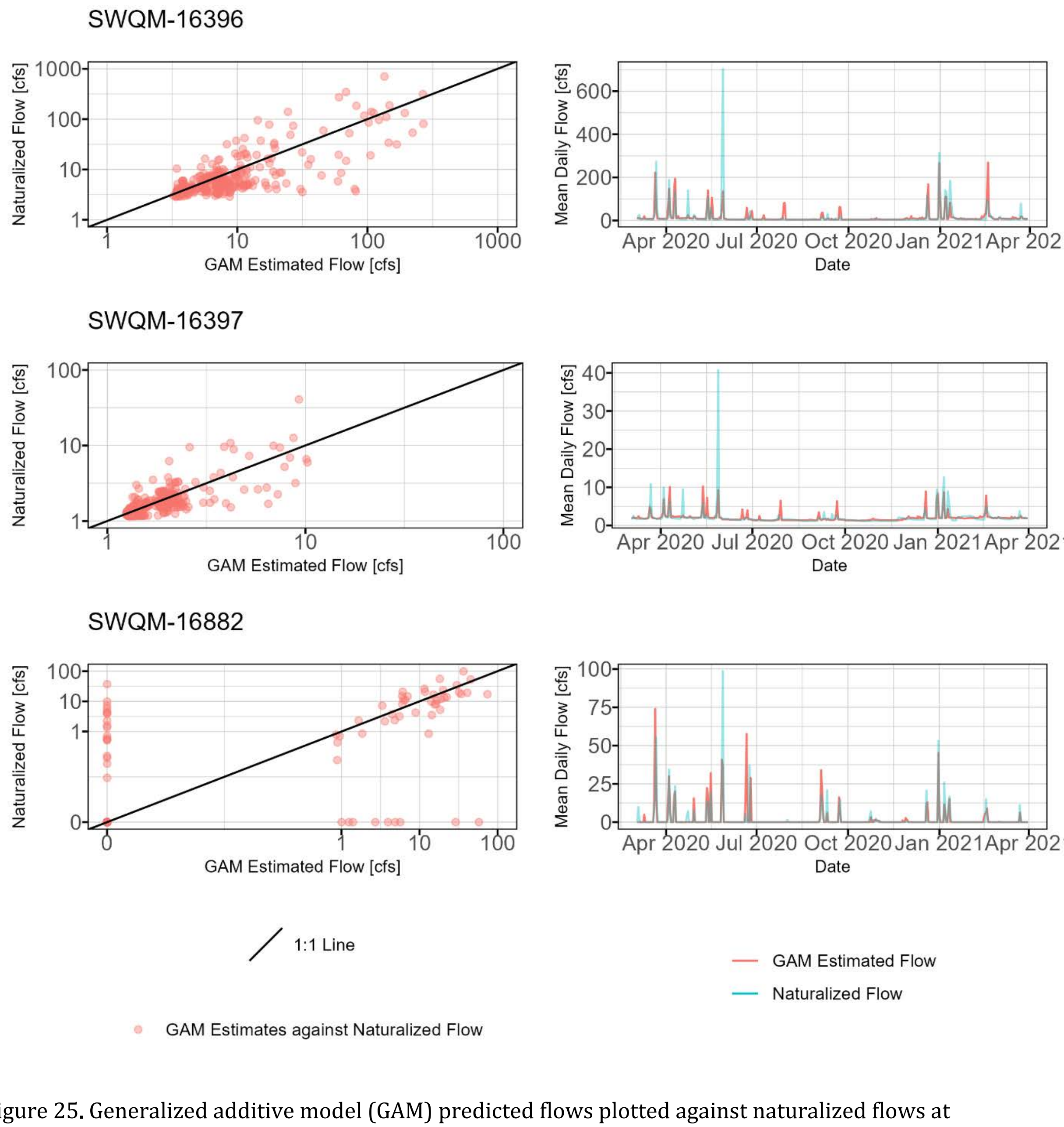

Figure 25. Generalized additive model (GAM) predicted flows plotted against naturalized flows at each surface water quality monitoring (SWQM) station. cfs = cubic feet per second.

Method Evaluation

Goodness-of-fit metric indicates that the DAR approach produces poor results in the Thompsons Creek watershed with NSE and KGE value ranging from -7.55 to 0.263 (Table 13). Visual inspection of plots also indicates that DAR produces systematically biased flow predictions at low and high flows (Figure 17, Figure 18, Figure 19). Goodness-of-fit metrics indicate linear regressions show improvement over the DAR approach. The low KGE score, which is less sensitive to the accuracy of high streamflows, at SWQM station 16882 is indicative of the linear regression's inability to model streamflow values of 0 cfs.

The NSE and KGE values ranged from 0.316 to 0.585 for the GAM approach. One should note that a 1:1 interpretation of KGE to NSE is not appropriate, as using the mean flow as a predictor would result in a KGE score of approximately -0.41, indicating values over -0.41 are an improvement over the mean model (Knoben et al. 2019). Overall, these simplified metrics indicate GAMs and linear regression have similar calibrated performance. The relatively higher KGE scores for GAMs are likely indicative of better performance at lower flow ranges. The two-step hurdle model process used for SWQM station 16882 was a substantial improvement over linear regression.

Table 13. Goodness-of-fit measures for results of the drainage area ratio (DAR), linear regression, and generalized additive model (GAM) methods applied at each surface water quality monitoring (SWQM) station.

Method	NSE			KGE		
	SWQM-16396	SWQM-16397	SWQM-16882	SWQM-16396	SWQM-16397	SWQM-16882
DAR 08065800	-0.265	-5.48	-2.50	-0.0841	-7.55	-6.48
DAR 08109800	0.249	-1.07	0.256	-0.364	0.005	0.0907
DAR 08110100	0.263	-1.42	0.0347	-0.215	0.031	0.233
Linear regression	0.518	0.389	0.470	0.209	0.298	0.025
GAM	0.425	0.316	0.478	0.458	0.386	0.585

Note - NSE = Nash-Sutcliffe efficiency, KGE = Kling-Gupta efficiency.

Validation

Cross validation was conducted with both linear regression and GAMs on all the sites to provide insight for how the methods perform on out-of-sample data. For both methods, the nRMSE was similar across sites (Figure 26, Figure 27, Figure 27). NSE was generally higher for linear regression than GAMs, although GAMs showed higher variance (the best fit GAMs had higher scores, and the worst fit GAMs had much lower scores). This suggests that GAMs were much more sensitive to the data used for fitting. Given the variance of the data used to fit the models, it is possible that more data would result in improved GAM performance. KGE scores indicate that linear regression was more suitable than GAMs for SWQM stations 16396 and 16397. The hurdle GAM model shows improved KGE scores at SWQM station 16882. This is attributable to the ability of the hurdle model to predict 0 cfs streamflows at the site.

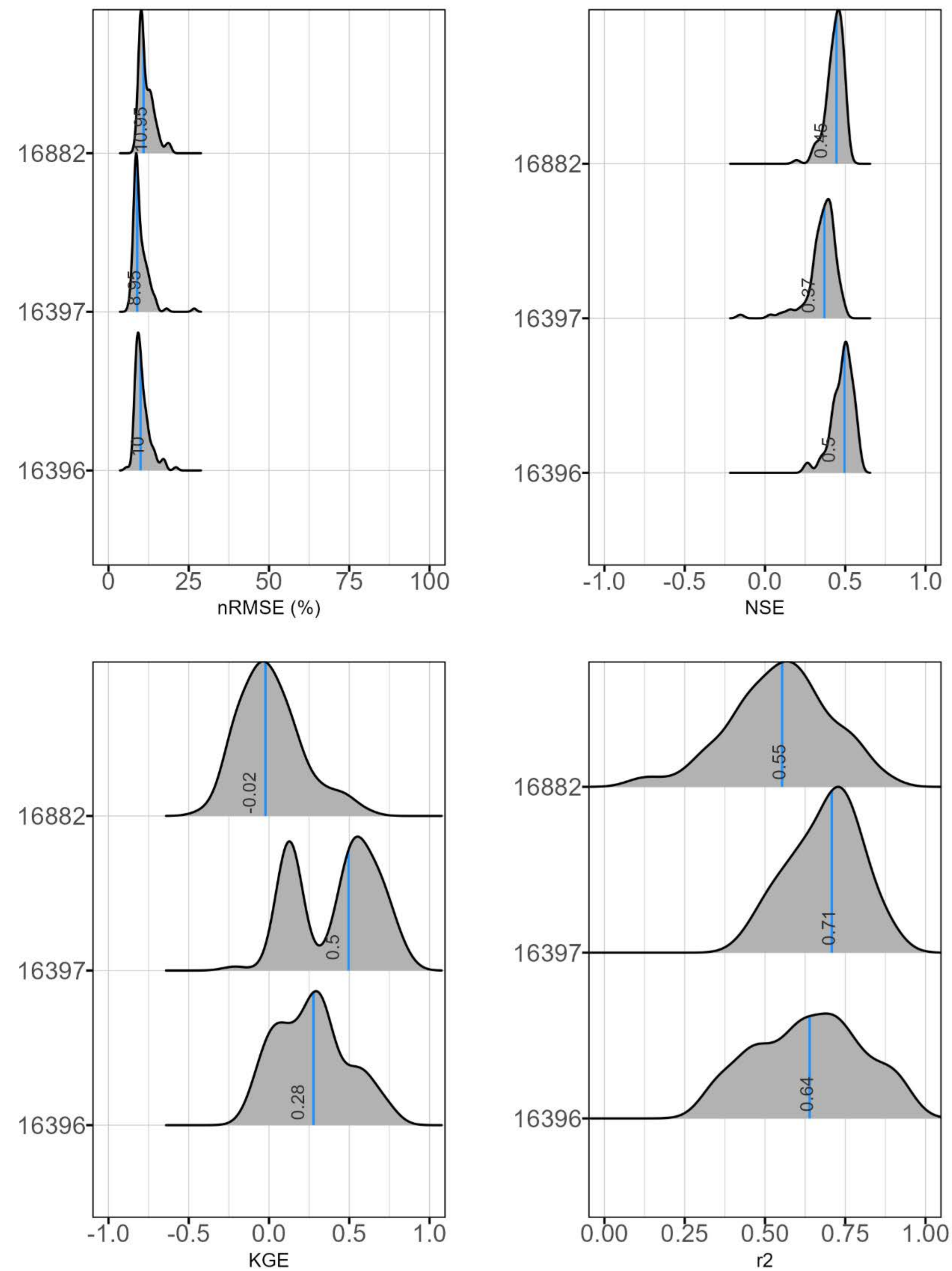

Figure 26. Goodness-of-fit metric distribution from Monte-Carlo cross validation of linear regression at each surface water quality monitoring (SWQM) station; blue line indicates the median value. nRMSE = normalized root mean square error, NSE = Nash-Sutcliffe efficiency, KGE = Kling-Gupta efficiency, r2 = r-square.

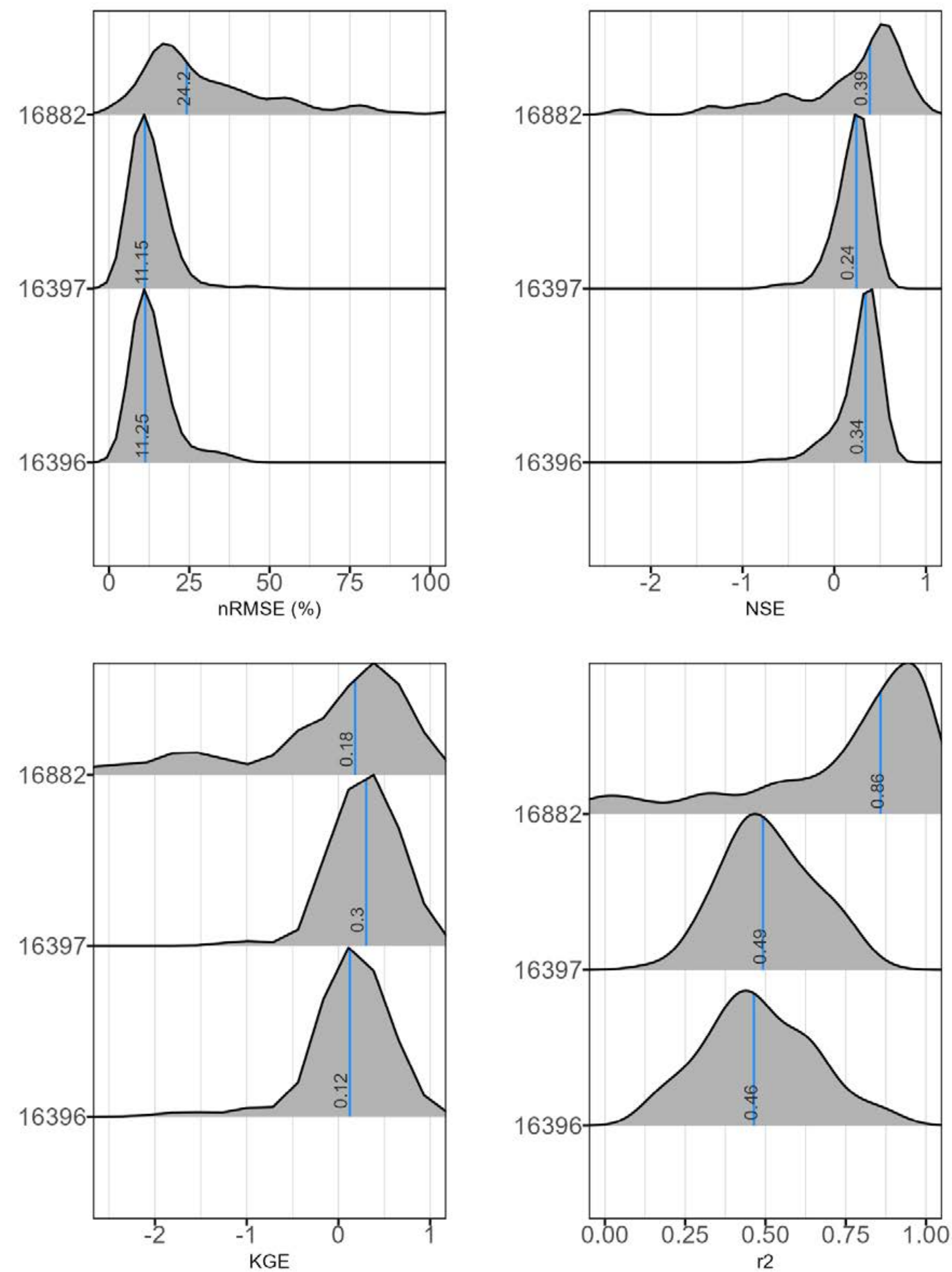

Figure 27. Goodness-of-fit metric distribution from Monte-Carlo cross validation of generalized additive model (GAM) at each surface water quality monitoring (SWQM) station; blue line indicates the median value. nRMSE = normalized root mean square error, NSE = Nash-Sutcliffe efficiency, KGE = Kling-Gupta efficiency, r2 = r-square.

Flow Duration Curves

Although daily streamflow prediction is an important component of TMDL development, load allocations require an accurate estimation of flow exceedance percentiles. For most TMDLs in Texas, the pollutant load allocation is calculated for flows at the median of high flow exceedances (typically at the 5% flow exceedance). Visual inspection of the measured and predicted FDCs provides additional information about the performance and appropriate methods for TMDL development.

At SWQM station 16396, the GAM-produced FDC is most consistent with the measured FDC at high flows (low proportion of days flow exceeded; Figure 28). However, the GAM generally overpredicts under moderate flow conditions. The linear regression model performs well from around 15% flow exceedance through 100% flow exceedance. The implication of these results is that if accuracy of predicting high flow volumes is desired, the GAM model might be more appropriate. However, if accuracy at the rest of the FDC is more valued, then the linear regression approach is appropriate. Similar results are observed for SWQM station 16397 (Figure 29). At SWQM station 16882, the GAM hurdle model clearly performs better than linear regression (Figure 30). It is notable that the GAM slightly overpredicts the proportion of days with no streamflow.

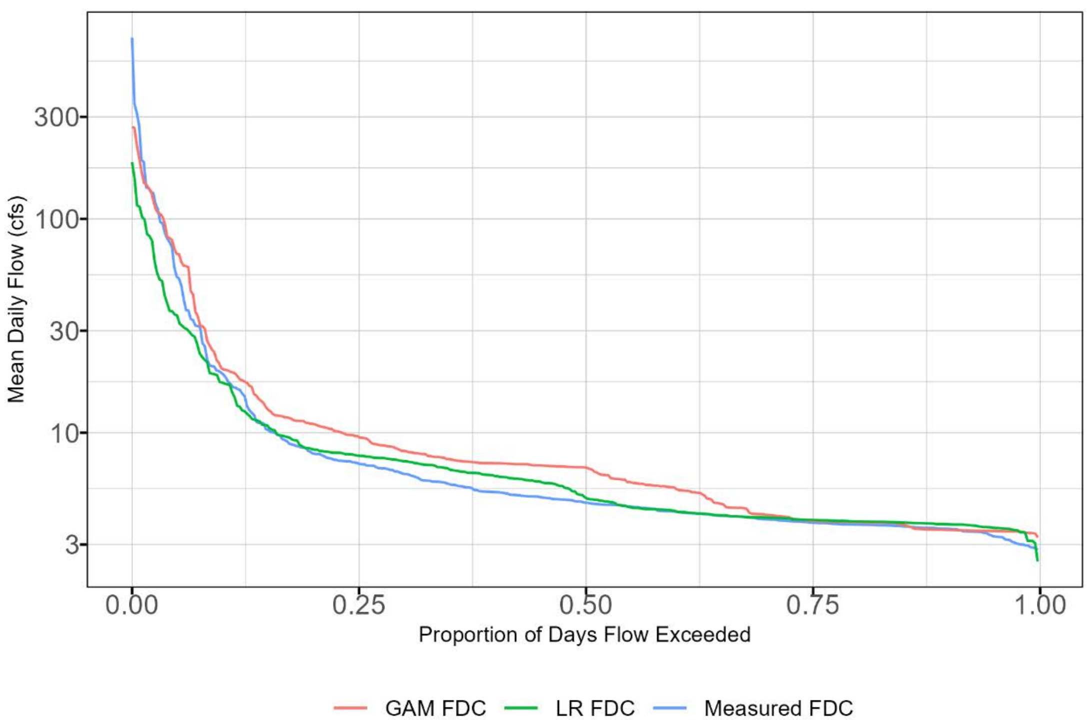

Figure 28. Measured and predicted flow duration curves (FDCs) at surface water quality monitoring (SWQM) station 16396. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

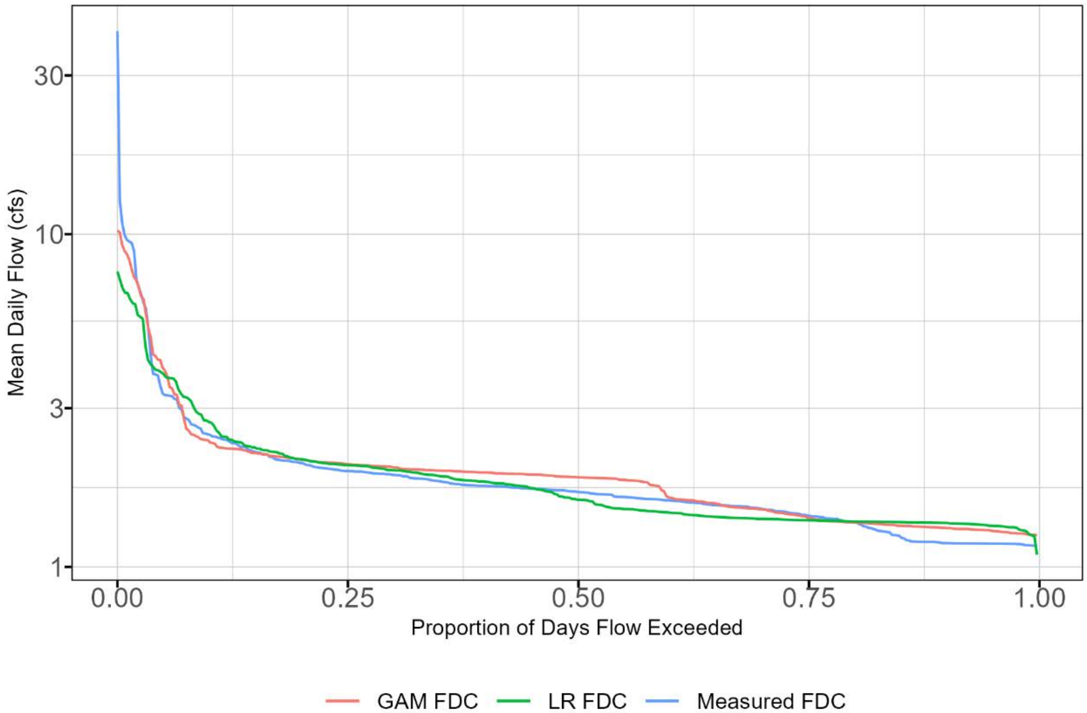

Figure 29. Measured and predicted flow duration curves (FDCs) at surface water quality monitoring (SWQM) station 16397. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

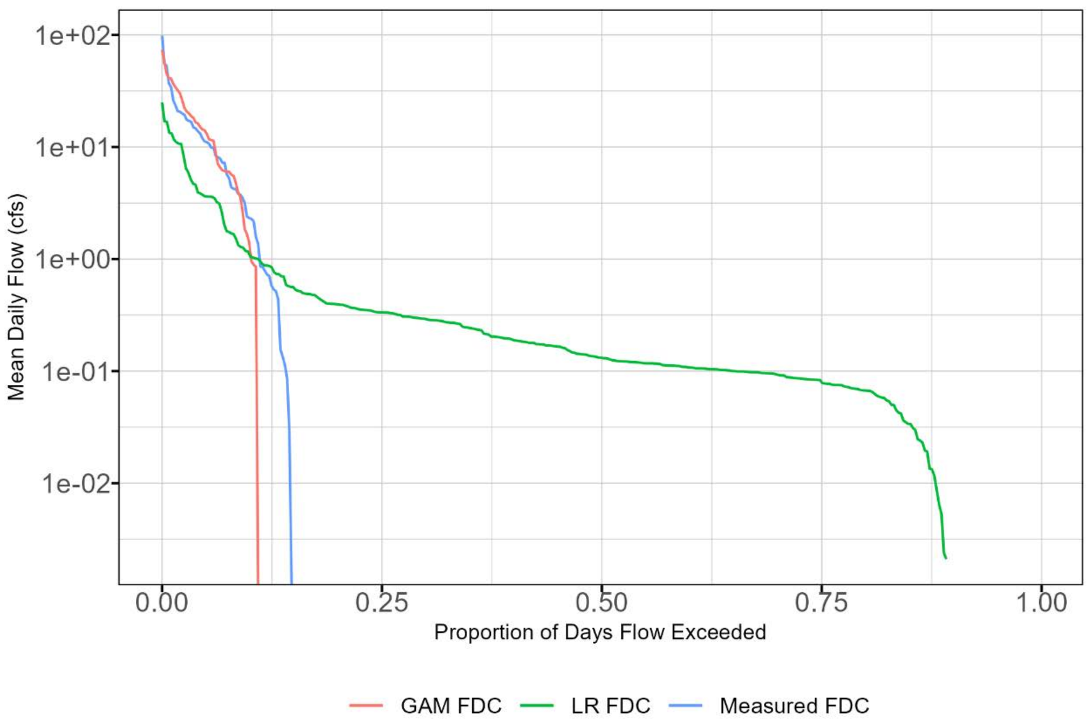

Figure 30. Measured and predicted flow duration curves (FDCs) at surface water quality monitoring (SWQM) station 16882. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

Period of Record Estimated Mean Daily Streamflow

Mean daily streamflow was predicted using linear regression and GAMs from January 1, 2011 through March 30, 2021. Although there is no way to validate the accuracy of the data produced outside of the period that streamflow data was collected, the data was plotted to ensure reasonableness of predictions. The hydrographs indicate that the period of collected flow data had relatively low precipitation compared to previous years (Figure 31). The lack of consistent high flow and high precipitation days might result in biased predictions by the GAM method across sites. At SWQM station 16396, the GAM predicts unrealistically high flows during three precipitation events (including the Hurricane Harvey rainfall event in 2017; Figure 31). Cross-checking these flows with flows at the USGS stream gage on the Brazos River (08108700) near Bryan indicates the predicted flows are higher than the measured flows on the Brazos. Considering the relative watershed sizes, this is a highly unlikely scenario. At SWQM station 16397, the GAM method again predicts higher streamflows than linear regression (Figure 32). The predicted flows are more reasonable when compared to measured flows. Both linear regression and GAM appear to predict flow peaks that align with precipitation events at SWQM stations 16396 and 16397 as would be anticipated. At SWQM station 16882, the relative magnitudes of linear regression and GAM predicted flows are similar, but the timing of the highest flow events is not aligned (Figure 33). It appears the GAM provides streamflow prediction peaks that are better aligned with precipitation events at SWQM station 16882.

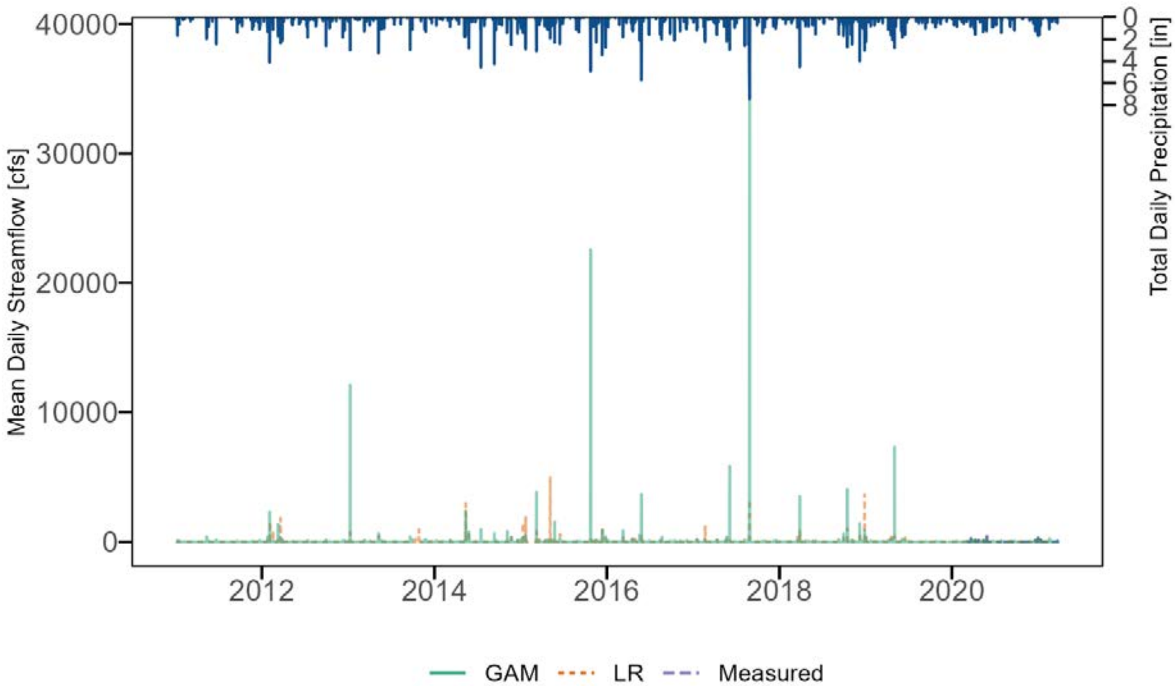

Figure 31. Predicted and measured hydrographs at surface water quality monitoring (SWQM) station 16396; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second, in = inches.

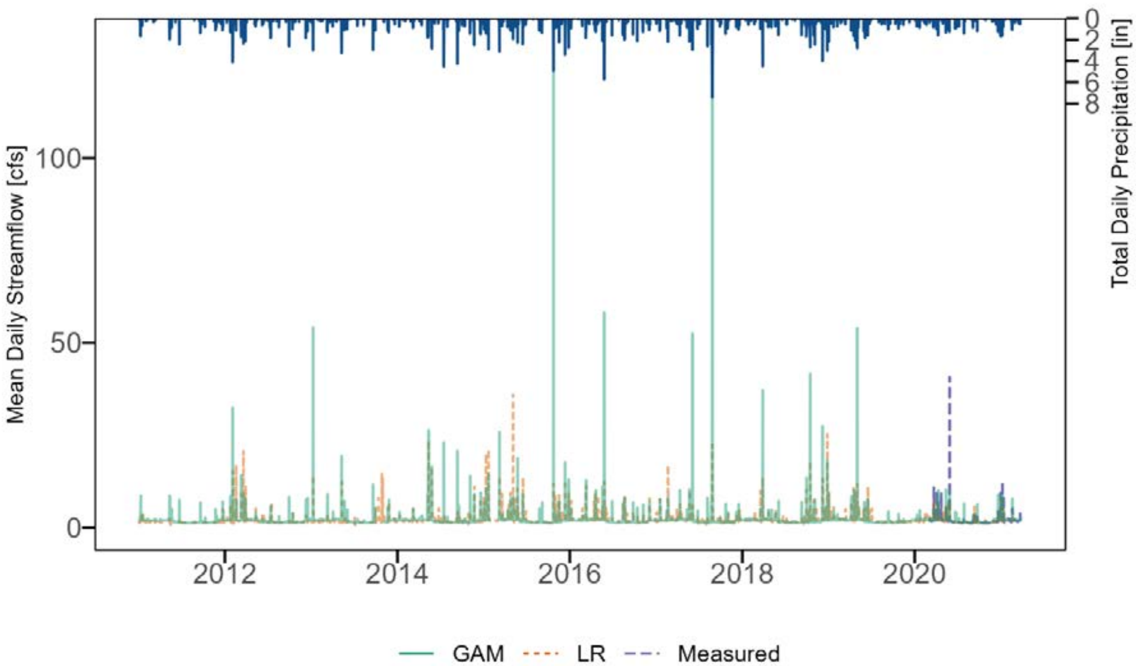

Figure 32. Predicted and measured hydrographs at surface water quality monitoring (SWQM) station 16397; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second, in = inches.

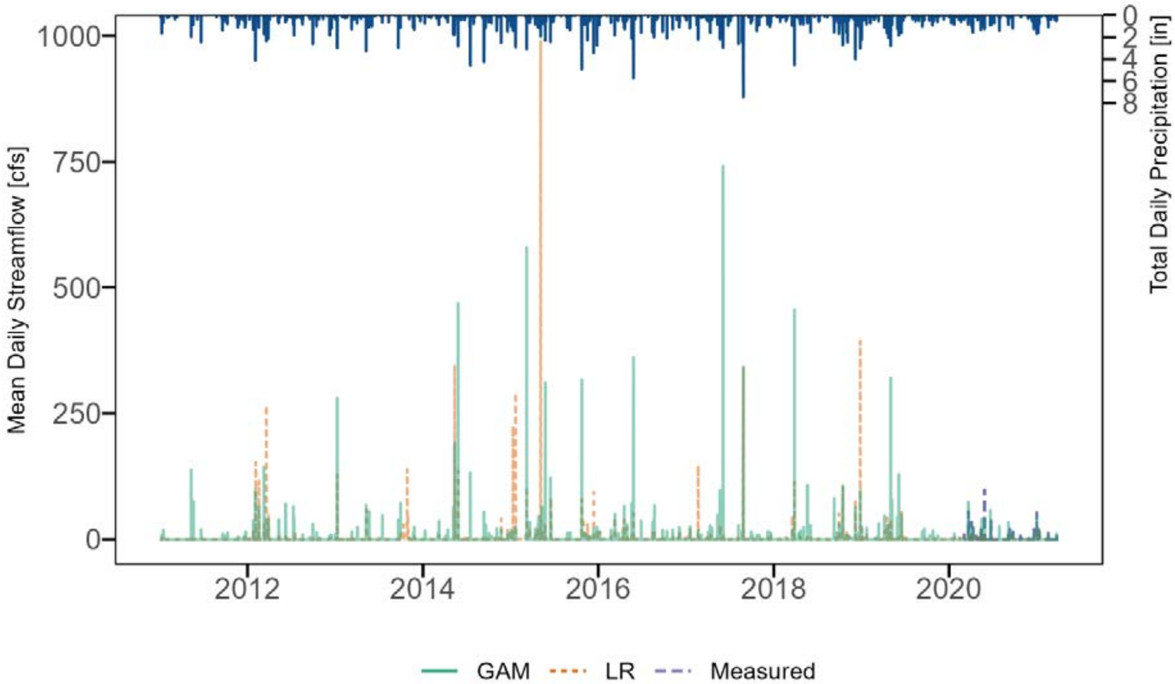

Figure 33. Predicted and measured hydrographs at surface water quality monitoring (SWQM) station 16882; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second, in = inches.

Naturalized FDCs were also plotted for each SWQM station using linear regression and GAM predicted flows. At SWQM stations 16396 and 16397, the difference between linear regression and GAM outputs is small from approximately 10% through 100% exceedance (Figure 34, Figure 35). At SWQM station 16396, the difference becomes more substantial from zero to 5% flow exceedance, with estimates of 74 cfs and 44 cfs by GAM and linear regression respectively at 5% exceedance. At SWQM station 16397, the FDCs are much closer together, with only the very extreme values at the end of the curve differing between the two methods. At SWQM station 16882, linear regression fails to predict no-flow days that are properly predicted by the GAM hurdle model (Figure 36). The GAM also appears much closer to the measured FDC for the time period that we measured flows, suggesting the GAM hurdle model is the appropriate method for SWQM station 16882.

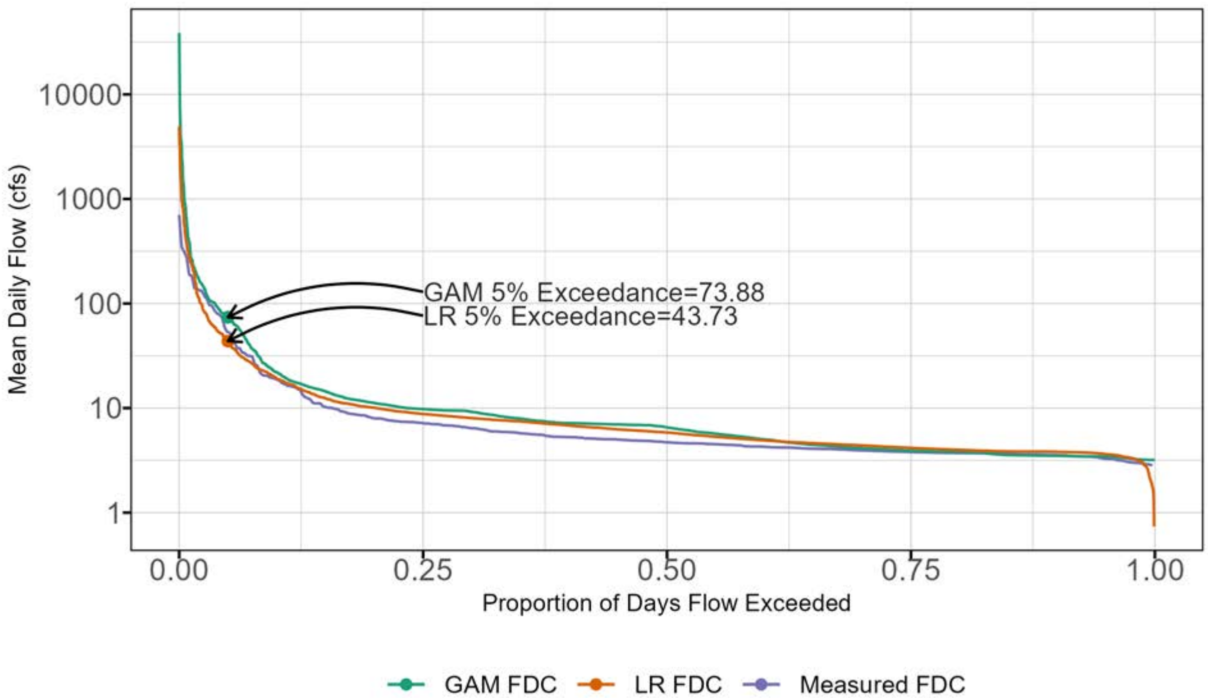

Figure 34. Predicted naturalized flow duration curves (FDCs) at surface water quality monitoring (SWQM) station 16396; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

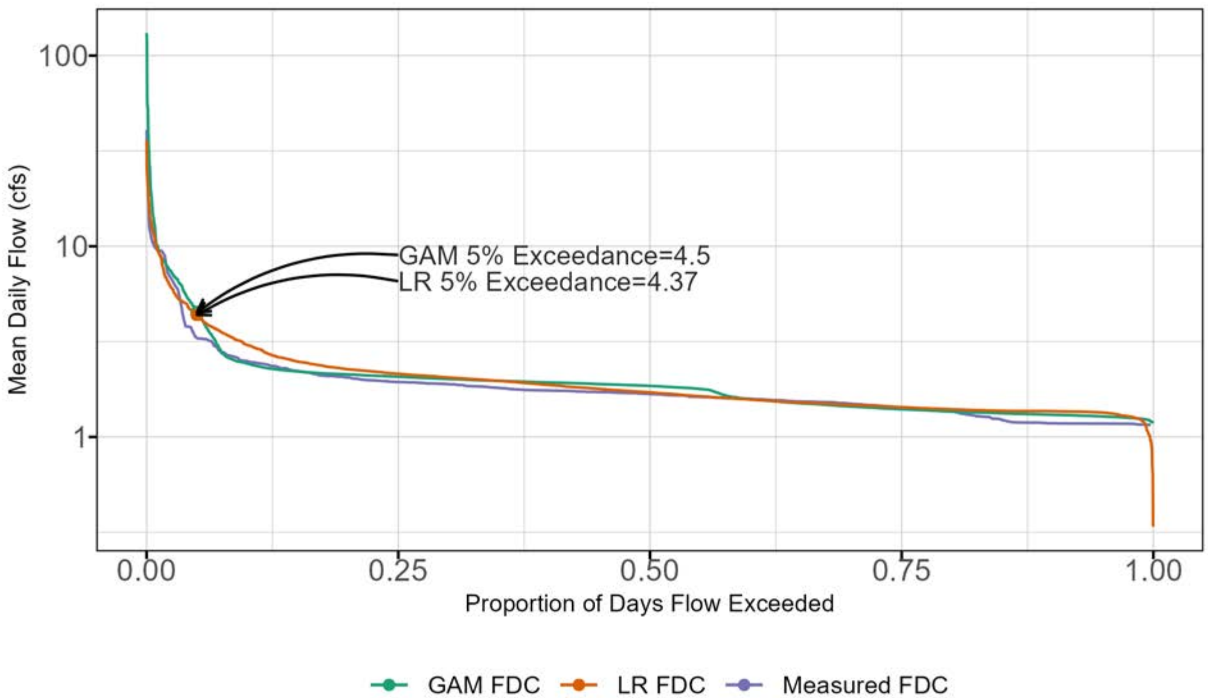

Figure 35. Predicted naturalized flow duration curves (FDCs) at surface water quality monitoring (SWQM) station 16397; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

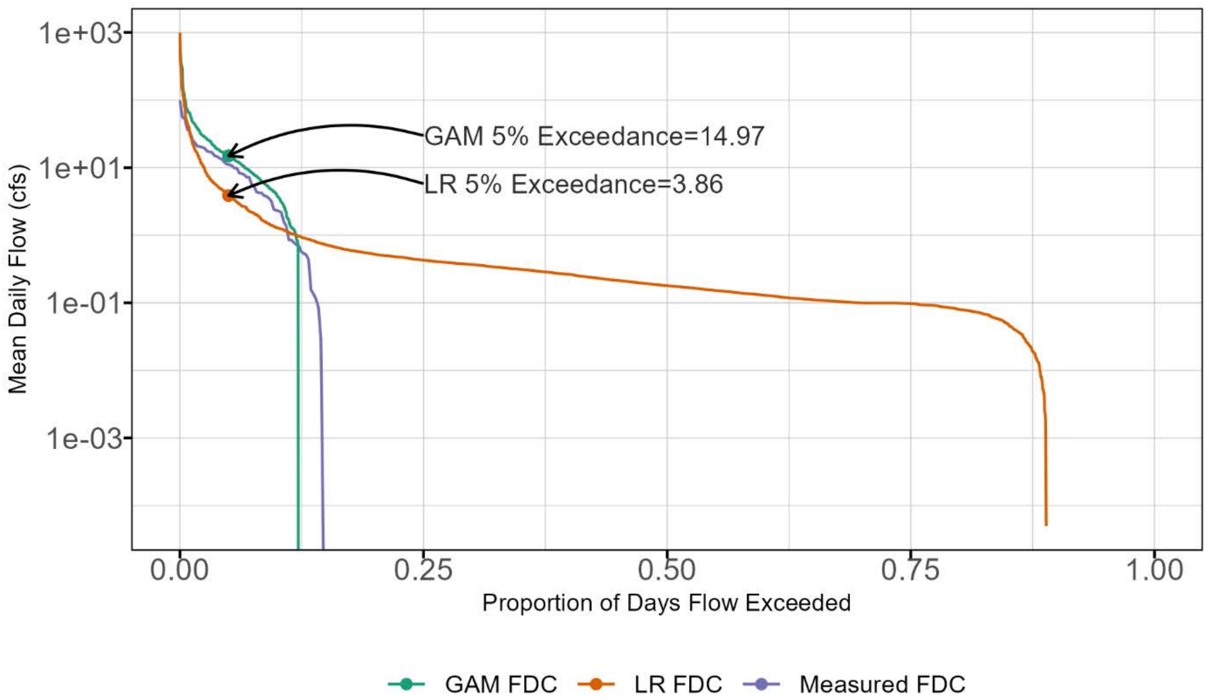

Figure 36. Predicted naturalized flow duration curves (FDCs)at surface water quality monitoring (SWQM) station 16882; January 1, 2011 through March 30, 2021. GAM = generalized additive model, LR = linear regression, cfs = cubic feet per second.

Discussion

This project consisted of two major parts: (1) fitting streamflow rating curves and developing estimates of mean daily streamflow at each SWQM station; and (2) utilizing the developed mean daily streamflow to evaluate the effectiveness of DAR and statistical models for predicting mean daily streamflow at each SWQM station. In many previous TMDLs in Texas, the DAR approach has been utilized without any way to validate the accuracy of the approach. In general, the academic literature has shown DAR to be reasonably effective at predicting streamflows between nearby locations. However, researchers often evaluate accuracy based strictly on flow exceedance percentiles as opposed to daily flow. For the purpose of allocating allowable pollutant loads, the flow exceedances are all that is necessary. For providing empirical estimates of historic pollutant loads, relatively accurate daily streamflow estimates are needed to estimate daily loadings within the stream. The disadvantage of the DAR approach becomes evident in this scenario if the timing of streamflow peaks does not align between the source gage and the ungaged basin.

The linear regression and GAM approach detailed in this report provide adequate prediction of daily streamflows and streamflow exceedances. Due to the simplicity of reproduction, the linear regression method may be preferred. However, for sites with a large proportion of no-flow days, the GAM hurdle model methodology is required. For simplicity, the logistic regression model might be combined with the linear regression approach and provide a more interpretable model for stakeholders and technical reviewers.

The approach detailed in this report provides a method for validating estimated streamflows and alternative empirical approaches for developing a streamflow period of record if locally available meteorological records are available and the DAR approach is not appropriate. While GAMs are used extensively for statistical modelling in academic literature, few studies have applied GAMs to hydrologic data. Shortridge et. al (2016) evaluated GAMs for predicting monthly streamflows. An extension of GAMs called generalized additive models for location, scale, and shape has also been applied to predict daily and monthly streamflows in intermittent streams (van Ogtrop et al. 2011; Rashid and Beecham 2019).

One remaining issue this report does not address is how to determine the amount of data required to develop a streamflow period of record for use in a TMDL. In the event of an extremely dry or wet year, the combination of covariates may not result in the needed empirical data to make good statistical predictions. The 95% confidence intervals produced by GAMs can provide some insight if the predictions are sufficient to proceed with TMDL development. Here we utilized 1 year of data due to project constraints, but further work to evaluate how much data to gather would be beneficial.

Rating-Curve Development

Rating curve development at SWQM station 16396 resulted in very accurate streamflow values at given stream heights. Measured streamflows at SWQM stations 16397 and 16882 exhibited higher variances at a given stream height resulting in less accurate rating curves. Visual inspection of the velocity measurements indicated substantial unexpected variances in measured velocity. Therefore, the high variance in reported streamflows at low stream height is most likely attributed to measurement error in velocity. The theoretical flow calculations used to report flow by the IQ Plus are reported to be accurate to 3-5% in regular channels but may be negatively affected by flow disturbances or if stream velocity varies substantially across the width of the stream (SonTek 2020). Additional error may be propagated if the measured stream cross section is inaccurate due to sedimentation or fouling of the doppler or extremely low water levels. Methods for quantifying this measurement error are not evident but would be useful for deciding what data to retain in rating curve development.

Sources of the measurement error may have been due to poor site characteristics. Both SWQM stations 16397 and 16882 were directly downstream from a major highway bridge with riprap-lined channels under the bridge. Immediately downstream of both sites were nearly 90-degree bends in the channel. Both sites were exposed to stormwater runoff from the highway that probably increased turbulence in the channel during storm events. The recommended installation methods call for a site with at least 10-channel widths upstream or downstream free of flow disturbances (SonTek 2020). With the proximity of the riprap and channel bend at SWQM stations 16397 and 16882, the two sites were not ideal for deployment of a bottom-mount acoustic flow meter. In retrospect, more frequent deployment of the flow meter at SWQM station 16396 would have probably produced more useful data.

Streamflow Prediction

Rating curve developed mean daily streamflows demonstrated that DAR performed poorly at the three SWQM sites. This result was anticipated due to the distance of candidate USGS gages and differences in land cover. Both linear regression and GAMs showed similar improved performance when fit to the full dataset. The GAM hurdle model performed best at SWQM station 16882, which included a high proportion of no-flow days. The daily estimates produced by linear regression were generally negatively biased on high-flow days. Since the linear regression was fit to log transformed streamflows, bias can be introduced when exponentiating the results back to the cubic feet per second scale. Additional consideration should be given to using a bias correction factor such as Duan smearing when back-transforming streamflow predictions (Duan 1983).

GAMs showed a particularly high variance in the cross-validation scores and were generally outperformed by linear regression. The high variance in GAM results showed that it is particularly sensitive to the amount of data used, and it may perform better with larger sample sizes.

One issue with all the models is streamflow prediction of the May 27, 2020 event (Figure 16**Error! Reference source not found.**, Figure 25). An enormous stream discharge event was measured coincident with a precipitation event that was not noticeably larger than other rainfall events. As noted earlier, the precipitation event captured at the Easterwood Airport weather station may be less than what occurred within the watershed based on local weather reports and anecdotal information. Smaller unnoticed deviances in measured and actual precipitation are not out of the question. Ideally an in-watershed weather station could be utilized, but a suitable weather station was not available. Potential alternative options are the use of gridded precipitation data developed from radar estimated rainfall (Jayakrishnan et al. 2005) or rainfall measurement loggers, such as tipping bucket gages, deployed at the stream station.

While the statistical methods used here are demonstrated as suitable methods for generating a streamflow record, there is still some uncertainty about the amount of streamflow and weather data that is required to generate streamflow records at a desired level of confidence. Methods for assessing the range of coverage of the dependent variable datasets over the desired period of record could be developed to assess the suitability of a particular year for predicting longer streamflow. While assessing performance of a physically based model such as SWAT was outside the scope of this project, it is assumed that physically based models do not face this disadvantage because parameter estimates should be reasonably based on some range of known relationships. Notable disadvantages of physically based models are the spatial and temporal data requirements and parameter calibration procedures (Jones et al. 2009). The approximately 1 year of daily streamflow data for calibration and validation would be a negative constraint for physically based models. This could be overcome by calibrating the models to a downstream gage on the mainstem reach of the Brazos River. The major disadvantage is that the calibration procedure would be carried out over a relatively large watershed that may or may not predict well at the desired SWQM stations because there is a considerably larger contribution area to the nearest downstream gage on the mainstem Brazos River.

Conclusion

Temporarily deployed acoustic streamflow meters and water level recorders provide an effective means for developing a short-term period of record for evaluating flow estimation methods. Site selection is critical for developing rating curves with high levels of confidence due to measurement errors associated with hydrologic properties at individual sites. Based on the results of this project and given equipment and budgetary constraints, future prioritization would generally be given to collecting as much data as possible at sites with good stream characteristics for measuring flow with the acoustic streamflow meter because it facilitates creating accurate and frequently updated rating curves. Consideration should also be given to collecting rainfall data alongside the streamflow measurements if there are plans to develop rainfall-runoff models and the nearest reliable weather station is outside of the watershed.

The collected streamflow data provided substantial evidence that the DAR method was not appropriate for the Thompsons Creek watershed. In consideration of cross-validation results and model simplicity, linear regression provides adequate but potentially biased daily streamflow and streamflow exceedance estimates in the watershed. For SWQM station 16882, the high number of no-flow days requires the use of a two-step or hurdle model that estimates the probability of flow occurrence and the actual streamflow value. Additional tuning of these statistical estimation methods through predictor variable tuning, bias estimators, or additional data collection may improve model performance.

References

Asquith, W.H., Roussel, M.C., Vrabel, J. 2006. Statewide Analysis of the Drainage-Area Ratio Method for 34 Streamflow Percentile Ranges in Texas, Reston, VA: U.S. Geological Survey. Scientific Investigations Report 2006–5286. https://pubs.usgs.gov/sir/2006/5286/pdf/sir2006-5286.pdf.

Beck, M.W., Murphy, R.R. 2017. Numerical and qualitative contrasts of two statistical models for water quality change in tidal waters. Journal of the American Water Resources Association. 53 (1): 197-219. https://doi.org/10.1111/1752-1688.12489.

Chamberlain, S., Hocking, D., Anderson, B., Erickson, A., Potter, N., Stachelek, J., Simmons, A., Ram, K., Edmund, H., rOpenSci. 2019. rnoaa: 'NOAA' Weather Data from R. https://cran.r-project.org/web/packages/rnoaa/index.html.

De Cicco, L.A., Hirsch, R.M., Lorenz, D.L., Watkins, W.D. 2018. dataRetrieval: R packages for discovering and retrieving water data available from Federal hydrologic web services. Reston, VA: U.S. Geological Survey. https://code.usgs.gov/water/dataRetrieval.

Duan, N. 1983. Smearing estimate: a nonparametric retransformation method. Journal of the American Statistical Association. 78 (383): 605-610. https://doi.org/10.1080/01621459.1983.10478017.

Fiedler, Rebecca. 2020. WATCH NOW: Storms bring rain, hail to Brazos Valley. The Eagle, May 28, 2020. https://theeagle.com/news/local/watch-now-storms-bring-rain-hail-to-brazos-valley/article_bfc05cd6-a097-11ea-8260-174a14937224.html.

Gitter, A., Nayal, M., Rambo, J., Yang, L., Gregory, L. 2020. Watershed Characterization of the Thompsons Creek Watershed, College Station, TX: Texas Water Resources Institute. Technical Report 56. https://twri.tamu.edu/publications/technical-reports/2020-technical-reports/tr-526/ .

Haddad, K., Rahman, A., A Zaman, M., Shrestha, S. 2013. Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression. Journal of Hydrology. 482: 119-128. https://doi.org/10.1016/j.jhydrol.2012.12.041.

Hagemann, M., Kim, D., Park, M.H. 2016. Estimating nutrient and organic carbon loads to water-supply reservoir using semiparametric models. Journal of Environmental Engineering. 142 (8): 04016036. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001077.

Jayakrishnan, R., Srinivasan, R., Santhi, C., Arnold, J.G. 2005. Advances in the application of the SWAT model for water resources management. Hydrological Processes. 19 (3): 749-762. https://doi.org/10.1002/hyp.5624.

Jones, C.A., Wagner, K., Di Giovanni, G., Hauck, L., Mott, J., Srinivasan, R., Ward, G., Wythe, K. 2009. Bacteria Total Maximum Daily Load Task Force Final Report, College Station, TX: Texas Water Resources Institute. Technical Report 341. http://hdl.handle.net/1969.1/86092.

Knoben, W.J.M., Freer, J.E., Woods, R.A. 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences. 23 (10): 4323-4331. https://doi.org/10.5194/hess-23-4323-2019.

Lasater, A., Austin, B., Scott, E., Haggard, B.E. 2019. Streamflow monitoring and load estimations in small-scale watersheds. 11th National Water Quality Monitoring Conference. Denver, Colorado. https://acwi.gov/monitoring/conference/2019/presentations/D6_Lasater_Secure.pdf.

Liu, L., Shih, Y-C.T., Strawderman, R.L., Zhang, D., Johnson, B.A., Chai, H. 2019. Statistical analysis of zero-inflated nonnegative continuous data: a review. Statistical Science. 34 (2): 253-279. https://doi.org/10.1214/18-STS681.

Murphy, R.R., Perry, E., Harcum, J., Keisman, J. 2019. A generalized additive model approach to evaluating water quality: Chesapeake Bay case study. Environmental Modelling & Software. 118: 1-13. https://doi.org/10.1016/j.envsoft.2019.03.027.

van Ogtrop, F.F., Vervoort, R.W., Heller, G.Z., Stasinopoulos, D.M., Rigby, R.A. 2011. Long-range forecasting of intermittent streamflow. Hydrology and Earth System Sciences. 15 (11): 3343-3354. https://doi.org/10.5194/hess-15-3343-2011.

Onset. 2018. HOBO U20L Water Level Logger (U20L-0x) Manual. https://www.onsetcomp.com/files/manual_pdfs/17153-G%20U20L%20Manual.pdf.

Ouali, D., Chebana, F., Ouarda, T.B.M.J. 2017. Fully nonlinear statistical and machine-learning approaches for hydrological frequency estimation at ungauged sites. Journal of Advances in Modeling Earth Systems. 9 (2): 1292-1306. https://doi.org/10.1002/2016MS000830.

Padfield, D., Matheson, G. 2020. nls.multstart: Robust Non-Linear Regression using AIC Scores. https://CRAN.R-project.org/package=nls.multstart.

Petersen-Øverleir, A. 2006. Modelling stage—discharge relationships affected by hysteresis using the Jones formula and nonlinear regression. Hydrological Sciences Journal. 51 (3): 365-388. https://doi.org/10.1623/hysj.51.3.365.

R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rashid, M.d.M., Beecham, S. 2019. Simulation of streamflow with statistically downscaled daily rainfall using a hybrid of wavelet and GAMLSS models. Hydrological Sciences Journal. 64 (11): 1327-1339. https://doi.org/10.1080/02626667.2019.1630742.

Richards, R.G., Tomlinson, R., Chaloupka, M. 2010. Using Generalized Additive Models to assess, explore and unify environmental monitoring datasets. Modelling for environment's sake: proceedings of the 5th biennial conference of the international environmental modelling and software society, Ottawa, Canada. https://scholarsarchive.byu.edu/iemssconference/2010/all/584/.

Schramm, M.P. 2018. echor: access EPA "ECHO" data. https://cran.r-project.org/package=echor.

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F. 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences. 20 (7): 2611-2628. https://doi.org/10.5194/hess-20-2611-2016.

SonTek. 2020. SonTek-IQ Series Intelligent Flow Featuring SmartPlusHD User's Manual.

State of Ohio Environmental Protection Agency. 2009. TMDLs for the White Oak Creek Watershed Appendix A: Streamflow Estimation Techniques. Ohio: State of Ohio Environmental Protection Agency. https://epa.ohio.gov/static/Portals/35/tmdl/WhiteoakCreekTMDL_final_dec09_appA.pdf.

Venetis, C. 1970. A note on the estimation of the parameters in logarithmic stage-discharge relationships with estimates of their error. International Association of Scientific Hydrology Bulletin. 15 (2): 105-111. https://doi.org/10.1080/02626667009493957.

Wood, S.N. 2008. Fast stable direct fitting and smoothness selection for generalized additive models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 70 (3): 495-518. https://doi.org/10.1111/j.1467-9868.2007.00646.x.

Zakwan, M. 2018. Spreadsheet-based modelling of hysteresis-affected curves. Applied Water Science. 8 (4): 101. https://doi.org/10.1007/s13201-018-0745-3.

Zuur, A.F., editor. 2009. Mixed Effects Models and Extensions in Ecology with R. New York, NY: Springer (Statistics for biology and health).