Thompsons Creek Watershed

Flow Estimation Methodologies

Texas Water Resources Institute | Texas A&M AgriLife Research

2021-11-23

Project Drivers

- DAR works well under most circumstances for predicting the FDC (Ries and Friesz, 2000; Asquith, Roussel, and Vrabel, 2006);
- Rarely more than a handful of instantaneous streamflow measurements to "validate" data (mean daily streamflow ≠ instantaneous streamflow);
- We also often want daily streamflow estimates, this works best with a gage in the same watershed where daily peaks are more likely to correspond;

Project Drivers

- SWAT is well established and accepted, especially in rural watersheds (many new extensions to incorporate groundwater, stormwater networks, etc.) (Arnold and Fohrer, 2005);
- Issue of validation in ungaged watersheds still exists;
- Often calibrated to downstream gages on mainstem reaches that may or may not be reflective of the subwatershed of interest;

Project Drivers

- Can we efficiently provide mean daily streamflow measurements to validate streamflow estimation methods/models?
- DAR is a desirable approach: simple, reproducible, acceptable;
- If DAR doesn't perform well do we have options to estimate streamflow other than the numerous numeric models that require substantial overhead and are somewhat difficult to interpret?

- Physical (numeric) models: lumped hydrologic models, distributed hydrologic models etc.
    - varying degrees of interpretibility, high data requirements, great for forecasting or predicting outside of calibration data range
- Statistical methods: Linear regression, semi-parametric regression, etc.
    - higher interpretibility, lower data requirements, not always great for forecasting
- Machine learning (statistical) methods: ANN, regression trees, ...
    - low interpretibility, lower data requirements (?), popular in forecasting and prediction

Methods

1: Develop mean daily streamflow period of record;

2: Evaluate DAR, linear regression, and semi-parametric regression for estimating streamflow records at site of interest.

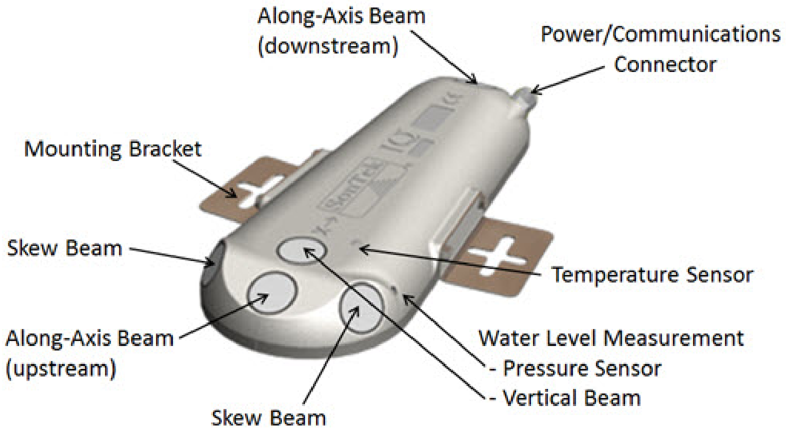
Methods

Develop 1-Yr Mean Daily Streamflow Period-of-Record

- Record 15-minute stream depths using HOBO Water Level Logger
- pressure transducer deployed instream and separate pressure transducers deployed to measure atmospheric pressure.

Methods

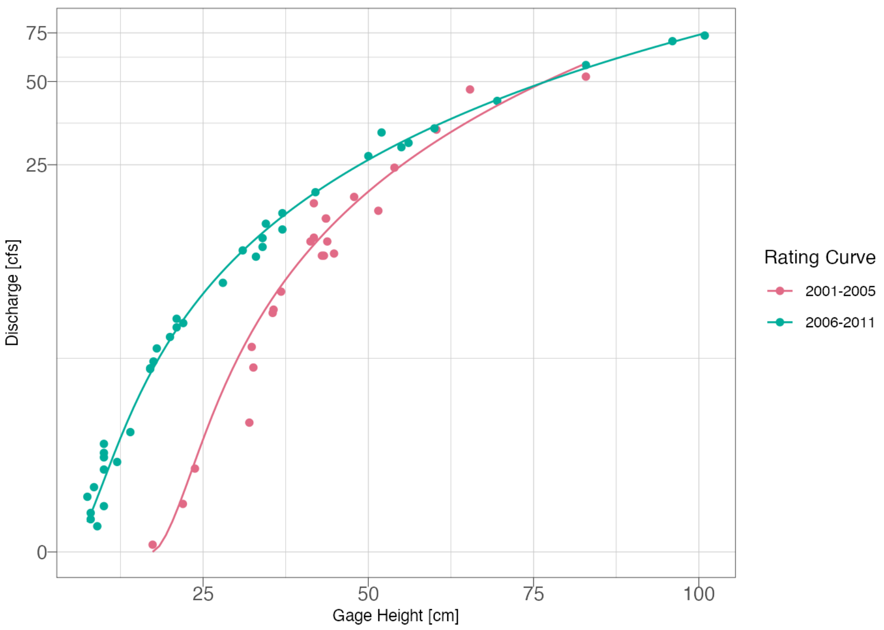
Develop 1-Yr Mean Daily Streamflow Period-of-Record



- Measure 15-minute streamflows using Sontek IQ Plus (Bottom mount acoustic doppler velocity meter);
- Periodic deployments, utilizes proprietary index velocity method to calculate streamflow over the chosen interval.

Methods

Develop 1-Yr Mean Daily Streamflow Period-of-Record



- Develop rating curves relating flow and depth over the course of the year
- Power function:

$$Q = K(H - H_0)^z$$

- parameterize K, H_0, and z using non-linear least square (minimize SSE).

Methods

Develop 1-Yr Mean Daily Streamflow Period-of-Record

Unsteady flow:

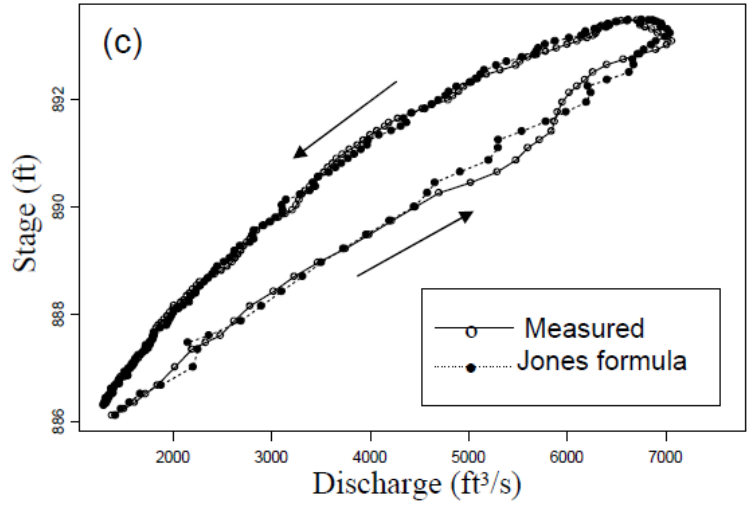


Figure from Petersen-Ã˜verleir (2006)

- For unsteady flows use the modified Jones Formula:

$$Q = K(h - a)^n \times \sqrt{1 + x\frac{\partial h}{\partial t}}$$

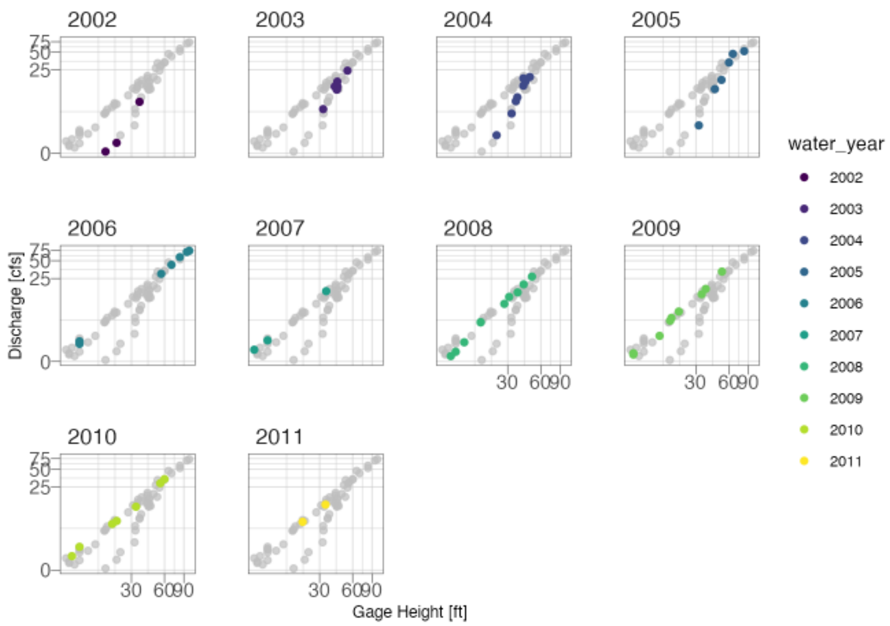
- parameterize K, a, n, x;
- $\frac{\partial h}{\partial t}$ is the "rate of change" in stream height as a given time.

(Petersen-Ã˜verleir, 2006; Zakwan, 2018)

Methods

Develop 1-Yr Mean Daily Streamflow Period-of-Record

- Fit one or more rating curves to the data based on visual inspection of the time series record;
- Changes in channel shape, vegetation, etc. will alter the rating curve and can necessitate updating the curve throughout the year;



Methods

Develop 1-Yr Mean Daily Streamflow Period-of-Record

- Use the rating curve to calculate flows from the HOBO measured depths;
- 15-minute streamflow record is aggregated to mean daily streamflow.

Methods

Evaluate methods for estimating daily streamflows

- Information transfer methods
  - Statistical or algebraic transfer of runoff data from one watershed to another
  - DAR

$$Q_y^t = Q_x^t \left(\frac{A_y}{A_x} \right)^{\phi}$$

- Linear regression between gaged site and ungaged site

$$Q_y = \beta_0 + \beta_1 X_1 + \varepsilon$$

Methods

Evaluate methods for estimating daily streamflows

- Information transfer methods
    - Assumes similar precipitation and rainfall-runoff characteristics
    - Generally performs well for FDC estimation, but estimating daily streamflow values require a well-chosen gaged site.

Methods

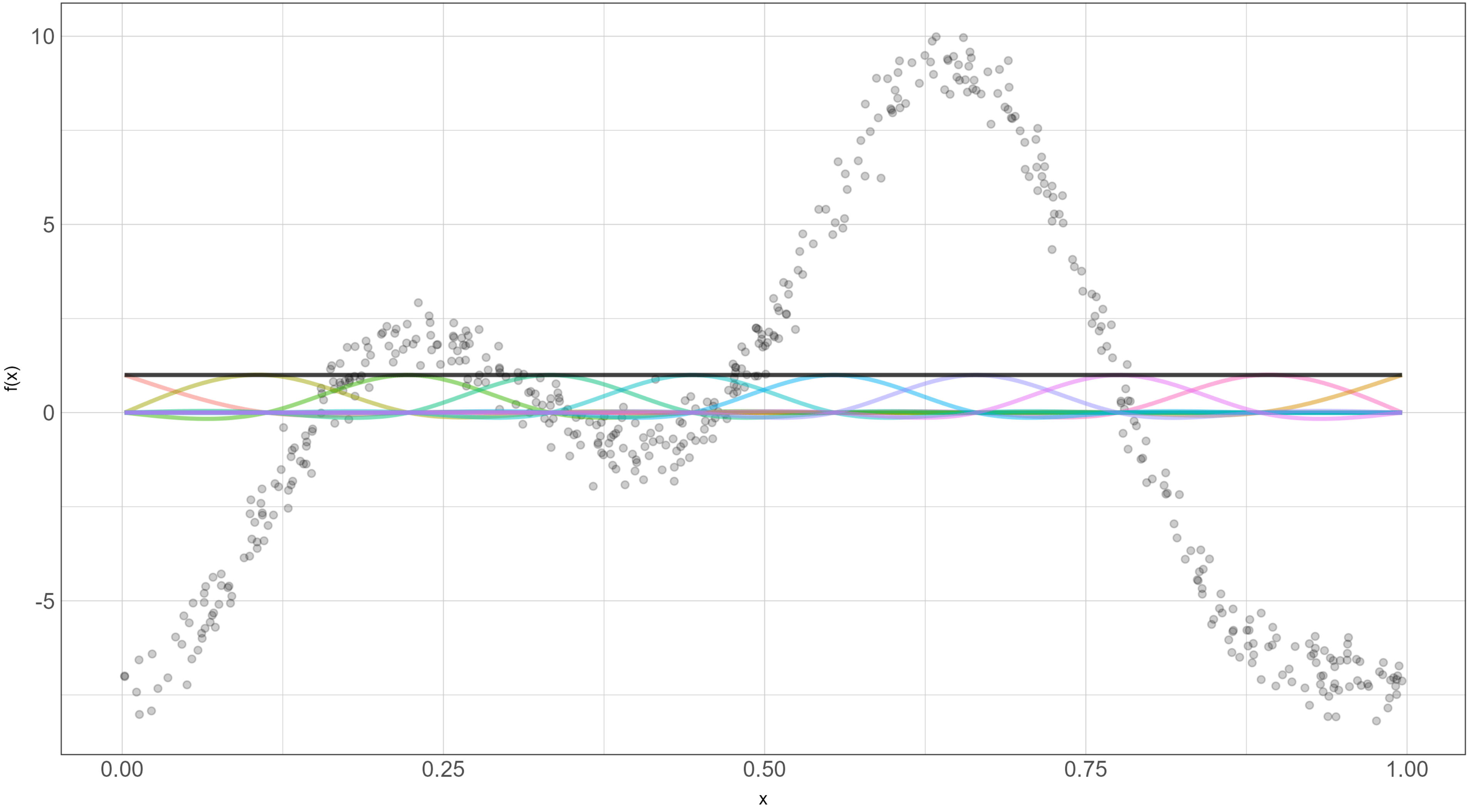
Evaluate methods for estimating daily streamflows

- Semi-parametric rainfall-runoff regression
  - Generalized Additive Model (GAM) is an extension of the Generalized Linear Model (GLM) that incorporates nonlinear forms of the predictor variables. The advantage of generalized models over linear regression models is the ability to incorporate different distributions in the error structure and a flexible *link function* that relates the mean of the response to the predictor variables. We know stream flow is restricted to >= 0 and has a extremely skewed distribution. By selecting the appropriate family and link structure we can restrict the response to the positive space.
  - GAMs also allow nonlinear relationships between the predictor and response variables.

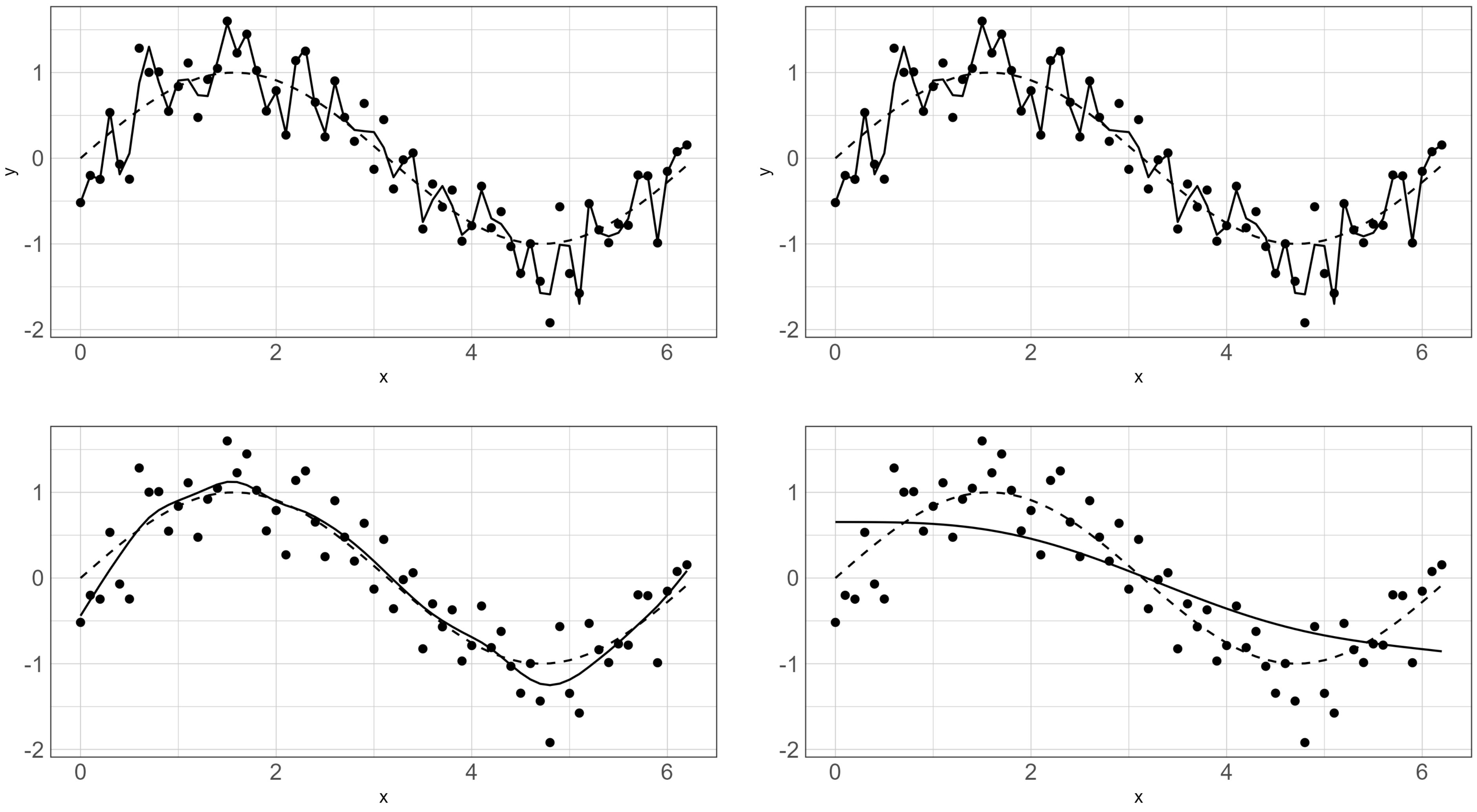
$$Q_y = \beta_0 + f(x_1) + \varepsilon$$

- **WHAT?!**

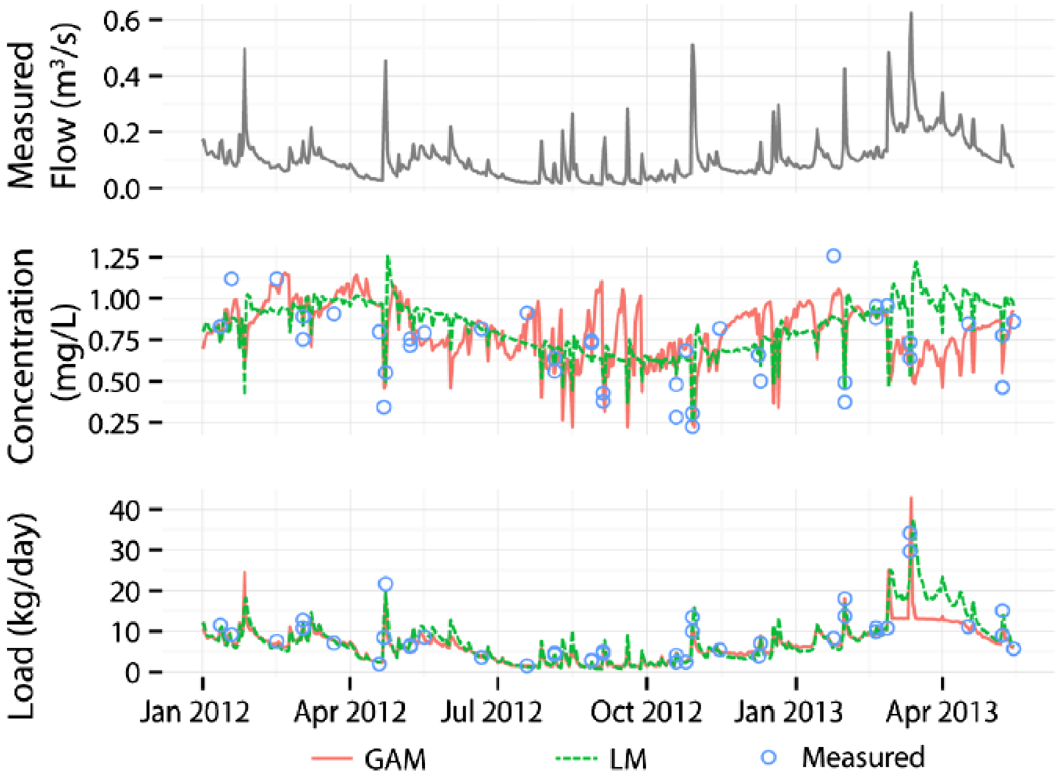
GAMs fit wobbly lines (smooth functions) to the data



Too wobbly or not wobbly enough?



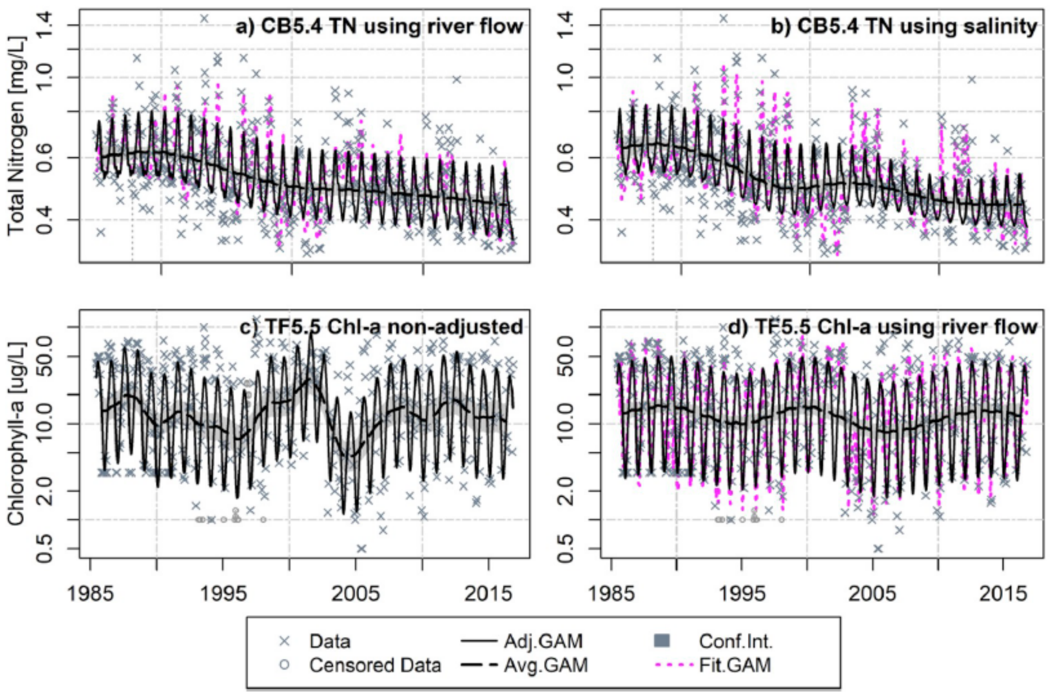
Nutrient Load Prediction



(Hagemann, Kim, and Park, 2016)

Nutrient Load Prediction

Chesapeake Bay Program

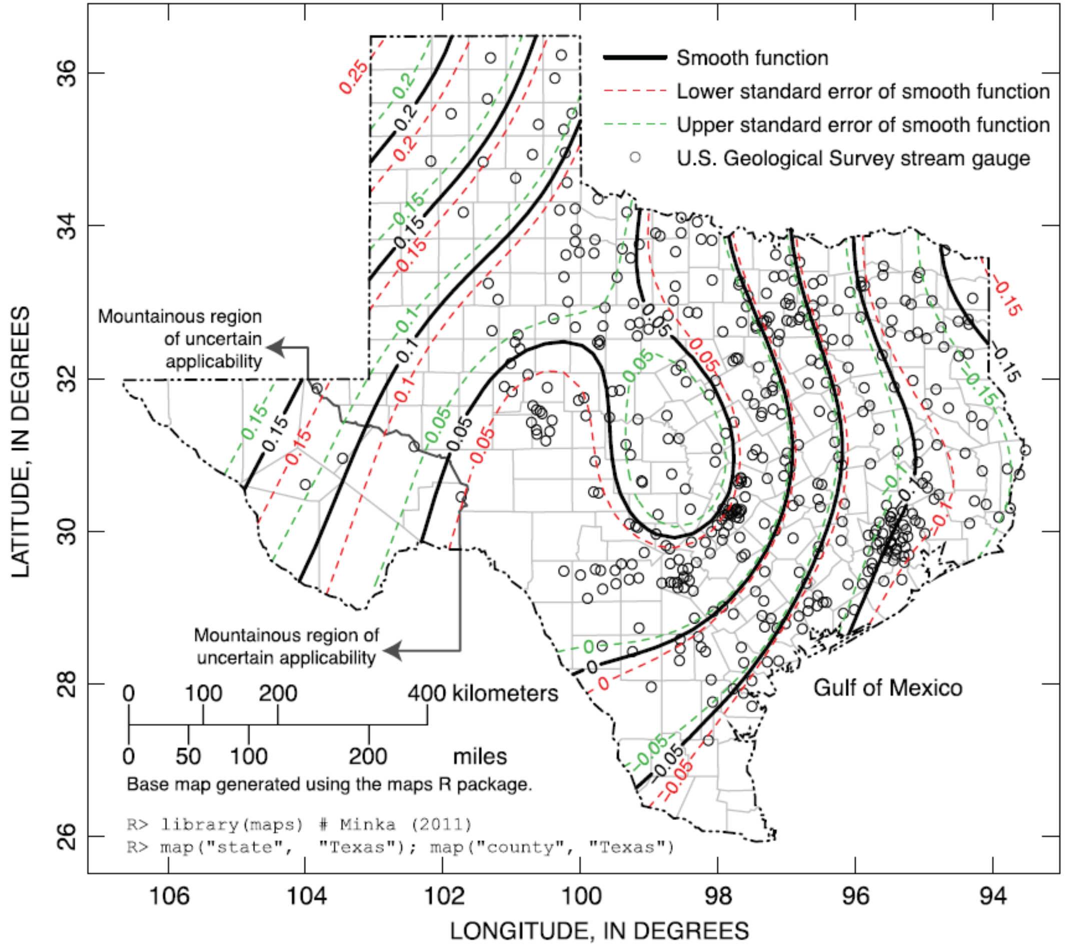


(Murphy, Perry, Harcum, and Keisman, 2019)

Nutrient Load Prediction

Chesapeake Bay Program

Discharge/Velocity Prediction



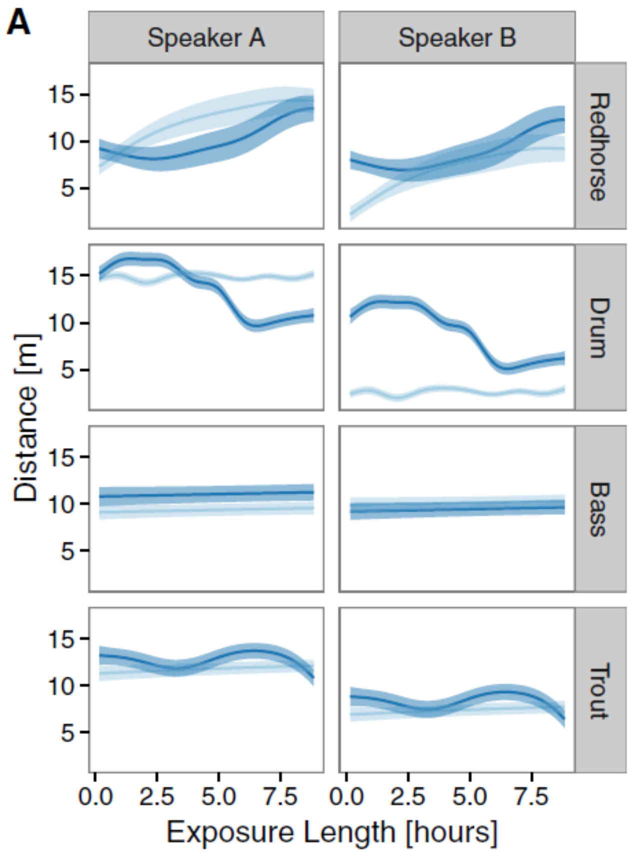
(Asquith, Herrmann, and Cleveland, 2013)

Nutrient Load Prediction

Chesapeake Bay Program

Discharge/Velocity Prediction

Environmental Mitigation



(Schramm, Bevelhimer, and Scherelis, 2017)

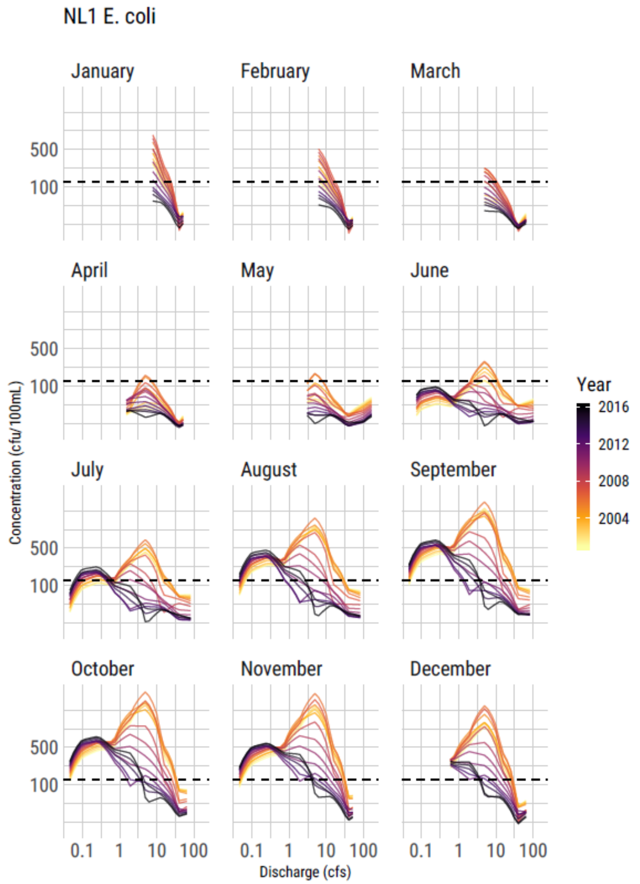
Nutrient Load Prediction

Chesapeake Bay Program

Discharge/Velocity Prediction

Environmental Mitigation

Upper Llano Watershed



(Schramm, Broad, and Arsuffi, 2018)

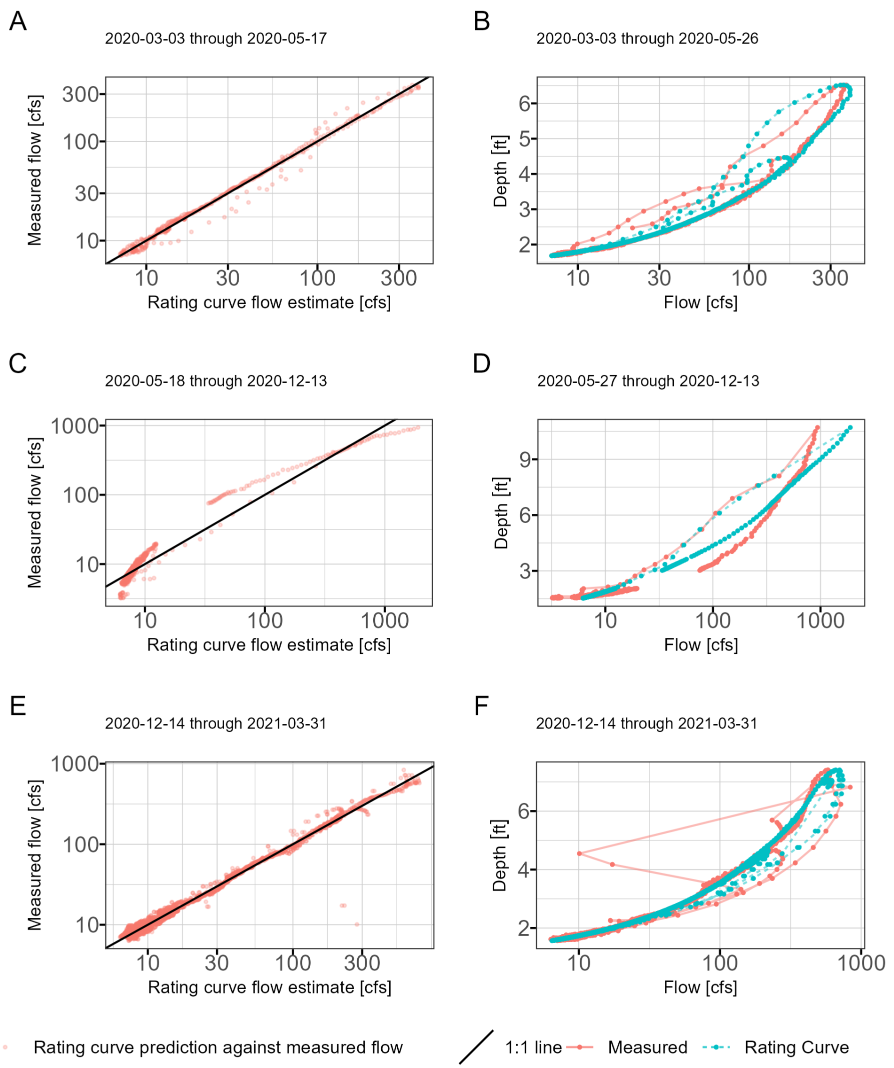
GAM

$$Q = f(P) + f(T) + f(P_{lag,1}) + f(P_{sum,3}) + f(T_{mean,5}) + f(H) + f(M)$$

- $f() =$ some unknown smoothing function
- $P =$ log(Precipitation + 1)
- $T =$ squared max temp
- $P_{lag,1} =$ 1 day lag P
- $P_{sum,3} =$ 3 day sum rainfall
- $T_{mean,5} =$ *5 day mean T_max*
- $H =$ Relative Humidity
- $M =$ Month

Results

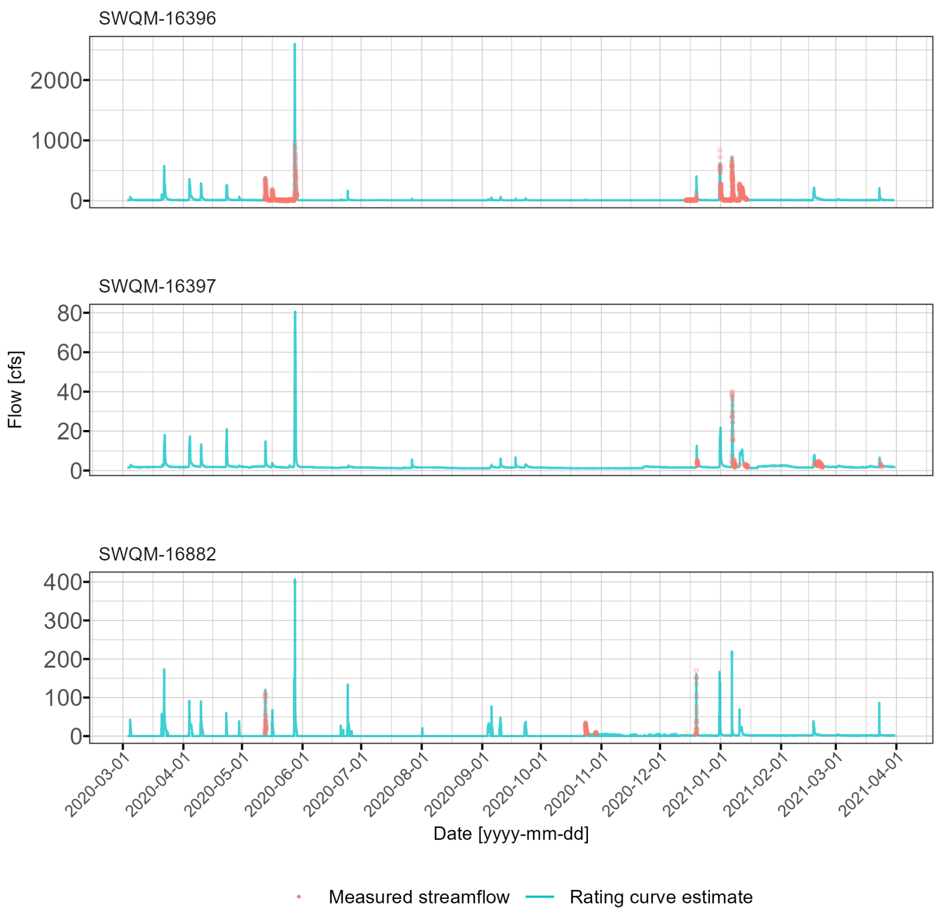
Rating Curve: 16396 Thompsons @ Silver Hill Rd



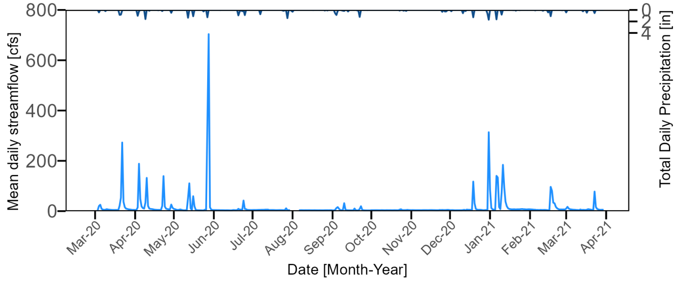
Rating Curve Parameters & Fit

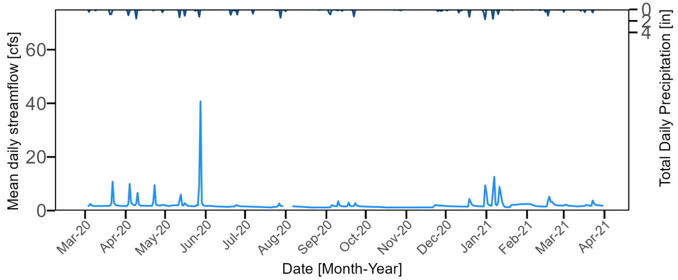
K	a	n	x	NSE	nRMSE
4.8077	0.366	0.4816	-0.1808	0.99	2.5
1.5574	-0.696	1.3786	-0.0786	0.73	6.8
4.3915	0.1785	0.6552	0.0808	0.97	1.8

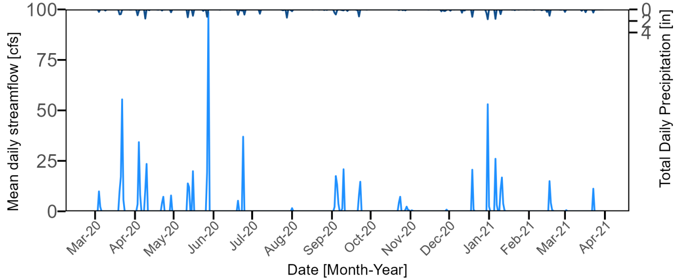
15-minute Streamflow



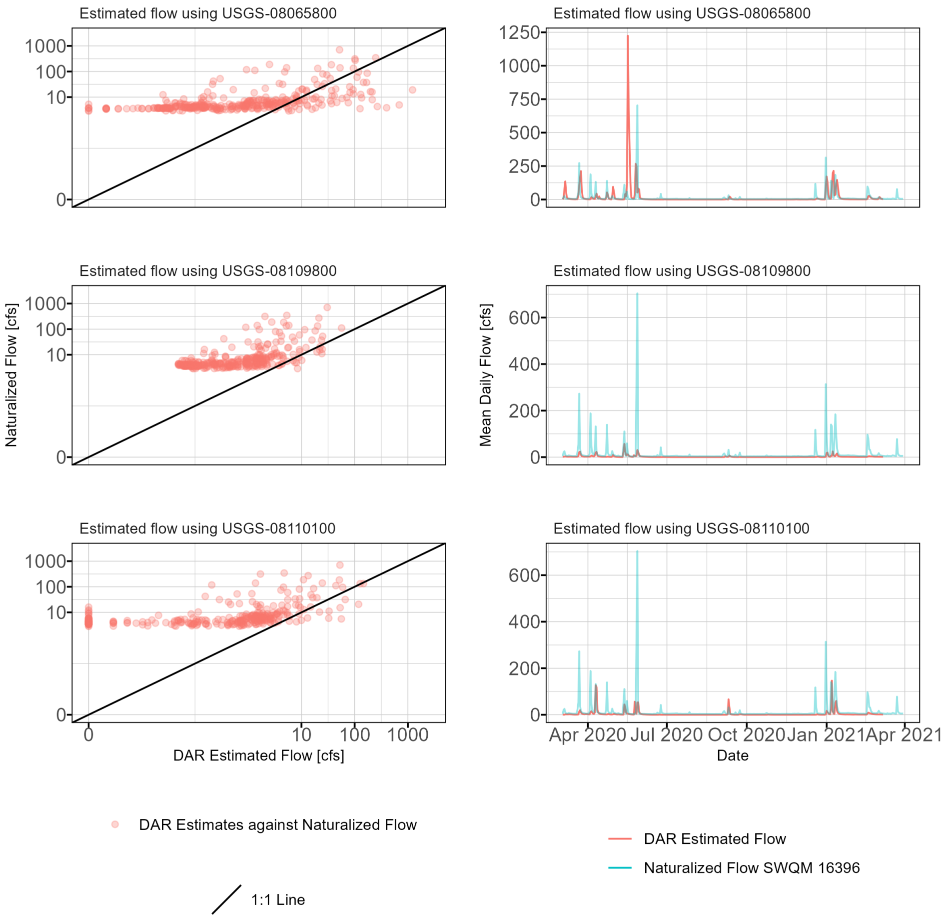
Naturalized Hydrograph





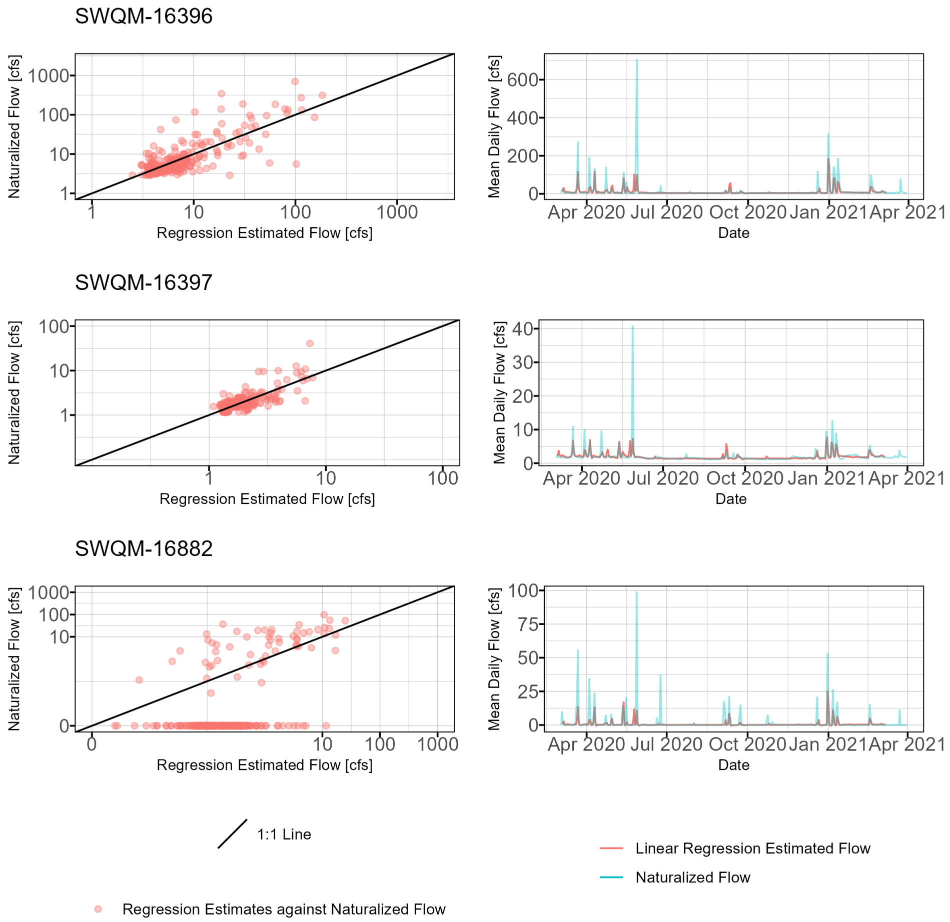


DAR results (Thompsons @ Silver Hill Rd.)



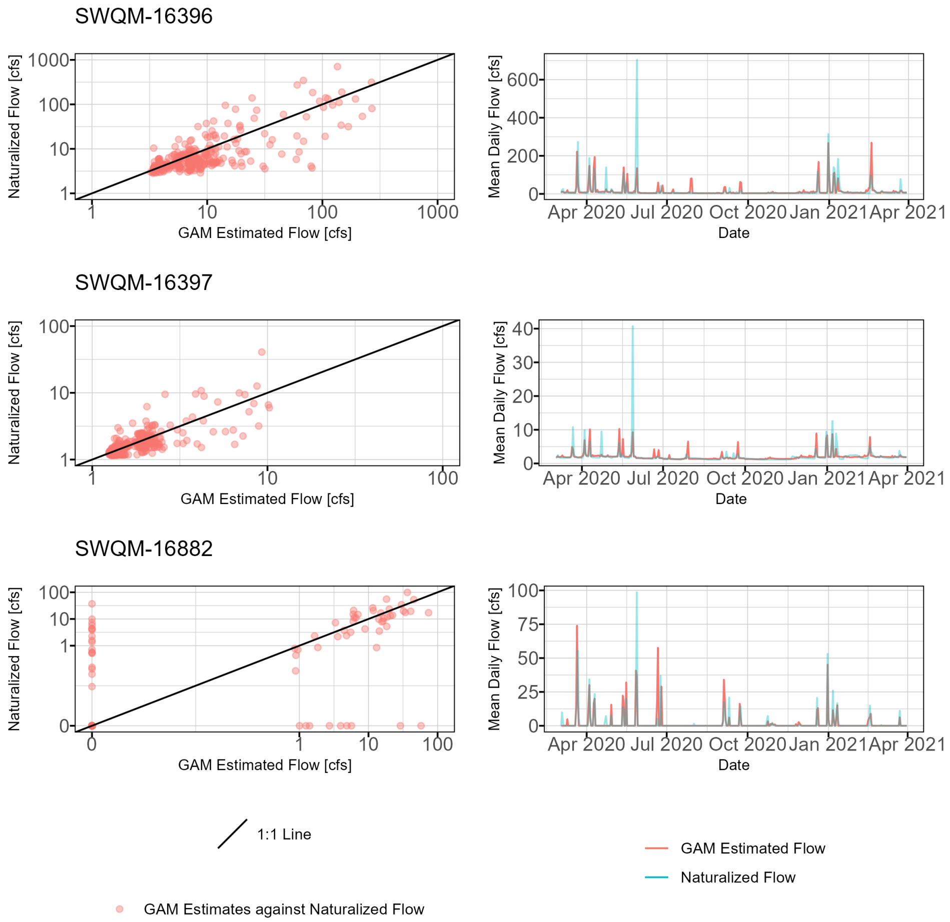
Method	NSE	KGE
DAR 08065800	-0.27	-0.08
DAR 08109800	0.25	-0.36
DAR 08110100	0.26	-0.22

Linear Regression Results



Method	NSE	KGE
Linear Regression	0.52	0.21

GAM Results



Method	NSE	KGE
GAM	0.425	0.46

Metrics

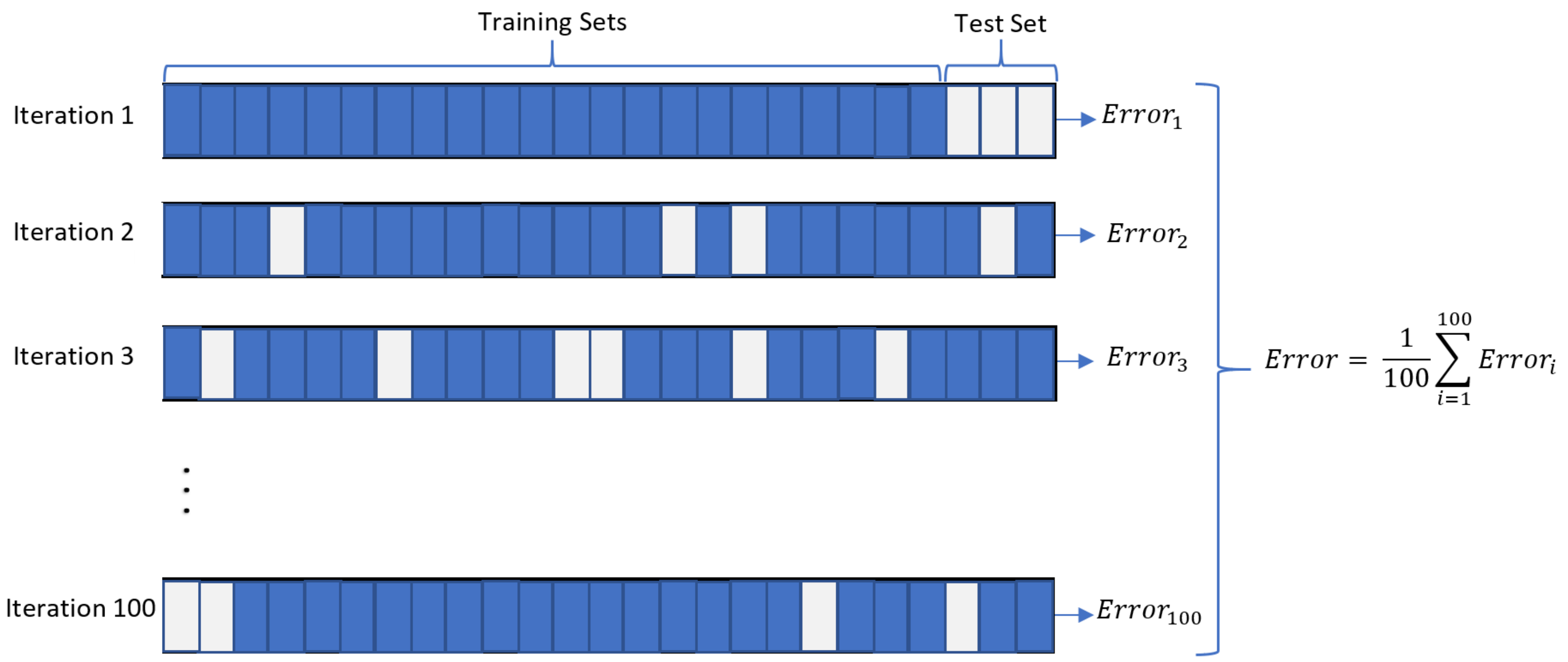
Site 16396

Method	NSE	KGE
DAR 08065800	-0.27	-0.08
DAR 08109800	0.25	-0.36
DAR 08110100	0.26	-0.22
Linear Regression	0.52	0.21
GAM	0.425	0.46

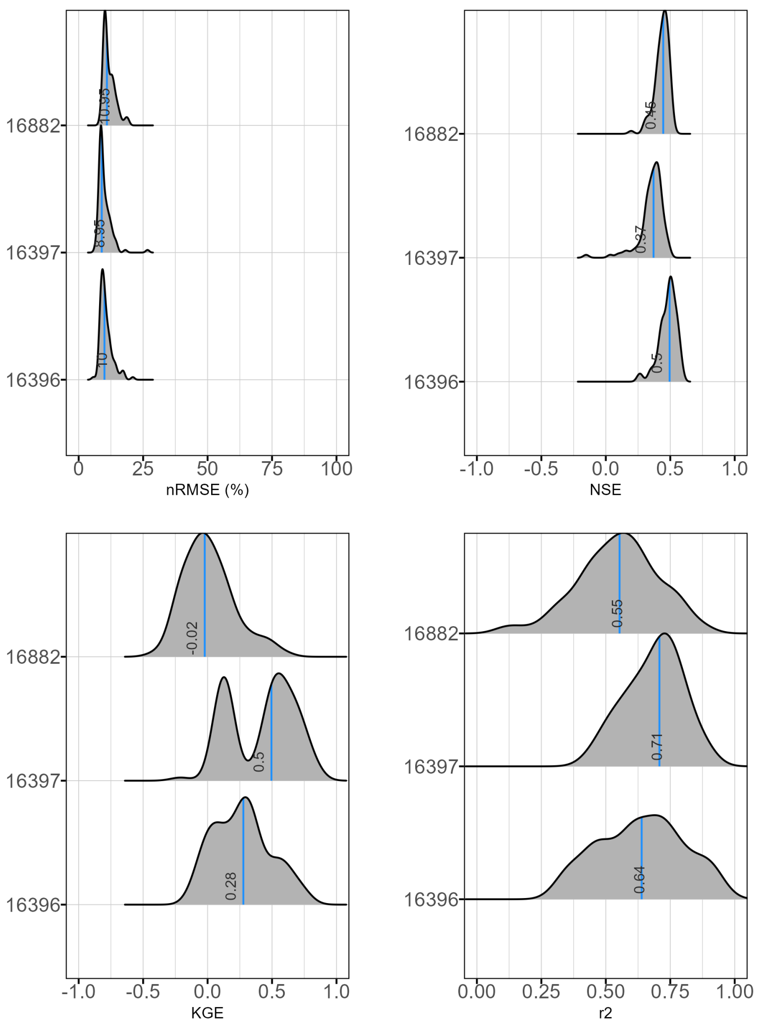
Cross-validation

We want to evaluate how this approach works on data outside of the data we fit the models to. Normally, we hold out a portion of data and use it as a test data set. However, we don't have much data.

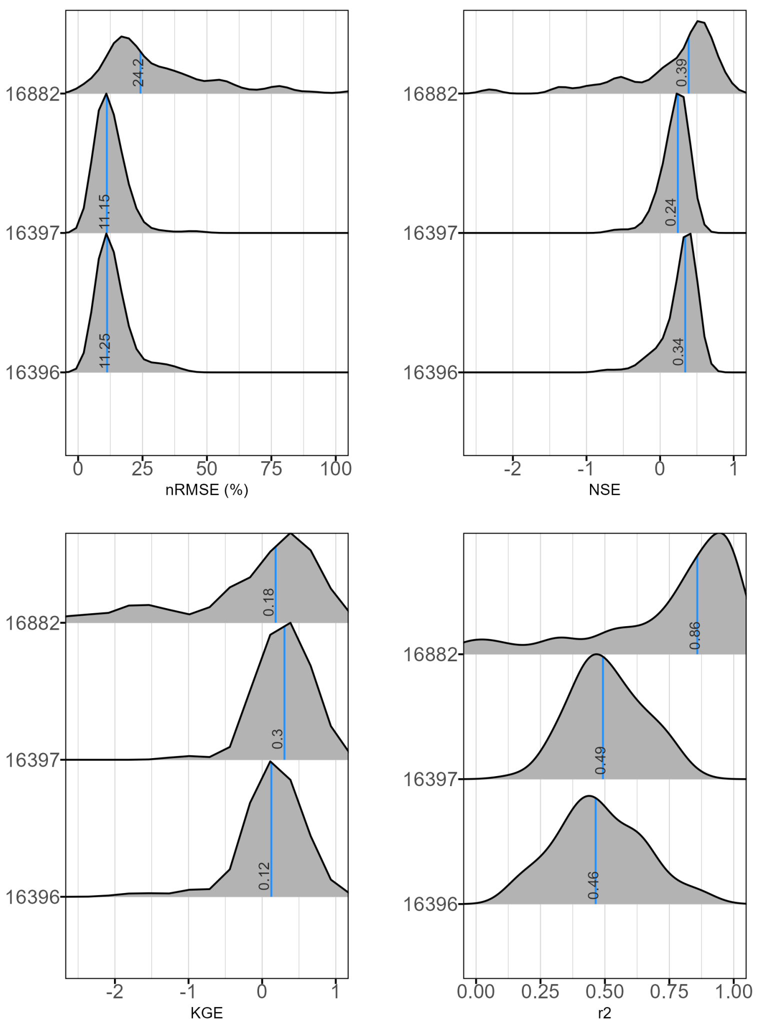
We use Monte-Carlo Cross Validation:



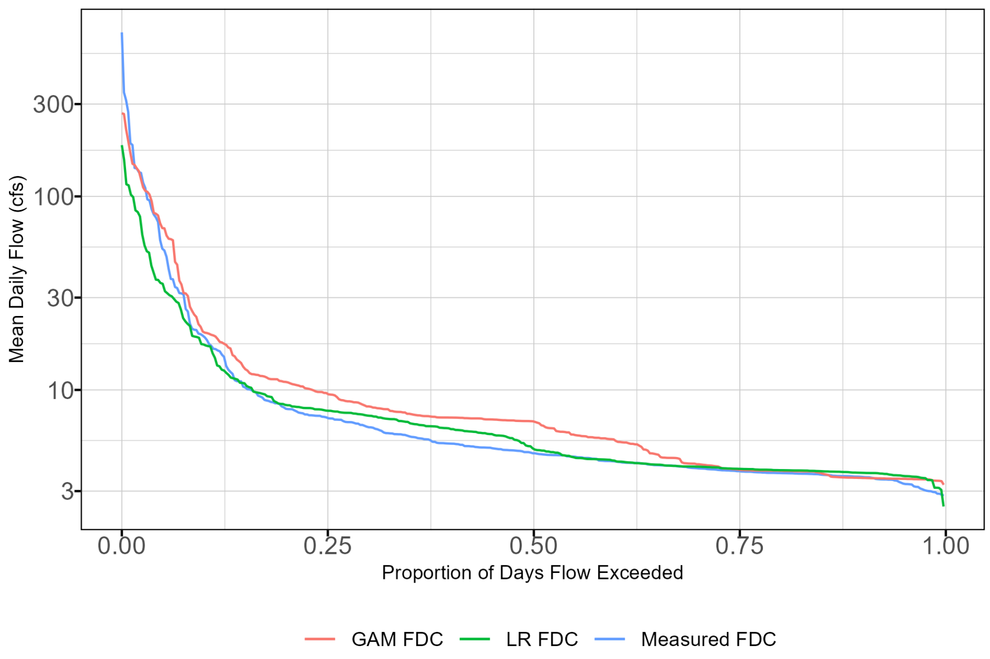
Linear regression

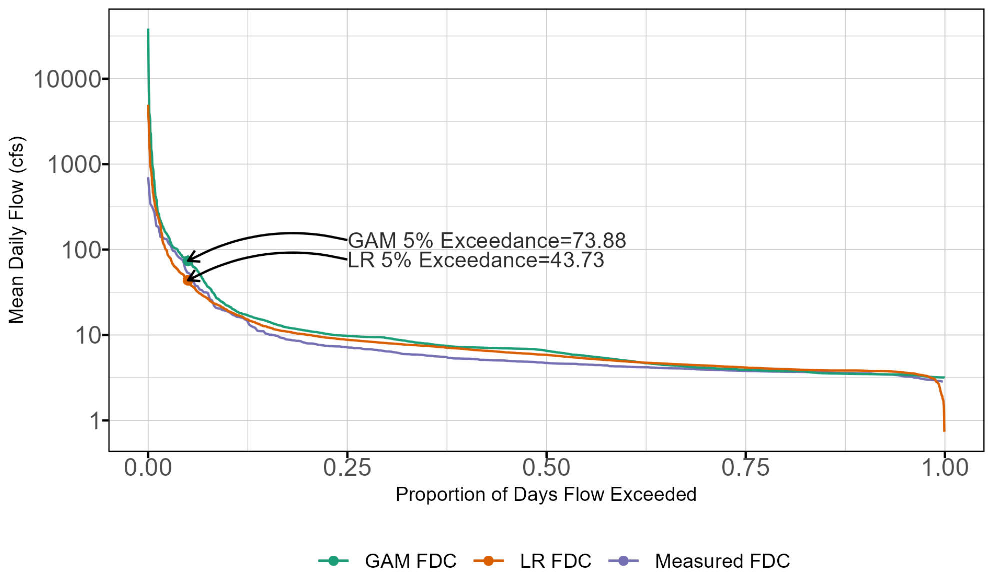


GAM



Flow Duration Curves





To Do:

- Revisit rating curves. Over prediction of high flow events might be an issue.
- Explore co-variates (PET) and effects of normalizing streamflows before fitting regression based models
- Compare precipitation driven GAMS to rainfall runoff approaches (HyMOD, SCS-CN, TWDB Rainfall-Runoff model, etc.)
- Explore period of record needed to confidently predict to out of sample data.

Some lessons:

- Site selection is difficult for deploying bottom mount dopplers
- More frequent data, possibly longer (>1 year) sampling would be useful (need to quantify tradeoffs between data collection costs and developing distributed models like SWAT)
- Flow data collected alongside routine data will avoid some of the challenges associated with recreating daily flow records, much easier to estimate flow exceedances and match flows instead of collection dates

References

Arnold, J. G. and N. Fohrer (2005). "SWAT2000: current capabilities and research opportunities in applied watershed modelling". In: *Hydrological Processes* 19.3, pp. 563-572. ISSN: 0885-6087, 1099-1085. DOI: 10.1002/hyp.5611. URL: http://doi.wiley.com/10.1002/hyp.5611.

Asquith, W. H., G. R. Herrmann, and T. G. Cleveland (2013). "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". In: *Journal of Hydrologic Engineering* 18.10, pp. 1331-1348. ISSN: 1084-0699, 1943-5584. DOI: 10.1061/(ASCE)HE.1943-5584.0000635. URL: [http://ascelibrary.org/doi/10.1061/

Asquith, W. H., M. C. Roussel, and J. Vrabel (2006). *Statewide Analysis of the Drainage-Area Ratio Method for 34 Streamflow Percentile Ranges in Texas*. Technical Report 2006-5286. U.S. Geological Survey. URL: https://pubs.usgs.gov/sir/2006/5286/pdf/sir2006-5286.pdf.

Hagemann, M., D. Kim, and M. H. Park (2016). "Estimating Nutrient and Organic Carbon Loads to Water-Supply Reservoir Using Semiparametric Models". In: *Journal of Environmental Engineering* 142.8, p. 04016036. ISSN: 0733-9372, 1943-7870. DOI: 10.1061/(ASCE)EE.1943-7870.0001077. URL: [http://ascelibrary.org/doi/10.1061/

Murphy, R. R., E. Perry, J. Harcum, et al. (2019). "A Generalized Additive Model approach to evaluating water quality: Chesapeake Bay case study". In: *Environmental Modelling & Software* 118, pp. 1-13. ISSN: 13648152. DOI: 10.1016/j.envsoft.2019.03.027. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364815218307801.

Petersen-Ã˜verleir, A. (2006). "Modelling stageâ€”discharge relationships affected by hysteresis using the Jones formula and nonlinear regression". In: *Hydrological Sciences Journal* 51.3, pp. 365-388. DOI: 10.1623/hysj.51.3.365.

References (cont.)

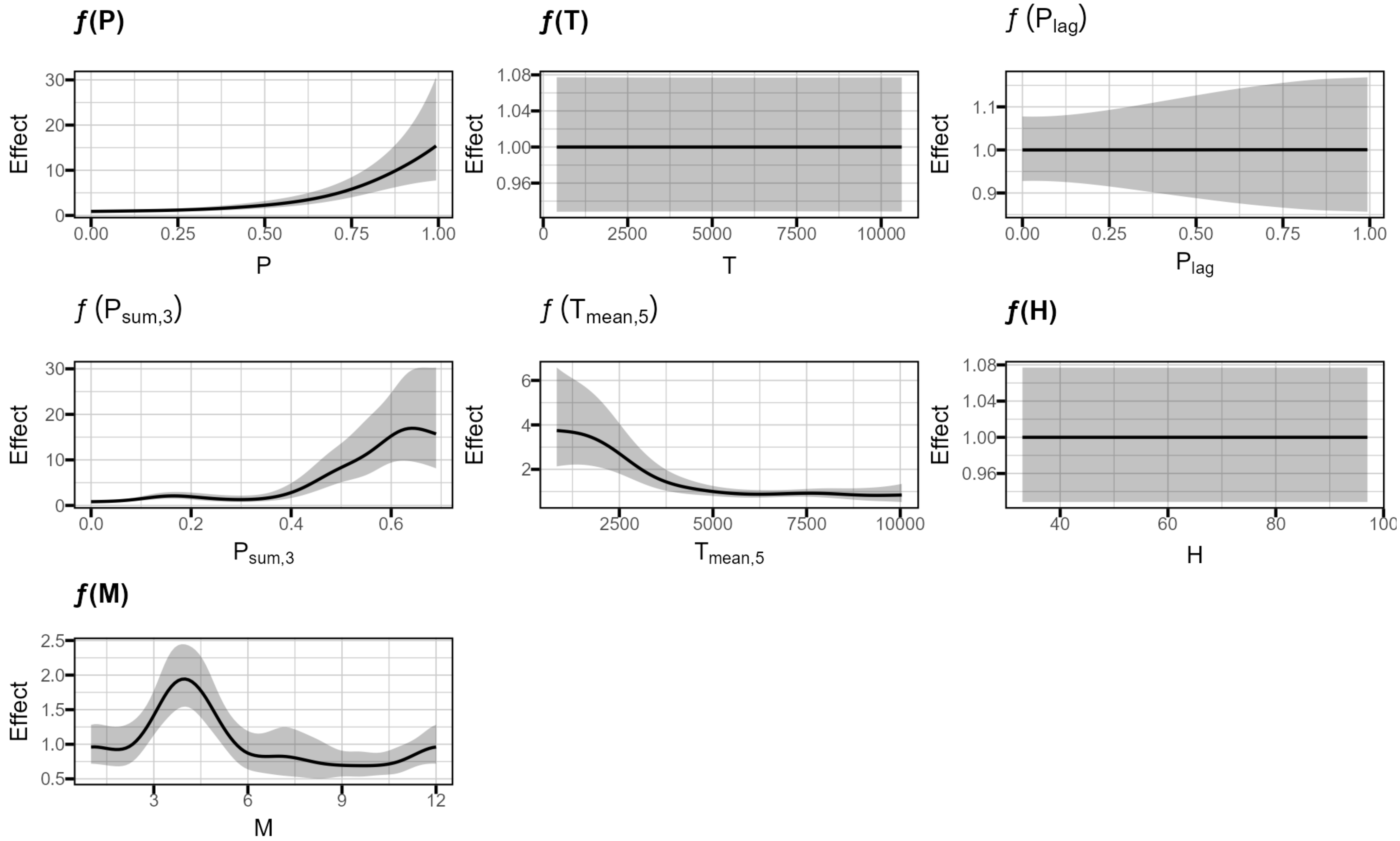
Ries, K. G. and P. J. Friesz (2000). *Methods for Estimating Low-Flow Statistics for Massachusetts Streams*. Technical Report 00-4135. U.S. Geological Survey, p. 81.

Schramm, M. P., M. Bevelhimer, and C. Scherelis (2017). "Effects of hydrokinetic turbine sound on the behavior of four species of fish within an experimental mesocosm". In: *Fisheries Research* 190, pp. 1-14. ISSN: 01657836. DOI: 10.1016/j.fishres.2017.01.012.

Schramm, M., T. Broad, and T. Arsuffi (2018). *Escherichia coli and Dissolved Oxygen Trends in the Upper Llano River Watershed, Texas (2001-2016)*. Technical Report TR-511. College Station, Texas: Texas Water Resources Institute, p. 27. URL: https://twri.tamu.edu/media/1458/tr-511.pdf.

Zakwan, M. (2018). "Spreadsheet-based modelling of hysteresis-affected curves". In: *Applied Water Science* 8.4, p. 101. DOI: 10.1007/s13201-018-0745-3.

Extra Slides


f(P)
Effect
30
20
10
0
0.00
0.25
0.50
0.75
1.00
P
f(T)
1.08
1.04
1.00
0.96
0
2500
5000
7500
10000
T
f (P_lag)
1.1
1.0
0.9
P_lag
f (P_sum,3)
0.0
0.2
0.4
0.6
P_sum,3
f (T_mean,5)
6
4
2
T_mean,5
f(H)
40
60
80
100
H
f(M)
2.5
2.0
1.5
1.0
0.5
3
6
9
12
M

GAM Summary 16396

Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	2.034	0.038	53.804	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(ewood_precip)	2.638	9.000	13.269	***
	s(ewood_tmax)	0.000	9.000	0.000	
	s(lagPrecip)	0.001	9.000	0.000	
	s(wetness)	5.372	9.000	22.621	***
	s(et)	4.383	9.000	3.662	***
	s(ewood_rh)	0.000	9.000	0.000	
	s(month)	5.959	8.000	6.926	***

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

Adjusted R-squared: 0.304, Deviance explained 0.801 -REML : 1114.800, Scale est: 0.553, N: 387