Estimation of Nonlinear Trends Using High-frequency Water-quality Monitoring Data

Guoxiang Yang*, Douglas Moyer

U.S. Geological Survey, Richmond, VA

*Natural Systems Analysts

11th National Monitoring Conference
March 27th, 2019, Denver, CO

Disclaimer: This information is preliminary or provisional and is subject to revision. It is being provided to meet the need for timely best science. The information has not received final approval by the U.S. Geological Survey (USGS) and is provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the information.

USGS
science for a changing world

Motivation

- ❖ Increasing collection of high-frequency water data
- ❖ A common question: is water quality improving over time
- ❖ Traditional trend estimation method
- ❖ Trend estimation method for high frequency data

Motivation

- Start with seasonal-trend decomposition procedure
  - Seasonal, trend and the remainder components
  - Atmospheric CO_2 concentration and temperature
- Need a universal approach to model different constituents
  - Ranging from water temperature to turbidity
- Generalized Additive Models (GAMs)

Study Area

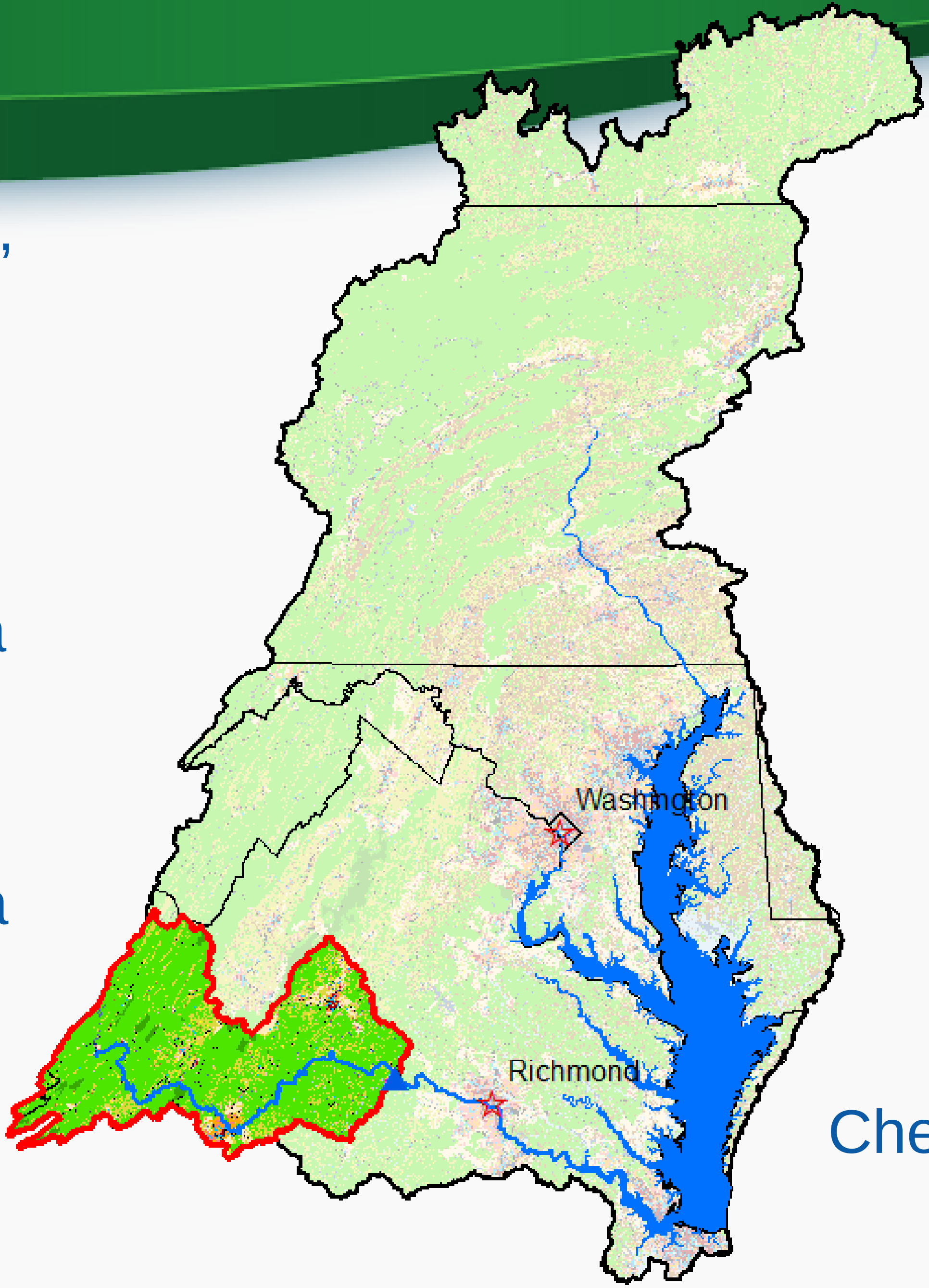
James River at Cartersville, VA (02035000)

Area: 16,190 km^2

The longest river in Virginia

15-min high frequency data

Mean daily value for trend estimation



2012 land use map

Urban: 7.5%

Ag: 16%

Forest: 73%

Chesapeake Bay

USGS
science for a changing world

Generalized Additive Model (GAM)

Generalized: many response distributions other than normal

Additive: terms add together

$$g\left(E(y_i)\right) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \ldots + \varepsilon_i \qquad (1)$$

g is a link function

$y_i \sim$ some exponential family distribution

f_1, f_2 … are unknown smooth functions

x_1, x_2 … are covariates

ε_i can have random effects

USGS
science for a changing world

GAM Theory - Smooth Function

Smooth function (spline): sum of basis functions *b*

and their corresponding regression coefficients β

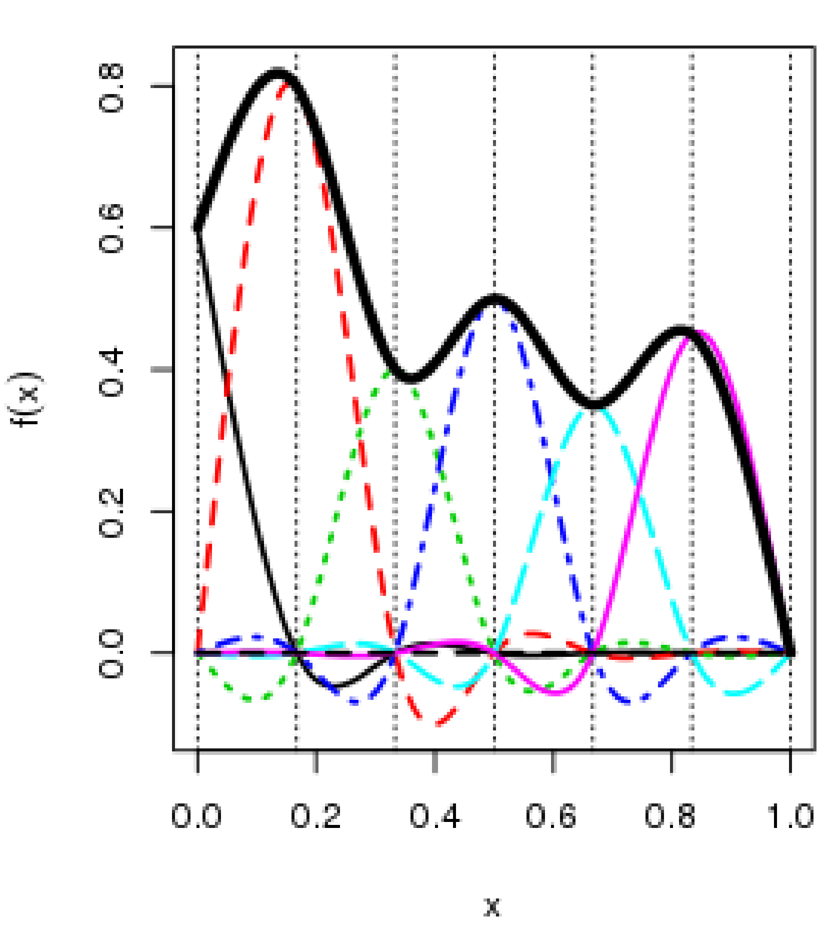
$$f(x) = \sum_{j=1}^{k} b_j(x)\beta_j$$

k: number of knots - basis dimensions used for the spline

The model (1) can be written in a linear way:

$$g\big(E(y)\big) = X\beta + \varepsilon \qquad (2)$$

GAM - a generalized linear model (GLM) with linear predictor involving a sum of smooth functions of covariates, subject to smoothing penalties



USGS
science for a changing world

GAM Theory – Correlation in Data

- GAMs assume that residuals are identically and independently distributed (i.i.d.)

  This assumption is too strong for daily time series

- Time-series data: sequential time points will be highly correlated
- Include autoregressive moving average (ARMA) model for errors in GAM model
- Generalized Additive Mixed Model (GAMM): to handle residual correlation

$$g\big(E(y)\big) = X\beta + \varepsilon \qquad (2)$$

$$g\big(E(y)\big) = X\beta + A + \varepsilon \qquad (3)$$

Water Temperature GAM Model

Water Temperature ~ s(season) + s(time)

```
Formula:
WT ~ s(nDay, bs = "cc", k = 20) + s(Time, bs = "cr", k = 20)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.2111     0.1055   153.7   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
           edf Ref.df       F p-value
s(nDay) 13.169 18.000 367.694 < 2e-16 ***
s(Time)  7.378  7.378   3.871 0.00025 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.94
  Scale est. = 5.1415    n = 4401
```

s(nDay) is the smoothing function of seasonal covariate.

s(time) is the smoothing function of time covariate.

"cr": cubic spline basis

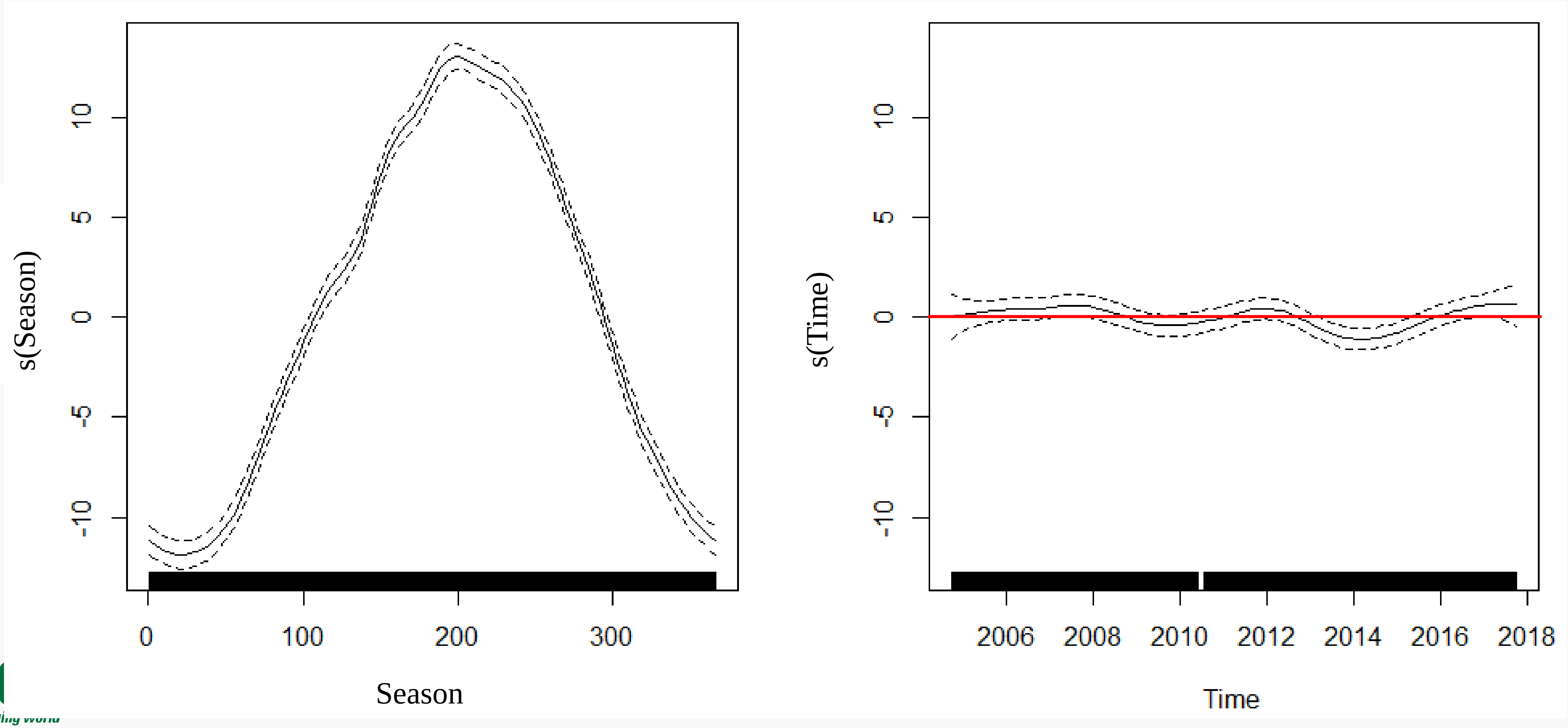
"cc": cyclic cubic spline

Adj-R^2

USGS
science for a changing world

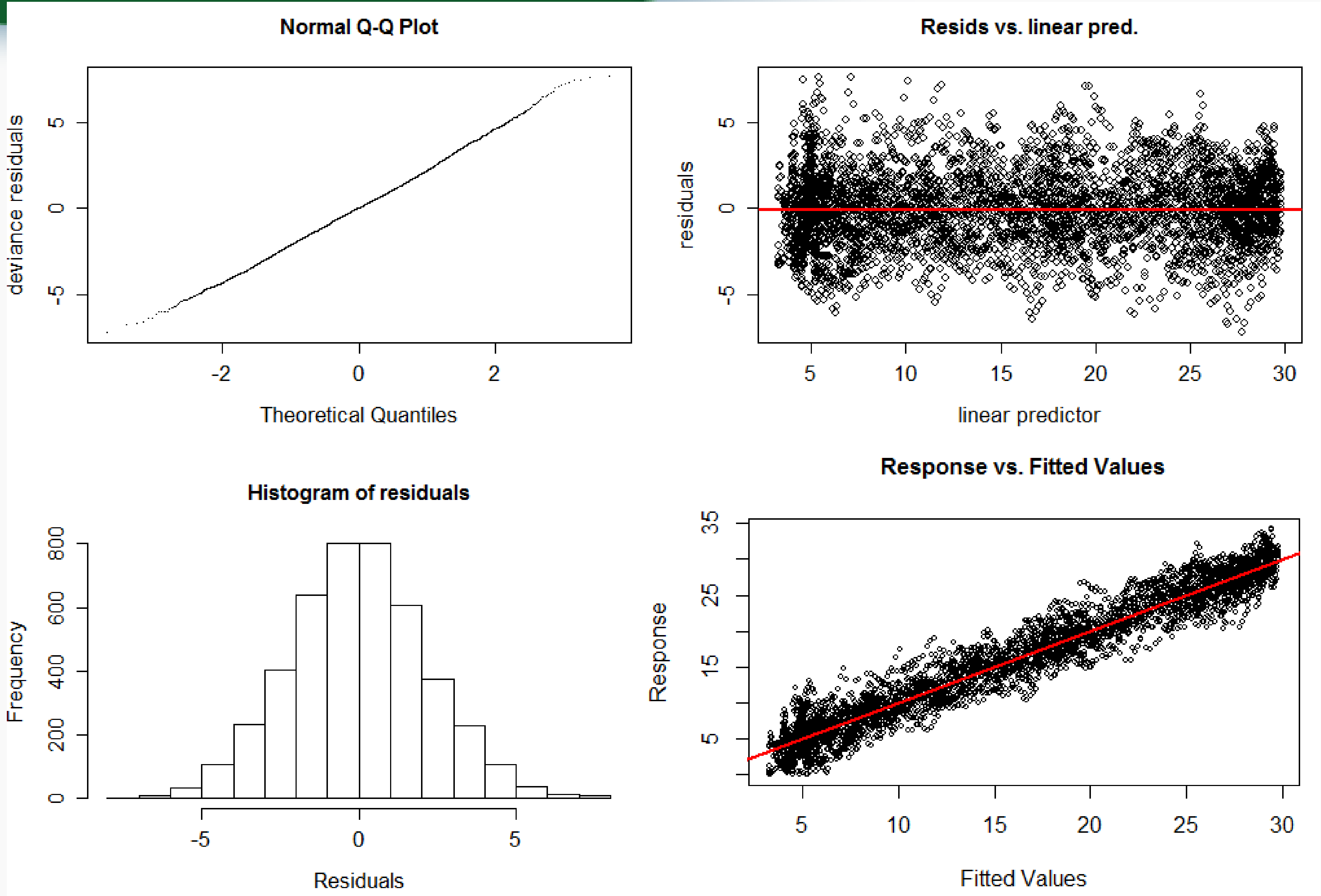
Water Temperature GAM Model

- Water Temperature ~ s(season) + s(time)
- The effect of each smooth function on water temperature

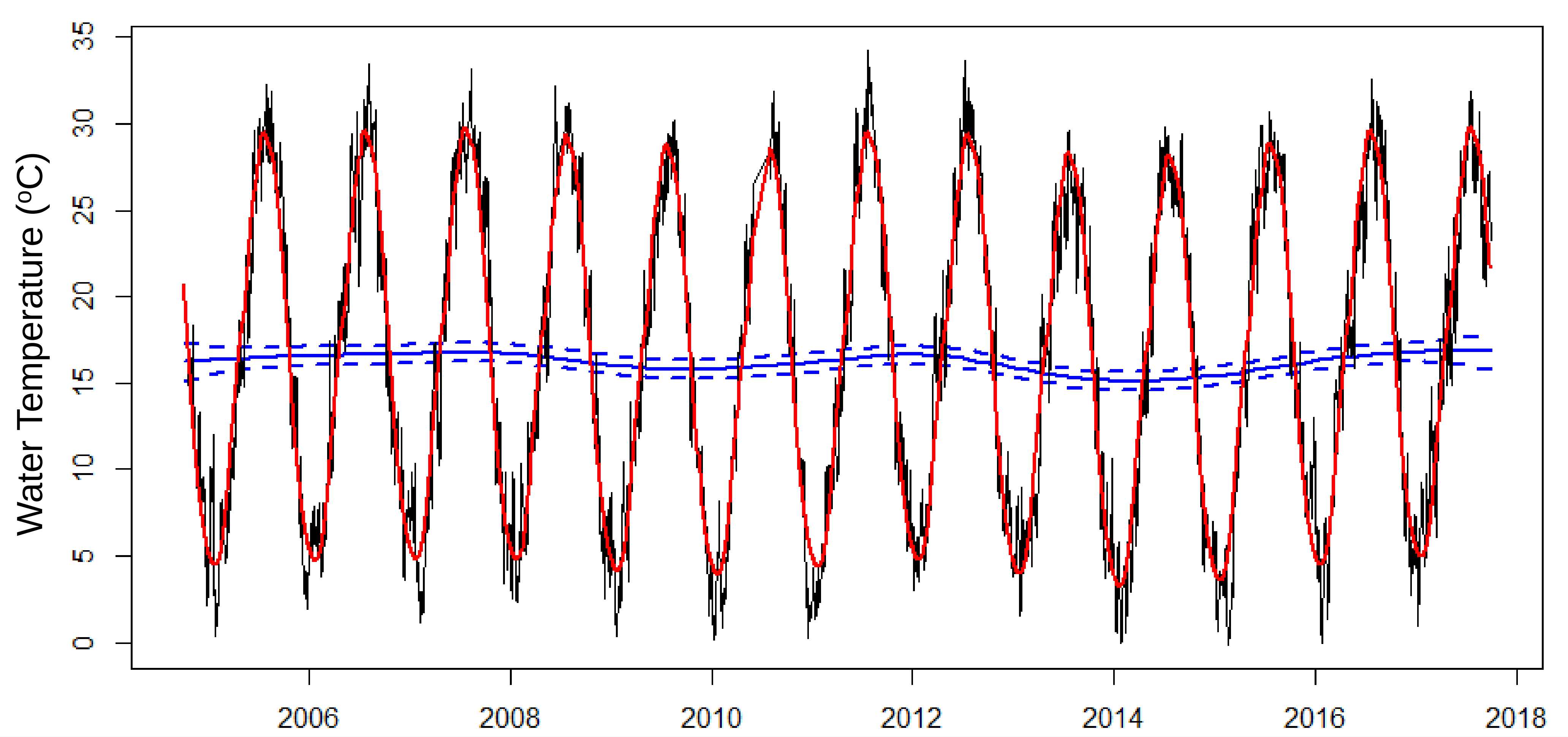


USGS
science for a changing world

Water Temperature GAM Model



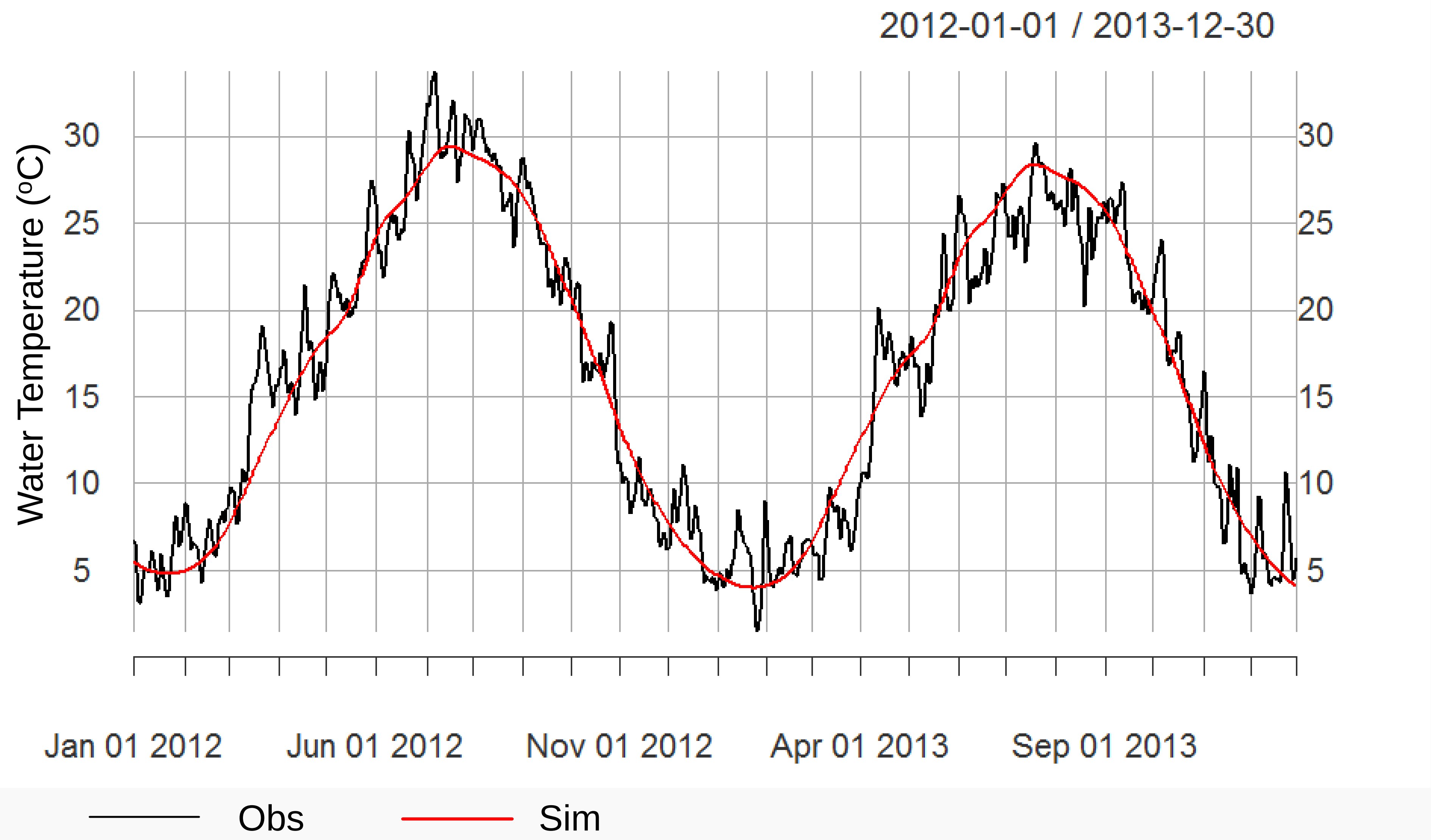
Water Temperature – Time Series



Obs Sim s(Time) Mean = 16.21 °C

USGS
science for a changing world

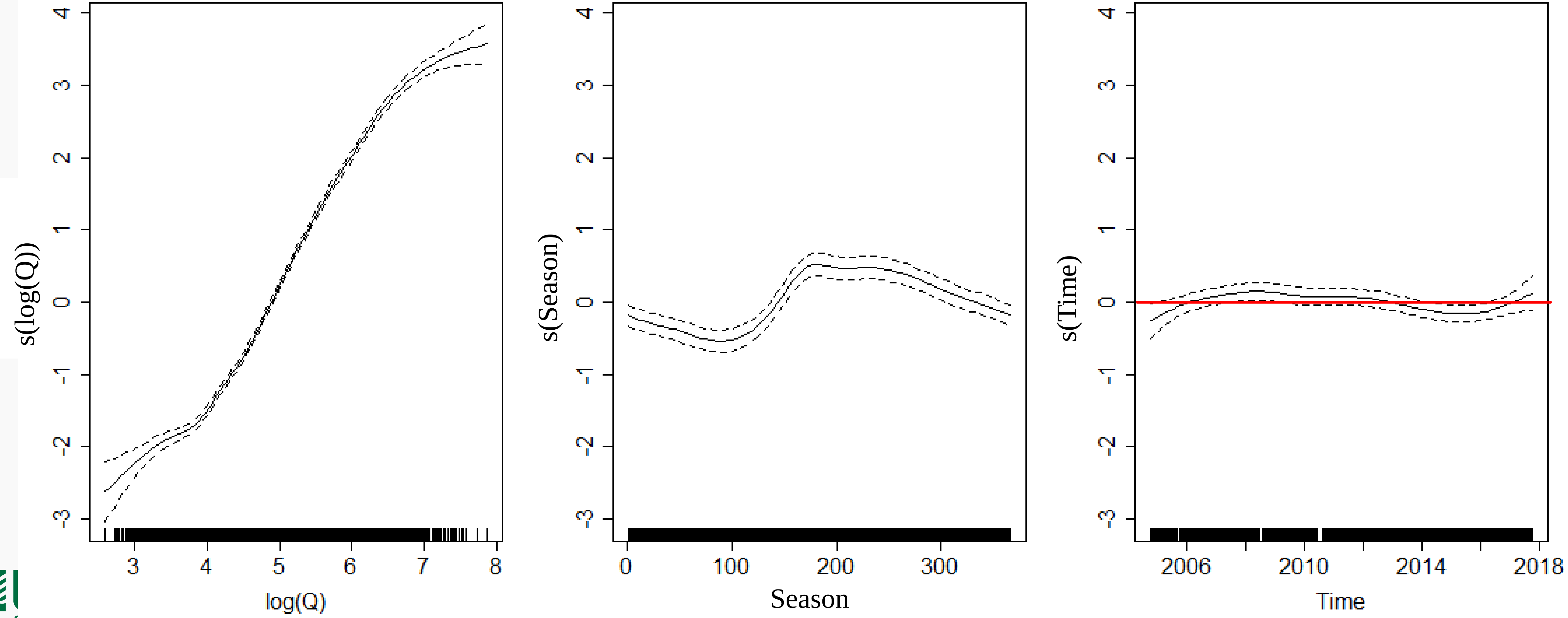
Water Temperature – Time Series



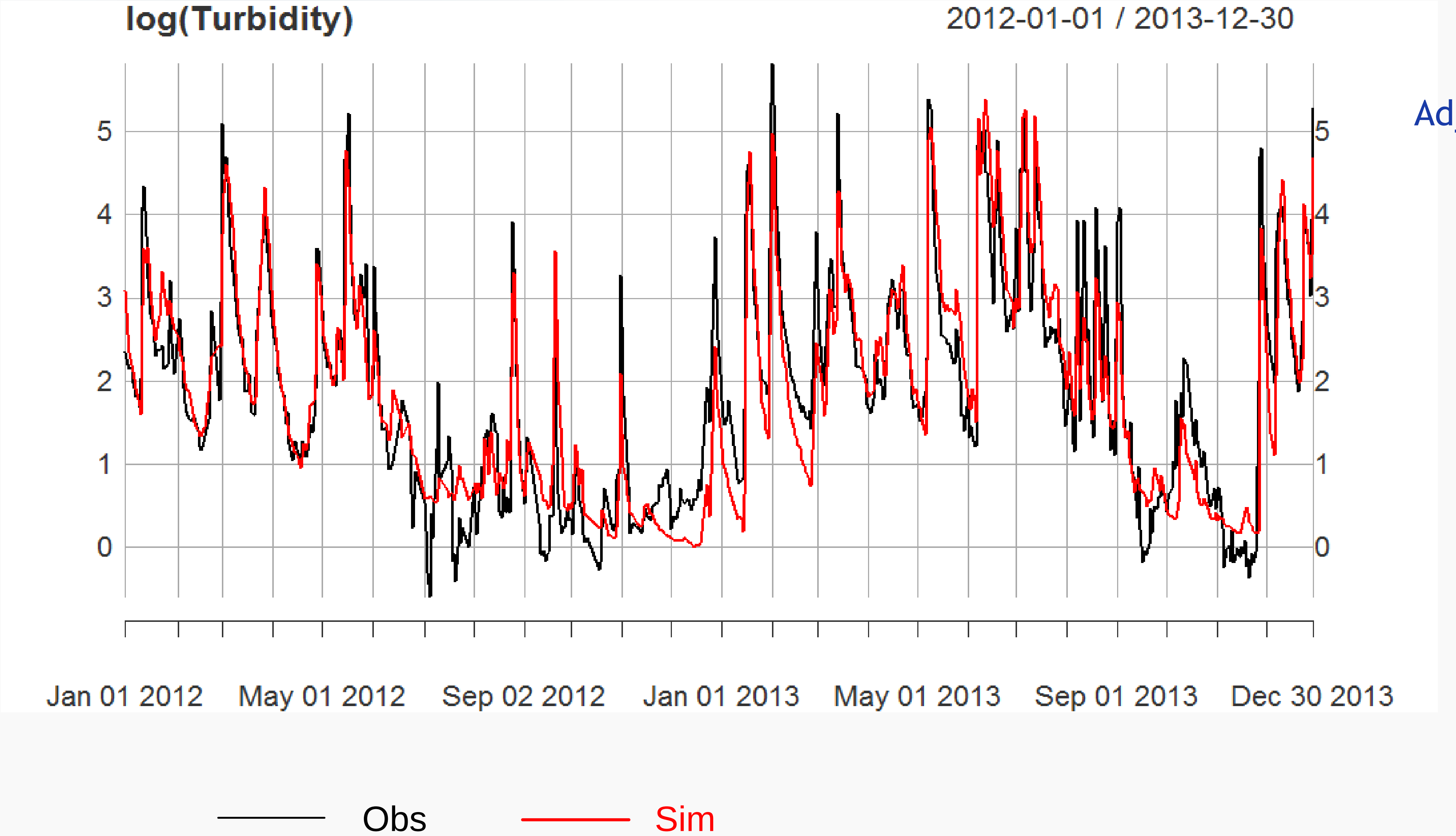
USGS
science for a changing world

Turbidity GAM Model

- Log(Turbidity) ~ s(log(Q)) + s(season) + s(time)
- The effect of each smooth function on turbidity



Log-Turbidity - Simulation vs. Observation



Adj-R^2 = 0.79

USGS science for a changing world

Trend & CI Estimation Between Two End-points

- GAM model has an underlying parametric representation

$$\hat{\eta}_p = f(time) = X_p\hat{\beta}$$

$\hat{\eta}_p$: smooth function of time component

X_p, β are model matrix and coefficient vector

- Derive the *mean* of the difference between the two endpoints (*trend*) and the standard error of this difference
- Conduct statistical inference using standard statistical theory
- Confidence interval and p-value

Water Temperature Trend Estimation

- Time period: 2007-01-01 ~ 2014-12-31
  - Trend over time period: -1.27°C

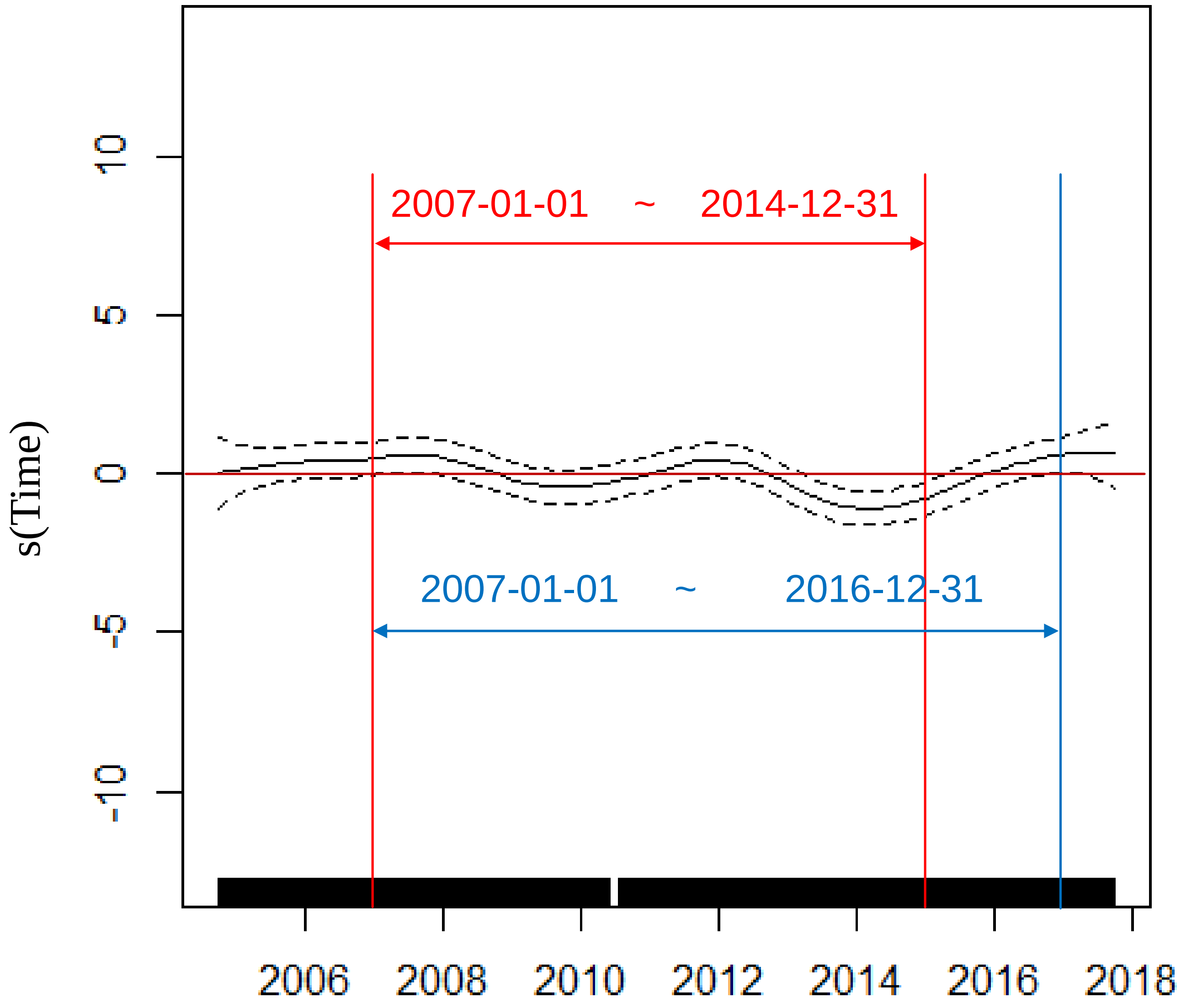
    95% confidence interval: -2.06 ~ -0.48

    p-value: < 0.01

- Time period: 2007-01-01 ~ 2016-12-31
  - Trend over time period: 0.12°C

    95% confidence interval: -0.7 ~ 0.93

    p-value: 0.78



Turbidity Trend Estimation

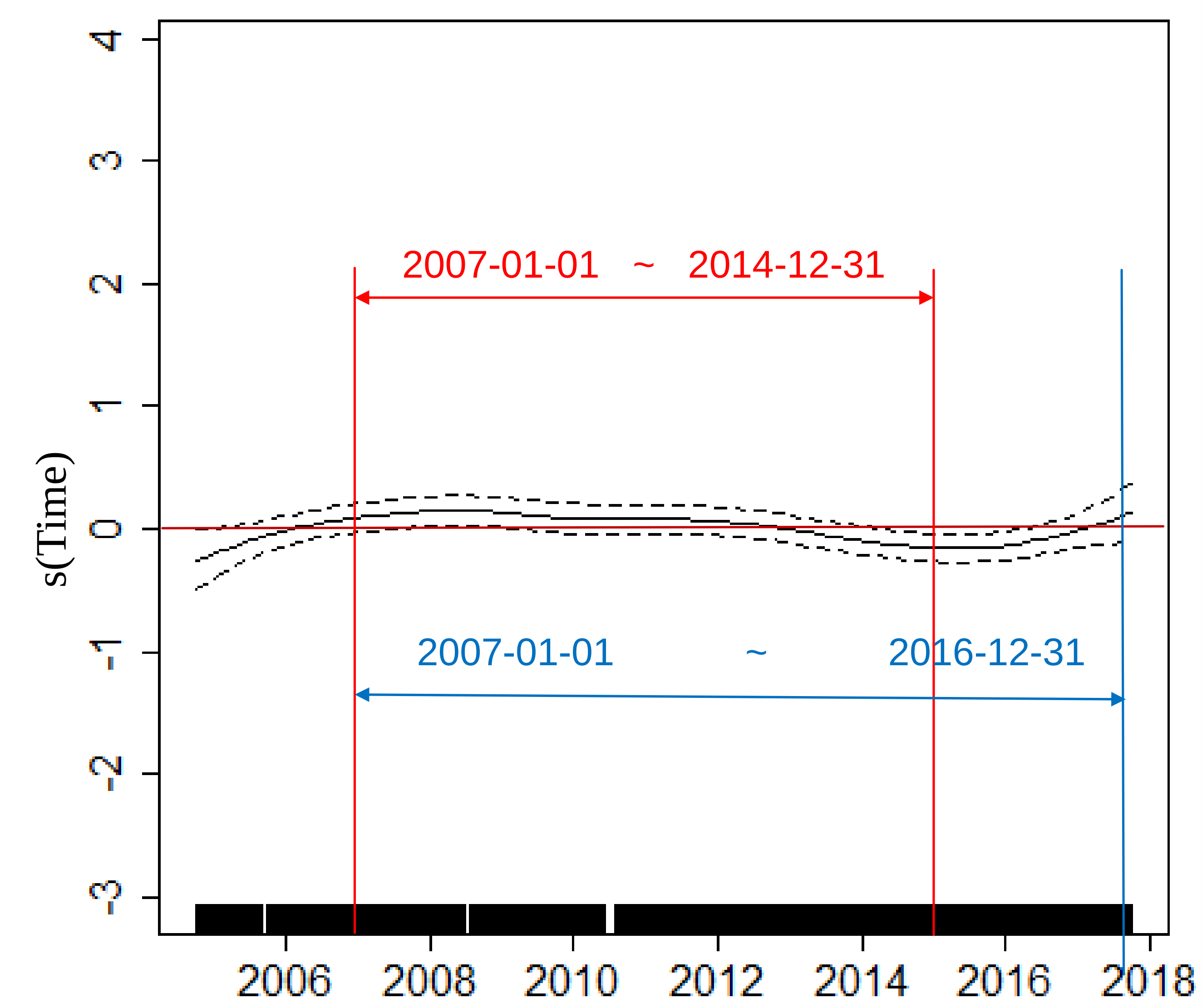
- Time period: 2007-01-01 ~ 2014-12-31
  - Trend (log scale): -0.24 FNU ($p < 0.01$)

    95% confidence interval: -0.42 ~ -0.06

- Time period: 2007-01-01 ~ 2016-12-31
  - Trend (log scale): -0.10 FNU ($p = 0.30$)

    95% confidence interval: -0.30 ~ 0.09

- Interpretation of the trend signal between estimation endpoints
  - How can we utilize the variability in the trend time series to better understand/link water-quality responses to episodic changes in the watershed (e.g. stream restoration, floods, or urbanization)?



GAM Trend Estimation Summary

- It is flexible to incorporate many possible predictors
- Decompose the signal into different components over time
- What does the trend pattern looks like over time
- Test if the trend is significant over a given time period
- Statistical inference, confidence interval calculation
- Can short-term trend signals be linked to restoration/disturbance events in the watershed?

Thank you!

Acknowledgments

Jeff Chanat, James Webber, John Jastram and Aaron Porter

Gavin Yang

gyang@usgs.gov

Water-quality monitors deployed on the James River at Cartersville, VA