An updateable statistical model for estimating future water quality exceedances and uncertainty

Casey Harris[1], AJ Reisinger[1], and Wendy Graham[2]

[1] University of Florida Soil, Water, and Ecosystem Sciences
[2] University of Florida Water Institute

Contact: caseyharris@ufl.edu

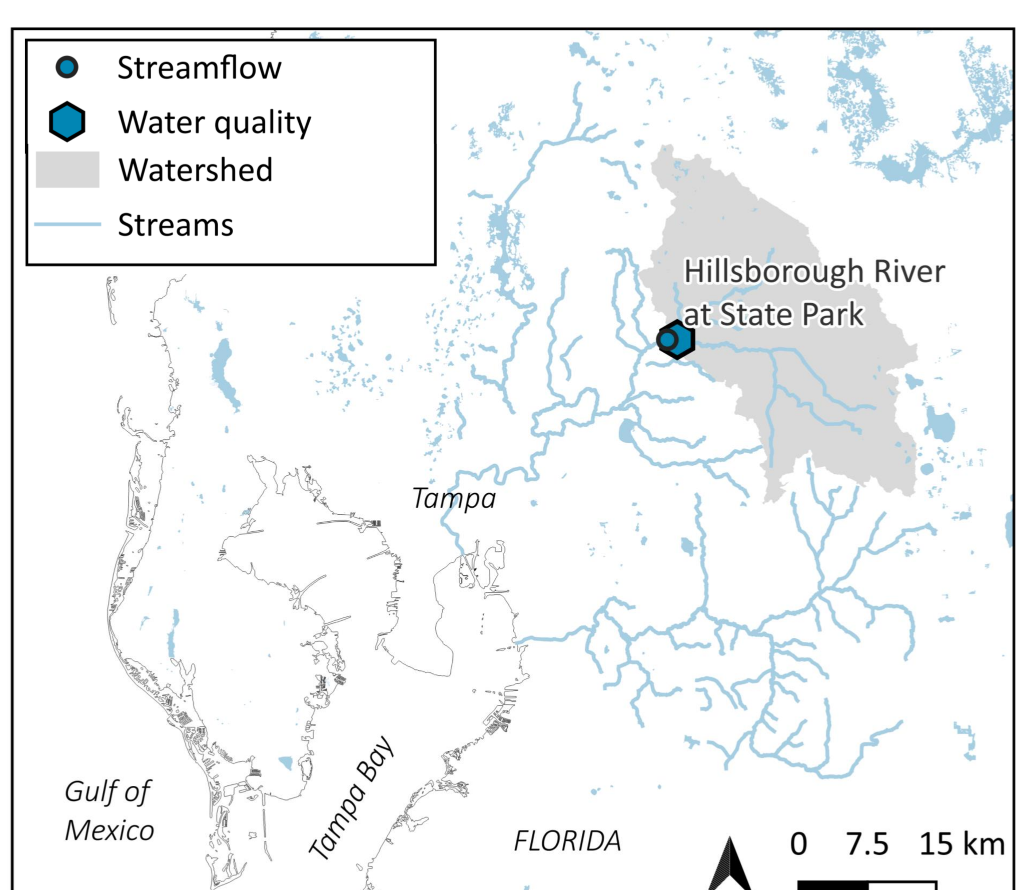
UF IFAS UNIVERSITY of FLORIDA | SOIL, WATER, AND ECOSYSTEM SCIENCES

Summary

- We show an example of projecting future water quality in a changing climate, using generalized additive models (GAMs) and a Bayesian approach for uncertainty estimation
- Past and future streamflow (Chang et al., 2018), along with day of year, are used as model predictors
- **What is the probability of total organic carbon (TOC) exceeding given thresholds in the Hillsborough River in the past and future?**
- Extending the analysis to other water quality constituents could be used to identify future water quality constituents of concern, highlighting the potential influence of changes in streamflow on future water quality

Location details

- Hillsborough River at State Park
- (USGS gage 02303000)
- Watershed includes natural areas, a spring, agricultural operations, rural development, some mining
- Designated Class I – Potable Water Supply, **water is withdrawn downstream for public supply** for the Tampa Bay area



Background

Generalized additive models (GAMs) flexibly describe relationships, so that the same approach may be suitable for different sites and responses

- Covariates include smoothing functions, allowing non-linear relationships

GAMs can be evaluated with a Bayesian approach, a helpful way of estimating uncertainty

- Markov Chain Monte Carlo (MCMC) simulations can be used to obtain posterior distributions of smoothing parameters and model coefficients
- Posterior distributions can be used to simulate responses (i.e., many predictions of total organic carbon (TOC) at a given streamflow and day of year)
- Inferences (i.e., means, percentiles, and measures of uncertainty) can be summarized from the simulated responses

Why total organic carbon (TOC)?

- TOC comes from natural humic substances from plants, sometimes from algae or animal waste
- TOC can react with water treatment disinfectants, forming harmful byproducts
- The EPA generally requires some percentage of TOC removal before water treatment, based on alkalinity and starting TOC:

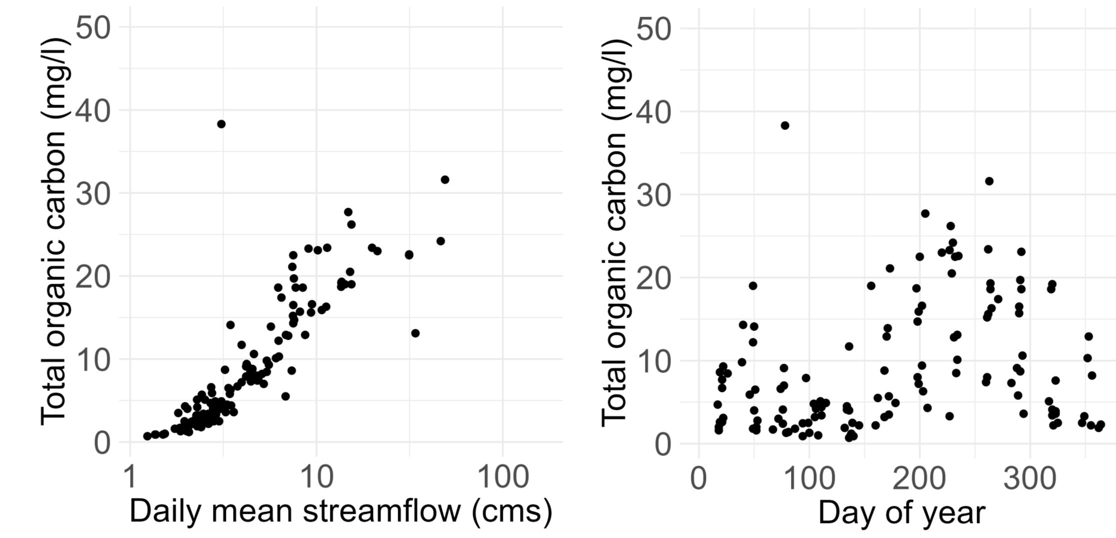
Percent removal of TOC required

Source water TOC (mg/l)	Source water alkalinity (mg/l as $CaCO_3$): 0 to 60	>60 to 120	>120
>2 to 4	35%	25%	15%
>4 to 8	45%	35%	25%
>8	50%	40%	30%

- Our example considers a threshold of **8 mg/l TOC**

Example

1) Input datasets (EPCHC, 2023; USGS, 2023)

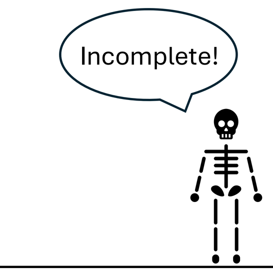


TOC generally increases with streamflow, and TOC is generally higher in summer/fall

2) Formulate model and define priors

$$\text{GAM:}\quad C \sim s(Q, bs = \text{"tp"}, k = 4) + s(d, bs = \text{"cc"}, k = 4)$$

C is log-transformed TOC concentration
s() are smooth terms
Q is log-transformed streamflow,
using a thin plate spline ("tp") with 4 knots
d is the Julian day of the year,
using a cubic spline ("cc") with 4 knots



We defined (uninformed) prior distributions

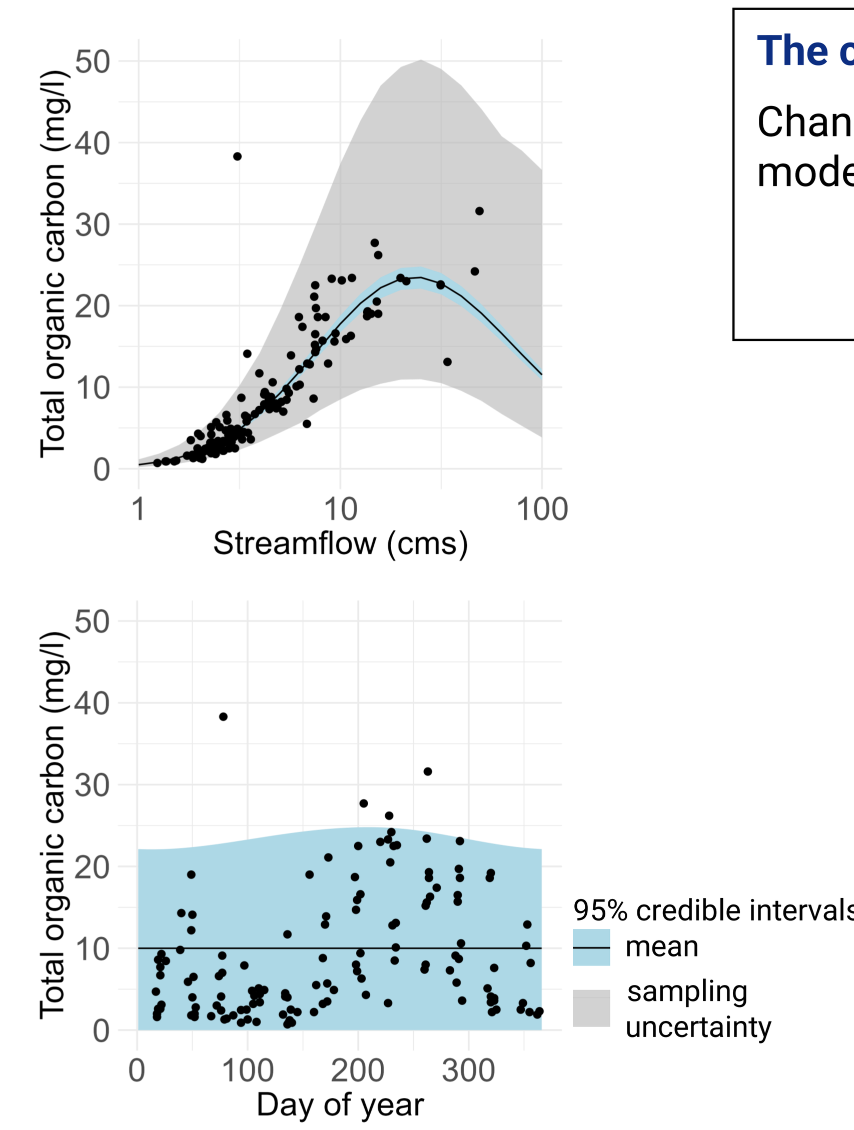
```
Model = model {
	mu <- X %*% b
		mu is the expected response
		X is the design matrix
		b is a vector of coefficients

	for (i in 1:n) {
			y[i] ~ dnorm(mu[i], tau) }
		prior for y, the estimated response
		normally distributed,
		with mean (mu) and precision (tau)

	…other priors for precision, parametric
	effects (normal), and smoothing
	parameters

	jagam() can help!
}
```

3) Obtain posterior distributions using MCMC

$$\text{jags(model, data)}$$

The R package JAGS analyses Bayesian models using MCMC

4) Summarize exceedance probabilities from the posterior

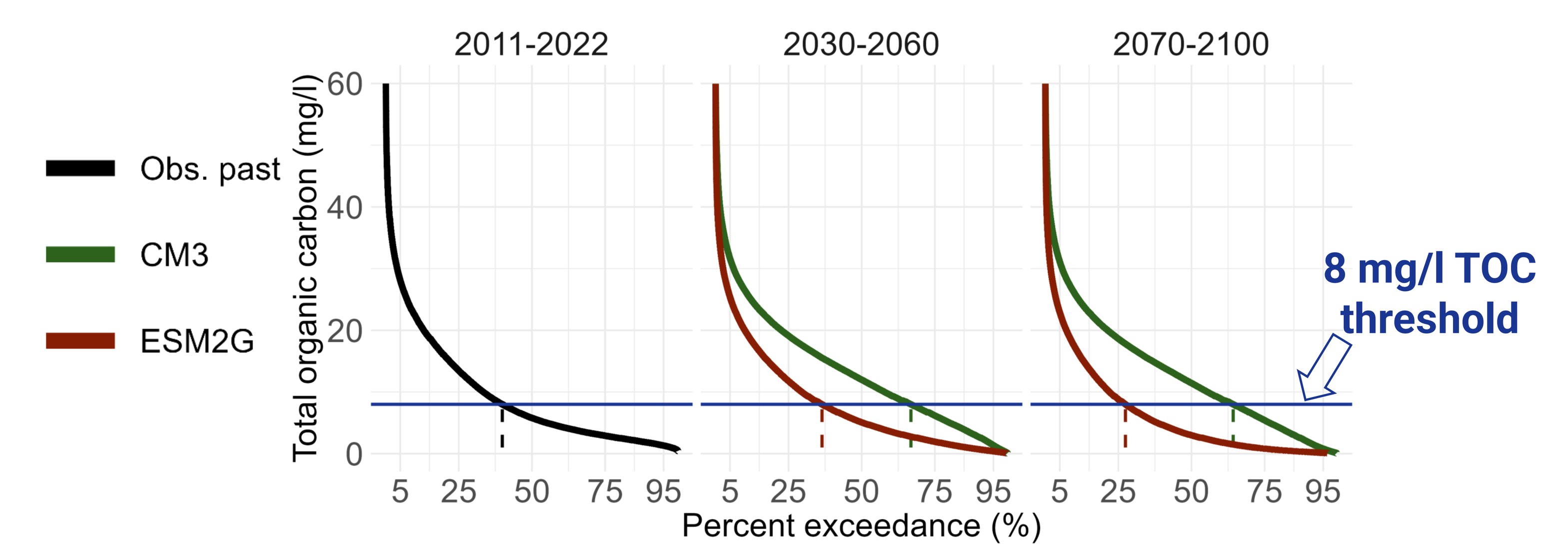


Model predictions show about the same amount of spread as the observed data, and most variation is due to streamflow rather than day of year

The climate models and future streamflow

Chang et al. (2018) estimated future streamflow in the Tampa Bay area based on climate models from the Coupled Model Intercomparison Project (CMIP5). Two of these were:

CM3 – predicts a wetter, hotter climate
ESM2G – predicts a drier, hotter climate



Percent of simulated values with TOC > 8 mg/l with 95% credible intervals

Dataset	Time period: 2011-2022	2030-2060	2070-2100
Obs. past	40% [38–43%]		
CM3		69% [66–72%]	66% [63–69%]
ESM2G		38% [36–41%]	28% [27–31%]

We expect more TOC exceedances in a wetter, hotter future climate, and fewer TOC exceedances in a drier, hotter future climate

Next steps

- Check and verify the process
- Repeat the analysis with additional climate models, and for additional water quality characteristics
- Include additional hydroclimatic variables in the GAM (i.e., preceding flow, precipitation, ET, wind speed, or solar radiation)
- Test the updateability!
  - Update hydroclimatic projections using CMIP6 climate models instead of CMIP5
  - Update streamflow projections using updated hydroclimatic projections
  - Update future water quality exceedance probabilities

References

(1) Environmental Protection Comm. of Hillsborough County (EPCHC). 2023. Online water quality data. Acc. Dec. 2023. https://www.epchc.org/divisions/water/water-monitoring-maps-and-data
(2) U.S. Geological Survey (USGS). 2023. National Water Information System data (USGS Water Data for the Nation). Acc. Dec. 2023. http://waterdata.usgs.gov/nwis/
(3) Chang, S., W. Graham, J. Geurink, N. Wanakule, and T. Asefa. 2018. Evaluation of impacts of future climate change and water use scenarios on regional hydrology. Hyd. and Earth Sys. Sci., 22(9).

Acknowledgements

Funding provided by Tampa Bay Water