

# Tracking San Francisco Bay water quality using generalized additive models in an R Shiny framework

Marcus W. Beck<sup>1</sup> Ian Wren<sup>2</sup> Rebecca Murphy<sup>3</sup>  
Perry de Valpine<sup>4</sup> David Senn<sup>5</sup>

<sup>1</sup>Tampa Bay Estuary Program

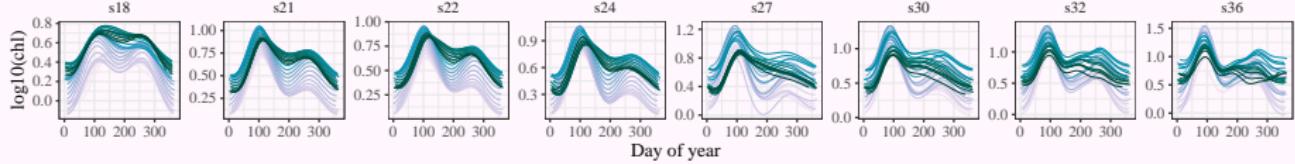
<sup>2</sup>San Francisco Baykeeper

<sup>3</sup>Chesapeake Bay Program

<sup>4</sup>University of California Berkeley

<sup>5</sup>San Francisco Estuary Institute

Nov. 4, 2019



# Why do we care about trends?

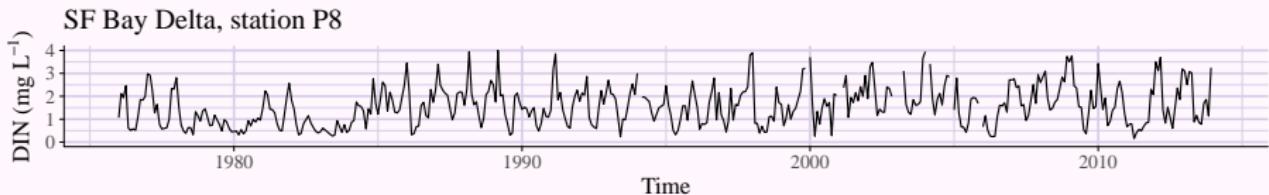


- Provide information on natural variation of water quality parameters - identify 1st order principles to understand a system
- Document historical changes in response to management actions - did investments make a difference?
- Anticipate future changes with proposed restoration or management - understand the past to predict the future

# Trends vary in space and time



*Observed data represent effects from many processes*



## Climate

precipitation  
temperature  
wind events  
ENSO effects

## Local

light/turbidity  
residence time  
invasive species  
trophic effects

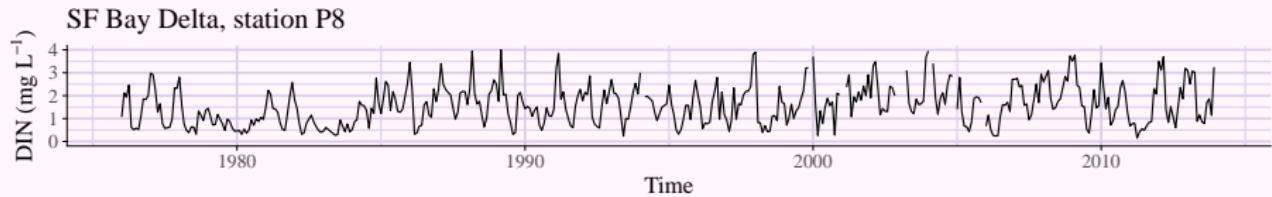
## Regional/historical

watershed inputs  
point sources  
management actions  
flow changes

# Must translate data into information



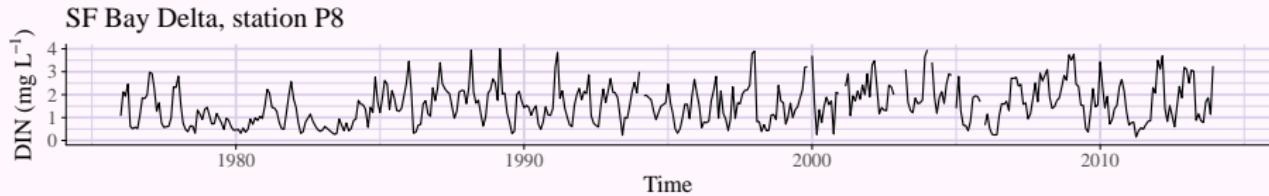
*Observed data represents effects of many processes*



*Models should describe components to evaluate effects*

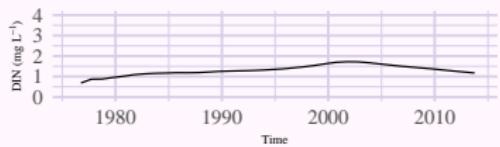
# Must translate data into information

*Observed data represents effects of many processes*

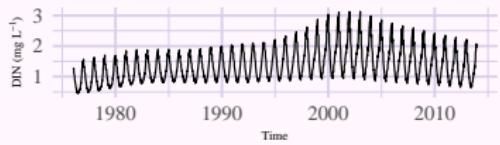


*Models should describe components to evaluate effects*

Annual



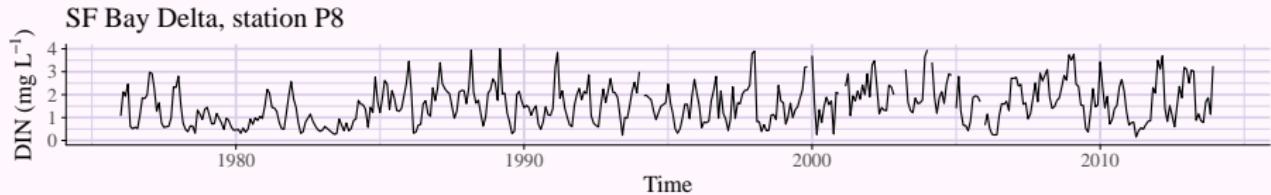
Seasonal



# Must translate data into information

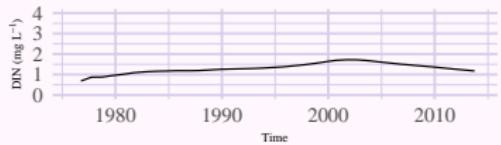


*Observed data represents effects of many processes*

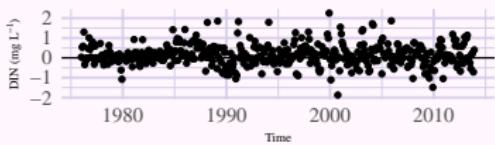


*Models should describe components to evaluate effects*

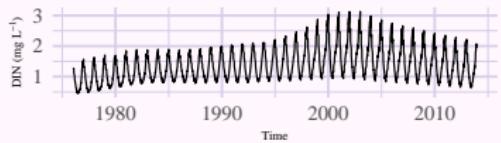
Annual



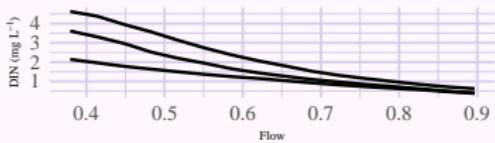
Residual



Seasonal



Flow effects



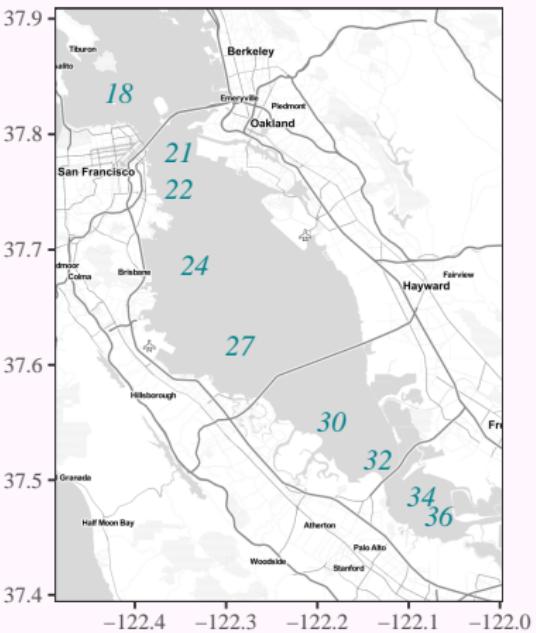
# South San Francisco Bay



- Historically a high-nutrients, high-turbidity, low-productivity system

[Cole and Cloern, 1984,  
Alpine and Cloern, 1988]

South San Francisco Bay  
Long-term monitoring stations



# South San Francisco Bay



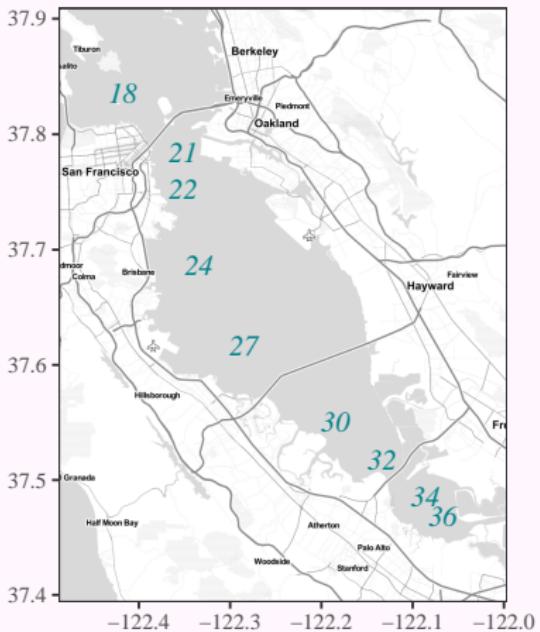
- Historically a high-nutrients, high-turbidity, low-productivity system

[Cole and Cloern, 1984,  
Alpine and Cloern, 1988]

- Recent increases observed in summer-fall chl-a concentrations

[Cloern et al., 2007, Cloern and Jassby, 2012]

South San Francisco Bay  
Long-term monitoring stations



# South San Francisco Bay



- Historically a high-nutrients, high-turbidity, low-productivity system

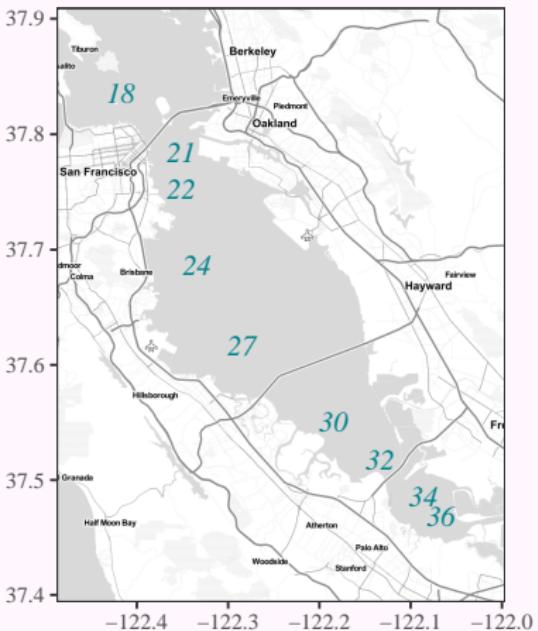
[Cole and Cloern, 1984,  
Alpine and Cloern, 1988]

- Recent increases observed in summer-fall chl-a concentrations

[Cloern et al., 2007, Cloern and Jassby, 2012]

- Lead to creation of a Nutrient Management Strategy (NMS) to characterize status/trends and management needs

South San Francisco Bay  
Long-term monitoring stations

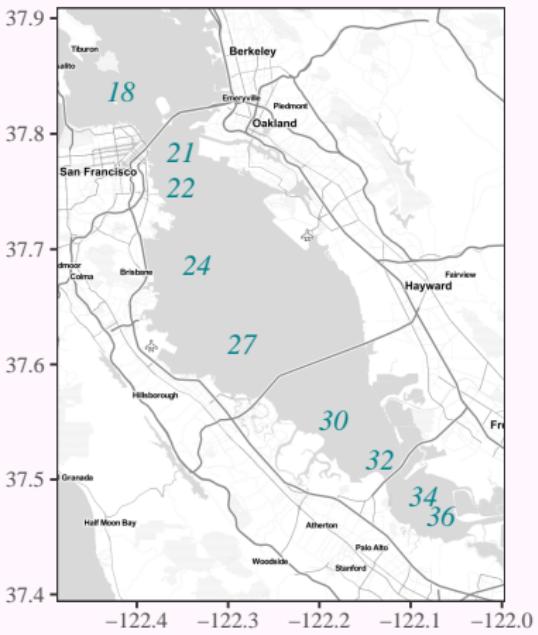


# South San Francisco Bay



Questions of concern:

## South San Francisco Bay Long-term monitoring stations



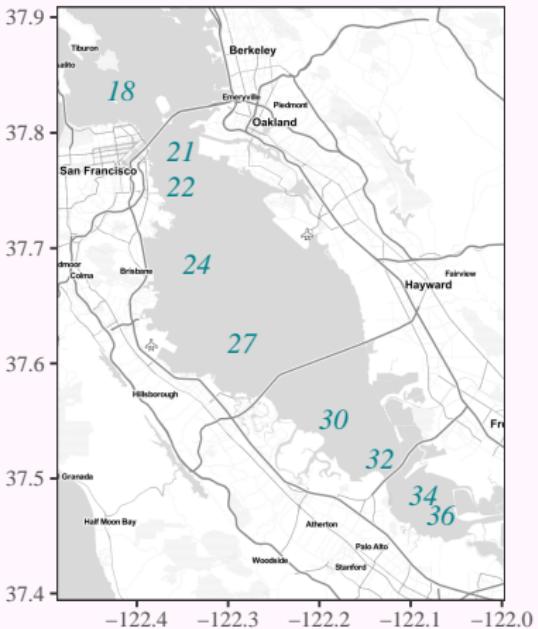
# South San Francisco Bay



## Questions of concern:

- Since changes are visually apparent, which are significant?

South San Francisco Bay  
Long-term monitoring stations

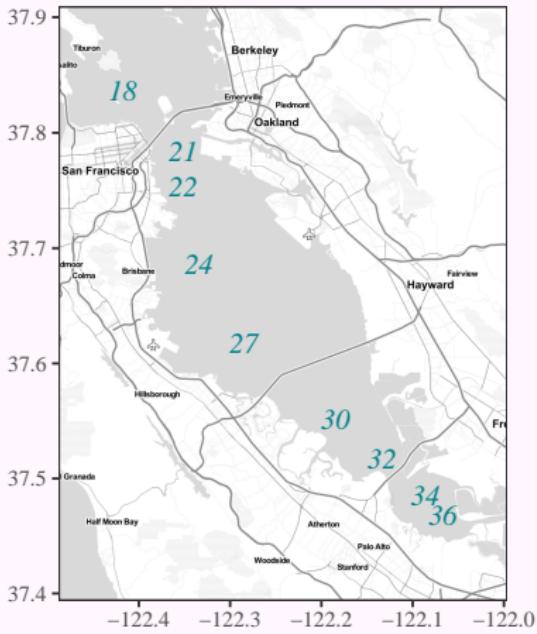


# South San Francisco Bay

Questions of concern:

- Since changes are visually apparent, which are significant?
- What has been the estimated rate and direction of any linear or non-monotonic change?

South San Francisco Bay  
Long-term monitoring stations



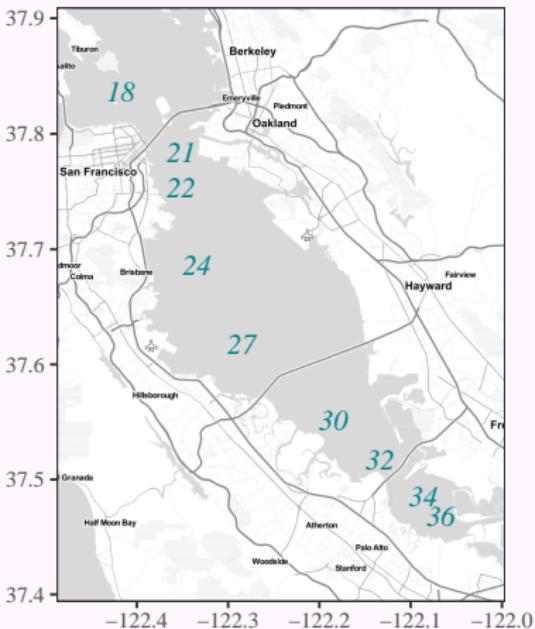
# South San Francisco Bay



Questions of concern:

- Since changes are visually apparent, which are significant?
- What has been the estimated rate and direction of any linear or non-monotonic change?
- Do any of these changes coincide with changes in other water quality parameters?

South San Francisco Bay  
Long-term monitoring stations



# Application of additive models



- The Chesapeake Bay Program (CBP) has been wrestling with similar issues, i.e., can a flexible statistical analysis method be applied to evaluate significant, non-linear changes in water quality parameters? [Beck and Murphy, 2017, Murphy et al., 2019b]
- We applied Generalized Additive Models (GAMs) developed by CBP to characterize long-term trends at nine stations over thirty years in South SF Bay
- An interactive website was also developed using R Shiny to explore trends and communicate results with stakeholders

# Application of additive models



For each station, chlorophyll was modelled as a function of annual and seasonal changes over time [Murphy et al., 2019a, baytrends R package]

Four GAMs were evaluated and compared using standard methods for model comparison ( $R^2$ , GCV)

# Application of additive models



For each station, chlorophyll was modelled as a function of annual and seasonal changes over time [Murphy et al., 2019a, baytrends R package]

Four GAMs were evaluated and compared using standard methods for model comparison ( $R^2$ , GCV)

- `gam0: chl ~ year + s(doy)`

# Application of additive models



For each station, chlorophyll was modelled as a function of annual and seasonal changes over time [Murphy et al., 2019a, baytrends R package]

Four GAMs were evaluated and compared using standard methods for model comparison ( $R^2$ , GCV)

- `gam0: chl ~ year + s(doy)`
- `gam1: chl ~ year + s(doy) + s(year)`

# Application of additive models



For each station, chlorophyll was modelled as a function of annual and seasonal changes over time [Murphy et al., 2019a, baytrends R package]

Four GAMs were evaluated and compared using standard methods for model comparison ( $R^2$ , GCV)

- `gam0`:  $\text{chl} \sim \text{year} + s(\text{doy})$
- `gam1`:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year})$
- `gam2`:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year}) + ti(\text{doy}, \text{year})$

# Application of additive models



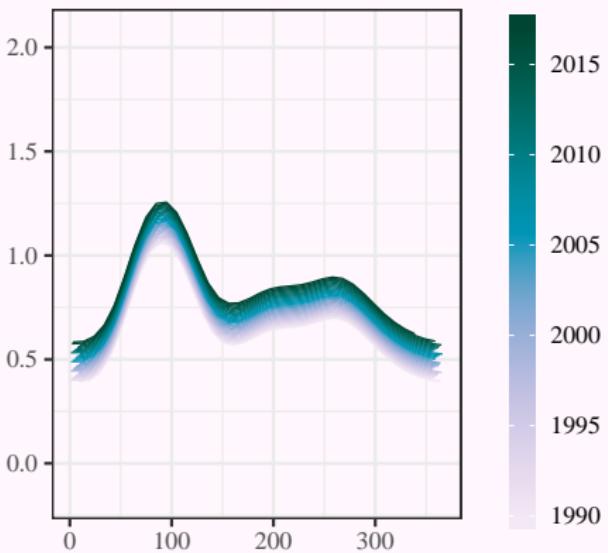
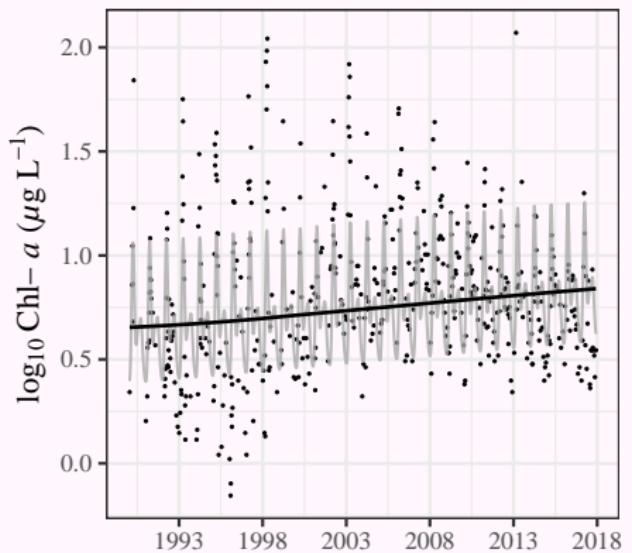
For each station, chlorophyll was modelled as a function of annual and seasonal changes over time [Murphy et al., 2019a, baytrends R package]

Four GAMs were evaluated and compared using standard methods for model comparison ( $R^2$ , GCV)

- `gam0`:  $\text{chl} \sim \text{year} + s(\text{doy})$
- `gam1`:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year})$
- `gam2`:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year}) + ti(\text{doy}, \text{year})$
- `gam6`:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year}, k = \text{large})$

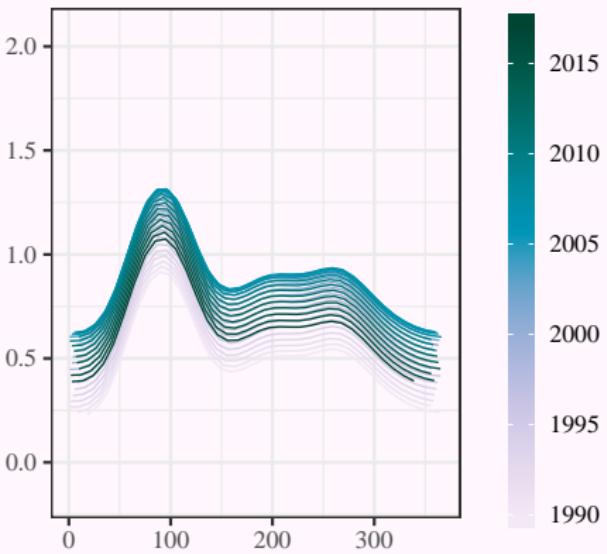
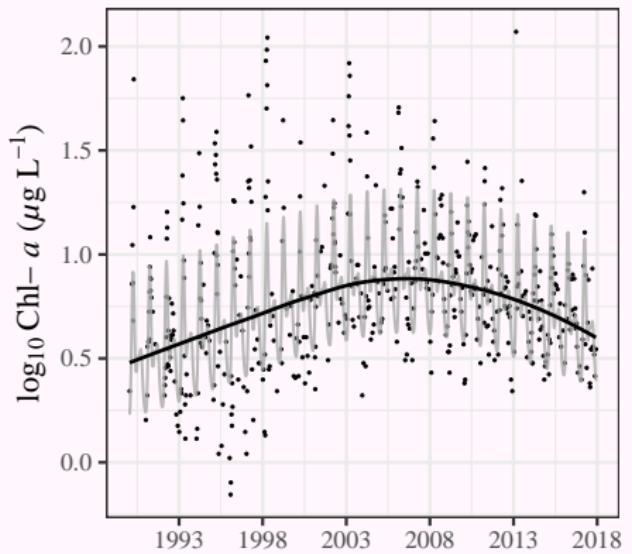
# Application of additive models

gam0: chl ~ year + s(doy)



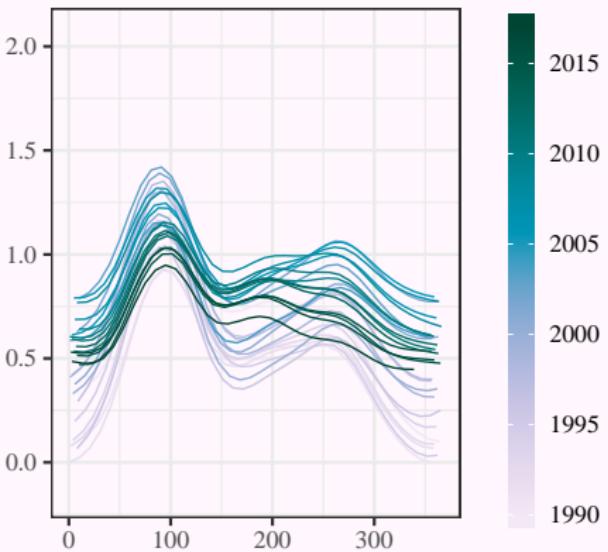
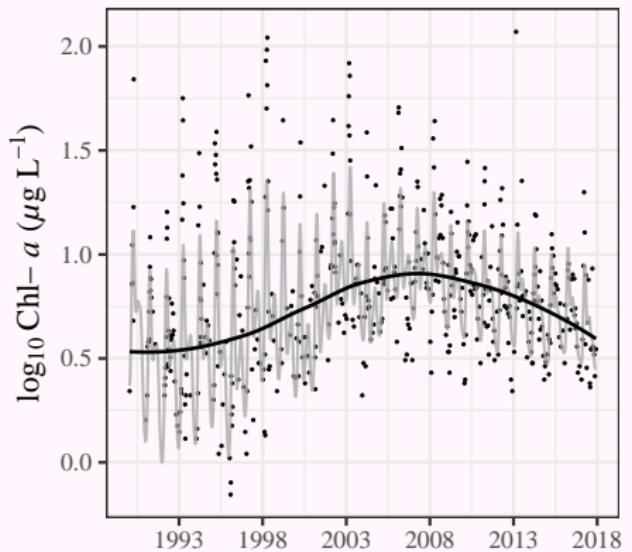
# Application of additive models

gam1: chl ~ year + s(doy) + s(year)



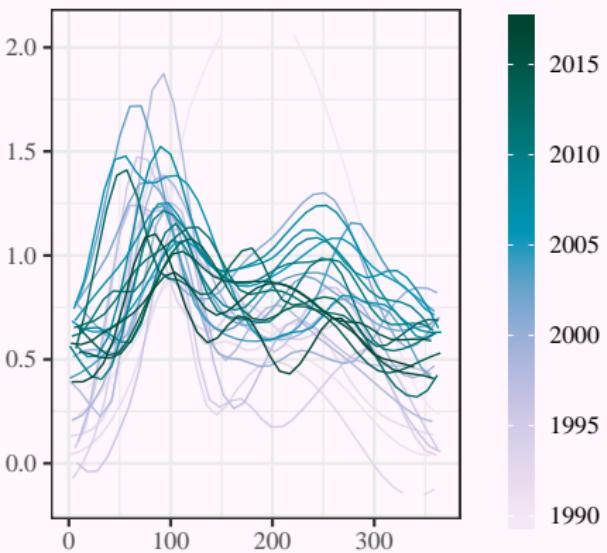
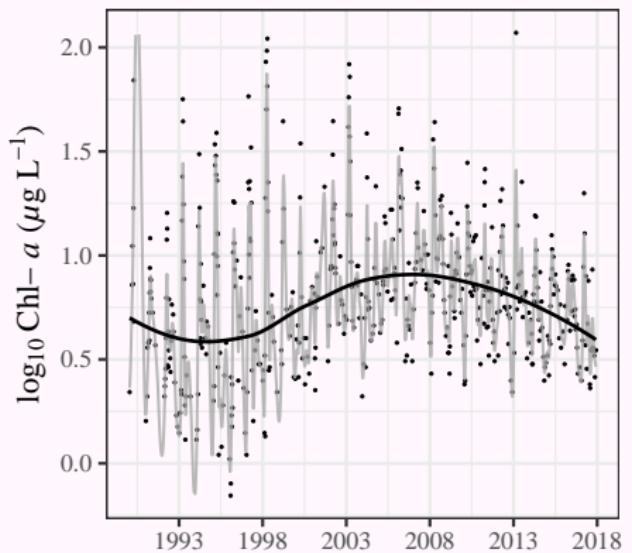
# Application of additive models

gam2:  $\text{chl} \sim \text{year} + s(\text{doy}) + s(\text{year}) + ti(\text{doy}, \text{year})$



# Application of additive models

gam6: chl ~ year + s(doy) + s(year, k = large)

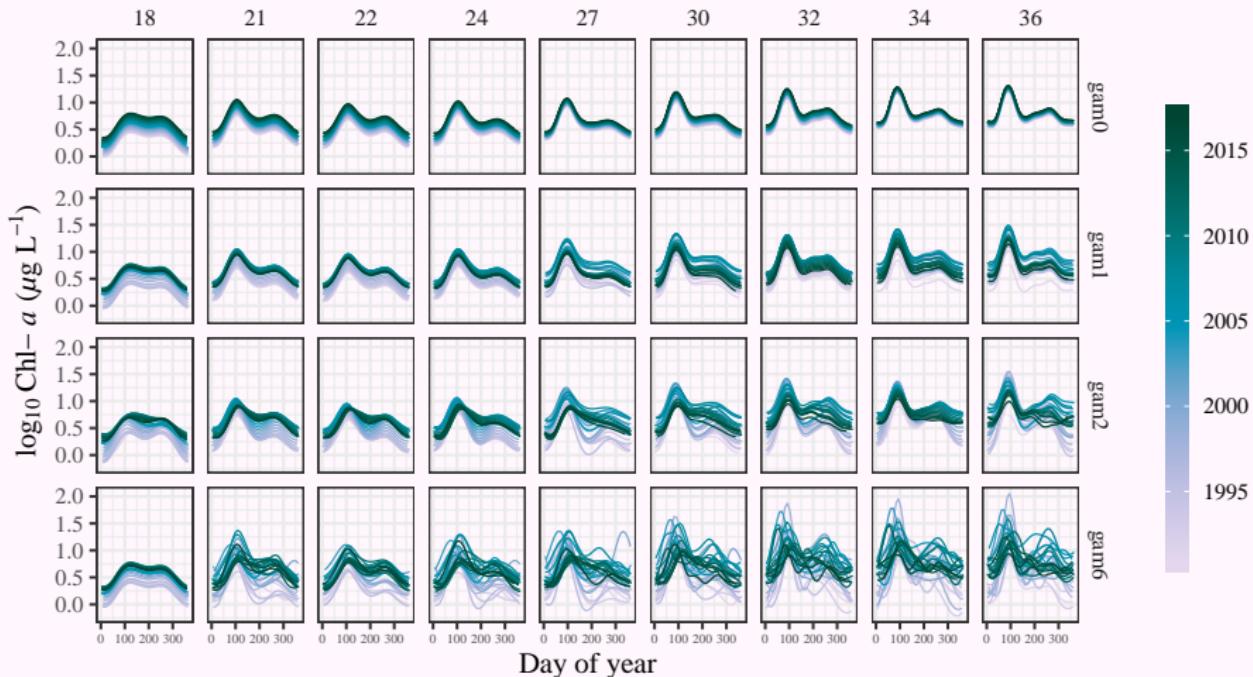


# Application of additive models



# Application of additive models

All years



# Application of additive models



Table: GCV summary statistics by station and model

Model	Stations (north to south)								
	18	21	22	24	27	30	32	34	36
gam0	-117.5	-30.6	-20.6	38.9	174.4	205.5	232.9	256.6	238.7
gam1	-138.8	-89.3	-70.3	-18.7	104.1	111.5	162.6	196	182
gam2	<b>-141.8</b>	-147.3	<b>-116.4</b>	-98.5	4	49.6	108.1	189.9	147.2
gam6	-139.5	<b>-235.2</b>	-116.3	<b>-176.8</b>	<b>-92.1</b>	<b>-115.9</b>	<b>-149</b>	<b>-3.3</b>	<b>-65.5</b>

# Application of additive models



Table: GCV summary statistics by station and model

Model	Stations (north to south)								
	18	21	22	24	27	30	32	34	36
gam0	-117.5	-30.6	-20.6	38.9	174.4	205.5	232.9	256.6	238.7
gam1	-138.8	-89.3	-70.3	-18.7	104.1	111.5	162.6	196	182
gam2	<b>-141.8</b>	-147.3	<b>-116.4</b>	-98.5	4	49.6	108.1	189.9	147.2
gam6	-139.5	<b>-235.2</b>	-116.3	<b>-176.8</b>	<b>-92.1</b>	<b>-115.9</b>	<b>-149</b>	<b>-3.3</b>	<b>-65.5</b>

Table: R-squared summary statistics by station and model

Model	Stations (north to south)								
	18	21	22	24	27	30	32	34	36
gam0	0.47	0.41	0.37	0.37	0.33	0.36	0.32	0.31	0.32
gam1	0.51	0.48	0.43	0.44	0.43	0.48	0.41	0.41	0.43
gam2	<b>0.53</b>	0.54	0.48	0.53	0.54	0.54	0.49	0.41	0.48
gam6	0.51	<b>0.68</b>	<b>0.54</b>	<b>0.66</b>	<b>0.68</b>	<b>0.72</b>	<b>0.75</b>	<b>0.69</b>	<b>0.75</b>

# Descriptive results of additive models



Extension to other response endpoints

# Shiny interactive web platform



Why do we need this? Synthesis of results in a communicable format  
Answer to specific questions clarity of description - how has seasonal component changed over time.  
Understand implications and limitations of different methods

# Shiny interactive web platform



## Example 1

# Shiny interactive web platform



## Example 2

# Shiny interactive web platform



## Example 3

# Summary and next steps



# References

- Alpine AE, Cloern JE. 1988.  
Phytoplankton growth rates in a light-limited environment, San Francisco Bay.  
Marine Ecology Progress Series, 44(2):167–173.
- Beck MW, Murphy RR. 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.
- Cloern JE, Jassby AD. 2012.  
Drivers of change in estuarine-coastal ecosystems: Discoveries from four decades of study in San Francisco Bay.  
Reviews of Geophysics, 50(4):1–33.
- Cloern JE, Jassby AD, Thompson JK, Hieb KA. 2007.  
A cold phase of the East Pacific triggers new phytoplankton blooms in San Francisco Bay.  
Proceedings of the National Academy of Sciences of the United States of America, 104(47):18561–18565.
- Cole BE, Cloern JE. 1984.  
Significance of biomass and light availability to phytoplankton productivity in San Francisco Bay.  
Marine Ecology Progress Series, 17(1):15–24.
- Murphy R, Perry E, Keisman J, Harcum J, Leppo EW. 2019a.  
baytrends: Long Term Water Quality Trend Analysis.  
R package version 1.1.0.
- Murphy RR, Perry E, Harcum J, Keisman J. 2019b.  
A Generalized Additive Model Approach to evaluating water quality: Chesapeake Bay case study.  
Environmenal Modelling & Software, 118:1–13.