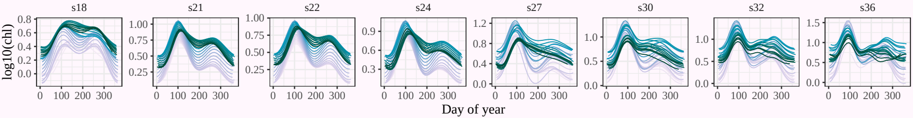
Tracking San Francisco Bay water quality using generalized additive models in an R Shiny framework

Marcus W. Beck[1] Ian Wren[2] Rebecca Murphy[3]
Perry de Valpine[4] David Senn[5]

[1]Tampa Bay Estuary Program
[2]San Francisco Baykeeper
[3]Cheseapeake Bay Program
[4]University of California Berkeley
[5]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

Why do we care about trends?

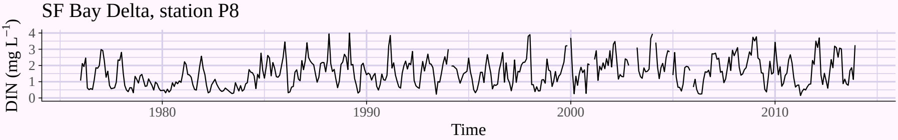
- Provide information on natural variation of water quality parameters - identify 1st order principles to understand a system
- Document historical changes in response to management actions - did investments make a difference?

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

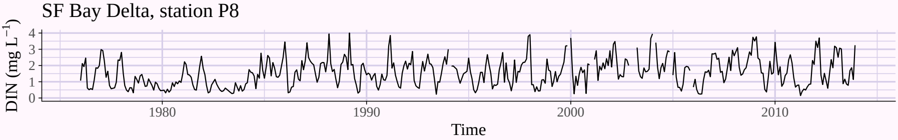
light/turbidity
residence time
invasive species
trophic effects

Regional/historical

watershed inputs
point sources
management actions
flow changes

Must translate data into information

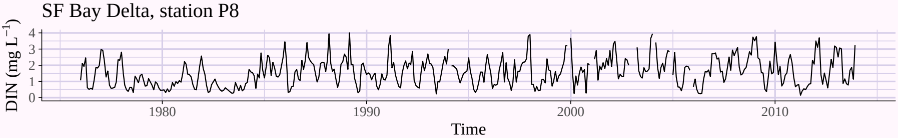
Observed data represents effects of many processes



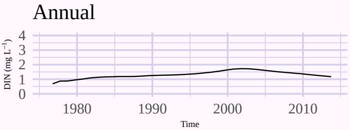
Models should describe components to evaluate effects

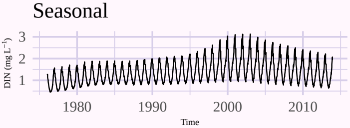
Must translate data into information

Observed data represents effects of many processes



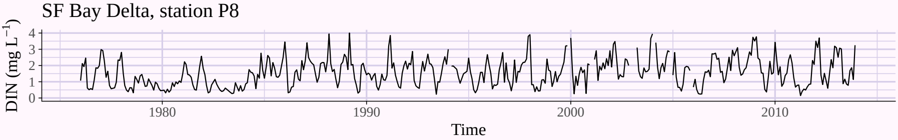
Models should describe components to evaluate effects



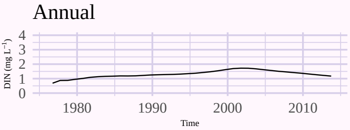


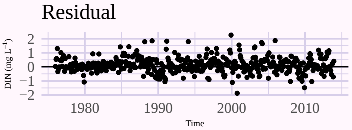
Must translate data into information

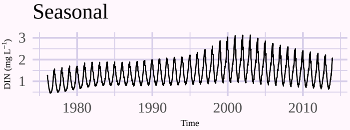
Observed data represents effects of many processes

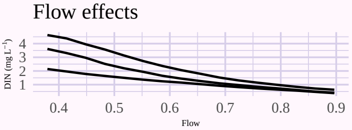


Models should describe components to evaluate effects







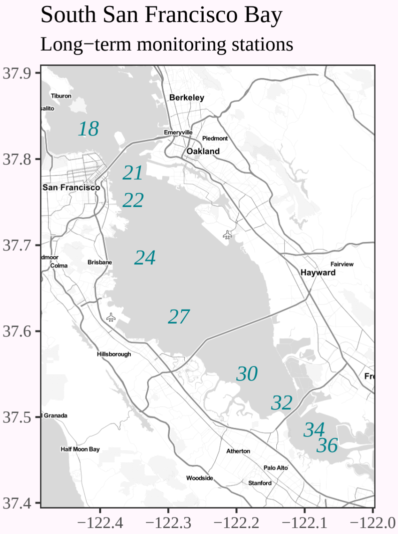


South San Francisco Bay

- A high-nutrient, high-turbidity, low-productivity system

  [Cole and Cloern, 1984,

  Alpine and Cloern, 1988]



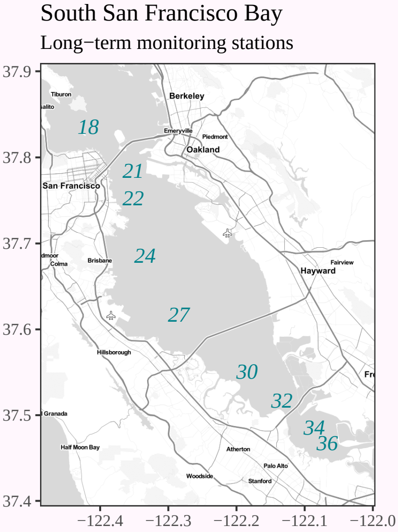
South San Francisco Bay

- A high-nutrient, high-turbidity, low-productivity system

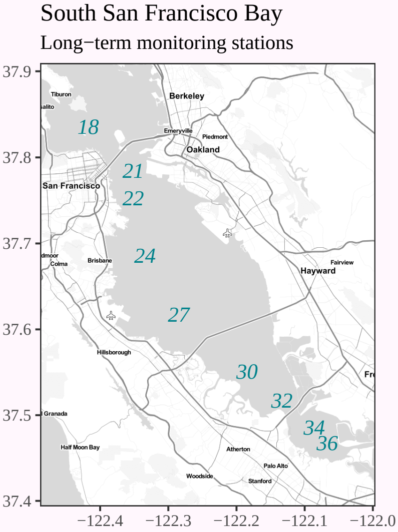
  [Cole and Cloern, 1984, Alpine and Cloern, 1988]

- Recent increases observed in summer/fall chl-a

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

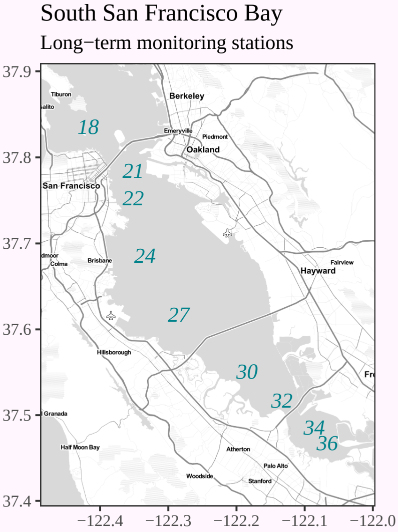


- A high-nutrient, high-turbidity, low-productivity system [Cole and Cloern, 1984, Alpine and Cloern, 1988]
- Recent increases observed in summer/fall chl-a [Cloern et al., 2007, Cloern and Jassby, 2012]
- Nutrient Management Strategy (NMS) to characterize status/trends and management needs



South San Francisco Bay

Questions of concern:



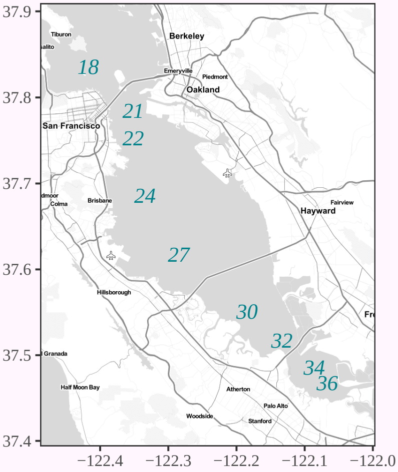
South San Francisco Bay

Questions of concern:

- Since changes are visually apparent, which are significant?

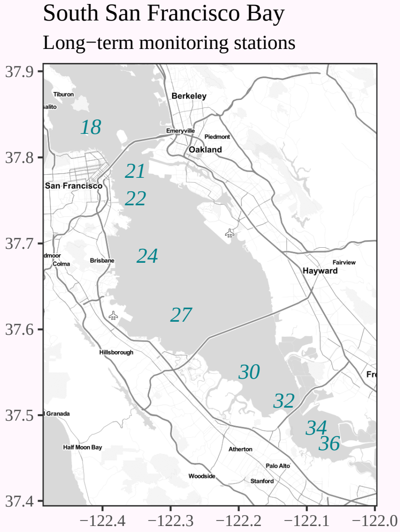
South San Francisco Bay

Long−term monitoring stations



Questions of concern:

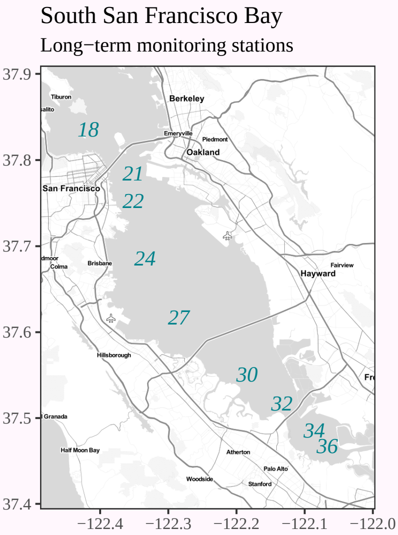
- Since changes are visually apparent, which are significant?
- What has been the estimated rate and direction of any linear or non-monotonic change?



South San Francisco Bay

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?



Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [Beck and Murphy, 2017, Murphy et al., 2019b]

Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [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

Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [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

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

    - `gam0`: chl $\sim$ year + s(doy)

Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [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

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

  - `gam0`: chl $\sim$ year + s(doy)
  - `gam1`: chl $\sim$ year + s(doy) + s(year)

Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [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

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

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

Application of additive models

- The Chesapeake Bay Program (CBP) has been wrestling with similar issues [Beck and Murphy, 2017, Murphy et al., 2019b]

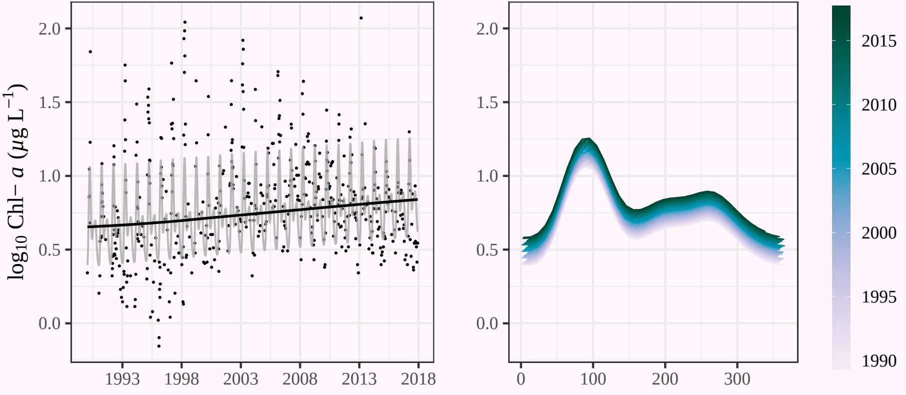
- We applied Generalized Additive Models (GAMs) developed by CBP to characterize long-term trends at nine stations over thirty years in South SF Bay

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

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

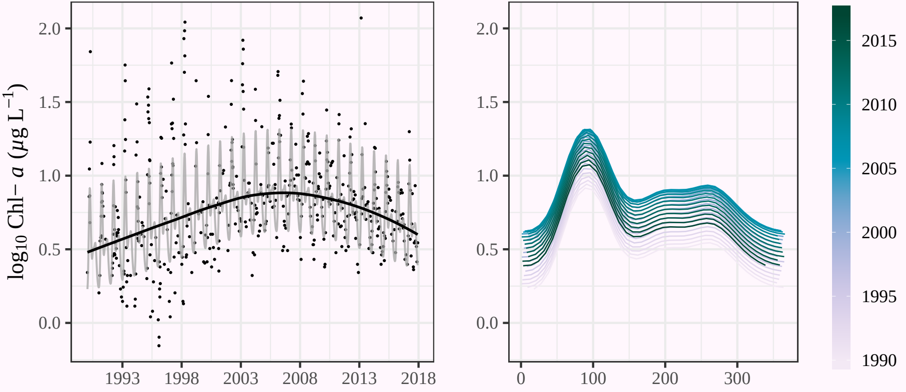
Application of additive models

`gam0`: chl $\sim$ year + s(doy)

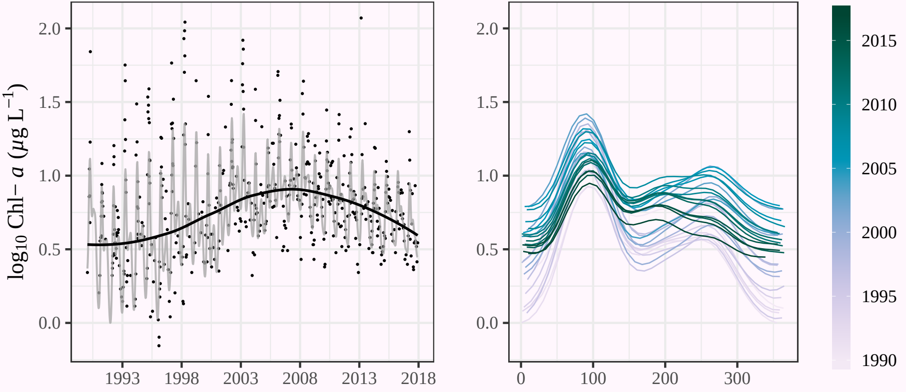


Application of additive models

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

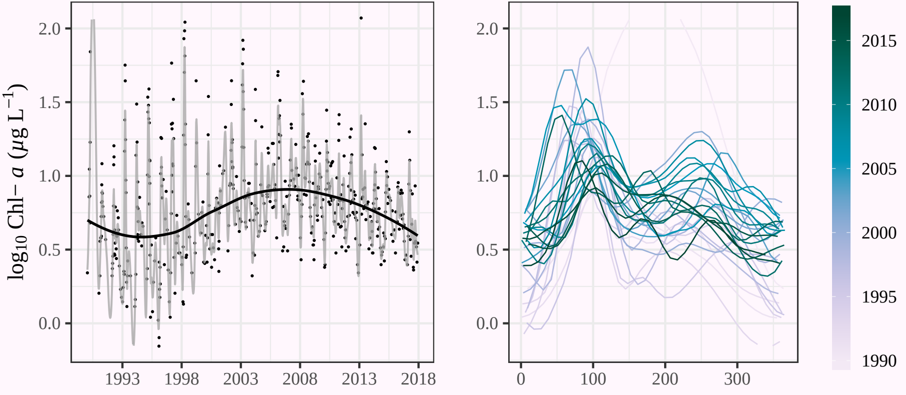


`gam2`: chl $\sim$ year + s(doy) + s(year) + ti(doy, year)

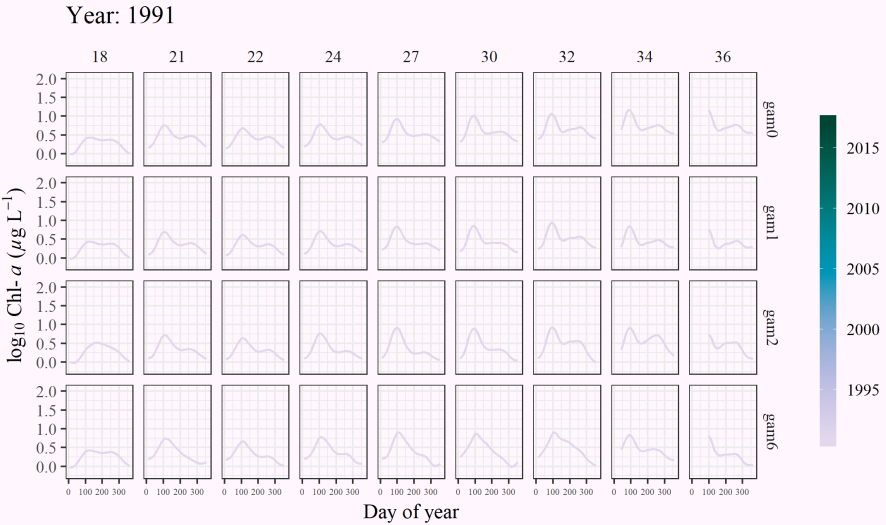


Application of additive models

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

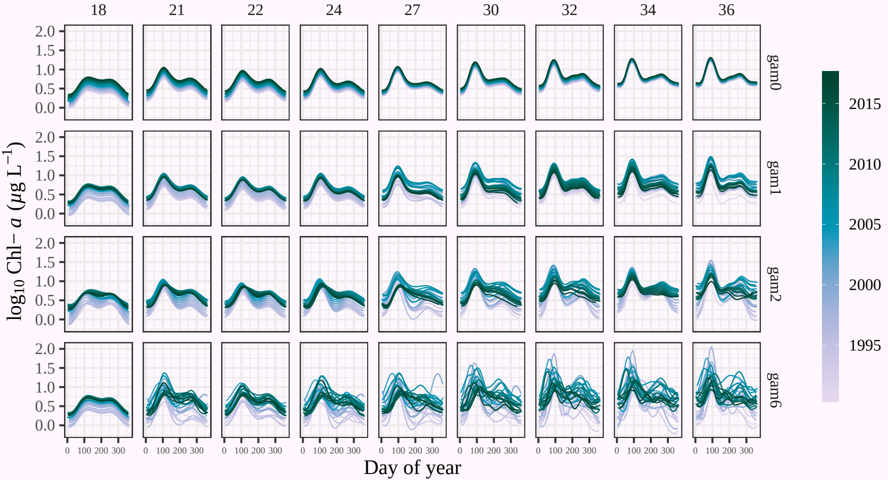


Application of additive models



Application of additive models

All years



Application of additive models

Table: GCV summary statistics by station and model

	Stations (north to south)								
Model	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	**-141.8**	-147.3	**-116.4**	-98.5	4	49.6	108.1	189.9	147.2
gam6	-139.5	**-235.2**	-116.3	**-176.8**	**-92.1**	**-115.9**	**-149**	**-3.3**	**-65.5**

Application of additive models

Table: GCV summary statistics by station and model

	Stations (north to south)								
Model	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	**-141.8**	-147.3	**-116.4**	-98.5	4	49.6	108.1	189.9	147.2
gam6	-139.5	**-235.2**	-116.3	**-176.8**	**-92.1**	**-115.9**	**-149**	**-3.3**	**-65.5**

Table: R-squared summary statistics by station and model

	Stations (north to south)								
Model	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	**0.53**	0.54	0.48	0.53	0.54	0.54	0.49	0.41	0.48
gam6	0.51	**0.68**	**0.54**	**0.66**	**0.68**	**0.72**	**0.75**	**0.69**	**0.75**

Conclusions from additive models

- More complex models provided more “parsimonious” fit
  - Non-monotonic trends were obvious
  - Interactions between year/season were common

Conclusions from additive models

- More complex models provided more “parsimonious” fit
  - Non-monotonic trends were obvious
  - Interactions between year/season were common
- Clear demarcation of annual and seasonal trends allowed a description of the separate components
  - Annual trends showed increase, with some decreases in recent period of record
  - Change in dominance between spring/fall peaks were observed

Conclusions from additive models

- More complex models provided more "parsimonious" fit
  - Non-monotonic trends were obvious
  - Interactions between year/season were common
- Clear demarcation of annual and seasonal trends allowed a description of the separate components
  - Annual trends showed increase, with some decreases in recent period of record
  - Change in dominance between spring/fall peaks were observed

How can this information support decision-making??

Shiny interactive web platform

https://sccwrp.shinyapps.io/sfbaytrends

GAM evaluation - SF South Bay

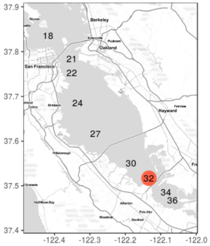
Exploratory plots

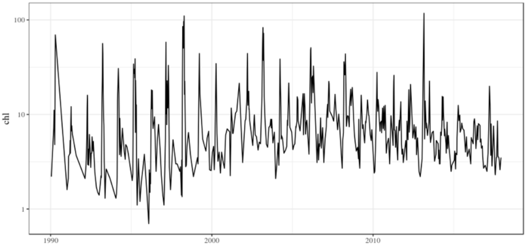
The following plots show the raw data for all monitoring stations and parameters in South Bay, 1990 - 2017. Select the parameter, plot type (total time series, by year, or by month), and variable transformation. The year and month plots are aggregated boxplots of all observations at a station for each selected time period. The variable transformation can be used to show the observations in arithmetic or logarithmic space.

Choose station: 32

Choose plot type: tot

Log-space: TRUE

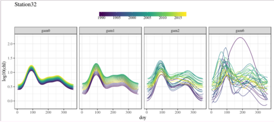




Explore results for each station, by model

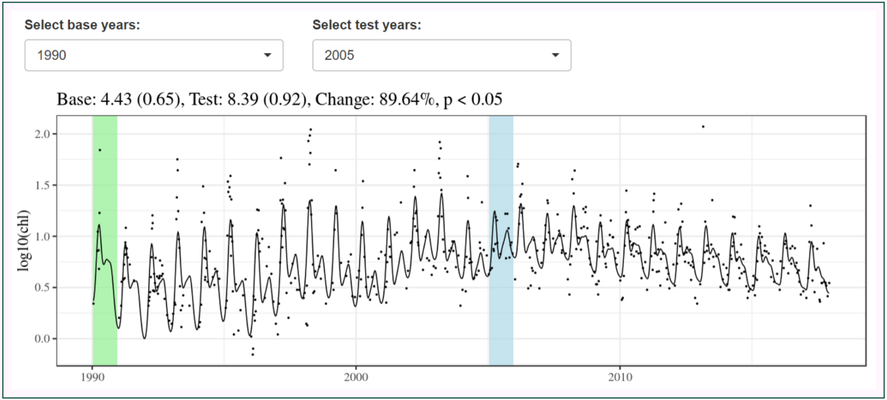
model	smoother	edf	Ref.df	F	p.value
gam0	s(doy)	7.10	8.00	31.64	0.00
gam1	s(dec_time)	2.45	27.00	2.74	0.00
gam1	s(doy)	7.17	8.00	37.97	0.00
gam2	s(dec_time)	21.40	27.00	5.05	0.00
gam2	s(doy)	7.11	8.00	42.39	0.00
gam2	ti(dec_time,doy)	8.98	12.00	3.98	0.00
gam6	s(dec_time)	218.84	334.00	3.06	0.00
gam6	s(doy)	5.07	8.00	2.24	0.00

model	k	AIC	GCV	R2
gam0	NA	232.88	0.09	0.32
gam1	28	162.57	0.08	0.41
gam2	28	108.06	0.07	0.49
gam6	336	-148.96	0.06	0.75



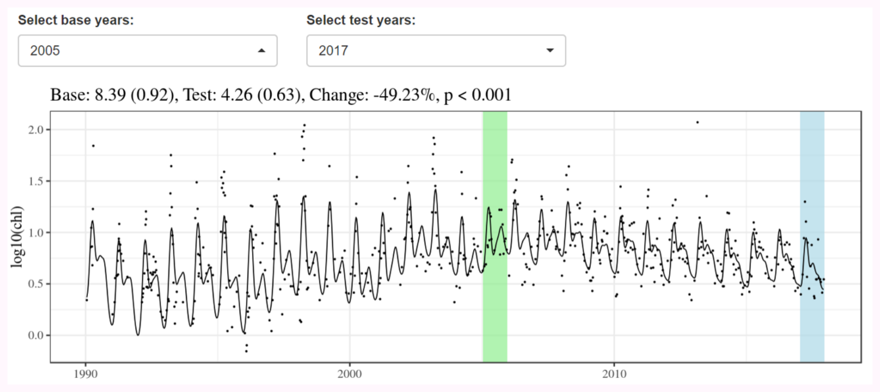
Shiny interactive web platform

Perform trend tests with selected years

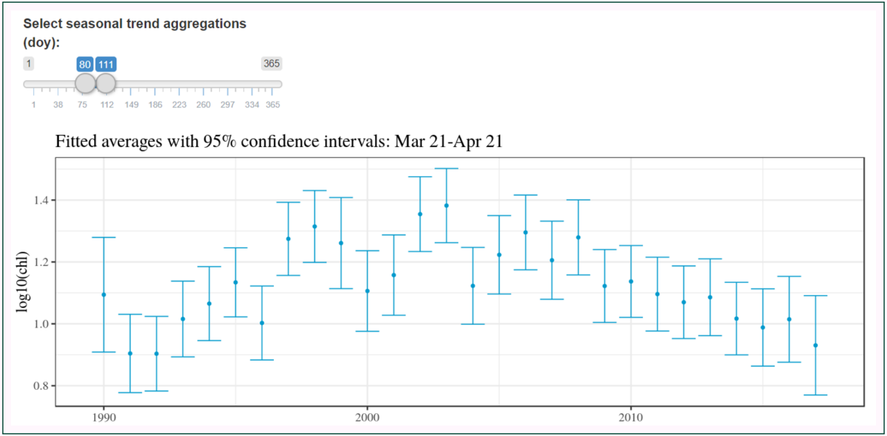


Shiny interactive web platform

Perform trend tests with selected years

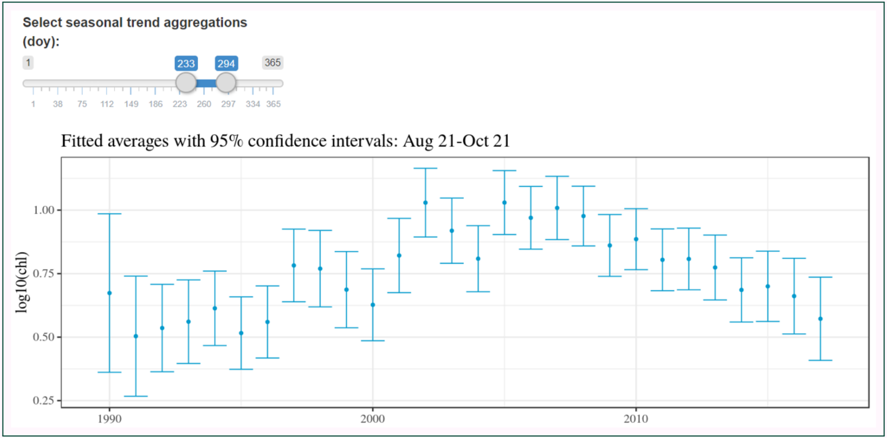


Evaluate trends between years, by season



Shiny interactive web platform

Evaluate trends between years, by season



Summary and next steps

- This analysis was a proof of concept
  - Does the GAM approach developed by the CBP transfer to SF Bay?
  - Do additional considerations need to be made?

Summary and next steps

- This analysis was a proof of concept
  - Does the GAM approach developed by the CBP transfer to SF Bay?
  - Do additional considerations need to be made?
- Shiny platform helps communicate results to stakeholder
  - Which station, model, and time period do I care about?
  - How can I understand limitations of the different models?

Summary and next steps

- This analysis was a proof of concept
  - Does the GAM approach developed by the CBP transfer to SF Bay?
  - Do additional considerations need to be made?
- Shiny platform helps communicate results to stakeholder
  - Which station, model, and time period do I care about?
  - How can I understand limitations of the different models?
- Follow-up work:
  - Extend to other locations in the Bay
  - Explore trend analysis of aggregated stations
  - Incorporate additional variables - as response or as explanatory

Acknowledgments

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.