Generalized Additive Model (GAM) Development Briefing: Application to Tidal Water Quality

Water Quality GIT

Oct. 13, 2015

Rebecca Murphy (UMCES at CBPO)

Elgin Perry (statistical consultant)

Jeni Keisman (USGS)

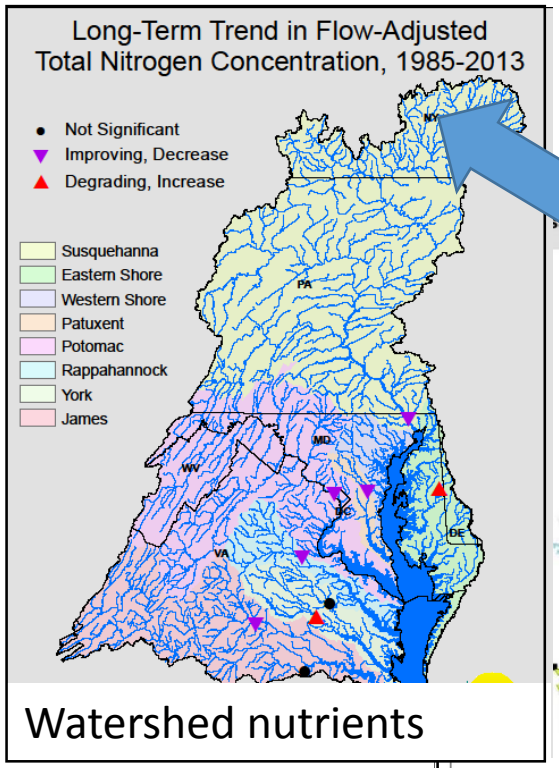
Chesapeake Bay Program

Science. Restoration. Partnership.

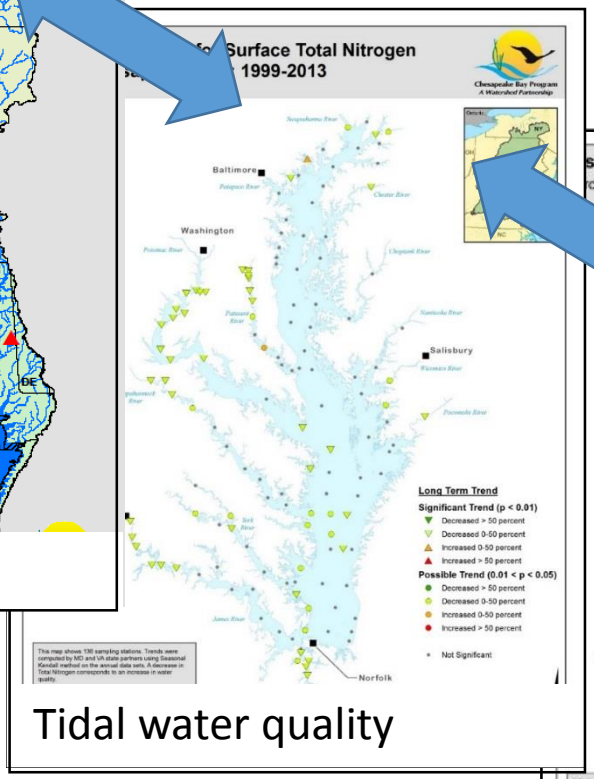
Tidal Data Analysis: Where it fits

Through tidal data analysis:

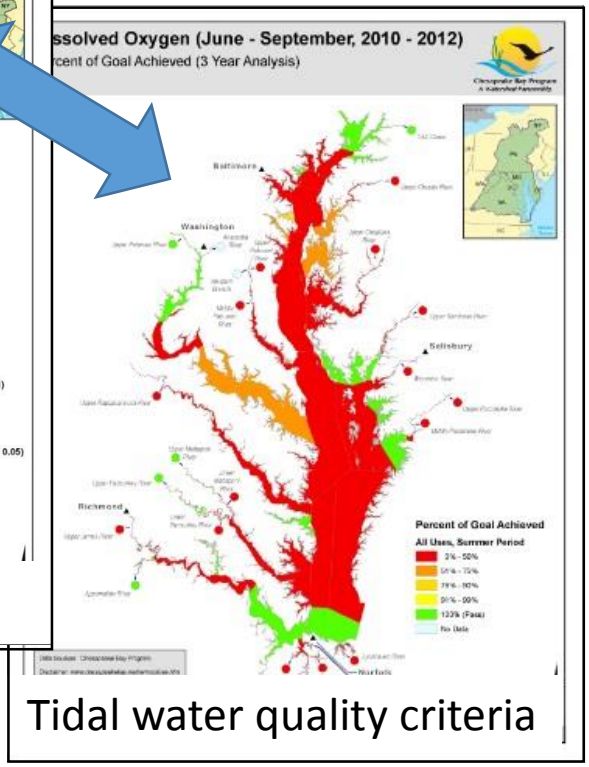
1. Will see trends before seeing them in attainment metrics
2. Can identify finer spatial and temporal periods of change
3. Can statistically link trends to watershed factors and loads



Watershed nutrients



Tidal water quality



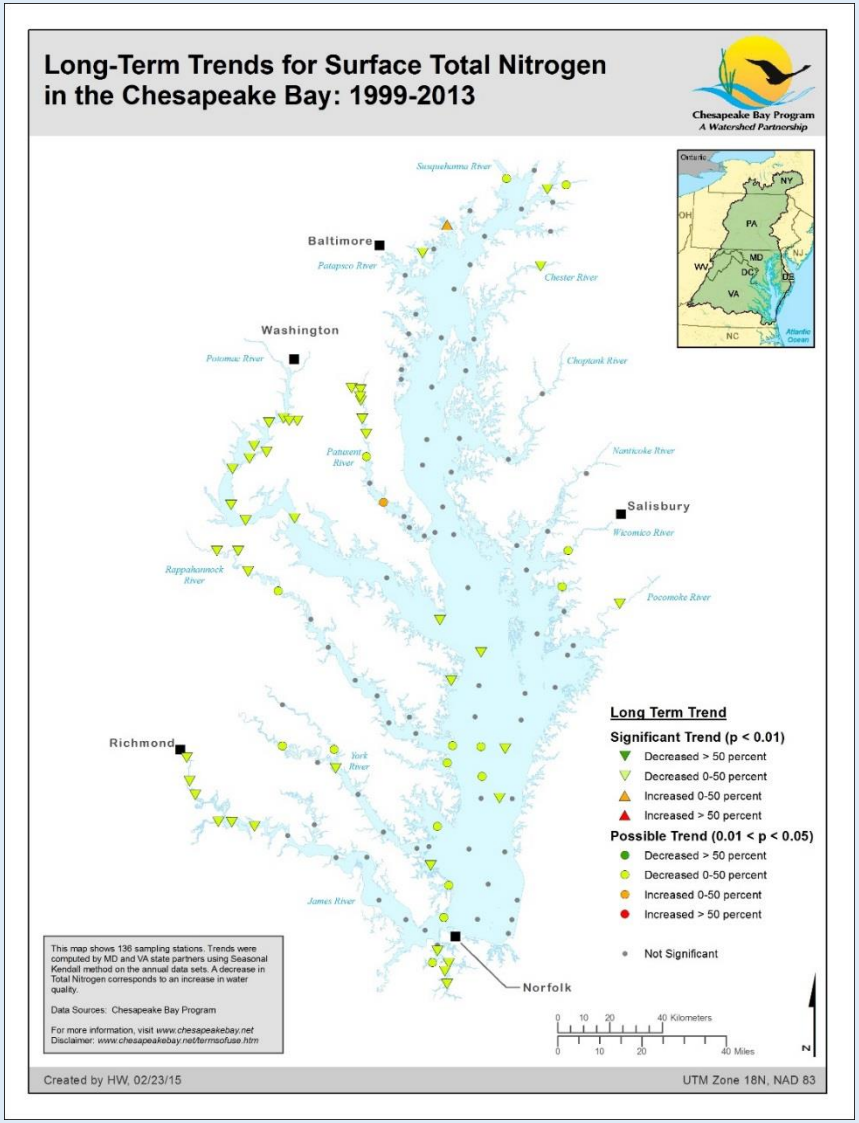
Tidal water quality criteria

Tidal Water Quality: Current Approach

- Seasonal Kendall used by CBP, MDDNR and VADEQ since 1990s for tidal water quality trend analysis

- Beneficial features:
  - Allows for identification of monotonic trends
  - Good for outliers
  - Does not require a distributional assumption

Seasonal Kendall-based trend maps (Presented to WQGIT March 2015)

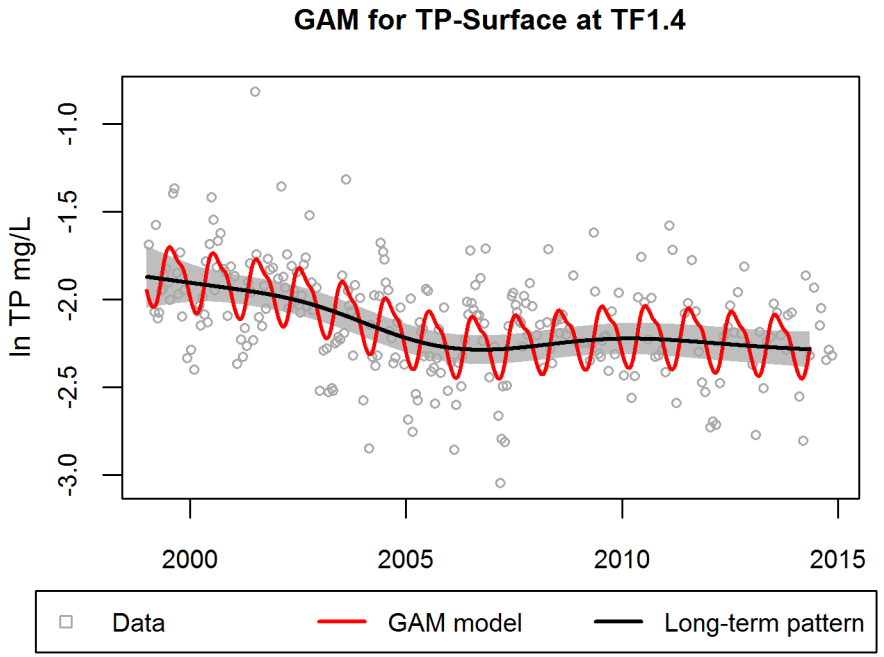


http://www.chesapeakebay.net/maps

Why a method change?

Based on lessons learned from current approach, looking for a method that:

- Is flexible enough to represent many possible patterns, *including trends that have changed direction over time*
- Is able to model non-linear relationships
- Generates a statistical confidence measure
- Can be used to test "factors affecting trends"



GAMs: Steps Towards Implementation to Tidal Chesapeake Bay

2013

- Evaluations of trend method options at CBP and select GAMs as a viable option

2014

- March: STAC workshop on Explaining Trends
- All year: Meetings with state partners who do current trend analysis and USGS-WRTDS team
- All year: Test GAM capabilities
- October: Introduce GAMs at WQGIT meeting

2015

- All year: Continue to test GAMs and meet with state partners
- Spring: Mainstem pilot application and compare to SK
- Summer: Tributary pilot application and refining uncertainty output
- October: Present Version 1 method (1999-2014) to WQGIT
- December: Version 1 available in R tool

2016

- Version 2 and 1985-2015 application
- Continue R&D applications to explaining trends project

GAM Version 1: Approach

Generalized Additive Model: A response variable is modeled as the sum of multiple functions of explanatory variables

Water quality =
linear(date) + s(date) + s(doy) + Interaction(date, doy)

s = spline smooth functions
doy = day of year

GAM Version 1: Approach

Generalized Additive Model: Represents a response variable as the sum of multiple functions of explanatory variables

Water quality =
linear(date) + s(date) + s(doy) + Interaction(date, doy)

Functions can be linear

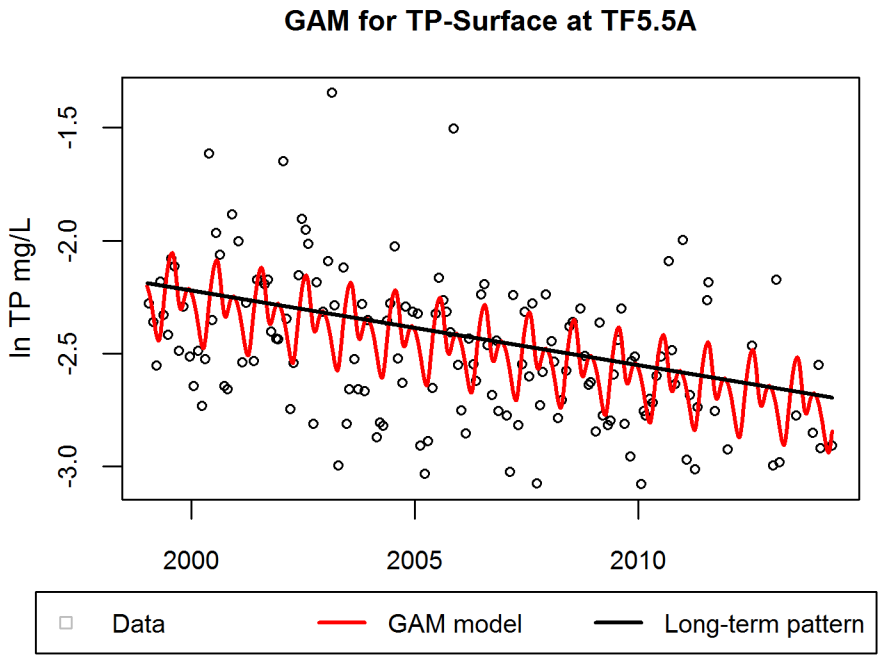
Smoothly-varying non-linear "spline" functions

And multi-dimensional smooth functions

s = spline smooth functions
doy = day of year

GAM Version 1: Approach

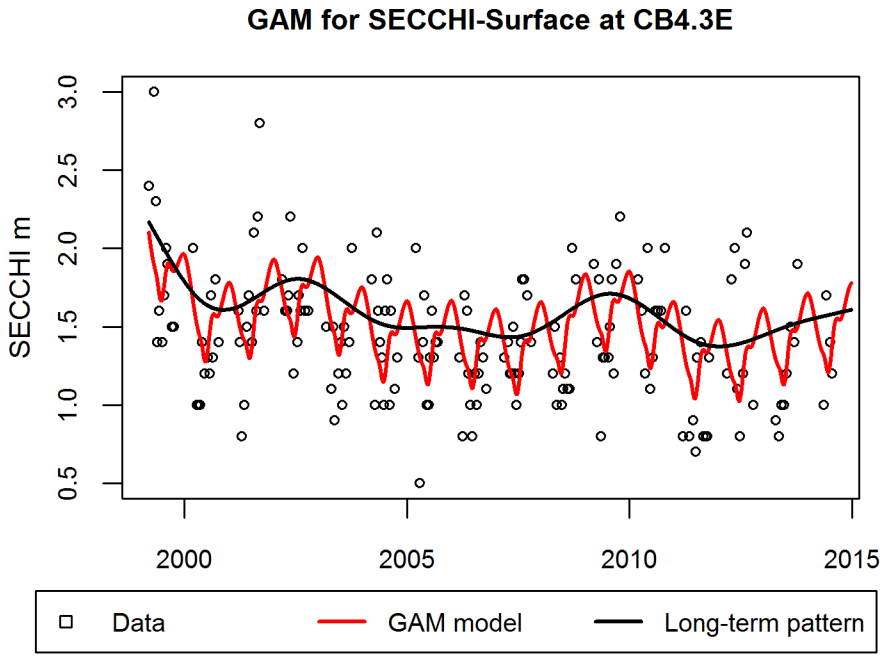
TP = linear(date) + s(date) **+ s(doy)** + Interaction(date, doy)



Example 1: A smooth seasonal cycle, but the overall trend is a linear decrease.

GAM Version 1: Approach

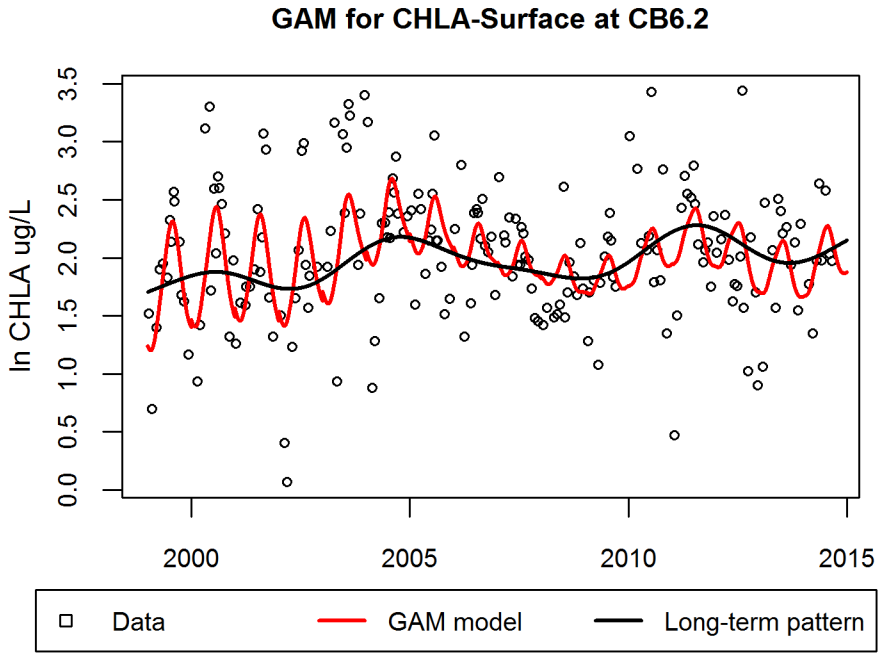
Secchi = linear(date) + **s(date) + s(doy)** + Interaction(date, doy)



Example 2: A significant, smoothly-varying pattern over time.

GAM Version 1: Approach

CHLA = linear(date) + **s(date) + s(doy) + Interaction(date, doy)**



Example 3: shape of the seasonal cycle is changing over time.

GAM Version 1: Full Tidal Application

- Fit GAMs to tidal data from 1999-2014
  - Both mainstem and tributary stations
  - Secchi disk depth; Surface and Bottom TN, TP, DO, and chlorophyll-a
- Conducted GAM/Seasonal Kendall comparison for mainstem
  - Are the overall trend results going to change with GAMs?: NO
  - Any systematic differences appear to be when the data is non-linear
- Developing ways to present and evaluate full set of output

GAM Version 1: Layers of Information

1. Is there a trend over a given time period?
2. What does that pattern look like over time?
3. Is there a seasonal difference in the temporal patterns?

Question: How can we most effectively share these layers of information without being overly complicated?

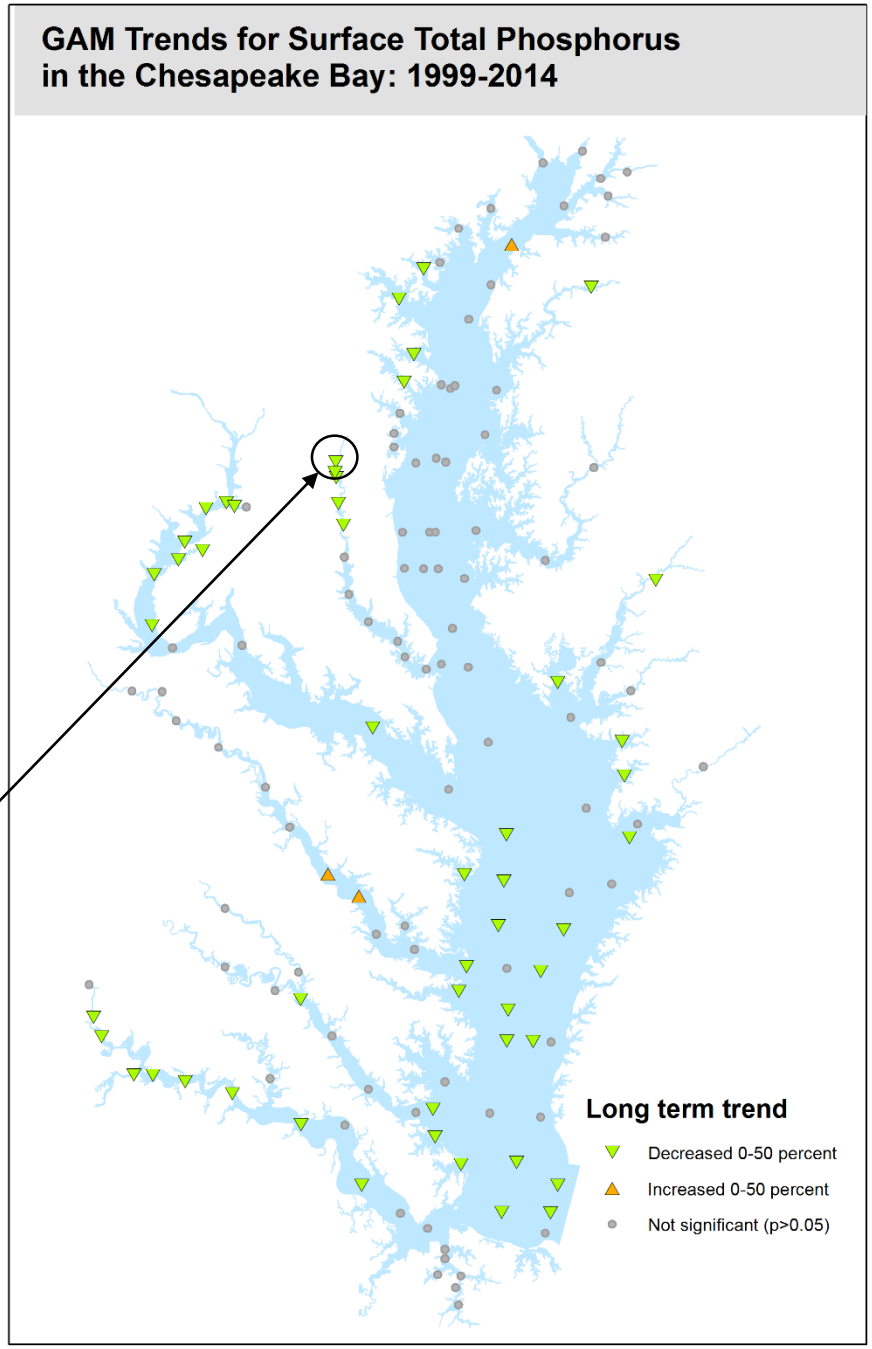
GAM Version 1: Results

Layers of output:

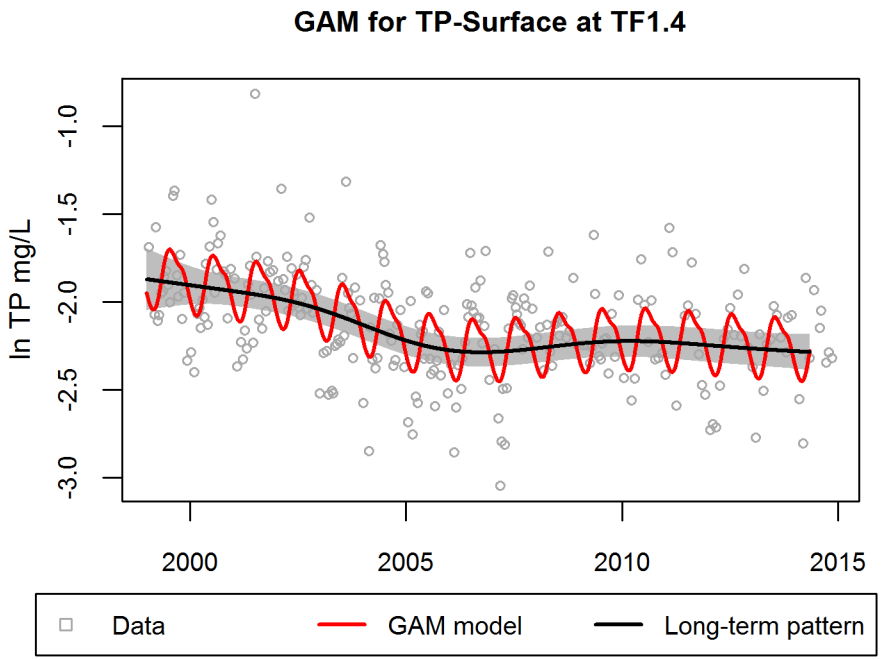
1. Is there a trend over a given time period?
   - Identification and significance of long-term trends
   - Slope and direction of a trend

Example: TF1.4 TP Surface 1999-2014

```
Baseline log mean = -1.90
Current  log mean = -2.26
Estimated log difference = -0.36
Std. Err. log difference = 0.060
Confidence interval for log difference =
(-0.48 , -0.25)
Difference p-value = <0.0001
Percent Change Estimate =  -30.5 %
```



GAM Version 1: Results



Layers of output:

2. What does the trend look like?
   - Pattern and confidence bounds on long-term temporal pattern
   - Significance of explanatory variables

Example: TF1.4 TP Surface 1999-2014

GAM output

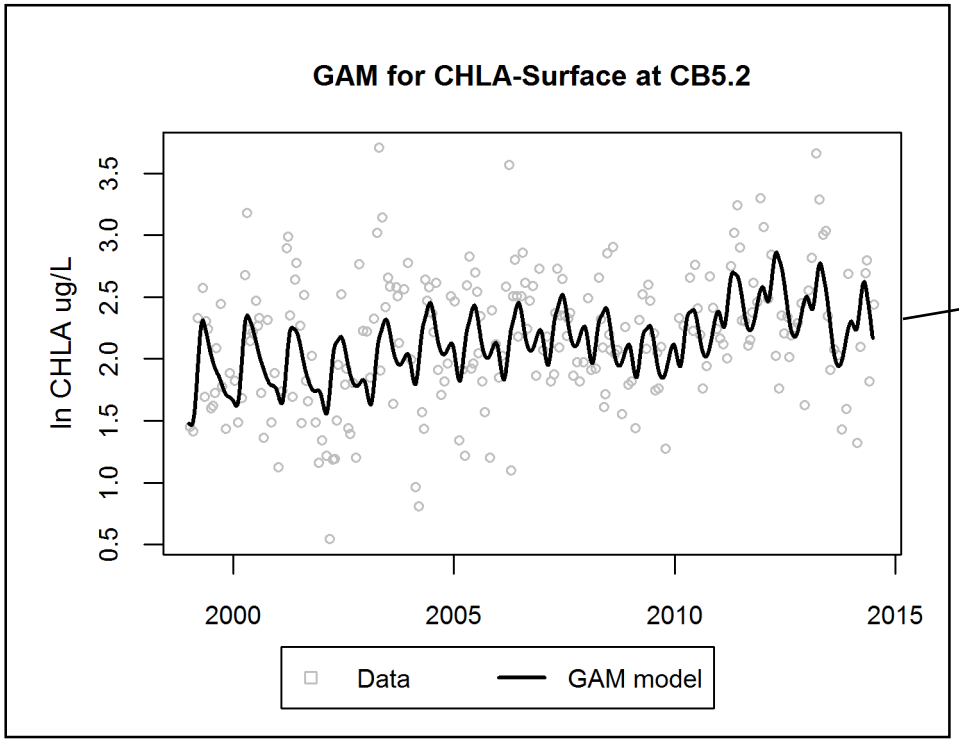
Source	edf	F-stat	p-value
linear(date)	1	5.71	0.018
s(date)	3.91	6.29	<0.0001
s(doy)	3.89	8.84	<0.0001

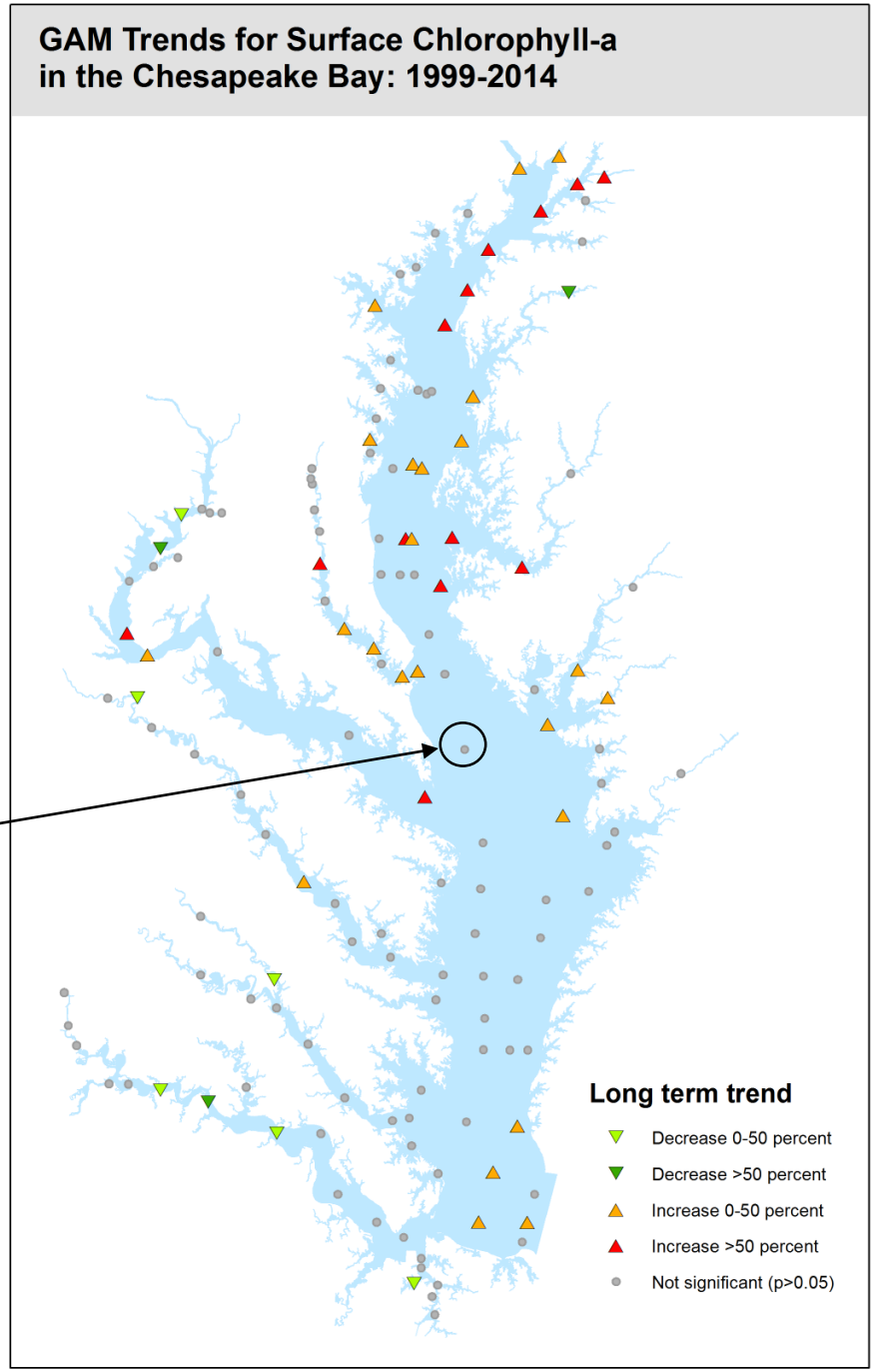
AIC 10.4
root mean-square error = 0.24
adjusted r-square = 0.36

GAM Version 1: Results

Layers of output:

3. Is there a seasonal difference in the temporal trend?

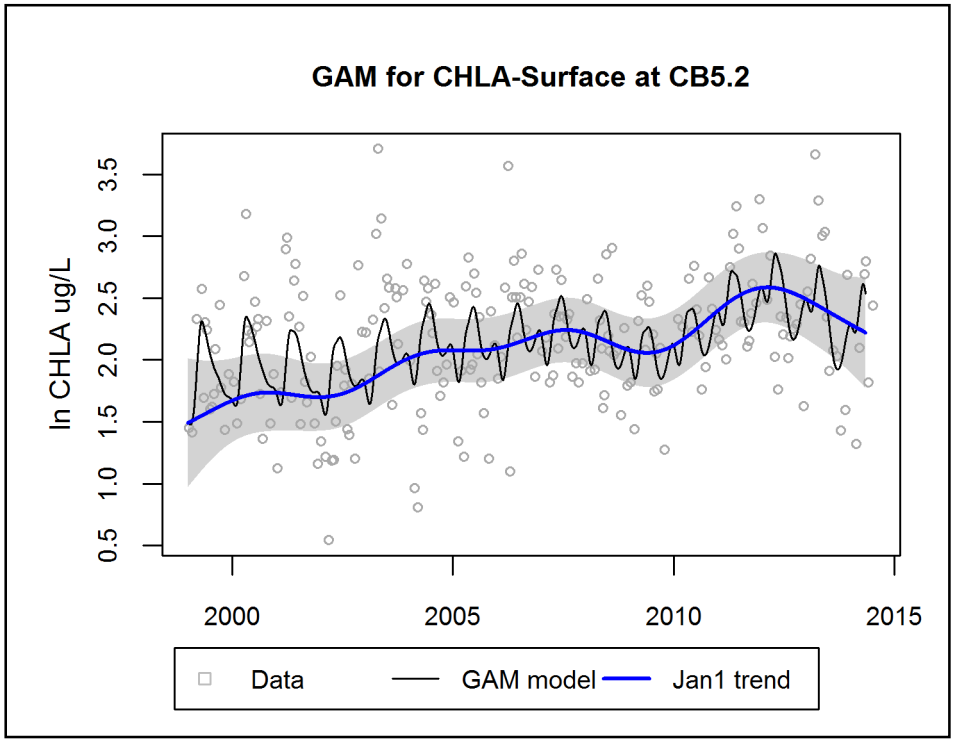


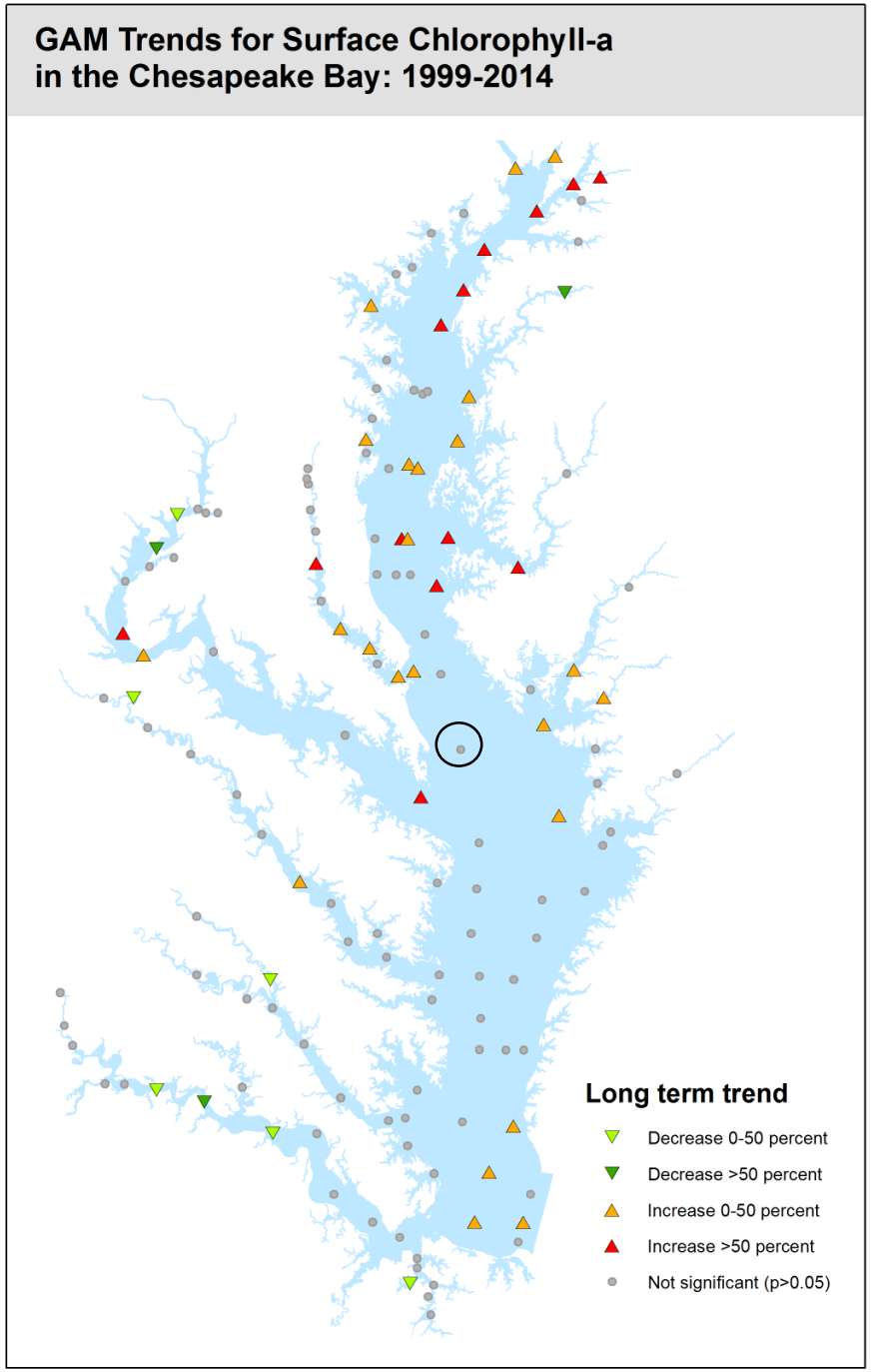


GAM Version 1: Results

Layers of output:

3. Is there a seasonal difference in the temporal trend?

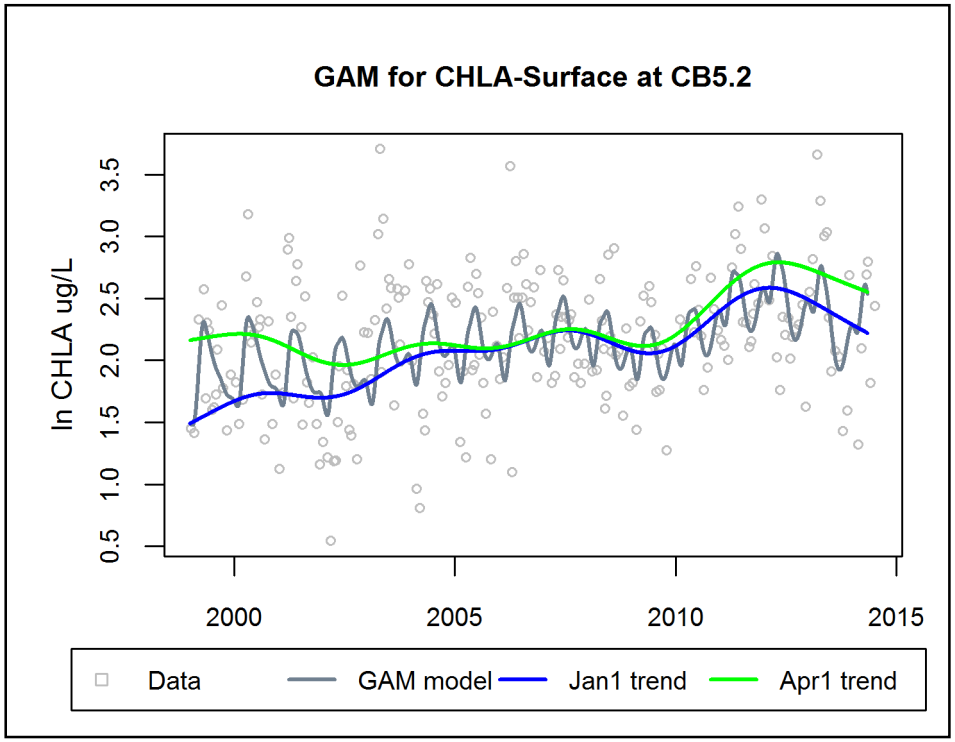


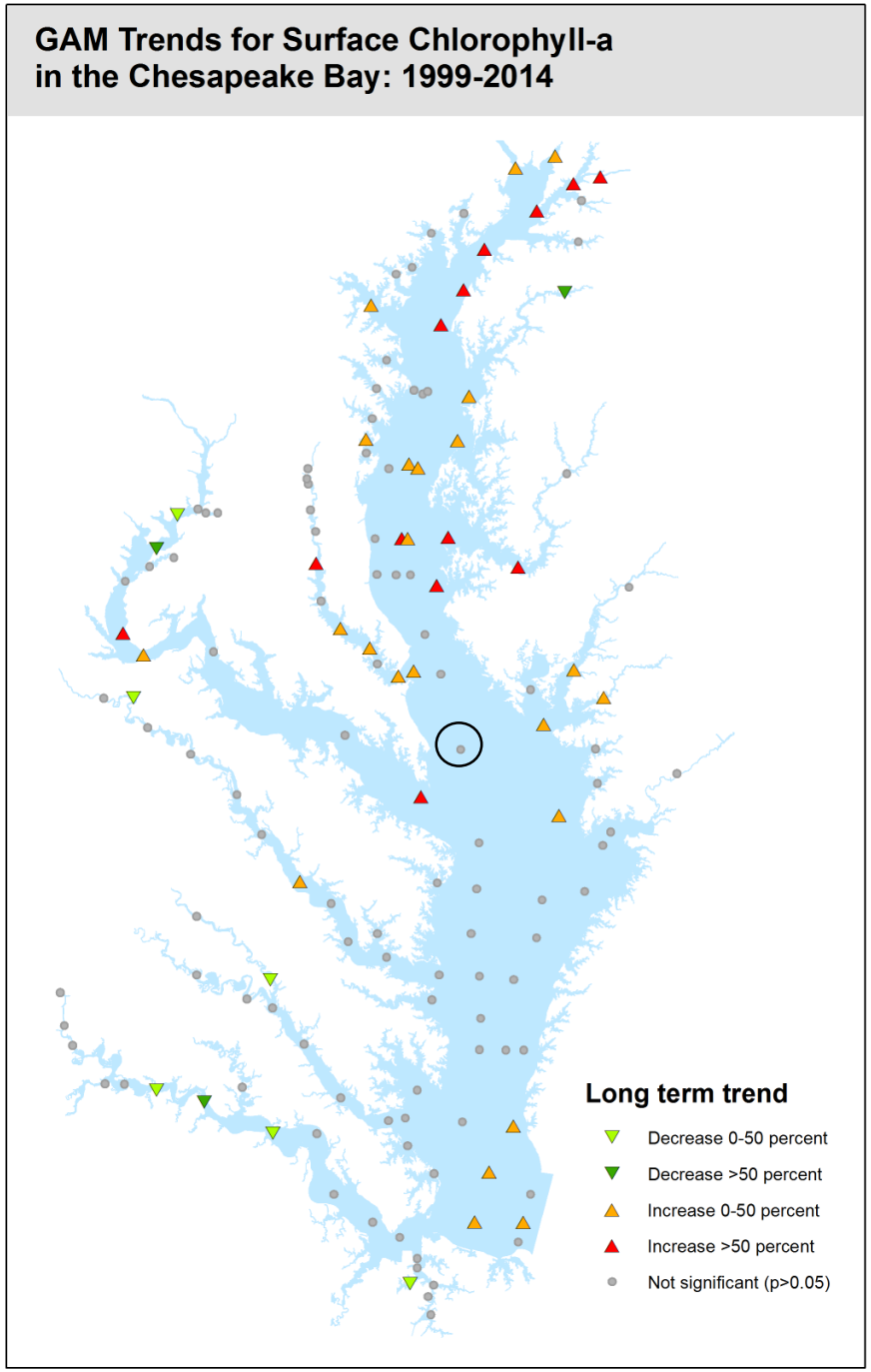


GAM Version 1: Results

Layers of output:

3. Is there a seasonal difference in the temporal trend?

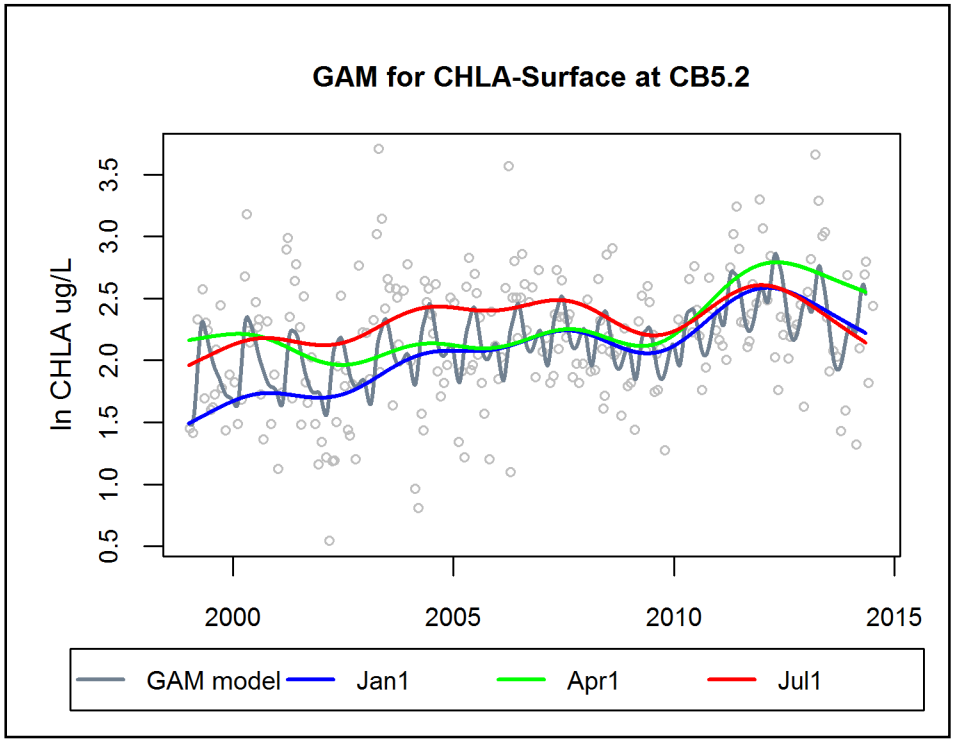


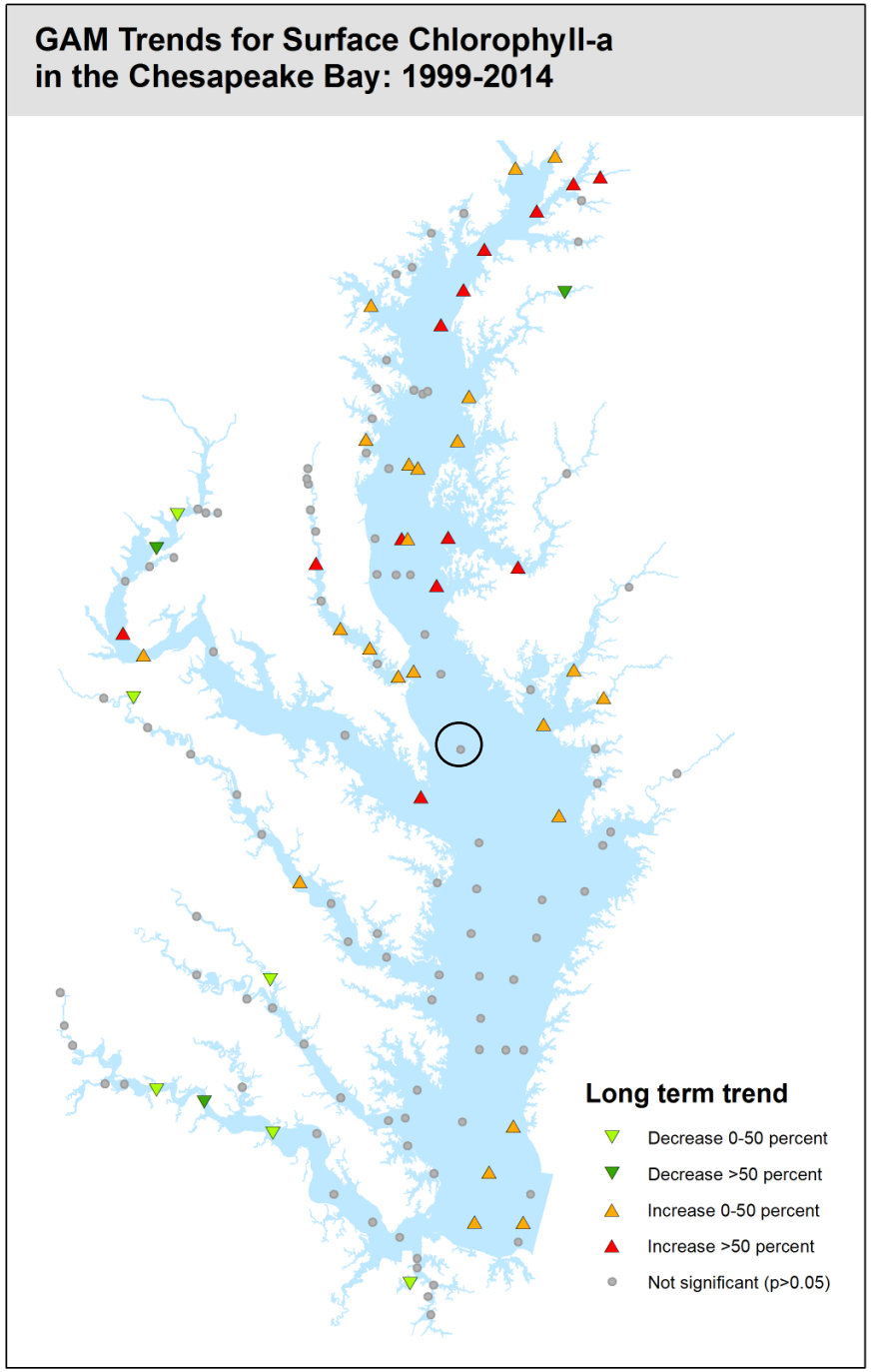


GAM Version 1: Results

Layers of output:

3. Is there a seasonal difference in the temporal trend?

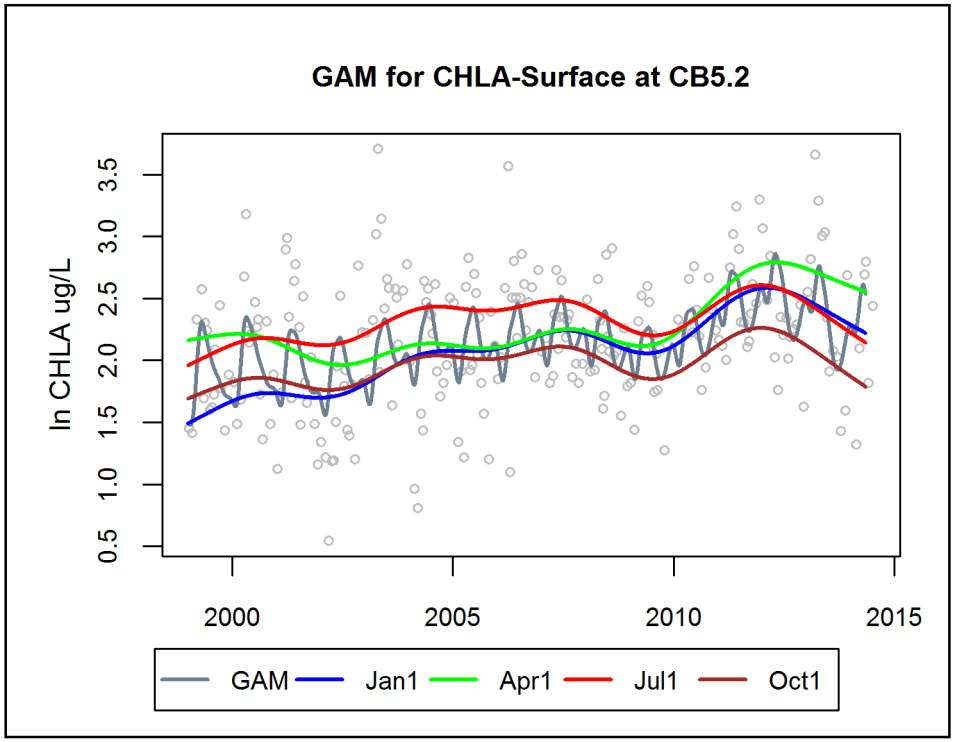


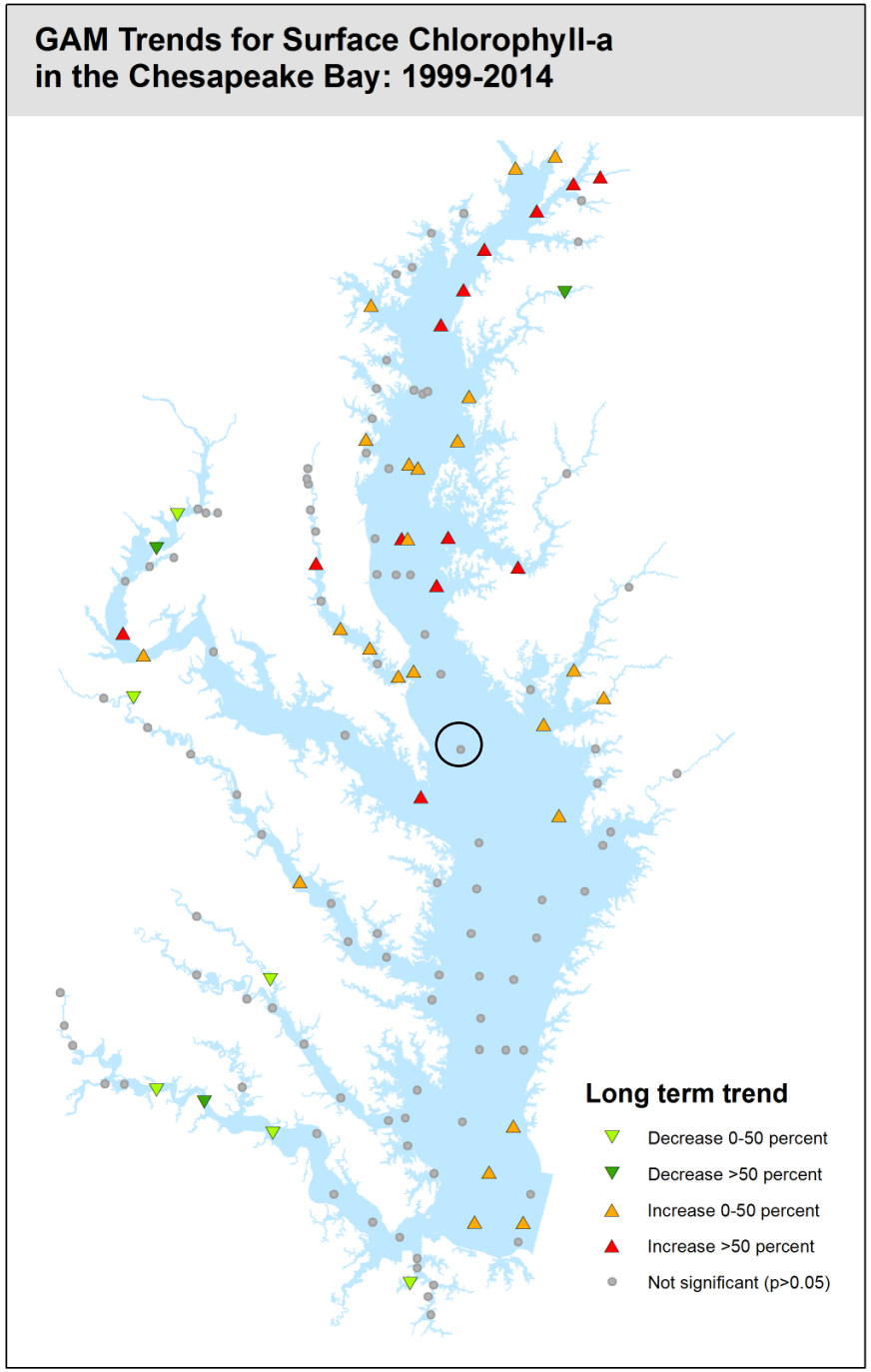


GAM Version 1: Results

Layers of output:

3. Is there a seasonal difference in the temporal trend?



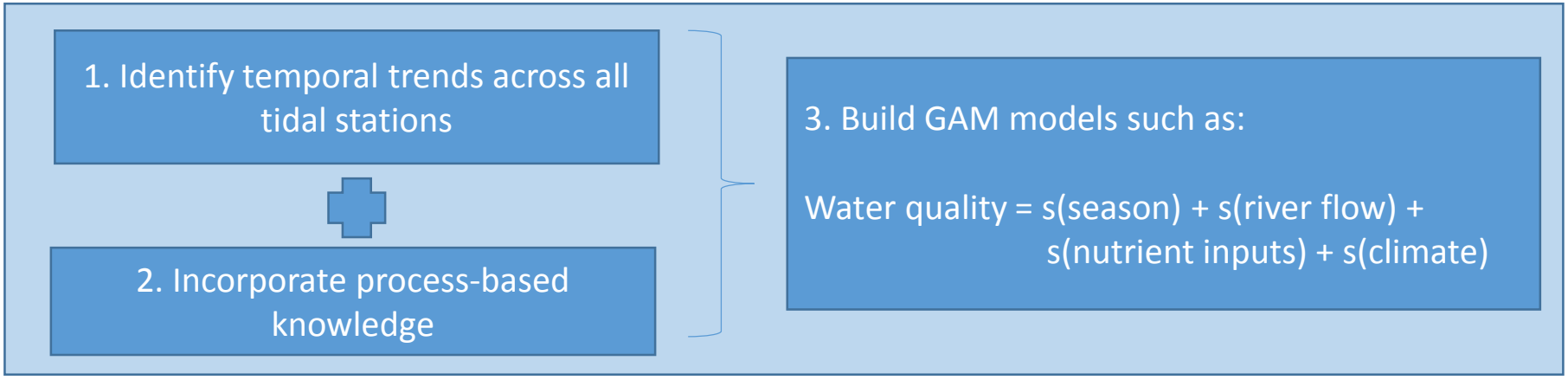


GAM Version 1: Layers of Information

1. Is there a trend over a given time period?
2. What does that pattern look like over time?
3. Is there a seasonal difference in the temporal patterns?

Next Steps

- Finish examining Version 1 results (2015)
- GAM tool in R (1st draft end 2015)
- Version 2 GAM approach for tidal stations (2016)
  - Finalize flow as explanatory variable
  - Application to 1985-present
- Applications for factors explaining trends (preliminary results 2016-'17):



Extra

Seasonal Kendall and GAM features/applications side-by-side	SK	GAM V1	Future GAM versions
Temporal trend identification			
Identification and significance of long-term trends	x	x	x
Slope and direction of a trend	x[a]	x	x
Pattern and confidence bounds on long-term temporal pattern		x	x
Significance of explanatory variables (e.g., date, season)		x	x
Incremental periods with significant trends		x	x
Accounting for residual temporal autocorrelation			x
Application			
Trends in mainstem and tributary 1999-2014 water quality data	x	x	x
Account for step changes and varied detection limits (i.e., use all data 1985-present)	x[b]		x
Flow as an explanatory variable (optional)		x[c]	x
Include other explanatory variables for hypothesis testing			x

[a] Sen slope test performs this for the SK approach
[b] SK is applied to pre-1999 using data censoring and block-approaches
[c] An approach is implemented, but some modifications are needed