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

Water Quality GIT

Sept. 14, 2015

Rebecca Murphy (UMCES at CBPO)

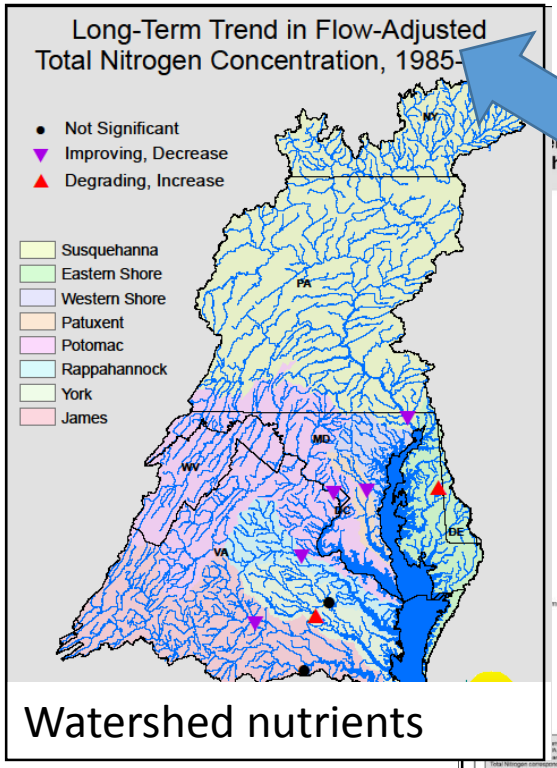
Elgin Perry (statistical consultant)

Jeni Keisman (USGS)

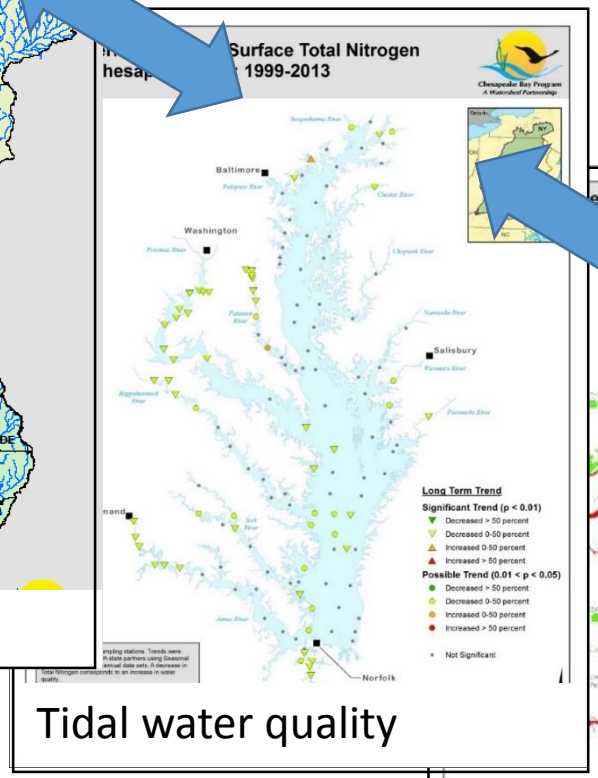
Chesapeake Bay Program

Science. Restoration. Partnership.

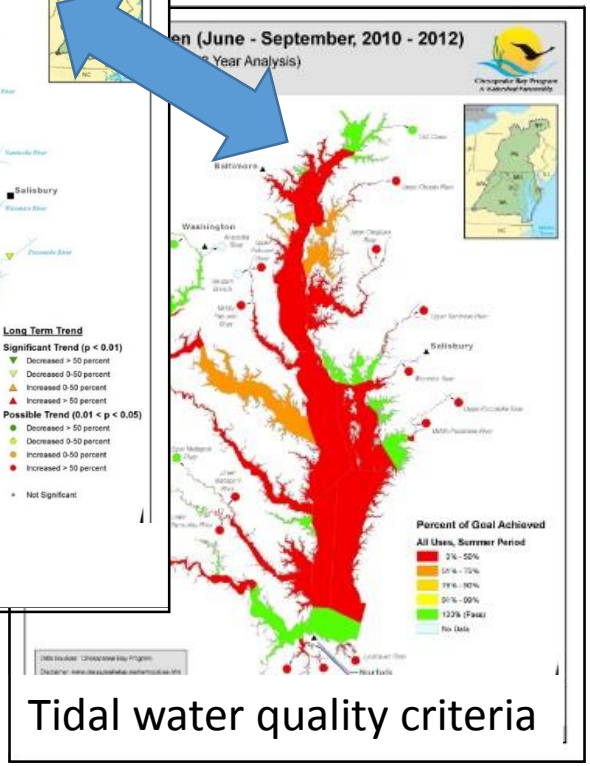
Using Monitoring Data To Measure Progress and Explain Change



Watershed nutrients



Tidal water quality



Tidal water quality criteria

Linked efforts underway to:

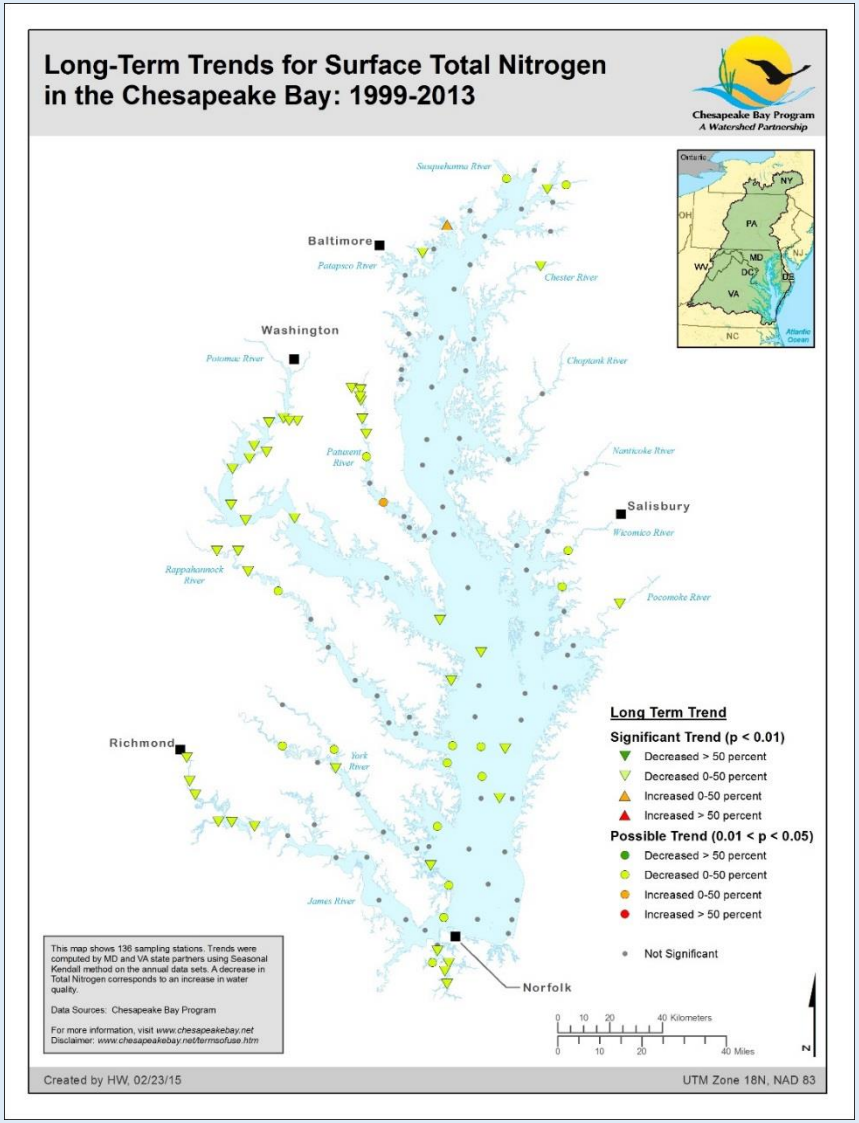
1. Measure progress
2. Explain water-quality changes
3. Enhance CBP models
4. Inform management strategies

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)

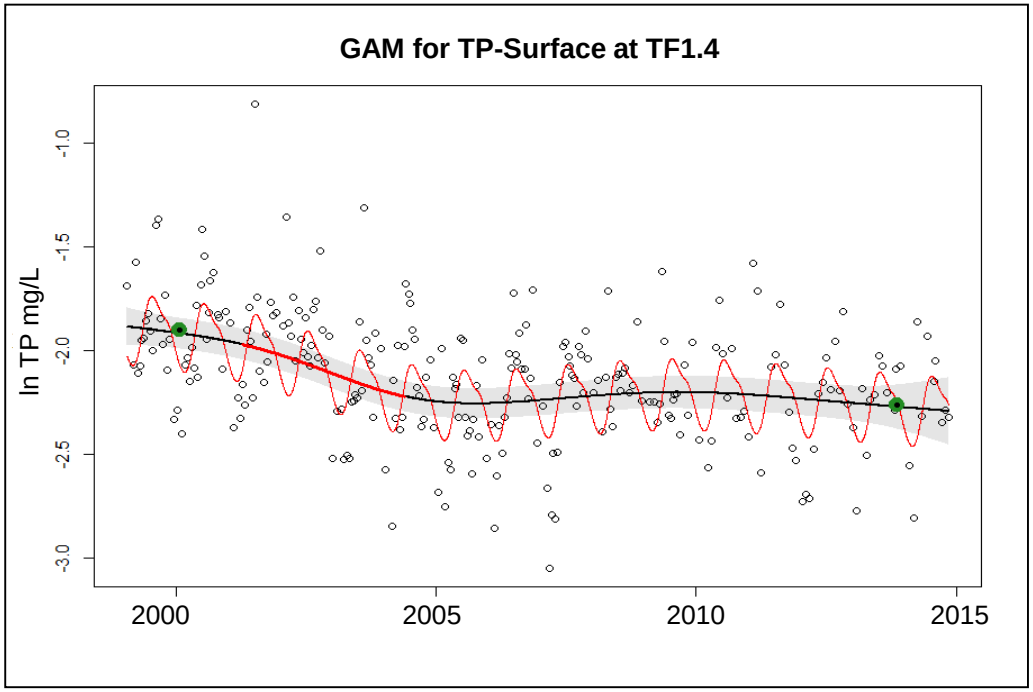


http://www.chesapeakebay.net/maps

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
- September: Present Version 1 method to WQGIT
- Fall: STAC review
- December: Version 1 available in R tool

2016

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

Version 1: Approach

With a GAM, a response variable is modeled as the sum of multiple nonlinear (or linear) functions of explanatory variables

GAM 1: Linear trend with seasonality

```
y = linear(date) + s(doy)
```

GAM 2: Smoothed trend with seasonality

```
y = linear(date) + s(date) + s(doy)
```

GAM 3: Seasonally-varying smooth trends

```
y = linear(date) + s(date) + s(doy)
    + Interaction(date, doy)
```

s = spline smooth functions
doy = day of year

Version 1: Approach

GAM 1: Linear trend with seasonality

```
y = linear(date) + s(doy)
```

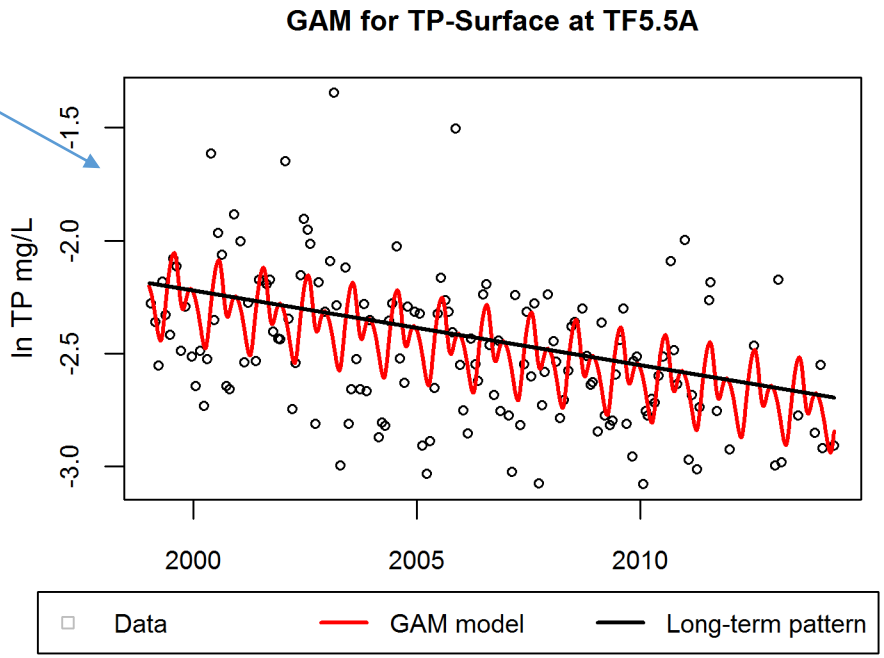
GAM 2: Smoothed trend with seasonality

```
y = linear(date) + s(date) + s(doy)
```

GAM 3: Seasonally-varying smooth trends

```
y = linear(date) + s(date) + s(doy)
    + Interaction(date, doy)
```

s = spline smooth functions
doy = day of year



GAM1 is a good fit because there is a smooth seasonal cycle, but the overall trend is a linear decrease.

Version 1: Approach

GAM 1: Linear trend with seasonality

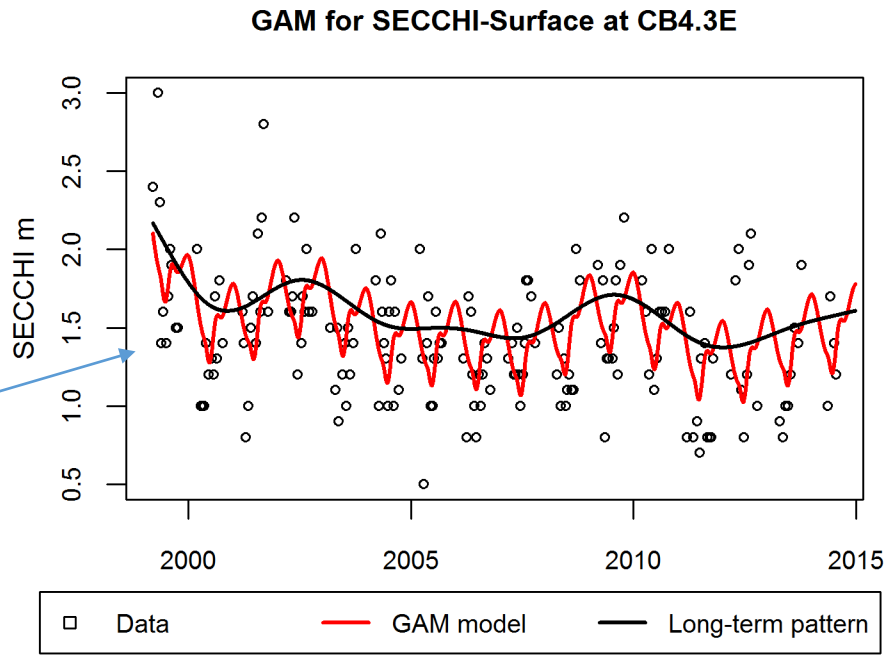
```
y = linear(date) + s(doy)
```

GAM 2: Smoothed trend with seasonality

```
y = linear(date) + s(date) + s(doy)
```

GAM 3: Seasonally-varying smooth trends

```
y = linear(date) + s(date) + s(doy)
    + Interaction(date, doy)
```



GAM2 is a useful because there is a significant, smoothly-varying pattern over time.

s = spline smooth functions
doy = day of year

Version 1: Approach

GAM 1: Linear trend with seasonality

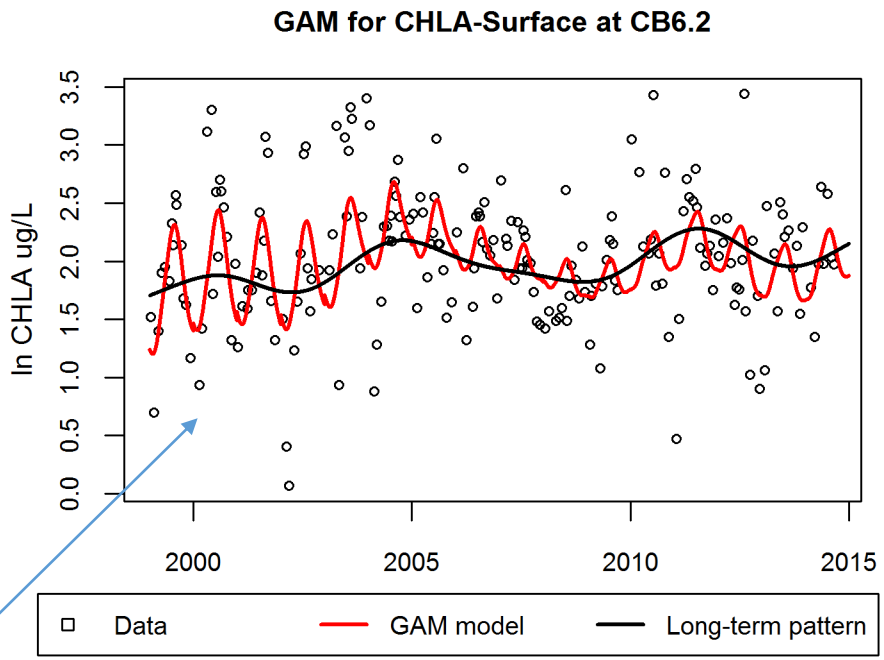
```
y = linear(date) + s(doy)
```

GAM 2: Smoothed trend with seasonality

```
y = linear(date) + s(date) + s(doy)
```

GAM 3: Seasonally-varying smooth trends

```
y = linear(date) + s(date) + s(doy)
    + Interaction(date, doy)
```



GAM3 is a good choice because the shape of the seasonal cycle is changing over time.

s = spline smooth functions
doy = day of year

Version 1: Full Tidal Application

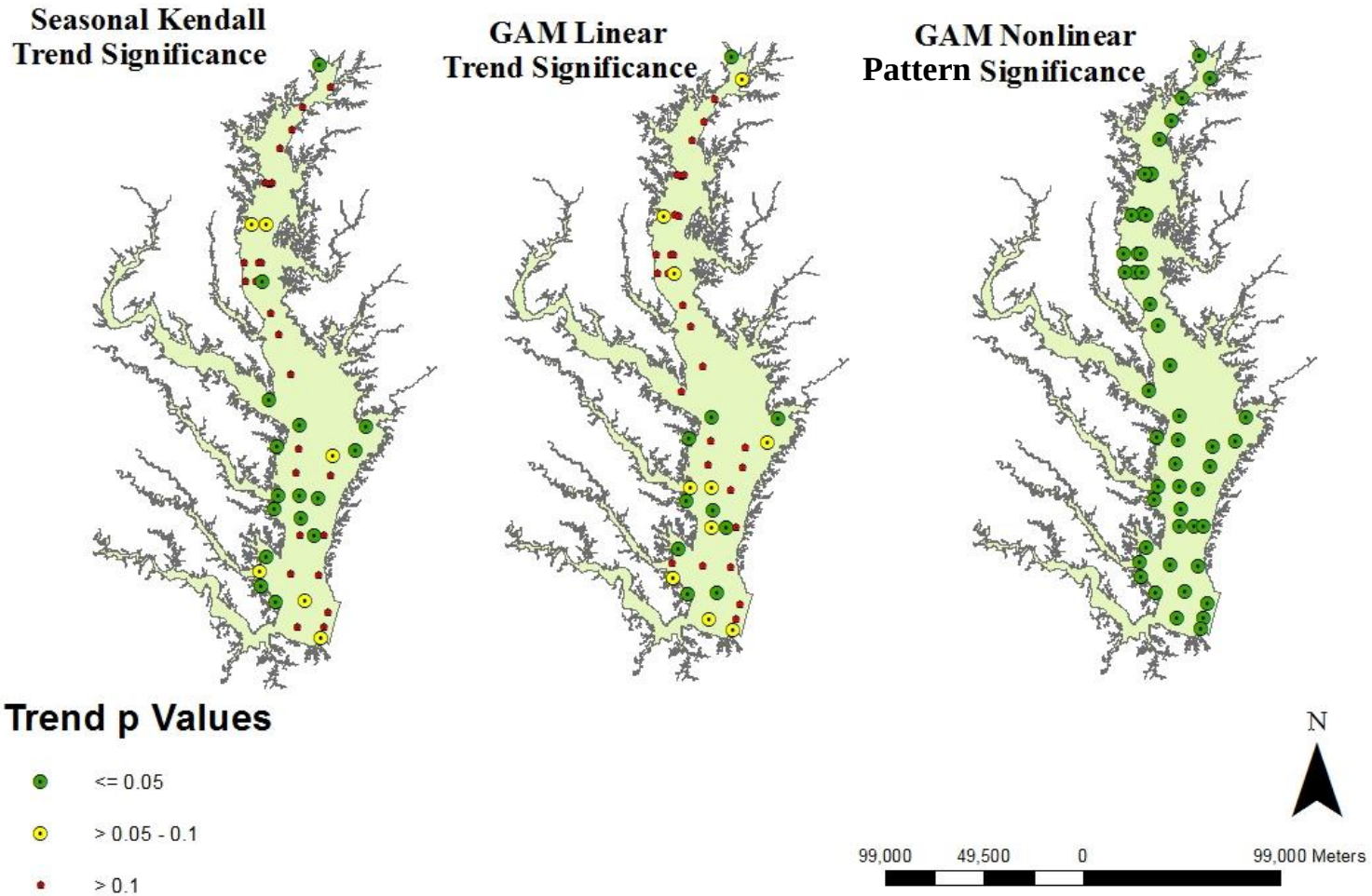
- Fit the GAMs to all mainstem and tributary stations using data from 1999-2014
  - Secchi disk depth; Surface and Bottom TN, TP, DO, and chlorophyll-a
  - Can compare the 3 models using model-fit statistics

- Conducted GAM/Seasonal Kendall comparison for mainstem to answer: "Are the overall trend results going to change with GAMs?"
  - Answer: No, because the linear components of the methods perform similarly
  - Any systematic differences appear to be when the data pattern is non-linear

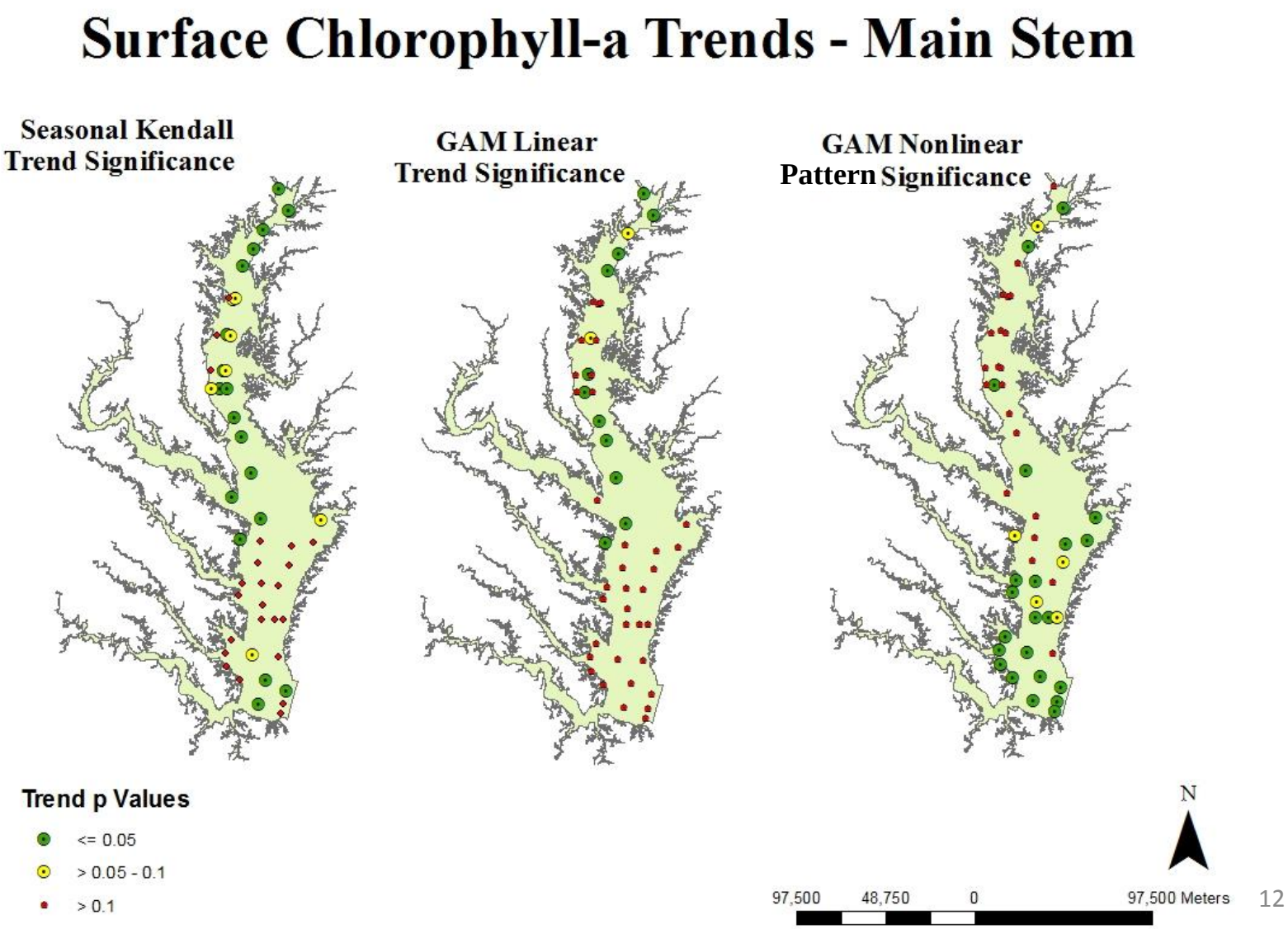
- Developing ways to present and evaluate full set of GAM output

Seasonal Kendall/GAM Comparison

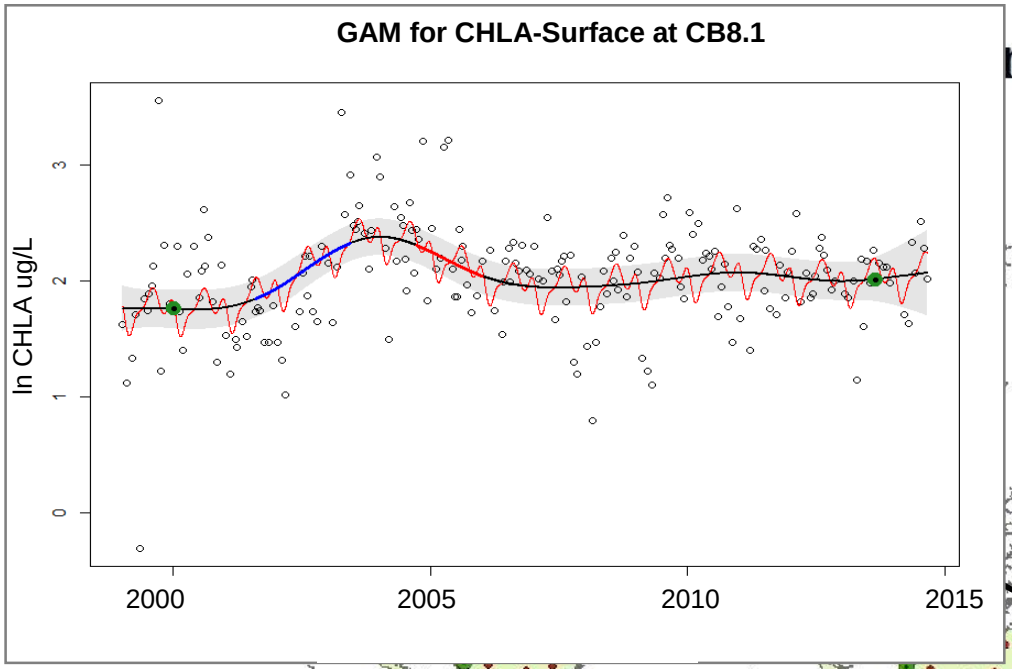
Surface Total Nitrogen Trends – Main Stem



Seasonal Kendall/GAM Comparison



Seasonal Kendall/GAM Comparison



rends - Main Stem

GAM Nonlinear
Pattern Significance

Trend p Values
<= 0.05
> 0.05 - 0.1
> 0.1

N

97,500 48,750 0 97,500 Meters

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

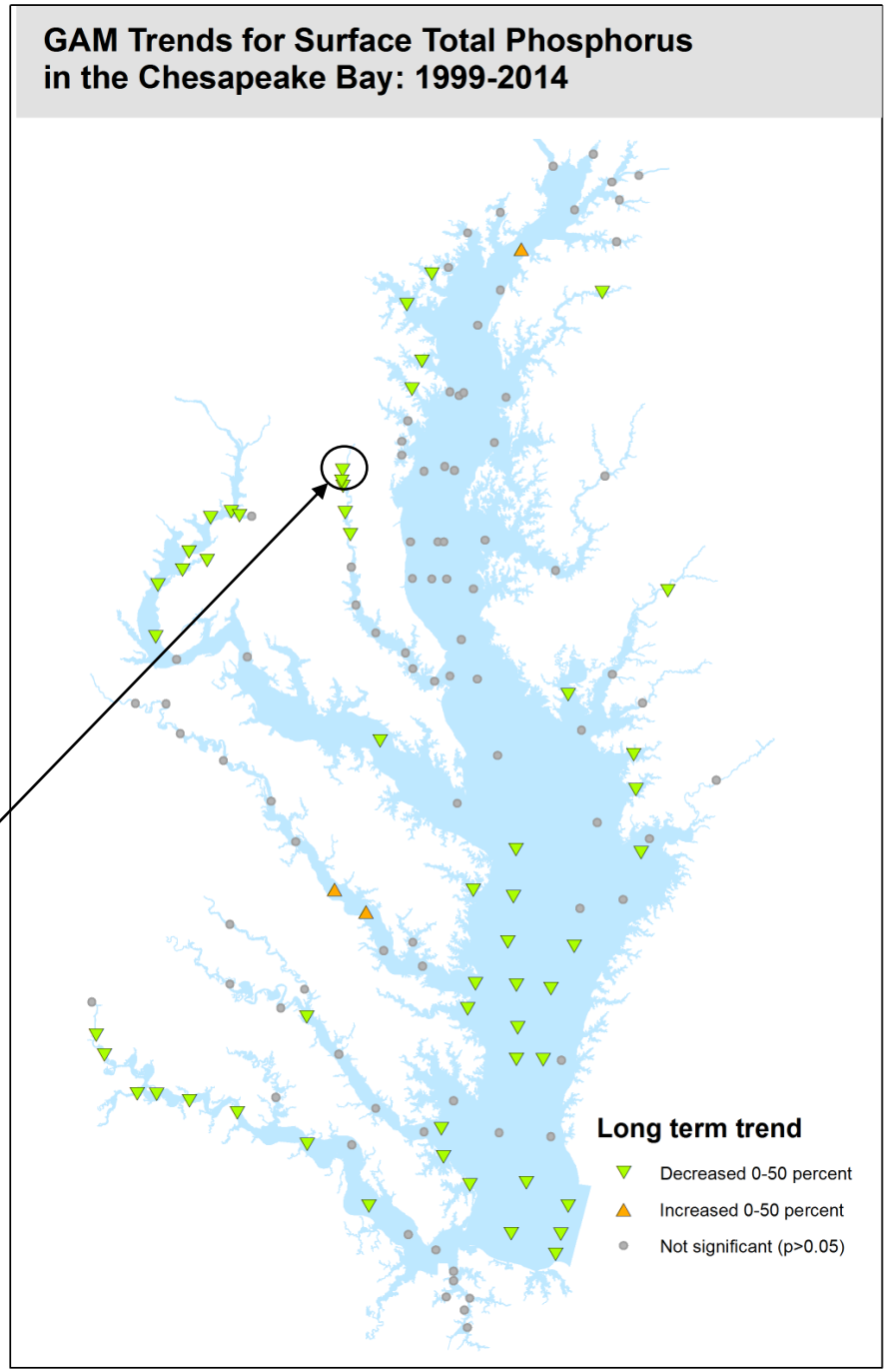
Version 1: Results

Layers of output:

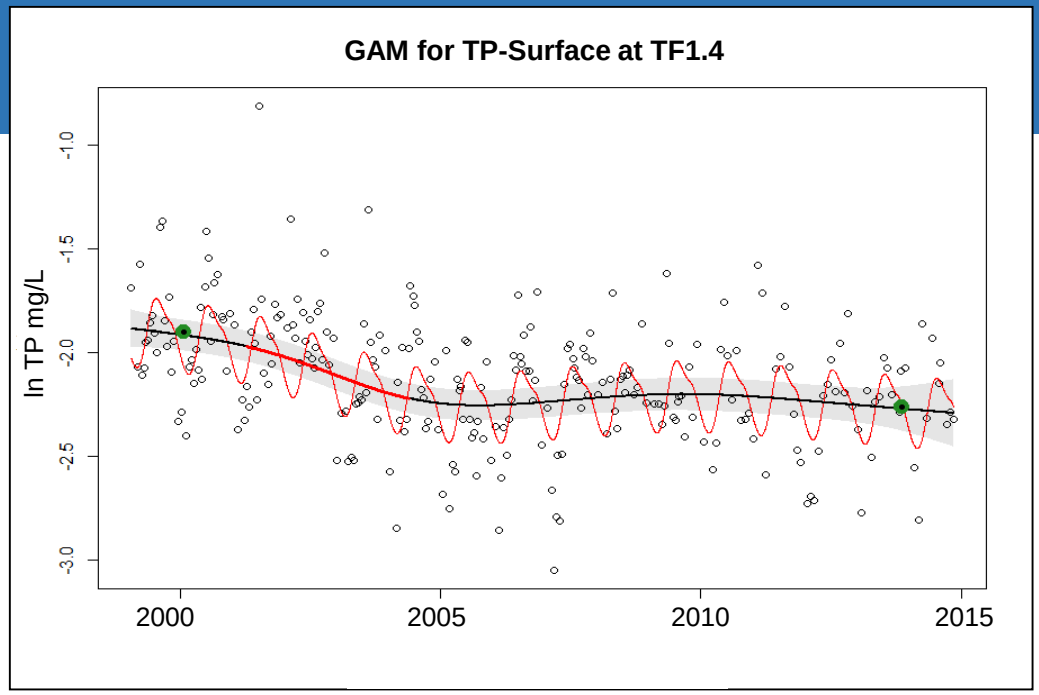
1. Is there a trend over a given time period?
   - Identification and significance of long-term trends
   - Slope and direction of a trend

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 =  -30.6 %
```



Version 1: Results



Layers of output:

2. What does the trend look like?

- Pattern and confidence bounds on long-term temporal pattern
- Incremental periods with significant trends
- Significance of explanatory variables

Example: TF1.4 TP Surface 1999-2014

GAM output

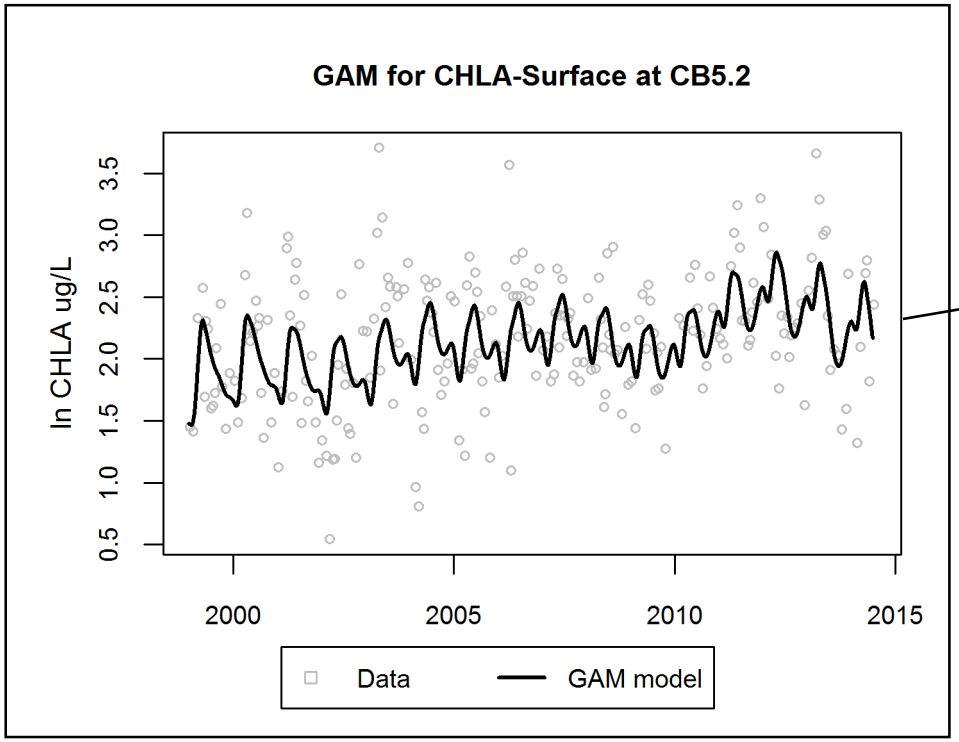
Source	edf	F-stat	p-value
linear(date)	1	0.69	0.40
s(date)	3.74	5.97	<0.0001
s(doy)	3.83	8.21	<0.0001

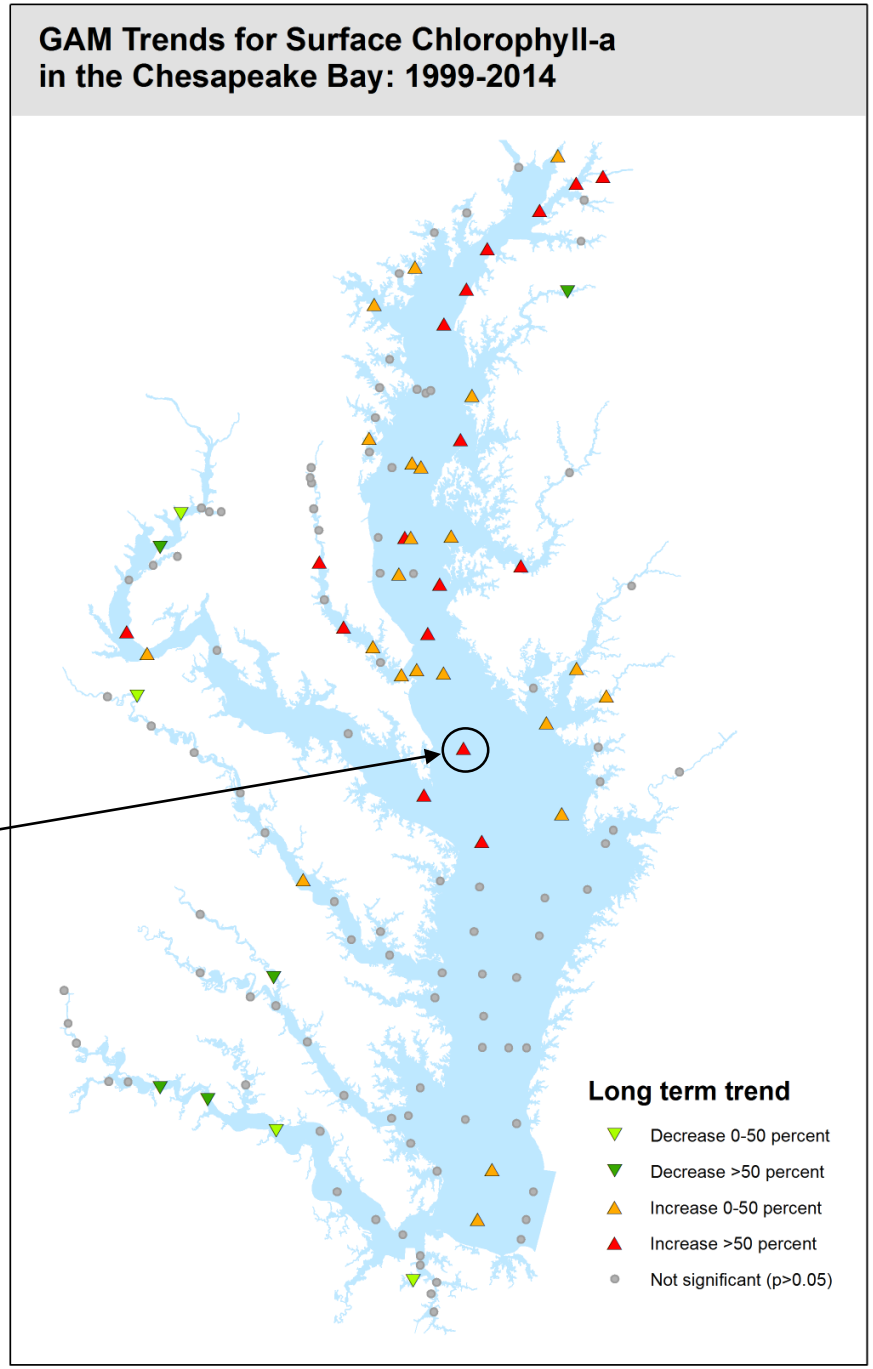
AIC 12.8
root mean-square error = 0.24
adjusted r-square = 0.36

Version 1: Results

Layers of output:

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

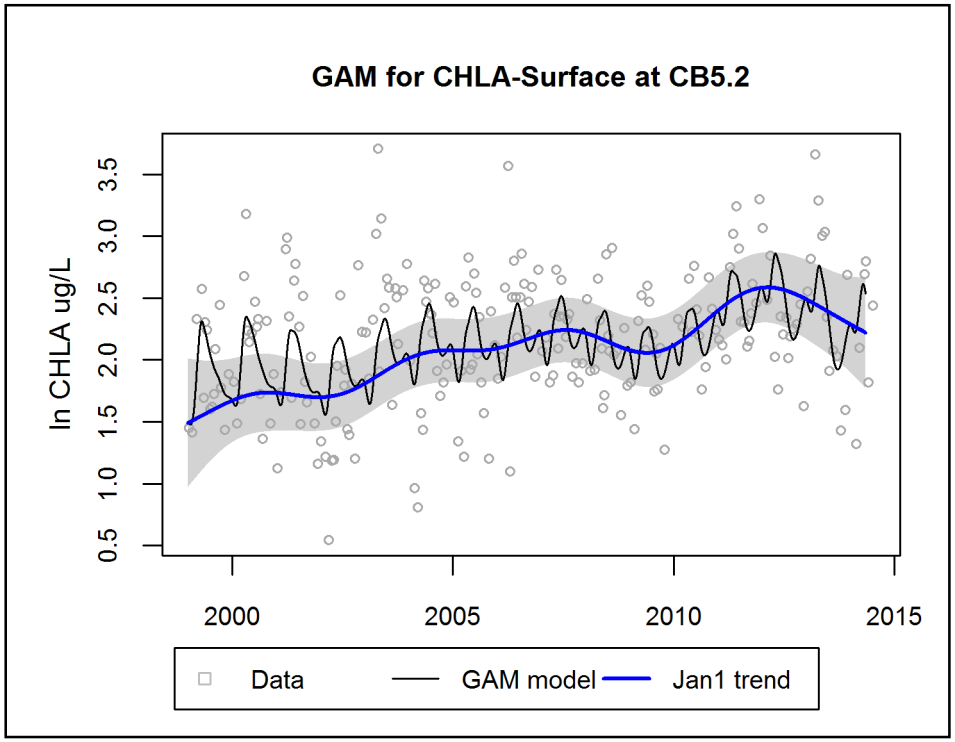


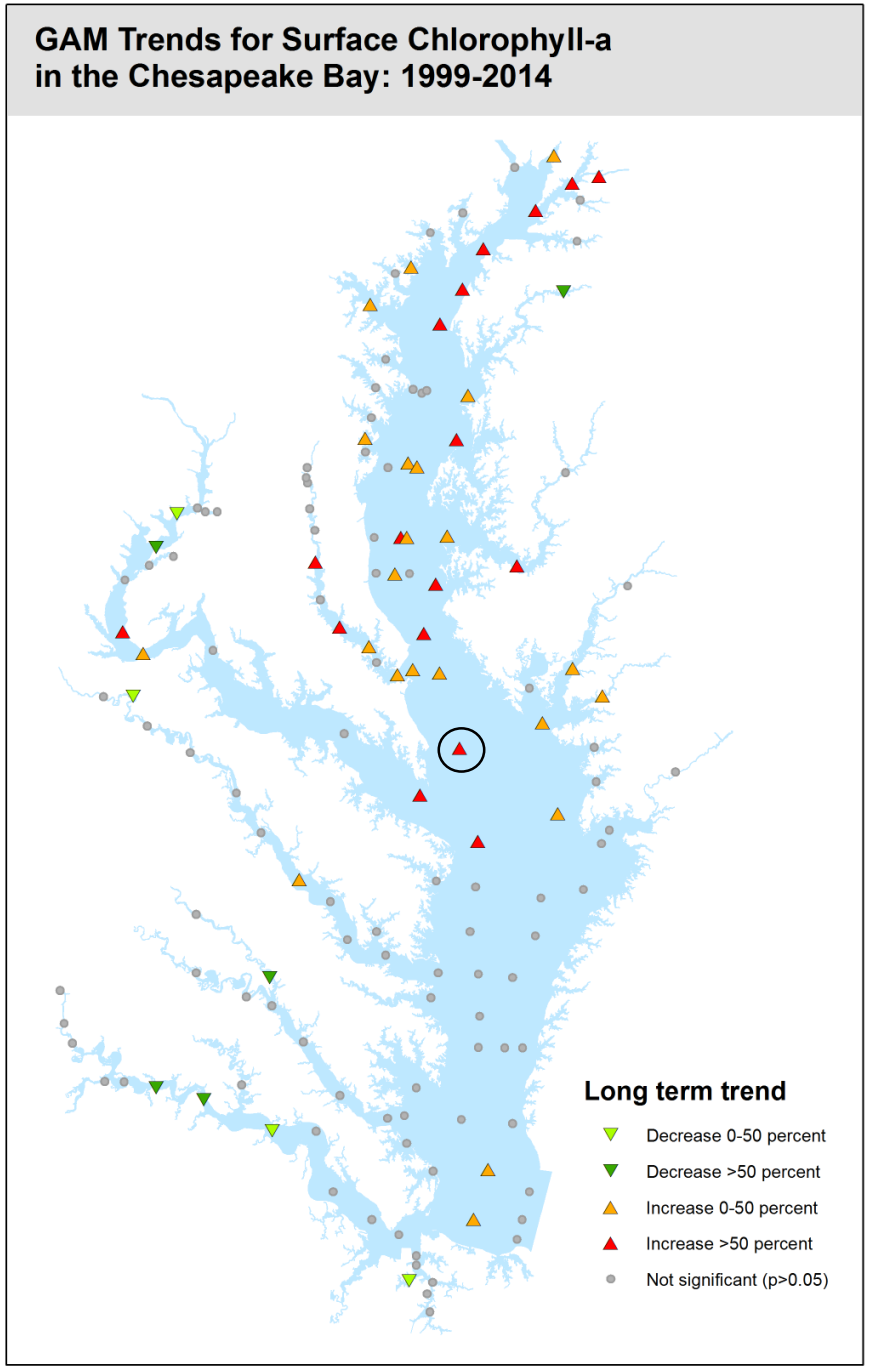


Version 1: Results

Layers of output:

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

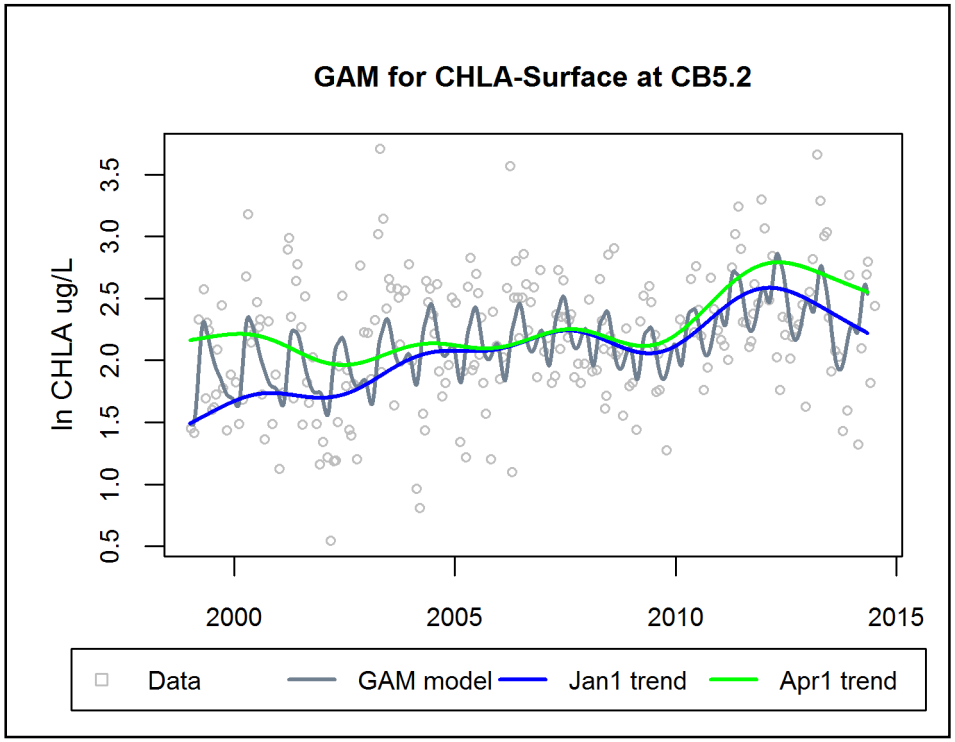


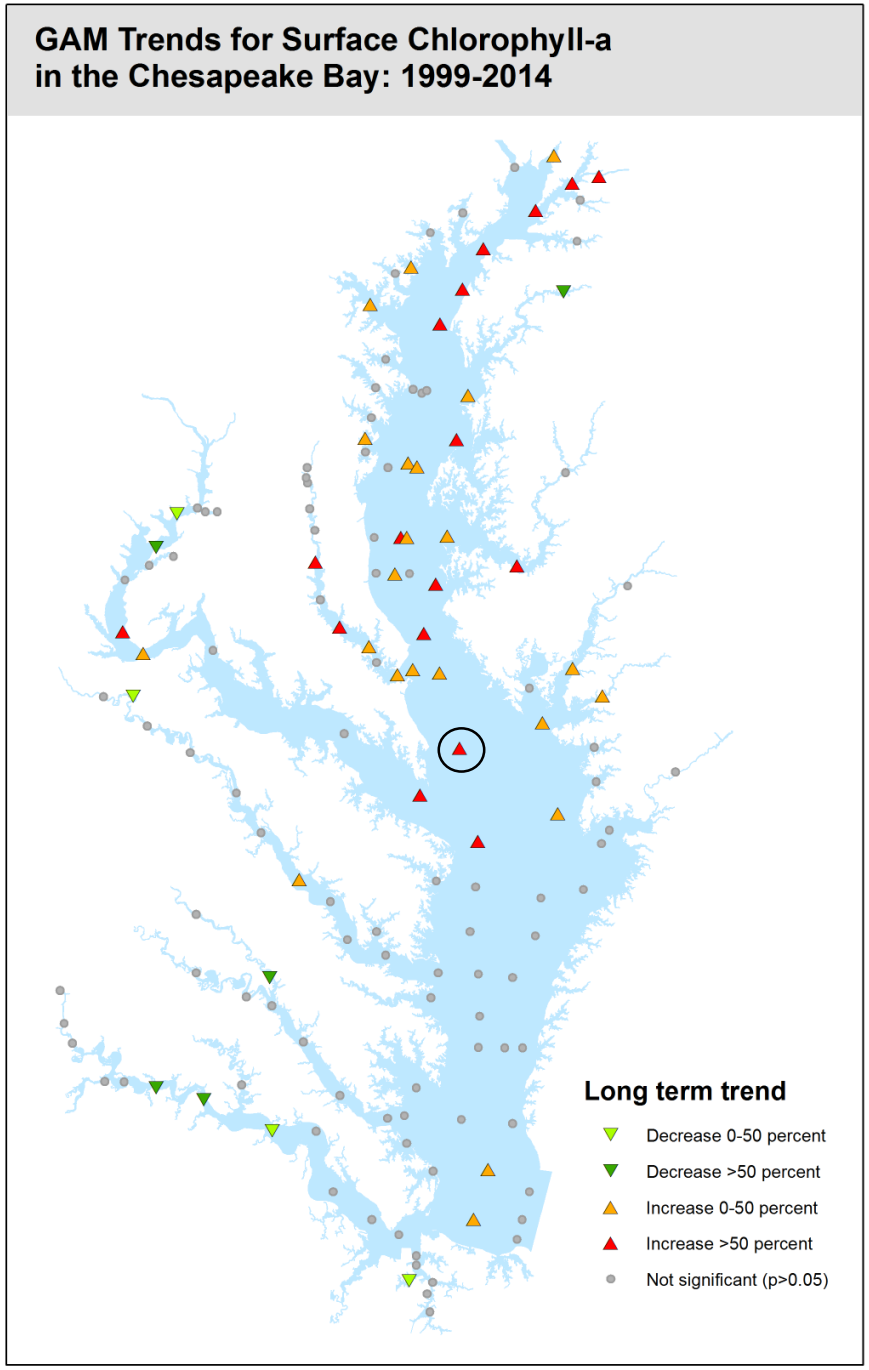


Version 1: Results

Layers of output:

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

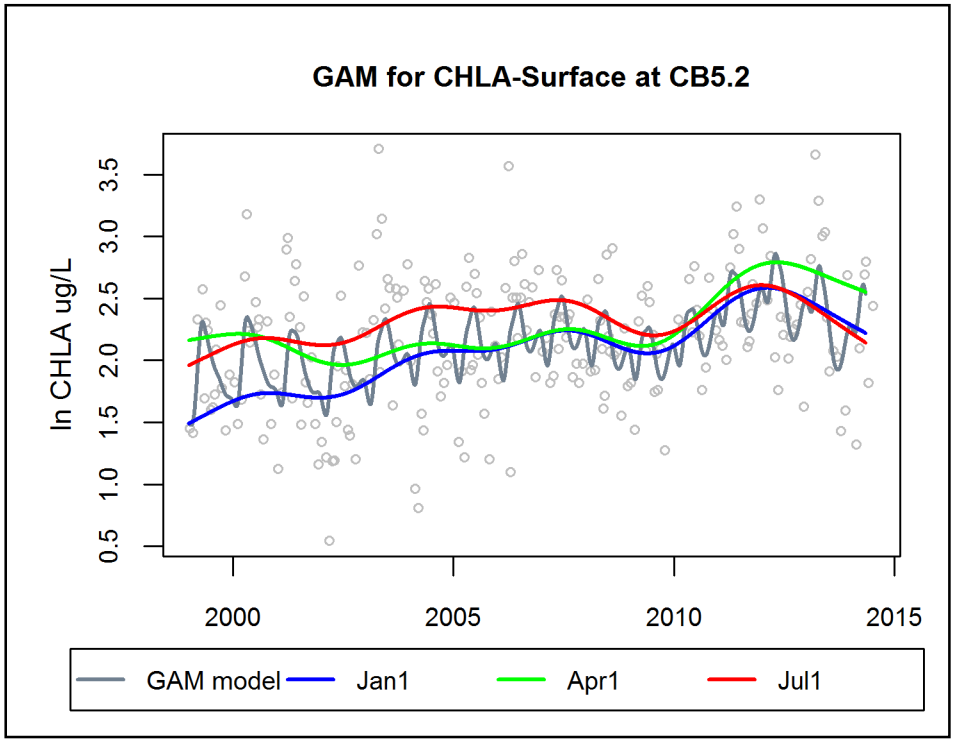


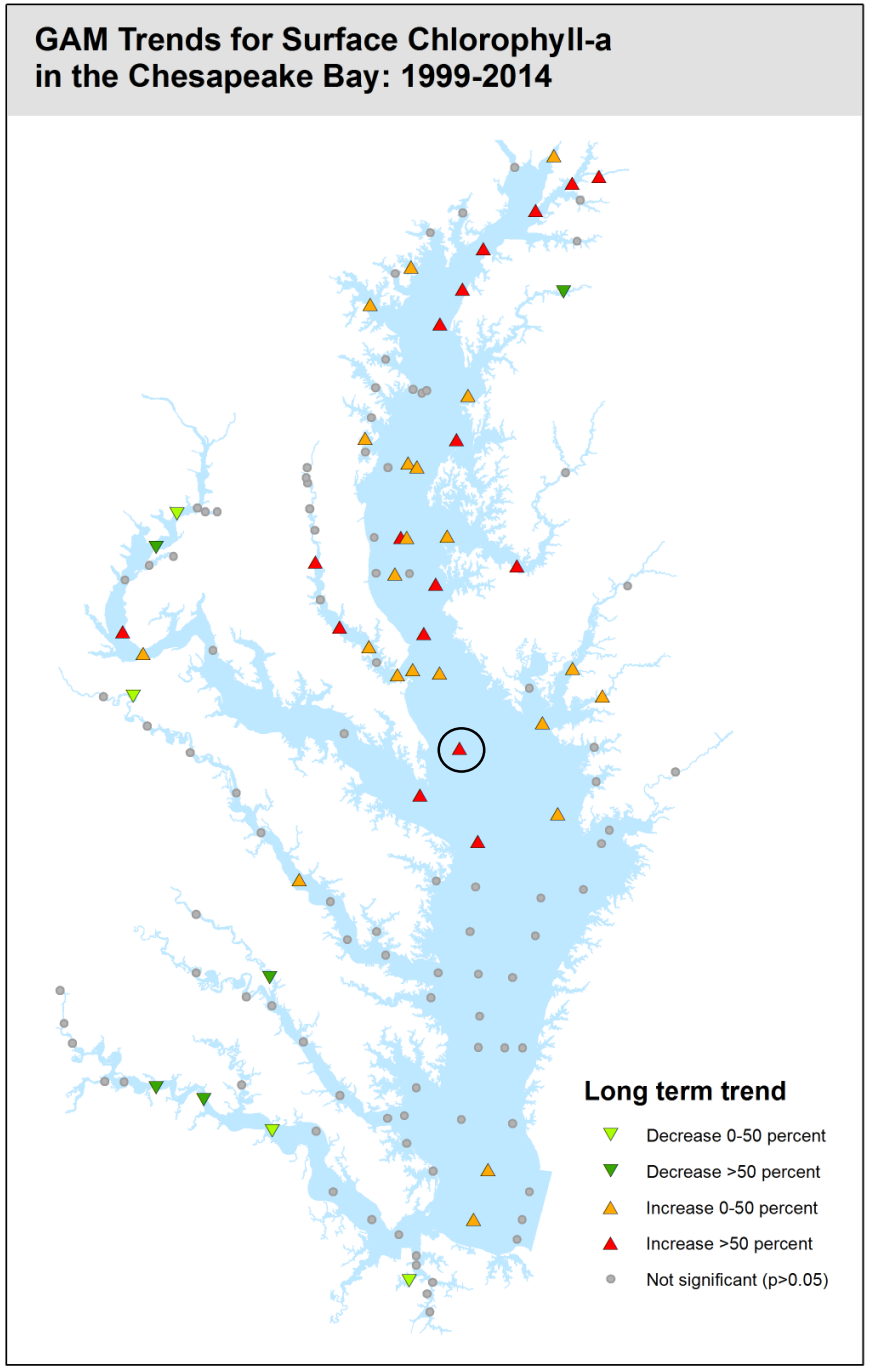


Version 1: Results

Layers of output:

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

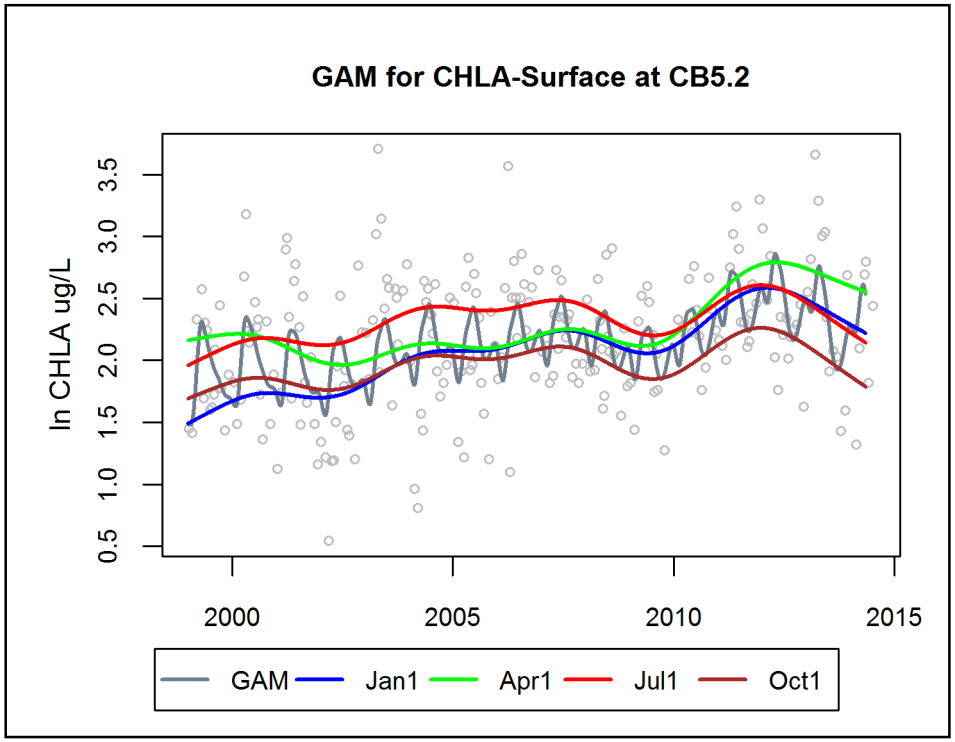


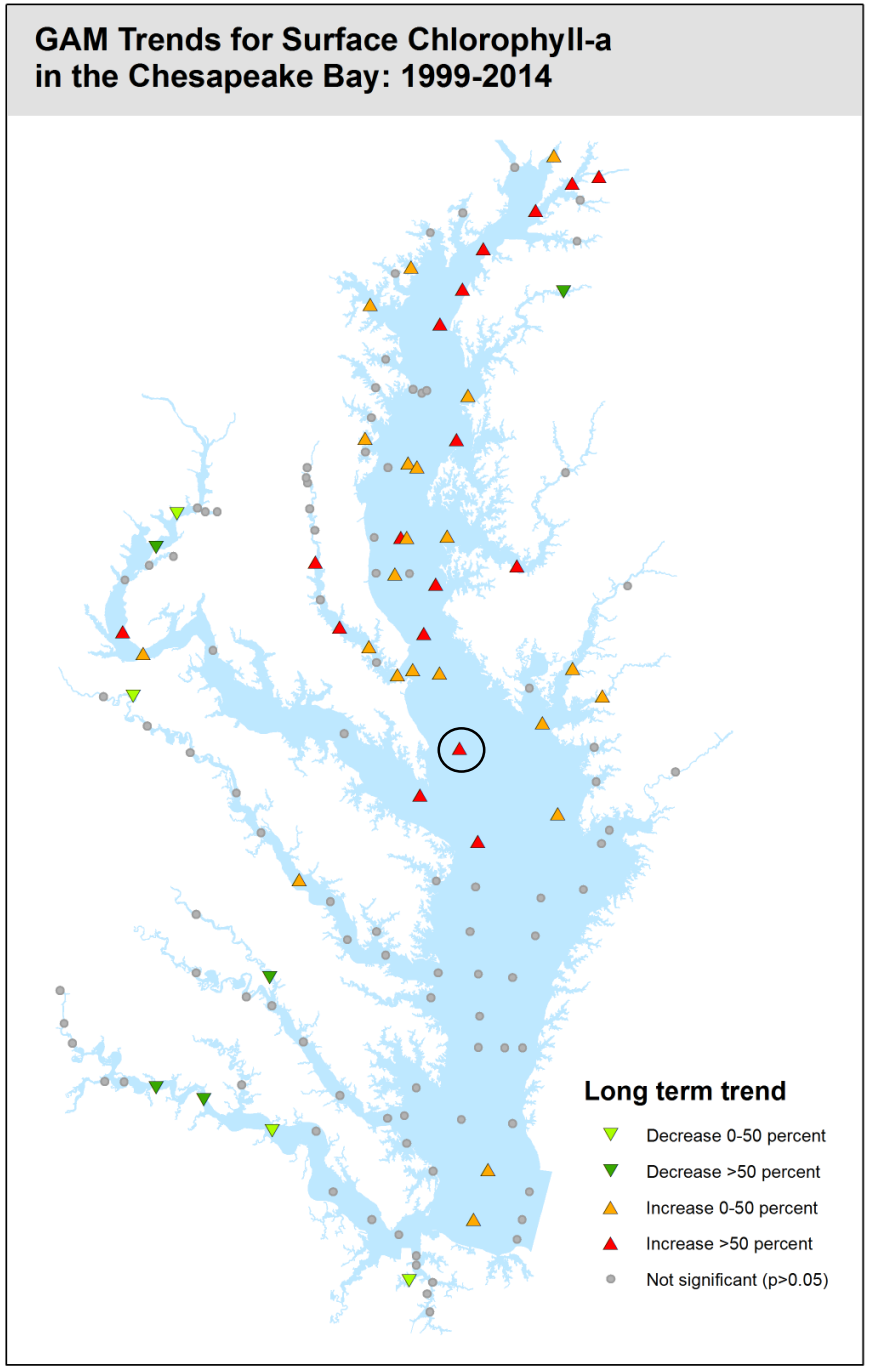


Version 1: Results

Layers of output:

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





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?

Next Steps

- GAM tool in R
  - First draft end of the week
  - Updated version end of November
- Version 2 GAM approach for tidal stations
  - Finalize flow as explanatory variable approach
  - Application to 1985-present
- Applications for factors explaining trends
  - Examining V1 results will help target further analyses
  - Work begun to use GAMs with nutrient inputs and climatic factors to explain trends
  - Doing this hand-in-hand with the research community (ITAT)