Status Update: GAMs implementation for estimating tidal water quality trends

Integrated Trends Analysis Team Meeting

Nov. 12, 2014

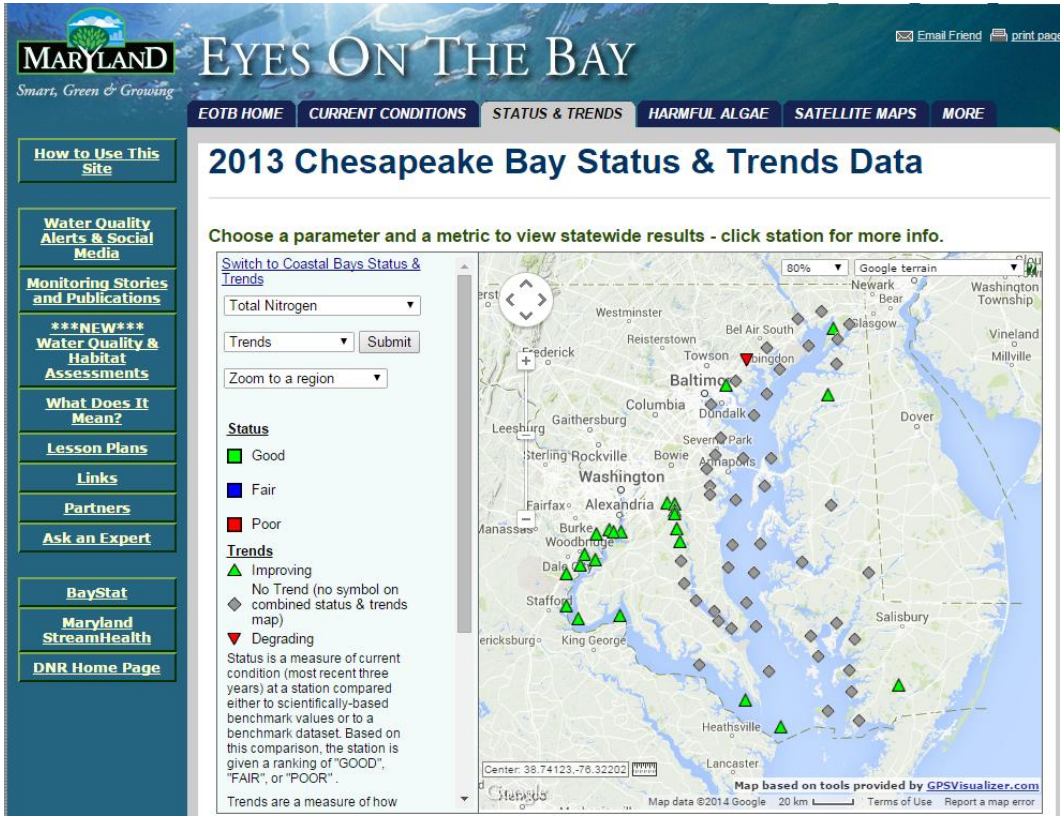
GAM Implementation Group Meeting Participants

- Jeff Chanat (USGS)
- Cindy Johnson (VADEQ)
- Renee Karrh (MDDNR)
- Jeni Keisman (USGS)
- Mike Lane* (ODU)
- Doug Moyer (USGS)
- Tish Robertson (VADEQ)
- Bill Romano (MDDNR)
- Rebecca Murphy (UMCES)
- Elgin Perry* (consultant)
- Amanda Pruzinsky (CRC)
- Peter Tango (USGS)

Trend Analysis in Tidal Waters

- Seasonal Kendall method currently used by MD and VA
- Recent increased focus by CBP and partners to explore trends in water quality that could be linked to TMDL efforts
- Therefore, this re-evaluation of the approach that can:
  - Analyze trends in DO, clarity, chl-a, and nutrients
  - Relate any trends to watershed loads and natural factors

Trends output for Maryland Tidal Waters using Seasonal Kendall Analysis of TN

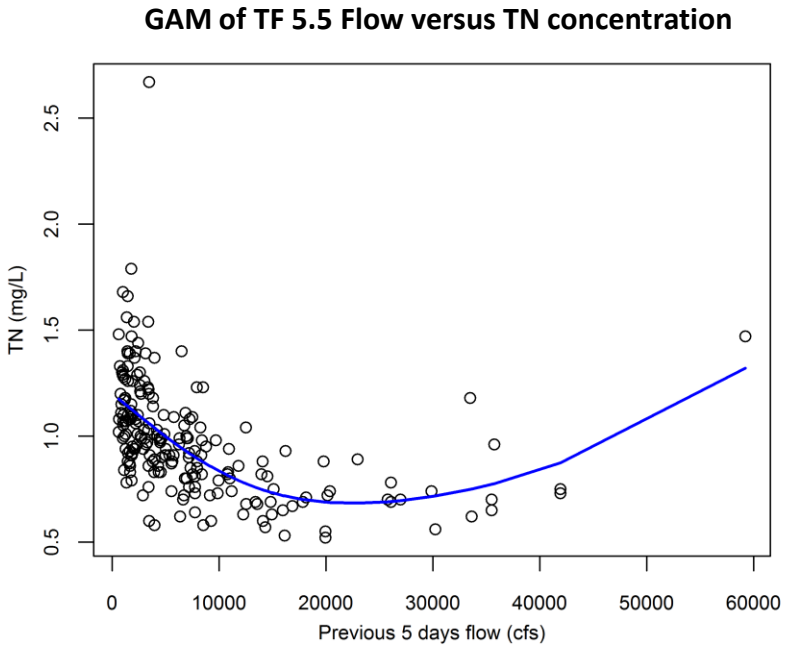


http://mddnr.chesapeakebay.net/eyesonthebay/statustrends.cfm

Why Generalized Additive Models?

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

- Is flexible enough to represent many possible patterns
- Is able to model non-linear relationships
- Generates a statistical confidence measure



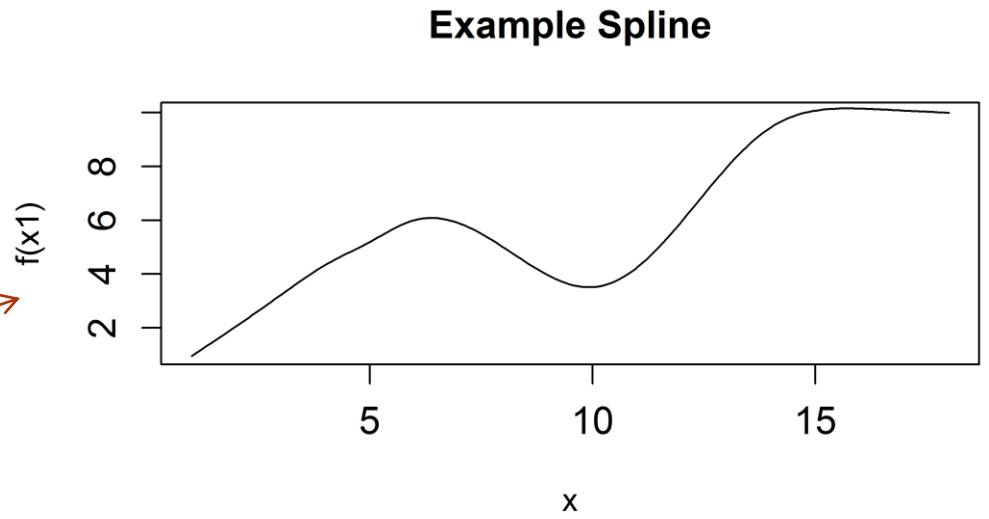
GAM Basics

A response variable is modeled as the sum of multiple nonlinear (or linear) functions of explanatory variables

$$Y = \beta + f(X_1) + f(X_2) + \ldots f(X_n) + \varepsilon$$

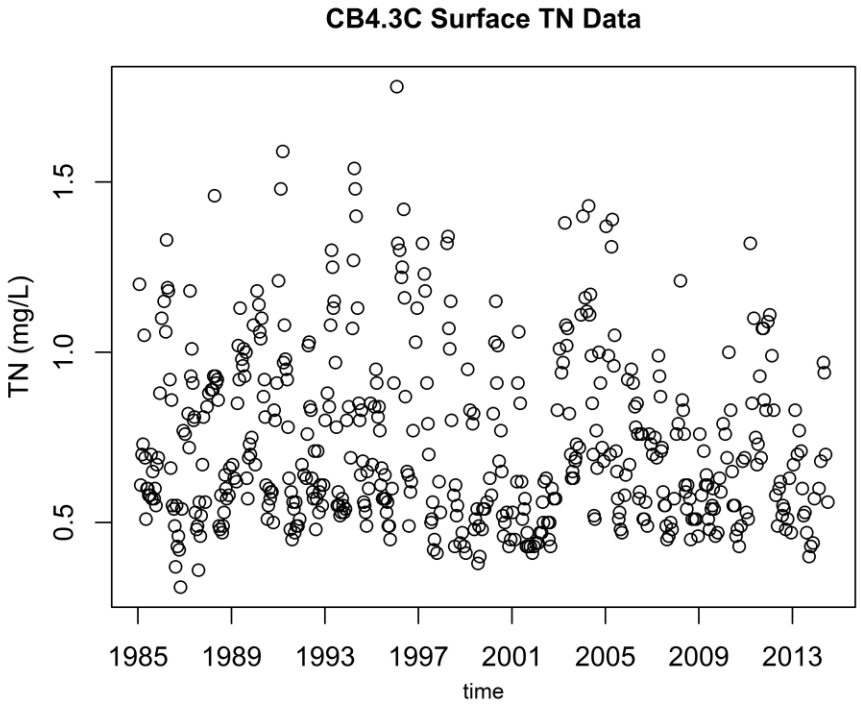
where $\varepsilon \sim N(0,\sigma^2)$

A spline: which can be a series of 3rd order polynomials on segments of the data, connected smoothly at knot locations



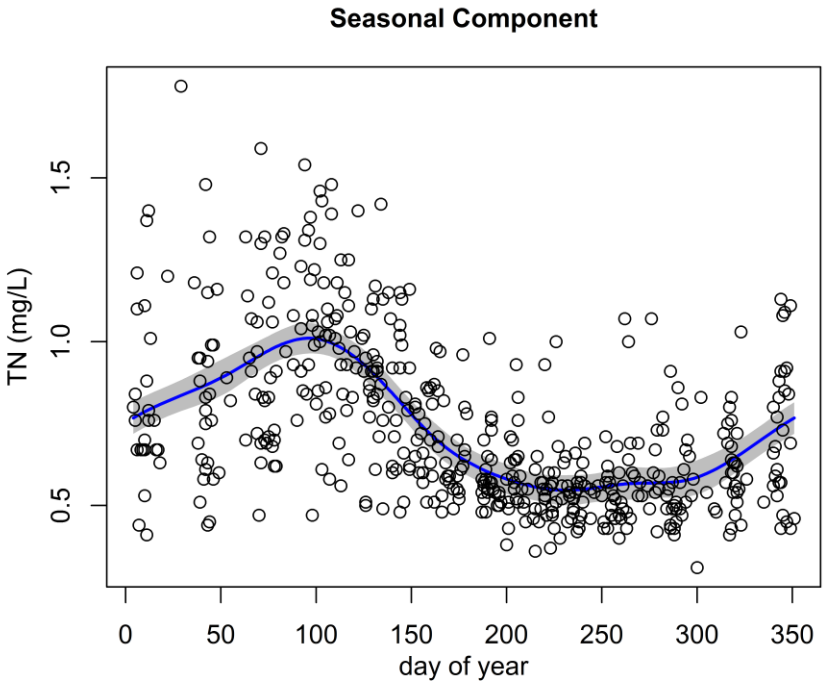
Example GAM Application: Daily Total Nitrogen Concentration at CB4.3C

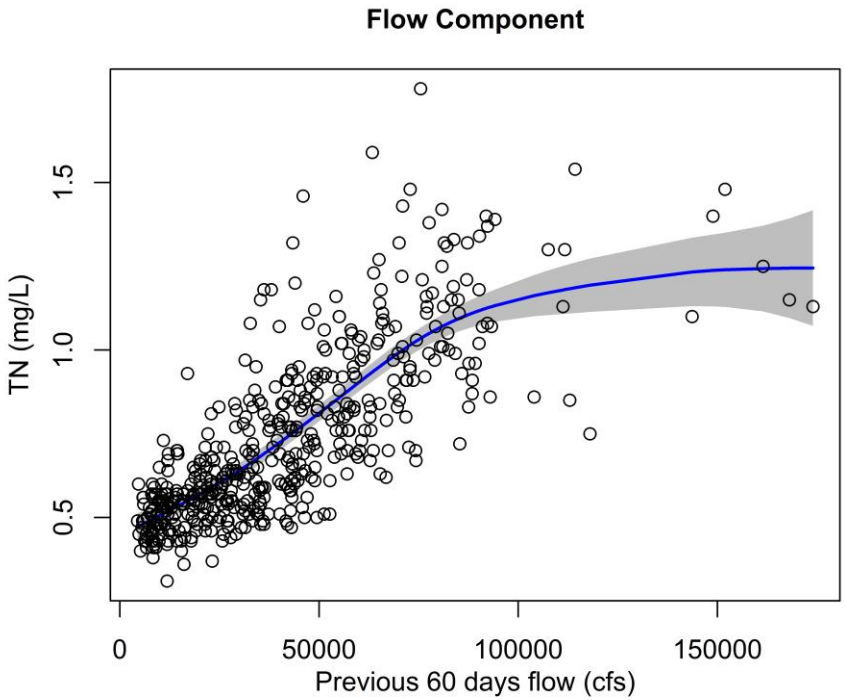
The data:



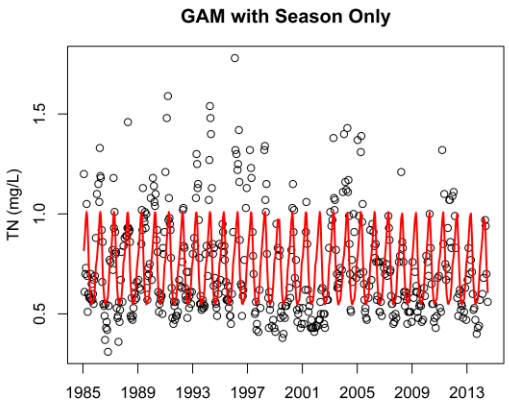
Proposed Model: $TN = \alpha + f(season) + f(year) + f(river\ flow) + \varepsilon$

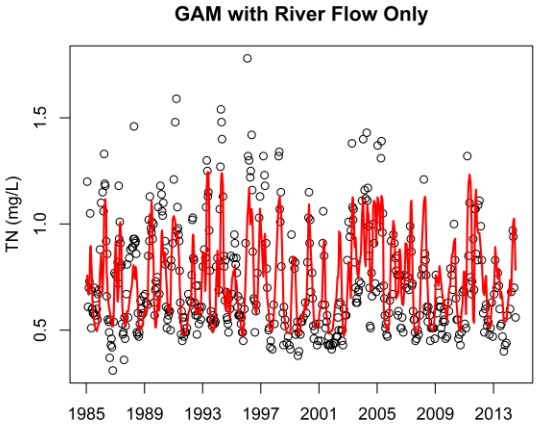
Fitting the GAM:
Explore Relevant Variables

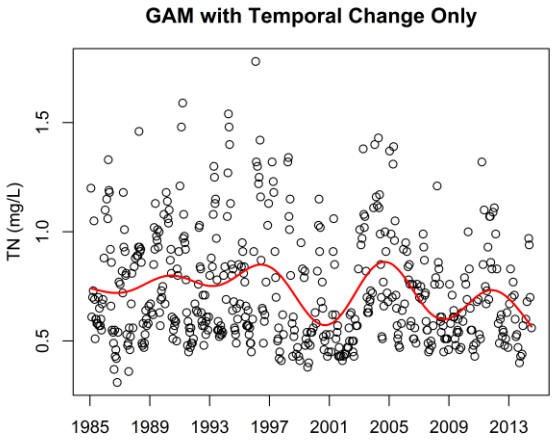


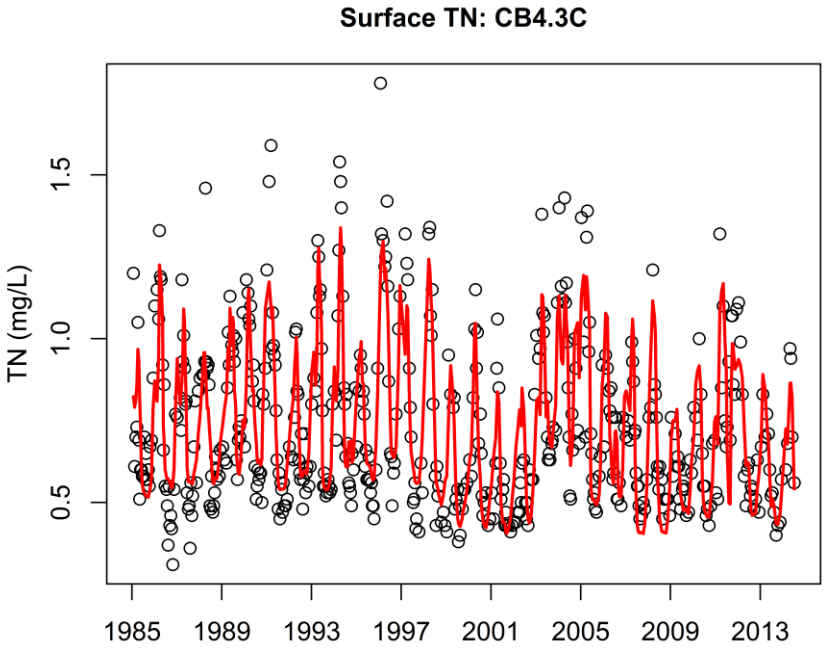


Combine the Variables





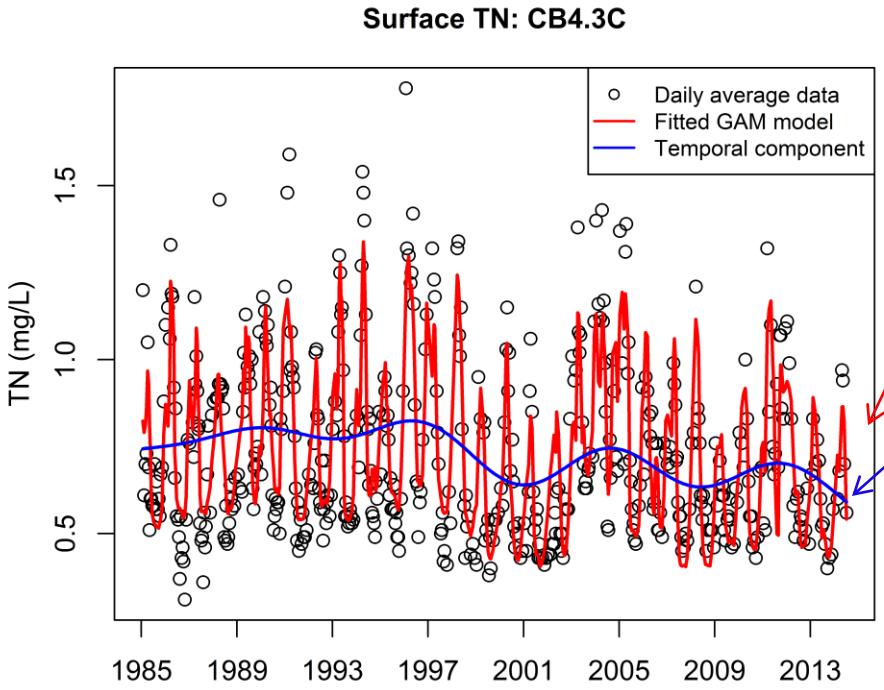




Some model information

variable	edf	F	p
season	5.1	16.2	<0.001
flow	7.2	53.6	<0.001
date	8.7	12.1	<0.001

Resulting Model (draft)



Red is all 3 factors

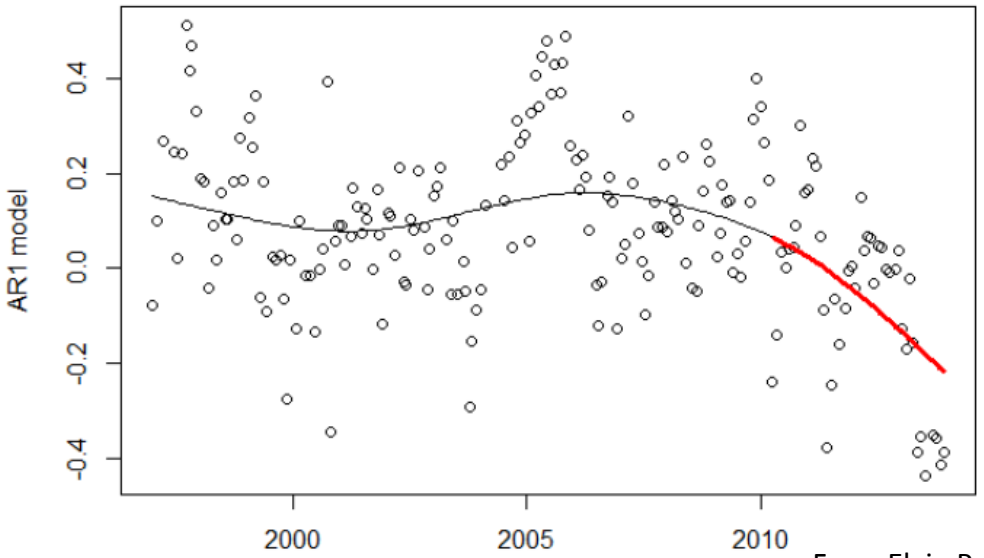
Blue is the annual temporal change after accounting for season and flow

Possible improvements

- Temporal autocorrelation (with a GAMM)
- Interaction between components (e.g., if trend is changing differently in different seasons)

Current Work

- Identifying periods of significant change
- Implementing GAM approach in SAS and R



From Elgin Perry

- Pilot side-by-side comparison with Seasonal Kendall trend analysis

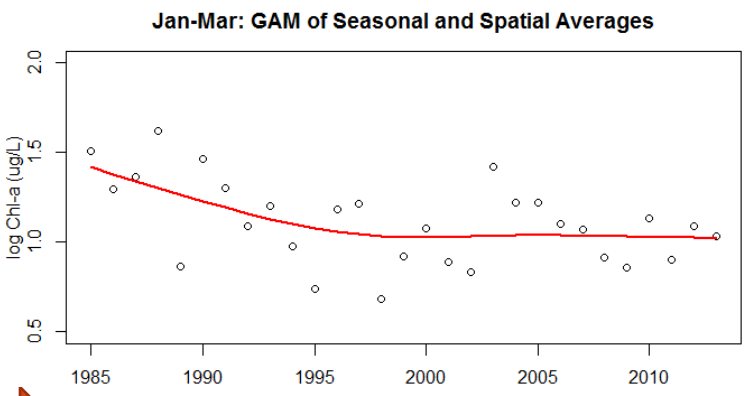
Where we are headed

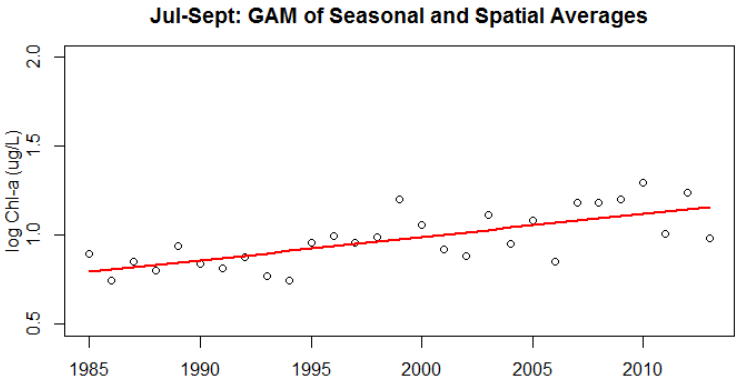
- Big picture tool:
  - Fit GAMs to many of the WQ datasets throughout the Bay
  - Look for overall patterns
- For specific pilot locations:
  - Build collaborations with researchers
  - Detailed spatial and temporal analyses
  - Incorporate nutrient loads and other explanatory variables in the GAMs to test hypotheses

Example:

$$CHLA = \alpha + f(light) + f(river\ flow) + f(TN\ load) + f(TP\ load) + \varepsilon$$

Example: Draft Chl-a trends with seasonal differences from 3 tributary stations





Questions?

Contact Info:

Rebecca Murphy

rmurphy@chesapeakebay.net

410-267-9837

GAM package in R used here: mgcv

http://cran.r-project.org/web/packages/mgcv/index.html