DRAFT: Subject to Revision

Methods for Application of Generalized Additive Models (GAMs) for Water Quality Trends in Tidal Waters of Chesapeake Bay

May 2018

Prepared by:
Rebecca Murphy, University of Maryland Center for Environmental Science
and
Elgin Perry, Statistics Consultant
For the Chesapeake Bay Program Integrated Trends Analysis Team

Contents

Appendices

1. Background and overview

Trend analyses of the Chesapeake Bay tidal water quality data have been conducted in partnership between the Chesapeake Bay Program (CBP) and the state agencies (Maryland Department of Natural Resources, MDDNR, and Virginia Department of Environmental Quality, VADEQ) since the early 1990s (Ebersole et al. 2002). These have been used by the CBP and the state agencies for multiple purposes including detailed water quality reports, communication tools to managers, politicians, and the public, and as input to scientific research studies (e.g., Dauer and Alden 1995).

Until recently, the Seasonal Kendall (SK) nonparametric trend technique (Hirsch et al. 1982) has been the primary tool to identify significant degrading or improving trends in the tidal water quality data. The SK test is a non-parametric test for a monotonic trend (Gilbert 1987). The application of the Seasonal Kendall test for Chesapeake Bay tidal trends has been described by Ebersole et al. (2002) and Marshall et al. (2009). For Chesapeake Bay tidal data, a strength of the approach is that it does not require any particular distribution, but a weakness is that it is only designed to test for trends that do not change direction over time. Soon after the SK approach was implemented, it became clear that many of the long-term patterns were not monotonic, and an additional quadratic test was added to the analysis in order to identify those types of patterns. Additional complications arose as the data sets lengthened, analytical techniques were updated, laboratories analyzing the samples changed, and detection limits improved for some constituents. Changes like these can often result in shifts, inconstancies, and even the potential for creating spurious trends in the data. As a result, piece-meal approaches have been implemented over time to deal with these changes in regards to analyzing long-term trends (Ebsersole et al. 2002). By 2015, trend computations conducted by the state agencies included:

- Blocked Seasonal Kendall tests from Virginia for nutrients to account for 1990s method changes,
- Two separate sets of trend results for data pre-and post-1999 for nutrients in Maryland, and
- Analyst-required pre-processing of the data for parameters and stations where detection limits have changed in order to censor to the highest detection limit.

Despite the challenges, examining the water quality trends over time has become more critical as the timeline for implementing the Total Maximum Daily Load (TMDL) by 2025 progresses, and there is a desire to identify impacts of management actions on Bay water quality. Therefore, the Chesapeake Bay Program analysts, state partners, and collaborators at USGS proposed the consideration of Generalized Additive Models (GAMs) as an alternative to the SK/quadratic model approach used for trends to-date. A GAM is a statistical model in which a response of interest can be modeled as the sum of multiple smooth functions of explanatory variables. These smooth functions can be constructed in many ways (Hastie and Tibshirani 1986, 1990), and GAMs allow for model shapes from linear to nonlinear – including patterns that change direction over time. Research applications of GAMs to water quality change are relatively new, but growing (e.g., Haraguchi et al. 2015), with some recent work in Chesapeake Bay (Harding et al. 2016; Beck and Murphy 2017).

The work on trends in tidal water quality is not being done in isolation from efforts to analyze trends in nontidal river and stream water quality conducted at USGS using the Weighted Regression on Time, Discharge, and Season (WRTDS) approach (Hirsch et al 2010, Moyer et al. 2012). In March 2014, STAC and the Harry R. Hughes Center for Agro-Ecology co-sponsored the workshop “Management Effects on Water Quality Trends” to bring together state and federal partners, academic, and federal researchers to examine updating methods for trends analysis in the Chesapeake Bay and its watershed. Part of the discussions were centered on recommendations related to analytical approaches for analyzing changes in water quality within the Chesapeake Bay (Keisman et al. 2014). On the topic of trend detection, one of the major findings from the workshop was that the use of GAMs for estuary tidal trends requires further development. The specific recommendation was: "The CBP should continue to develop and apply GAMs to the appropriate response variables in tidal waters, and should develop a process of 'artificial intelligence' that enables automated application of GAMs." A peer review was recommended of the GAM technique as applied to the Chesapeake Bay before establishing it as the primary tool for examining trends in tidal water quality. A review of the GAM approach by STAC was conducted in 2017 (Ellis et al. 2017).

Currently, we have incorporated the GAM approach from the 'mgcv' R package (https://cran.r-project.org/web/packages/mgcv/mgcv.pdf) into our R package called baytrends which has been designed to fit GAMs for the tidal Chesapeake Bay water quality data over time. This documentation covers the GAM implementation for Chesapeake Bay (Section 2) and briefly describes future plans (Section 3). This documentation is not a User's Guide for the baytrends package, but instead describes this implementation of GAMs in detail, including the statistical and analysis decisions made and built into the baytrends package. Instructions for getting started on the baytrends package are included as Appendix 1A, and additional user guidance is available in the help files of the R package itself.

2. Current GAM application to Chesapeake Bay

2.1 Overview of baytrends functionality and output

The CBPO team has worked with contractor support from Tetra Tech to develop the baytrends package for the statistical software R to run the Chesapeake Bay tidal trends analyses with GAMs. The package has already been used by MDDNR and VADEQ (via analysts at Old Dominion University, ODU), and can be shared with others interested in these analyses. It runs in R Studio, which is a free, commonly used, interface for the open source R statistical software (https://www.rstudio.com/home/). We are using R markdown (Allaire et al. 2015) to allow for easy report generation in Microsoft Word. The process for a user involves opening an R markdown script in R studio; making modifications to input and output file locations, selection of parameters, stations, layers, and which analyses to conduct; matching of river flow stations and identification of interventions if applicable; and then running the script to generate both the Word file report and Excel tabular output.

The baytrends R package requires other published R packages to run, specifically 'mgcv' for the GAM functionality (see Section 2.2). See Appendix 2.1A for a copy of an R output report for a GAM analysis run for one station and parameter, and an example tabular output file for the same run. Details of the statistical functionality behind the baytrends package are described in the rest of this documentation.

2.2 The 'mgcv' package for GAM computations

As mentioned in Section 2.1, the GAM function used in baytrends is from the 'mgcv' R package, whose author, Dr. Simon Wood, has published the details of this approach in multiple publications (cited in this section) as well as the documentation for the R package (https://cran.r-project.org/web/packages/mgcv/mgcv.pdf). Features of this GAM implementation that are particularly valuable for Chesapeake Bay tidal trends analysis are: (1) automatic smoothness selection (Wood 2004), (2) smooths of more than one variable (Wood 2006), and (3) uncertainty calculations using a Bayesian approach (Marra and Wood 2012; Wood 2013). These features have all been published in the peer-reviewed literature and are not presented in detail here. Instead, they are described briefly to provide a background understanding of this GAM application for users.

In practice, a user of 'mgcv' inputs a proposed model structure into the function 'gam' where the independent variable is a function of smooth functions of predictor variables. For example a common model structure we are using is:

$$y = s(x_1) + s(x_2) + ti(x_1,x_2) \qquad \text{(Eqn 1)}$$

Where y is the dependent variable, x_1 and x_2 are predictor variables, s() indicates an isotropic smooth function, and ti() represents a scale-invariant smooth function of two parameters. The model is fit with a call to the 'gam' function, predictions with estimates of standard error are generated using 'predict.gam', and a summary of the model fit and estimated p-values on components are generated with a call to 'summary.gam.'

Automatic smoothness selection

For each s(), ti(), or similar component in Eqn. 1, a smooth class must be defined. For our proposes, we usually use the default penalized thin plate regression spline (selected with: bs='tp') (Wood 2003). The option of penalized thin plate regression splines in 'mgcv' helps overcome many of the obstacles to fitting GAMs in an application like ours. Specifically, these splines are a computationally efficient way to fit a somewhat "ideal" smoother called a thin plate spline, which avoids the need to manually place knots (Wood 2006).

This "ideal" smoother, the thin plate spline (Duchon 1977) is a smooth function of predictor variables that balances model fit vs. wiggliness of the function. This is accomplished with a smoothing parameter, λ, which controls that tradeoff, and mathematically removes the need to explicitly select knot locations for the function. It is useful that these "ideal" thin plate splines remove the need to select knot locations, but they also require as many parameters as there are data to define the spline. Therefore, fitting a full version of this spline is a computationally expensive process (Wood 2006). That is why the *regression* version of this spline is useful in the 'mgcv' package. Thin plate *regression* splines are low-rank approximations to full thin plate splines (Wood 2003). This regression approach was derived to essentially minimize the worst possible change in shape and fitted value, while retaining the useful properties of thin plate splines (Wood 2006).

In 'mgcv' penalized likelihood maximization, via penalized iteratively re-weighted least squares, is used to fit the model coefficients of the thin plate regression spline while generalized cross validation is used to fit the smoothing parameter in an iterative manner. The user has to specify a basis dimension, or accept the default (Wood 2006), which we discuss further in Section 2.4.

In our GAM structure, we use one other smooth class, the cyclic penalized cubic regression spline (selected with: bs='cc'). This structure is used on the seasonal parameter (day of year, doy) so that the spline function is smooth and equal in value from the transition between December 31 and January 1 of the next year. A cubic regression spline is a spline that is continuous to the second derivate at each knot location and has zero second derivative at each knot as well (Wood 2006). Unlike the thin plate regression spline, the number of knots are explicitly set for this spline. However, a default of 10 is included in the 'mgcv' package which we use to set these knots evenly throughout the parameter space (https://cran.r-project.org/web/packages/mgcv/mgcv.pdf). This is reasonable for the purposes of a seasonal smooth because 10 even knots should be sufficient to specify an annual seasonal cycle.

Smooths of more than one variable

A thin plate spline can be a function of more than one variable, with the assumption of isotropy in all directions. However, if isotropy cannot be assumed, which is generally our case with variables including date, day, and flow, then a tensor product bases can be used instead. For a tensor product bases (ti() in Eqn. 1) each of the terms has its own low-rank bases specified, for example by two separate thin plate regression splines. The two 'marginal smooths' are combined in the tensor product construction to represent how the variables vary jointly (Wood 2006).

Uncertainty calculations

Estimates of uncertainty from the GAM model fits are needed for both testing the relevance of individual predictor variables as well as to generate confidence in the overall model predictions. The p-values on model terms are estimated using the Bayesian estimated covariance matrix of the parameter estimators and are useful for testing for inclusion or exclusion of model terms (Wood 2013). Notably, however, these are computed without consideration for the uncertainty in the smoothing parameters estimates, meaning they are not precise p-values and may be too low when smoothing parameters are fairly uncertain (mgcv documentation). That consideration needs to be kept in mind as we use them to weigh the influence of factors explaining water quality trends (Section 3). To generate confidence bounds on the complete model fit, we rely on the computed standard errors, which are based on the Bayesian posterior distributions of the coefficients, and are shown in simulation studies to be reliable for the 'whole model', but with problems on a component-wise basis (Wood 2006). Our approach will be to use the 'whole model' confidence bounds to show results graphically as well as estimate percent change over the entire model (Section 2.4), and we will acknowledge that component-specific confidence bounds are more uncertain.

An additional feature was added to our implementation to compute periods of significant change. This is based on work by Dr. Gavin Simpson (http://www.fromthebottomoftheheap.net/2017/03/21/simultaneous-intervals-for-derivatives-of-smooths/), and will be more fully documented in a future version of this documentation. The result of this computation allows us to highlight periods of change in the smoothed output (Figure 1, yellow highlights) which are significantly changing over time.

2.3 Data format and processing required

Any water quality data could be entered and evaluated with baytrends as long as all the required files are available (see Appendix 1A). To fully account for censoring in the dataset, we have selected a format for the data (called qw) originally from the package smwrQW (https://github.com/USGS-R/smwrQW). This package is no longer being fully supported, so we have built in the functionality needed to transform and use data in the qw format. See Appendix 2.3A. An example of this format can be seen by examining the built-in dataset named 'dataCensored.' More details on the methods to account for censored data is in Section 2.6.

Another important data decision is whether or not to transform a data set before analyzing it. The standard decisions we have made for Chesapeake Bay tidal analysis are that all nutrient parameters and chlorophyll-*a* are log-transformed by taking the natural logarithm of the observed values before conducting the GAM analysis. Of the suite of parameters discussed here, dissolved oxygen and Secchi disk depth are not log-transformed. There is certainly precedent for these transformation decisions in water quality analysis, and we have examined the data distributions. We do not back-transform the results using a bias correction (e.g., Bradu and Mundlak 1970; Duan 1983) as is done with WRTDS for watershed load estimates (Hirsch et al. 2010). Based on the type of analyses we are conducting of the GAM results, it does not seem necessary to obtain estimates of the arithmetic mean. Any conclusions about trend direction, shape, or change over time are on the geometric mean of the distribution for these parameters. Graphical results are presented with a log-scale for the y-axis (see Fig 1a and 1c). These transformation decisions can be modified by the user in the parameter input table.

2.4 Model structures

Currently six progressively more complex GAM structures are available in the baytrends package. These models are shown in Table 1.

Table 1. Temporal GAM structures in baytrends

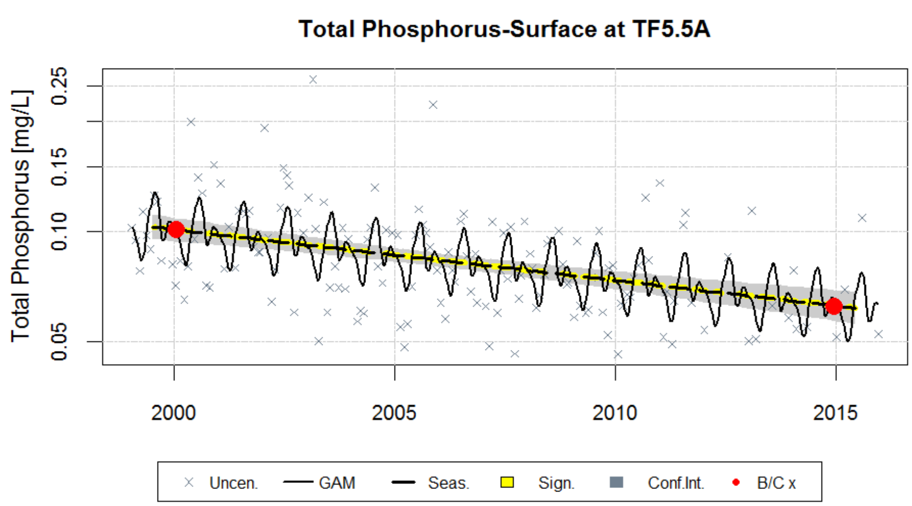
Model name	**Description**	**Structure of right hand side of equation**
gam0	Linear trend with seasonality	cyear + s(doy,bs='cc'), knots = list(doy = c(1,366)), select=TRUE
gam1	Nonlinear trends with seasonality	cyear + s(cyear, k=gamK1) + s(doy,bs='cc'), knots = list(doy = c(1,366)), select=TRUE *where: gamK1=c(10,2/3) means that the maximum of 10 or 2/3*number of years is selected*
gam2	Nonlinear trend with seasonality (plus interaction)	cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')), knots = list(doy = c(1,366)), select=TRUE *where: gamK1=c(10,2/3) means that the maximum of 10 or 2/3*number of years is selected*

gam3	Nonlinear trend with seasonality (plus interaction) and intervention	intervention + cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')), knots = list(doy = c(1,366)), select=TRUE *where: gamK1=c(10,2/3) means that the maximum of 10 or (2/3*number of years) is selected*
gam4	Nonlinear trend with seasonality (plus interaction) and hydrology effect	cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')) + s(flw_sal,k=gamK2) + ti(flw_sal,doy,bs=c('tp','cc')) + ti(flw_sal, cyear,bs=c('tp' ,'tp')) + ti(flw_sal,doy,cyear, bs=c('tp','cc','tp')), knots = list(doy = c(1,366)), select=TRUE *where: gamK1=c(10,1/3) means that the maximum of 10 or (1/3*number of years) is selected, and gamK2=c(10,2/3) means that the maximum of 10 or (2/3*number of years) is selected*
gam5	Nonlinear trend with seasonality (plus interaction), hydrology effect, and intervention	intervention + cyear + s(cyear, k=gamK1) + s(doy,bs='cc') + ti(cyear,doy,bs=c('tp','cc')) + s(flw_sal,k=gamK2) + ti(flw_sal,doy,bs=c('tp','cc')) + ti(flw_sal, cyear,bs=c('tp' ,'tp')) + ti(flw_sal,doy,cyear, bs=c('tp','cc','tp')), knots = list(doy = c(1,366)), select=TRUE *where: gamK1=c(10,1/3) means that the maximum of 10 or (1/3*num years) is selected, and gamK2=c(10,2/3) means that the maximum of 10 or (2/3*num years) is selected*

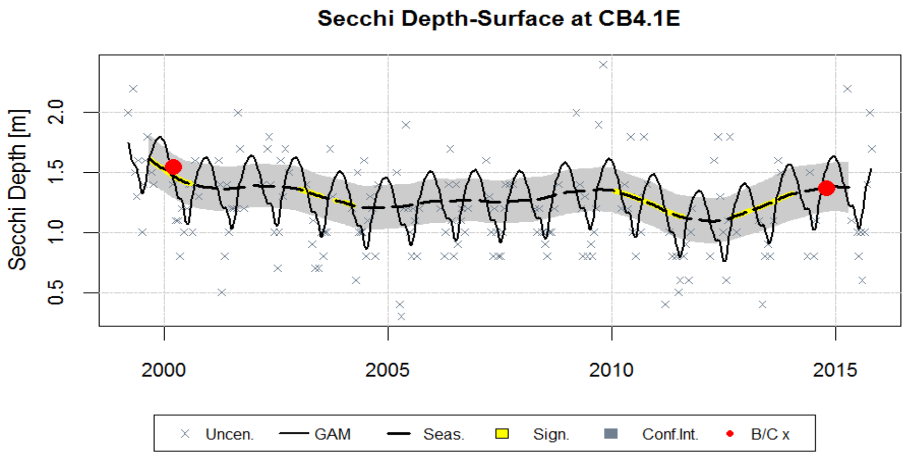
In the equations, *cyear* is a centered decimal date, meaning that a date is turned into a decimal (i.e., 2002.41), and then a time series is centered so that the middle date in a record becomes zero. The variable *doy* is day of year (e.g, 1, 2,… 366), with each year adjusted to 366 days to account for leap years. The *s()* indicates a spline function with the smooth class defined by *bs='cc'* or *'tp'* (see Section 2.2). The *ti()* indicates a tensor product interaction between two variables. The parameter "intervention" refers to an indicator variable that changes during the time series if a method or lab changed occurred that the analyst wants to test as a potentially significant indicator of a change in the values of the observations. And the parameter *flw_sal* indicates either a pre-processed average river flow, or salinity measured at the same place and time. Other items in the equations include specification of knots for the *doy* parameter to include days 1 and 366 so that the seasonal models do not have an artificial jump from one year to the next. The *select=TRUE* specification allows for individuals splines to be completely removed from the GAM during model fitting if they provide no benefit (Wood 2018). An upper limit on the number of knots for each spline can be specified (the basis dimension), and in model development we found that this k-value needed to be set to allow for enough flexibility in the *cyear* function over time. It is set to the maximum of 10 or 2/3 times the number of years for gam2 and gam3. For gam4 and gam5, it was found that the concurvity (Buja et al., 1989) between the spline bases for *cyear* and *flw_sal* was an issue. So an approach based on Peng et al., (2006) was used to limit the flexibility of the smooth on *cyear* so that more of the variability can be modeled with the smooth on *flw_sal*.

This set of models was developed based on the accumulated knowledge from the team that time series for different parameters and locations could follow very different patterns. If there is only a simple linear pattern over time with seasonal variation, then gam0 would be sufficient and no additional benefit would be gained from a more complicated pattern (see, for example Fig 1a). However, experience shows that there are frequently non-linear changes in water quality over time, so gam1 will be needed (e.g., Fig 1b). The model gam2 was included to represent time series for which the seasonal cycles is changing over time (e.g., Fig 1c). This structure in gam2 is most similar to the functionality in WRTDS used for non-tidal trend analysis (without including flow). The gam3 was included because over 30 years of the monitoring program, in some cases laboratory or method changes have resulted in shifts in the data. Section 2.7 describes the intervention approach in more detail, and Fig 1d shows and example. The effect of river flow fluctuations on water quality was an important component to model, and adjust for, in order to see what the long-term patterns would look like if extreme fluctuations in river flow had not happened. Examples in Fig 1e and 1f show this and more details are in Section 2.8. Draft baytrends detailed output for these examples is provided in Appendix 2.4A and Table 2 describes the abbreviations in the figure legends.

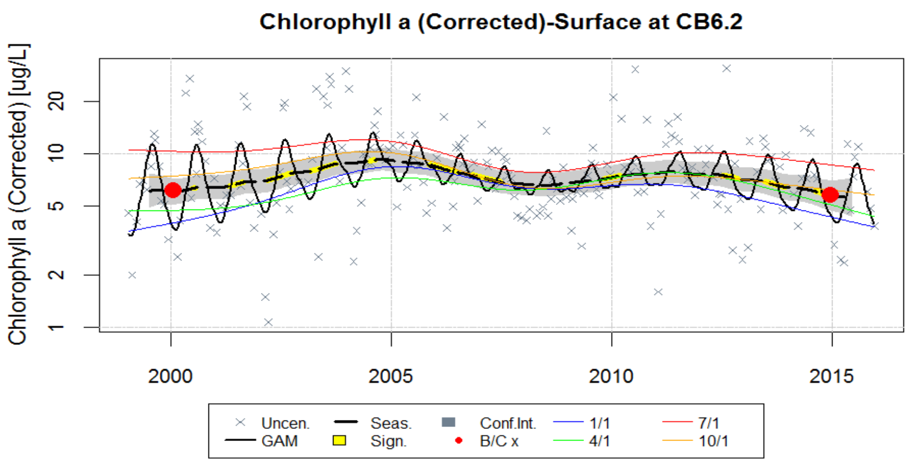
1a. gam0 example



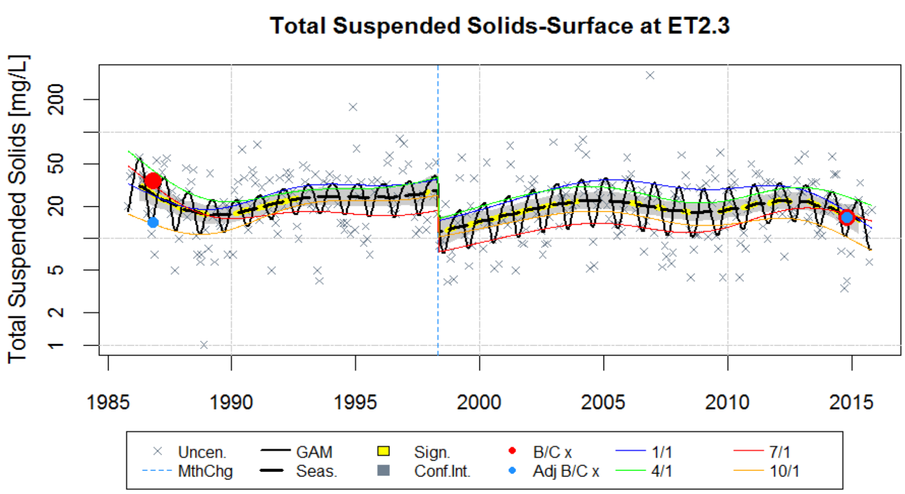
1b. gam1 example



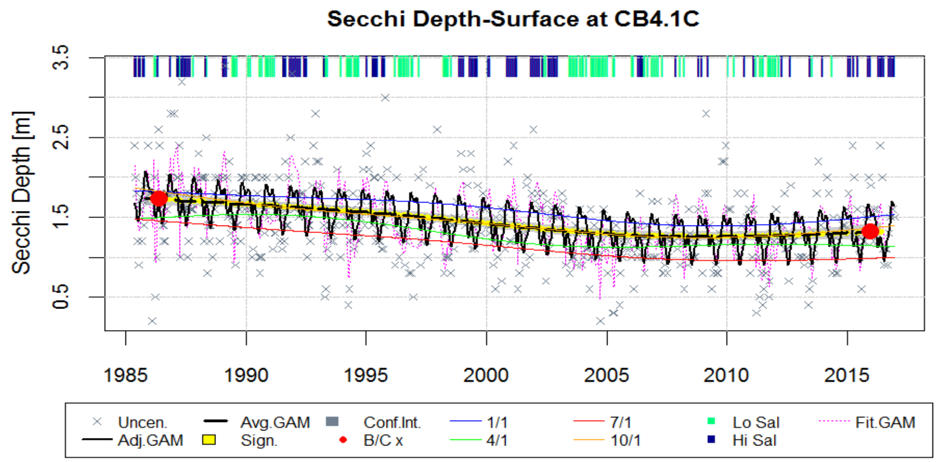
1c. gam2 example.



1d. gam3 example.



1e. gam4 example



1f. gam5 example

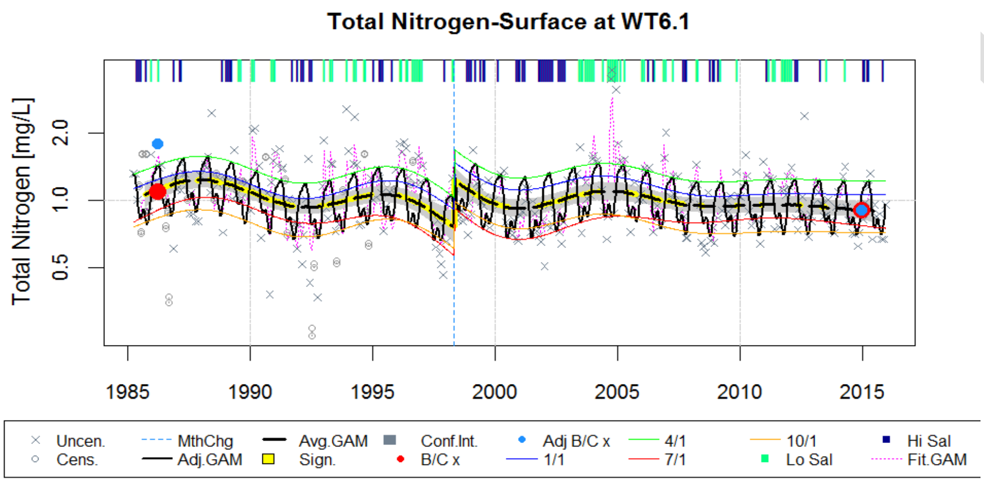


Figure 1. Example graphic output for different data sets demonstrating the types of conditions that would be best represented with each of the GAM model options in Table 1.

Table 2. Legend output from baytrends

Legend item	**Description**
Uncen.	Observed data that was not censored
Cens.	Censored observation shown at its detection limit (none in these examples)
MthChg	Date of method change tested in intervention
GAM	Full GAM fit
Seas	Seasonally-adjusted long-term mean estimates
Sign.	Period of significant change
Conf Int	95% confidence interval on the seasonally-adjusted mean estimates
B/C x	Baseline and current estimate of the mean
Adj B/C x	In gam3 or 5, and adjusted estimate of the baseline and current mean for the intervention

1/1, 4/1, 7/1, 10/1	Mean model estimates for each of these 4 days of the year over time
Adj. GAM	For gam4 and gam5, adjusted model results to show the long-term predictions as if flw_sal had been average over the record
Lo Sal	Periods with salinity values in the lowest 25th percentile
Hi Sal	Periods with salinity values in the highest 25th percentile
Fit.GAM	For gams 4 and 5, the magenta line shows the full model prediction with flw-sal incorporated, before adjustment.

When we began evaluating the first three temporal models, we thought it would be necessary to select the most representative of these for an individual data set by examining the Akaike information criterion (AIC) or other model performance statistics. However, in an analysis of total nitrogen (TN), TP, Chlorophyll-*a*, and dissolved oxygen (DO) surface and bottom data, and Secchi disk depth, for all mainstem and tidal stations from 1999-2014, the AIC results for gam2 were lowest (i.e., suggesting the best model) for 98% of the data sets. These results were consistent when root mean squared error (RMSE) and R^2 were analyzed also. Table 3 presents a summary of the differences between the AIC for gam2 and the next best model (either gam0 or 1). A negative difference means that the gam2 AIC was lower, therefore suggesting a "better" model. A positive difference means that either gam0 or gam1 had a lower AIC, suggesting gam2 is not the "best" model. It is clear that the gam2 models frequently have much lower AICs than the other models, and when gam0 or gam1 performs better, it is a very small difference in performance.

Table 3. Statistics on AIC ranges ($AIC_{gam2} – AIC_{gam0or1}$)

			Quantiles of differences			
parameter	**layer**	**min diff**	**25**	**50**	**75**	**max diff**
Chl-a	B	-29.94	-14.235	-4.33	-0.525	0.8
Chla	S	-31.67	-7.335	-3.12	-0.87	0.35
DO	B	-23.86	-4.39	-1.54	-0.035	0.36
DO	S	-26.97	-4.715	-1.8	-0.14	0.66
Secchi	S	-30.56	-6.56	-2.065	-0.22	0.04
TN	B	-30.93	-8.82	-4.6	-1.46	0.01
TN	S	-72.71	-10.28	-4.54	-0.895	0
TP	B	-16.22	-5.945	-1.66	-0.125	0.06
TP	S	-25.44	-7.81	-2.89	-0.635	0

These findings led to the use of the gam2 model structure as the basis for the more complex models (gam3-5) and in large-scale analysis as the temporal model. This is the approach we will use when generating trends to make bay-wide maps, for instance. In cases where a simpler GAM would have performed better, the p-values on the different smooth terms from the gam2 model demonstrate this. For example, in Appendix 2.4A, the results for CB4.3E Secchi for gam2 show a p-value of 0.77 on the ti() interaction term, telling the analyst that there is no significant interaction between season and date for this parameter. The percent change and p-value on the change for gam1 and gam2 are nearly identical in this case, suggesting no change in overall conclusion if the more complicated model is selected when it actually is not needed. If a targeted study is performed on just a few stations, it may be valuable to consider running the simpler models since they would be easier to explain. The functionality for all three models will continue to be retained in the baytrends package.

2.5 Percent change computation

One of the basic types of information we need for tidal trends analysis is whether the observations at a station are increasing or decreasing, and if we have confidence that that the change is different from zero. These results are needed to generate overall trend maps that serve as a first-view of the trend results in presentation, reports, and summaries (e.g., Figure 2). Percent change over time is not explicitly available from the summary of any of the GAM models. A low p-value on the cyear or s(cyear) terms is not the same as testing the hypothesis that the change over time is significantly different from zero. So an additional test on the GAM output was added to provide this information.

The percent change approach is to average the GAM-generated estimates over the first two years, or baseline period, and last two years, or current period. Estimates of percent change are computed based on the difference of these estimates relative to the baseline. The estimate of percent change is a linear function of the GAM's parameter vector, which is in turn a linear function of the data. Thus its variance is readily available through a quadratic form defined by the covariance matrix of the GAM-estimates. This computation is documented in Appendix 2.5A, as well as in Wood (2006).

With the estimate and standard error, we can obtain tests of significance and confidence intervals. An example is provide in Appendix 2.5A, and the output generated from the baytrends package is shown in Table 4. When the GAM model is fit to the log of the data values, this computation is conducted on the model predictions without back-transforming them, thereby providing an estimate of the percent change of the geometric mean. However, the tabular results also present the estimates of the baseline and current mean back-transformed. This approach assumes that the spline fit used in the model is correct, or at least a reasonable approximation.

When gam3 or 5 is used and an intervention is present, we want to compute a percent change the removes the effect of that intervention. The approach is therefore to adjust the model predictions before the intervention based on the step computed on the intervention term. The percent change is then computed based on these adjusted values, as well as the original values, and both are reported.

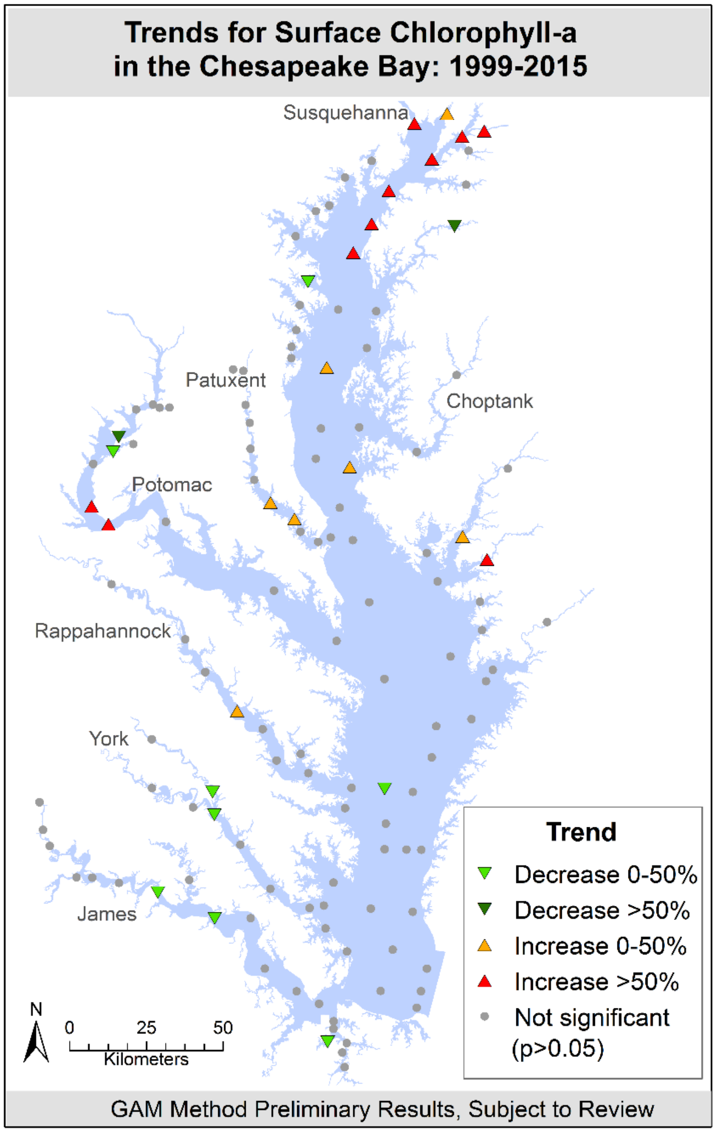


Figure 2. Draft map of Percent Change in surface chlorophyll-a concentrations from 1999-2015.

Table 4. Draft Estimates of Change from 1999-2014 for TF5.5A, Surface TP

Calculation	Estimate
Baseline log mean (geometric mean)	-2.2885 (0.1014)
Current log mean (geometric mean)	-2.7503 (0.0639)
Estimated log difference	-0.4618
Std. Err. log difference	0.0758
95% Confidence interval for log difference	(-0.6104 , -0.3132)
Difference p-value	<0.0001
Period of Record Percent Change Estimate (%)	-36.99%

2.6 Approach for censored data

Analytical method detection limits for the tidal water quality parameters have not be static over the more than 30 year monitoring program. Because methods, laboratories, and equipment types have varied both spatially and temporally throughout the monitoring program, there are frequently different method detection limits for the same parameter at different stations over time (see Table A3-3a in the CBP Water Quality Monitoring Data guide, Olson 2012). When an observation is below a detection limit, there is still is a value for the observation, it is just too small for a single observation to be differentiated from zero with high probability. The current process, since the late 1990s, has been for the state agencies to retain these "below the detection limit" (BMDL) values in a separate column in the database, while still placing a qualifier on the value. Before the late 90s, the only information retained in the database for a value below the detection limit is the detection limit itself and a qualifier. This data is referred to as "censored." Therefore, it is possible to do trend analysis on the data using the BMDL data for the later part of the record, but in the first half of the record, varying censoring levels need to be considered when computing trends. Many of the parameters in our data set are "left-censored" meaning that there is a minimum value below which the results are censored. However, some parameters, such as the total nutrients, are computed as sum of multiple measured constituents. These summed parameters have the potential to be "interval censored" if one or more of the constituents used to compute the total are below the detection limit. In very few cases, Secchi disk depth is "right-censored" meaning there is a maximum value above which the values are censored. These three types of censoring need slightly different consideration. More details of censoring in general for CBP data is discussed in Section 1 of Appendix 2.6A.

Approach for Data Sets with Censored Data

As described in Appendix 2.6A, the approach for Seasonal Kendall trend analysis in Chesapeake Bay was to re-censor all the records to the maximum detection limit used throughout an individual data set's record. This re-censoring eliminates spurious trends due to improving detection limits, but may lose information about some detected observations as data are re-censored. Other more rigorous approaches exist for dealing with data with multiple censoring levels, and it was reasonable to consider and test options as we propose switching to conducting trend analyses with GAMs. The Expectation Maximization (EM) approach was selected for baytrends, and the rest of this Section 2.6 describes a comparison with a Monte Carlo approach and scenario tests to understand the strengths and weaknesses of the EM method. Figure 3 shows an example of output from baytrends using a dataset with censored values. Circle symbols represent censored data, while x's represent uncensored observations. In this dataset, the level of censoring was too high (>50%) before 1991 to estimate the GAM model. However, there still was censoring after 1991, but enough other observations to start the GAM fit in that year.

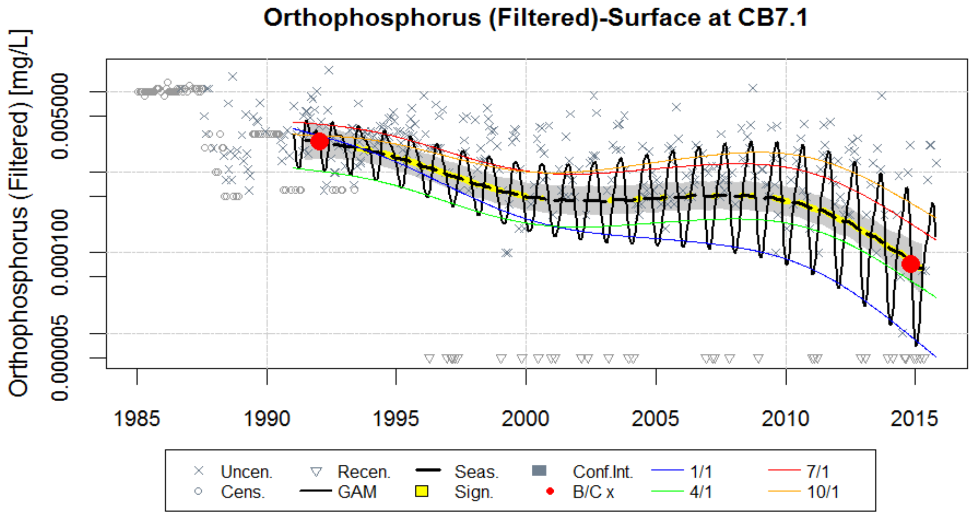


Figure 3. Example dataset with censored observations.

Options Considered for Censored Data

In understanding the impact of censored data and evaluating approaches, the two methods compared were:

1. Monte Carlo approach (MC): Replace censored data with a random value selected from the assumed data distribution below the detection limit (or within the censored interval for interval censoring). Repeat many times to capture the range of possibilities.
2. Expectation Maximization (EM): Produces maximum likelihood estimates (MLEs) by repeatedly replacing censored data with the conditional expectation of the censoring interval computed under the current model.

The Monte Carlo (MC) method was implemented by considering each censored data point in a time series, imputing a value for it from a censored lognormal distribution for it, and doing this for each of the censored data points in the time series to create one simulation set. This process was repeated 500 times to generate 500 possible data sets with imputed values. Distributional parameters of the censored data were estimated using fitdistcens() function from the 'fitdistrplus' R package (Delignette-Muller and Dutang 2015). A GAM was fit to each of those 500 data sets, the resulting statistics were summarized, and all prediction lines were plotted together to show the range of results. Some tests were conducted for one station/parameter to determine an appropriate number of simulations, and results showed no statistical difference between results sets generated with 1000 or 100 simulations. To be conservative, 500 was selected.

The Expectation Maximization (EM) approach uses an initial estimate of the censored values to start an iterative process that converges to a solution. For left-censored data, one-half the detection limit was used as a starting condition, and the mid-value of the censoring interval was used for interval-censored data. Like the MC approach, the EM approach also imputes values for the censored observations; however, the EM approach fits a GAM model based on the initial imputation and uses this model to generate an expected value for each censored data point and variance to inform the next imputation.

Equations for the conditional expectation and variance of censored data are then used to generate the next set of expected values (equations in Section 2 of Appendix 2.6A, based on logic in Liu et al. 1997). The process continues until the coefficients of the GAM model converge. The EM algorithm does not seem to be sensitive to the start values chosen.

After comparing these approaches, EM approach was selected and implemented into baytrends for dealing with censored data in GAMs. The EM approach has appeal because generating MLE's for censored data would be defensible and take much less computation time than the MC approach. We also have concern about the interpretability of hundreds of MC simulation results. However, the EM approach in an application like this is relatively new, and we will still be considering modifications as research evolves. Research (Green 1990; Segal et al. 1994; Wang et al. 2012; Silverman et al. 1990) on the use of the EM algorithm in constrained maximum likelihood estimation make it clear that the approach is acceptable in terms of getting the estimates, but there remains the problem of how to test hypotheses about these estimates. Right now we are essentially assuming that the expected values for censored data in the last iteration of the EM algorithm are actually observed values. Then we just let 'mgcv' give us the usual F-values and P-values. However, these are based on underestimates of variance because they assume specific values for the censored data when in fact there is additional uncertainty due to censored data being known only within an interval. For the purposes of identifying long-term patterns and trends in situations where censoring is modest, we think that this approximation is okay. In cases where censoring is extreme (>50% of the data), we already are truncating our data sets to avoid analyzing that highly censored part of the record.

Censoring Approach Comparison

Using uncensored data from the latter part of the CBP record, we created a series of censored-data test cases based on the typical patterns of censoring in the early CBP data record and used it to test both EM and MC approaches on censored data in GAMs. First, the extent of censoring in the database was summarized (Appendix 2.6B), with results clearly showing extensive censoring of nutrients pre-1995. This summary aided us in identifying typical censoring scenarios that exist in the CBP data record for testing the MC and EM approaches. Appendix 2.6C documents the selection of these scenarios and provides a description of the patterns of censoring observed in the data records. Three scenarios were selected, each with 3-5 cases of individual stations/parameters. For each of those cases, the data from 1999-2014 was artificially censored to be able to test the performance of the censoring methods against the true data result and the other method. The three censoring scenarios captured conditions where the data are censored (1) modestly at a 10-25% level in the first half of the record, (2) highly at a 25-75% level in first third or half of record, and (3) step improvements in the detection limit throughout the record. Note that the most extreme scenarios (>50% censoring for half the record) are more extreme than reality because censoring only occurs in the first 12-14 years of our currently 31 year record.

Each of the censored scenarios was run with the different methods used to impute the censored values. The basic statistical output was compared, and the graphs were examined (Section 2 of Appendix 2.6C), to determine under which treatments of the censored data we would draw a different conclusion than the conclusion that would be generated from the uncensored data set. Table 1 of Appendix 2.6C summarizes the take-home messages from the censoring scenarios. For censoring 10-25% of the first half of the data record, or for interval censoring at any level, the EM and MC methods perform very well and result in GAM models and conclusions about percent change over time that are consistent with the results generated from GAMs fit to the uncensored data. This is promising since the majority of the data censoring falls in this category. When left censoring is 50% or greater, the results vary. Scenarios 2g (50% left censoring, 25% interval censoring) and 3c (high level of step-censoring changes) are good examples of how the EM approach works well under a fairly high level of censoring. Scenarios 2j and 3d demonstrate that there are limits to that performance when the data is simply censored too much during a period of the record (>75% for the first half in 2j), or too much at the beginning of the record (100% for 2 years in 3d).

These scenarios capture some very extreme situations because we wanted to test the limits of the methods. For the amount of censoring found in the majority of data sets used for trend analysis, the EM approach appears to perform well. Additionally, these results suggest that some portions of datasets should be excluded when censoring is >50% for extended periods of time, and that has been implemented in baytrends. This decision is consistent with an upper limit of 50% censoring for the WRTDS survival regression routine used for nontidal trends by USGS (Chanat et al. 2016).

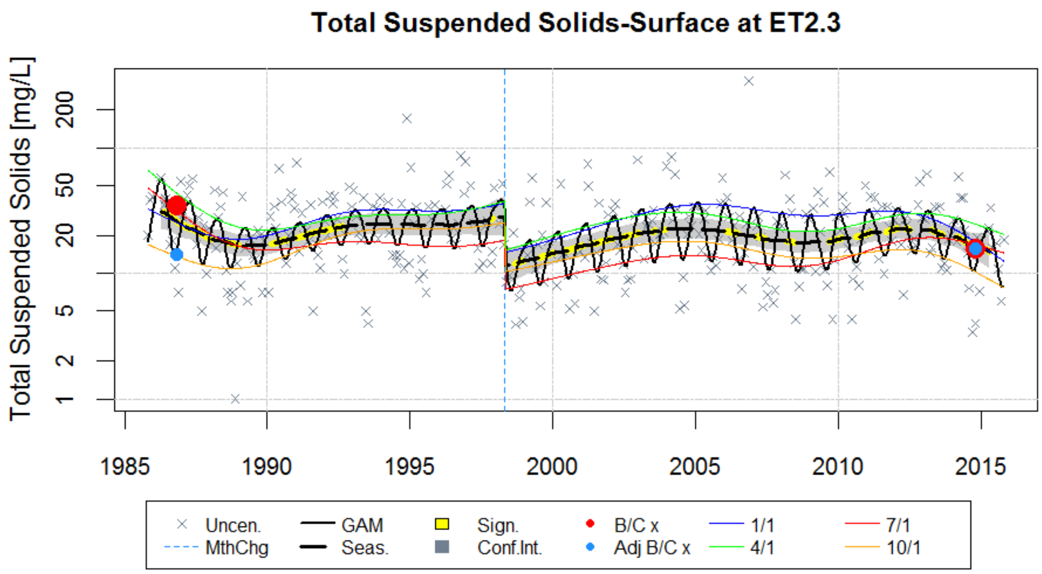
2.7 Approach for laboratory and method changes

In addition to changes in detection limits for water quality observations over time, there have also been changes in analytical laboratories and methods for some of the parameters. These types of changes are documented in multiple places: major method changes are described in the sections for each parameter in the CBP Water Quality Monitoring Data Guide (Olson 2012), dates for each laboratory are listed in Table A3-3a of the guide (Olson 2012), and the state agencies track their changes specifically (e.g., MDDNR 2016). Some detailed statistical analyses on changes in the data record are documented in the CBP's Data Analysis Issues Tracking System (CBP 2010).

In many cases, when it was suspected that a change impacted the values of observations, a paired sampling study was conducted to compare the two methods side-by-side (CBP 2010) and often an adjustment factor was computed that could be used to account for the change. Some of the major identified method/lab changes have been for TP, TN, and total suspended solids (TSS). Specifically, the method for TN computations in the VA tributaries changed in 1994 from Total Kjeldahl Nitrogen plus nitrate+nitrite (TKN+NO23) to total dissolved nitrogen plus particulate nitrogen (TDN+PN). A detailed analysis of an apparent step decrease in values associated with this change was conducted and the conclusion was that for trend analysis, a blocked Seasonal Kendall approach was needed to split the time periods (CBP 2010, see DAITS041). Similar step changes were identified in the VA tributaries for TP when the method changed in 1994 from a direct measure to the sum of particulate phosphorus plus total dissolved phosphorus (PP+TDP) (CBP 2010, see DAITS042). TSS records were also found to have major step changes due to a laboratory change for some of the VA stations in 1996 (CBP 2010, see DAITS045) and a change in the amount of filter rinsing of samples in MD tributaries.

These and other changes not described here have resulted in challenges for long-term trend analysis. With Seasonal Kendall, the work-arounds involved using blocked Seasonal Kendall tests (VA tributaries for TN, TP, DIN, PO_4^{3-} and mainstem for TSS) or shortening the length of the trend analysis (MD tributaries for TP, DIN, PO_4^{3-}, TSS). Neither of these approaches are completely satisfactory, and adjustment factors that would allow for a long-term analysis have been found to vary by station and environmental parameters (see Appendix 2.7A). Using baytrends, we plan to fit GAMs to generate both long term (1985 to present) and short term (last 10 year) trend results in an operational mode every year. So an approach to account for method and lab changes is needed.

The selected approach for accounting for these changes is intervention analysis (e.g., Box and Tiao 1975). Appendix 2.7A provides details of the method and some testing results. The basic approach involves adding the “intervention” term to the model (gam3 in Table 1). If there is one known lab/method change, this variable is set to 0 before the change and 1 after the change. The results from the gam3 model fit will indicate if the intervention term is significant (with a low p-value) and provide a value for the estimated change due to the lab. Figure 4 shows this for surface TSS at ET2.3.



Step-trend analysis for ET2.3

Type	Source	edf	F-stat	p-value
parametric terms	intervention	1.00	8.6706	0.0035
" "	cyear	1.00	13.4087	0.0003
smoothed terms	s(cyear)	6.96	3.2053	0.0005
" "	s(doy)	2.95	8.4719	<0.0001
" "	ti(cyear,doy)	8.36	2.0938	0.0010

Step-trend coefficients for ET2.3

parameter	estimate	Std. Err.	t value	p-value
(Intercept)	3.539225	0.184741	19.1578	<0.0001
interventionB	-0.900829	0.305927	-2.9446	0.0035
cyear	-0.287998	0.078649	-3.6618	0.0003

Figure 4 Example application of intervention analysis to surface TSS at ET2.3.

One major benefit of the intervention approach is that the importance and the magnitude of the laboratory or method change impact is determined uniquely for each data set. In addition to the structure shown in Table 1 with gam3, we have considered allowing the magnitude of a method or lab change to vary by season. Preliminary results suggest that fitting a model structure as follows allows for this (see results in Appendix 2.7A). Including "by=LabChange" in the ti() statement allows for a unique tensor product smooth for each value of the LabChange parameter.

wq = LabChange + cyear + s(cyear) + s(doy,bs='cc')+ ti(cyear,doy,by=LabChange,bs=c('tp','cc'))

Research is ongoing as to whether to implement this more complicated version of the intervention in baytrends, but the current gam3 structure is available as of spring 2018 and is being used to identify significant interventions. Specifically, state partners are testing the approach with additional parameters and stations, testing the approach for data records with more than one change, and making sure we can clearly interpret the results. At the same time, we are working with MDDNR and VADEQ to be sure all method and lab changes are tracked within the data record in order to be able to identify them automatically when conducting a GAM analysis.

2.8 Flow/salinity-adjustment procedure

Water quality in the tidal waters is frequently correlated to the amount of freshwater flowing in the bay. There are very high fluctuations in freshwater river flow measured at the fall-line river input monitoring stations, which can influence many factors that impact water quality including the amount of nutrients flowing into the bay, the degree of vertical density stratification, and the overall circulation. Flow has not been incorporated into the most recently used Seasonal-Kendall-based tidal trends assessments; however, flow effects have frequently been considered in scientific studies on water quality patterns and trends in Chesapeake Bay (e.g., Harding et al. 2016; Hagy et al. 2004). These studies and other like them have been valuable in their approaches to identify what the response to nutrient loads is after removing the very large, but uncontrollable, effect of river flow.

Similarly, the nontidal water quality trends are currently being reported as "flow-normalized" loads or concentrations (Moyer 2016), with a flow-normalization procedure built into the WRTDS computations. As described by Hirsch et al. (2010), evaluating concentrations or fluxes after removing the variation due to streamflow alone can be helpful in answering questions about how the watershed is responding to management changes, i.e., "for a given set of hydrologic conditions, are water-quality conditions getting better or worse over time, and how much better or worse?" Likewise, in the estuary if we wish to identify whether a long-term water quality record is responding to a long-term management effort in the watershed, it will be helpful to remove the year-to-year variability due to river flow. However, as also described by Hirsch et al. (2010), there are good reasons to evaluate water quality time series without removing the year-to-year variability due to flow. Ecological conditions in the estuary will be responding to the true load of nutrients, which tends to be much higher in high flow years than low flow years. To understand the year-to-year variations in phytoplankton concentrations, hypoxia conditions, and SAV area, we cannot forget the important role that extremes in river flow play.

Therefore, a flow-adjustment procedure for the GAM implementation in Chesapeake Bay tidal waters is an option in the baytrends R package. We expect to routinely run both flow-adjusted and non-flow-adjusted trends on all water quality data sets which will be useful for comparison as well as to evaluate different trend-related questions.

The gam4 and gam5 models in Table 1 show the structure of the models with flow-effects incorporated. The variable *flw_sal* is flexible because it is not usually straight-forward what values should be used for the river flow effect. The nine USGS river gages at the fall-lines of the major tributaries provide daily flow values. The 'baytrends' package allows for the user to match each tidal station to an individual USGS gage. It also allows the user choices for how to pre-process the river flow. A range of averaging periods is available so that the daily river flows will be averaged for the preceding *n* days before each observation. The package has the functionality to compute a correlation coefficient to help determine which averaging period may be most explanatory for the water quality data. Another option is for the user to do testing and set *a priori* what averaging period to use for river flow matched to an individual tidal station. This pre-processed river flow is then used as the *flw_sal* variable after it is log-transformed and the seasonal cycle is removed via a simple GAM.

Another option for the *flw_sal* variable is to use salinity measured at the same place and time as the water quality observations. This approach has precedent (Beck and Hagy 2015; Beck and Murphy 2017), and may be advantageous to use in locations where it is unclear which USGS gage to match to a monitoring station. When salinity is used, its seasonal cycle is also removed via a simple GAM as a pre-processing step.

Once gam4 or 5 is fit with *flw_sal*, the model results can then be "adjusted" to examine what the long-term pattern would look like if fluctuations in freshwater inputs had not occurred. This is done by setting *flw_sal* variable to the average for that day of the year. Because the seasonal cycle is removed, this value is zero. However, the standard deviation of this value does vary through the year, so these standard deviations are computed and retained for use in the flow adjustment predictions (see Appendix 2.8A).

4. Future plans

The main focus of this documentation is on the implementation of the GAM approach for identifying temporal patterns and trends over time throughout the tidal waters. Trend identification is just the first step in the explaining trends efforts for tidal waters. We are simultaneously experimenting and applying GAMs to multiple research projects in a hypothesis-testing framework to help explain what factors may be influencing the trends we have observed. One general approach being applied follows directly from the GAM development explained in this document:

Step 1. Fit temporal GAM model, possibly with and without flow adjustment
wq = s(date) + s(doy) + s(flow) + interactions

Step 2. Work with research teams including ecologists, physical scientists, and climate scientists to examine and identify temporal patterns of change that may be explained by anthropogenic factors (management actions, land use change, etc), climate factors (changing climate, cyclic climate patterns, etc), habitat and ecosystem factors, and other forces.

Step 3. Use GAMs for hypothesis testing of those process-based hypotheses, essentially replacing the "date" and sometime "doy" terms by explanatory factors to test whether the factor can explain the temporal change.

wq = s(factor) + s(doy) + s(flow) + interactions

A key component to this effort is to work side-by-side with researchers who have process-based knowledge about the estuarine ecosystem.

5. References

Allaire, J.J., J. Cheng, Y. Xie, J. McPherson, W. Chang, J. Allen, H. Wickham, A. Atkins and R. Hyndman. 2015. rmarkdown: Dynamic Documents for R. R package version 0.8.1. http://CRAN.R-project.org/package=rmarkdown

Beck, M.W. and J.D. Hagy. 2015. Adaptation of a Weighted Regression Approach to Evaluate Water Quality Trends in an Estuary. *Environ Model Assess* 20:637. doi:10.1007/s10666-015-9452-8.

Beck, M.W., Murphy, R.R., 2017. Numerical and Qualitative Contrasts of Two Statistical Models for Water Quality Change in Tidal Waters. JAWRA J. Am. Water Resour. Assoc. 53, 197–219. https://doi.org/10.1111/1752-1688.12489

Box, G.E.P. and G.C Tiao 1975. Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association* 70(349):70-79.

Bradu, D. and Y. Mundlak. 1970. Estimation in Log-Normal Linear Models. *Journal of the American Statistical Association* 65:198-211.

Buja, A., Hastie, T., Tibshirani, R., 1989. Linear Smoothers and Additive Models. Ann. Stat. 17, 453–510. https://doi.org/10.1214/aos/1176347115

Chanat, J.G., D.L. Moyer, J.D. Blomquist, K.E. Hyer, and M.J. Langland. 2016. Application of a weighted regression model for reporting nutrient and sediment concentrations, fluxes, and trends in concentration and flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, results through water year 2012: U.S. Geological Survey Scientific Investigations Report 2015–5133, 76 p. http://dx.doi.org/10.3133/sir20155133.

Chesapeake Bay Program (CBP). 2010. Data Analysis Issues Tracking System (DAITS). Online at: http://www.chesapeakebay.net/publications/title/data_analysis_issues_tracking_system_daits

Dauer, D.M. and R.W. Alden. 1995. Long-Term Trends in the Macrobenthos and Water Quality of the Lower Chesapeake Bay (1985-1991). *Marine Pollution Bulletin* 30(12): 840-850.

Delignette-Muller, M.L., and C. Dutang. 2015. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* 64(4):1-34.

Duan, N. 1983. Smearing estimate: a nonparametric retransformation method. *J. Am. Stat. Assoc*. 78, 605–610.

Ebersole, E., M. Lane, M. Olson, E. Perry and B. Romano. 2002. Assumptions and Procedures for Calculating Water Quality Status and Trends In Tidal Waters of the Chesapeake Bay and its Tributaries: A cumulative history. Available online at http://mddnr.chesapeakebay.net/eyesonthebay/documents/stat_trend_hist.pdf.

Ellis, H., P. Du, C. Friedrichs, V. Lyubchich. (2017). Scientific and Technical Advisory Committee Review of the Generalized Additive Model (GAM) Approach for Water Quality Trends in Tidal Waters. STAC. http://www.chesapeake.org/stac/stac_rev_pubs.php.

Hagy, J.D., W.R. Boynton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in the Chesapeake Bay, 1950–2001: Long-term changes in relation to nutrient loading and river flows. *Estuaries* 27: 634–658. doi:10.1007/BF02907650.

Haraguchi, L., J. Carstensen, P. C. Abreu, and C. Odebrecht. 2015. Long-term changes of the phytoplankton community and biomass in the subtropical shallow Patos Lagoon Estuary, Brazil. *Estuarine, Coastal and Shelf Science* 162(SI):76–87.

Harding, L. W., C. L. Gallegos, E. S. Perry, W. D. Miller, J. E. Adolf, M. E. Mallonee, and H. W. Paerl. 2016. Long-term trends of nutrients and phytoplankton in Chesapeake Bay. *Estuaries and Coasts* 39:664–681. doi:10.1007/s12237-015-0023-7.

Hastie, T. and R. Tibshirani. 1986. Generalized additive models (with discussion). *Statistical Science* 1: 297–318.

Hastie, T. and R. Tibshirani. 1990. *Generalized Additive Models*. Chapman & Hall/CRC.

Hirsch, R.M., J.R. Slack, and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18:107-121.

Hirsch, R.M., D.L. Moyer, and S.A. Archfield. 2010. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs. *Journal of the American Water Resources Association* 46(5):857–880.

Green, P.J. 1990. On Use of the EM for Penalized Likelihood Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(3): 443-452.

Keisman, J., J. Blomquist, S. Phillips, G. Shenk, E. Yagow. 2014. Estimating land management effects on water quality status and trends. Proceedings of the March 25-26, 2014 workshop. STAC Publication Number 15-002, Edgewater, MD. 33 pp.

Liu, S., J. Lu, D.W. Kolpin, and W.Q. Meeker. 1997. Analysis of Environmental Data with Censored Observations. *Environmental Science and Technology* 31:3358-3362. http://digitalcommons.unl.edu/usgsstaffpub/71

Marra, G and S.N. Wood. 2012. Coverage Properties of Confidence Intervals for Generalized Additive Model Components. *Scandinavian Journal of Statistics* 39(1):53-74.

Maryland Department of Natural Resources (MDNR). 2016. Quality Assurance Project Plan for the Maryland Department of Natural Resources Chesapeake Bay Water Quality Monitoring Program Chemical and Physical Properties Component for the Period July 1, 2016 - June 30, 2017. Prepared by Ben Cole and Thomas Parham.

Marshall, H.G., M.F. Lane, K.K. Nesius, and L. Burchardt. 2009. Assessment and significance of phytoplankton species composition within Chesapeake Bay and Virginia tributaries through a long-term monitoring program. *Environ Monit Assess* 150:143–155.

Moyer, D. L., R.M. Hirsch, and K.E. Hyer. 2012. Comparison of two regression-based approaches for determining nutrient and sediment fluxes and trends in the Chesapeake Bay Watershed, Technical Report Scientific Investigations Report 2012-544, US Geological Survey, US Department of the Interior, Reston, Virginia.

Moyer, D.L. 2016. Nitrogen, Phosphorus, and Suspended-Sediment Loads and Trends measured in Nine Chesapeake Bay Tributaries: Water Years 1985-2015: U.S. Geological Survey data release, http://dx.doi.org/10.5066/F7Q81B5N.

Murphy, R.R., W.M. Kemp, and W.P. Ball. 2011. Long-Term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading. *Estuaries and Coasts* 34:1293-1309.

Olson, M. 2012. Guide to Using Chesapeake Bay Program Water Quality Monitoring Data. (EPA Publication No. 903-R-12-001). Annapolis, MD: United States Environmental Protection Agency, Chesapeake Bay Program. Online at:
http://www.chesapeakebay.net/documents/3676/wq_data_userguide_10feb12_mod.pdf

Peng, R.D., Dominici, F., Louis, T.A., 2006. Model choice in time series studies of air pollution and mortality. J. R. Stat. Soc. Ser. A Stat. Soc. 169, 179–203.

Segal. M.R., P. Bacchetti, and N.P. Jewell. 1994. Variances for Maximum Penalized Likelihood Estimates Obtained via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 56(2): 345-352.

Silverman, B.W., M.C. Jones, J.D. Wilson, and D.W. Nychka. 1990. A Smoothed EM Approach to Indirect Estimation Problems, with Particular, Reference to Stereology and Emission Tomography. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(2): 271-324.

Testa, J.M., W.M. Kemp, W.R. Boynton, and J.D. Hagy. 2008. *Estuaries and Coasts* 31:1021-1037. doi:10.1007/s12237-008-9095-y.

Wang, L., P. Du, and H. Liang. 2012. Two-Component Mixture Cure Rate Model with Spline Estimated Nonparametric Components. *Biometrics* 68: 726-735.

Wood, S.N. 2003. Thin-plate regression splines. *Journal of the Royal Statistical Society (B)* 65(1):95-114.

Wood, S.N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Amer. Statist. Ass.* 99:673-686.

Wood, S.N. 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Wood, S.N. 2013. On p-values for smooth components of an extended generalized additive model. *Biometrika* 100:221-228.

Wood, S., 2018. mgcv. https://cran.r-project.org/web/packages/mgcv/index.html