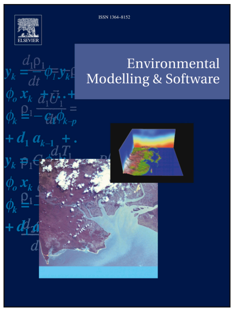
Accepted Manuscript



A Generalized additive model approach to evaluating water quality: Chesapeake Bay case study

Rebecca R. Murphy, Elgin Perry, Jon Harcum, Jennifer Keisman

PII: S1364-8152(18)30780-1

DOI: https://doi.org/10.1016/j.envsoft.2019.03.027

Reference: ENSO 4431

To appear in: *Environmental Modelling and Software*

Received Date: 6 August 2018

Revised Date: 15 March 2019

Accepted Date: 28 March 2019

Please cite this article as: Murphy, R.R., Perry, E., Harcum, J., Keisman, J., A Generalized additive model approach to evaluating water quality: Chesapeake Bay case study, *Environmental Modelling and Software* (2019), doi: https://doi.org/10.1016/j.envsoft.2019.03.027.

A Generalized Additive Model approach to evaluating water quality: Chesapeake Bay Case Study

Rebecca R. Murphy[a], Elgin Perry[b], Jon Harcum[c], and Jennifer Keisman[d]

[a] Corresponding author; University of Maryland Center for Environmental Science, Chesapeake Bay Program Office, 410 Severn Ave., Suite 112, Annapolis, MD 21403; rmurphy@chesapeakebay.net

[b]Statistics Consultant; 377 Resolutions Rd., Colonial Beach, VA 22443; eperry@chesapeake.net

[c]Tetra Tech, Inc., 3475 East Foothill Boulevard, Pasadena, California 91107; jon.harcum@tetratech.com

[d]United States Geological Survey, 5522 Research Park Drive, Baltimore, MD 21228; jkeisman@usgs.gov

Abstract

Nutrient-reduction efforts have been undertaken in recent decades to mitigate the impacts of eutrophication in coastal and estuarine systems worldwide. To track progress in response to one of these efforts we use Generalized Additive Models (GAMs) to evaluate a diverse suite of water quality constituents over a 32-year period in the Chesapeake Bay, an estuary on the east coast of the United States. Model development included selecting a GAM structure to describe nonlinear seasonally-varying changes over time, incorporating hydrologic variability via either river flow or salinity, and using interventions to model method or laboratory changes suspected to impact data. This approach, transferable to other systems, allows for evaluation of water quality data in a statistically rigorous way, while being suitable for application to many sites and variables. This enables consistent generation of annual updates, while providing a tool for developing insights to a range of management- and research-focused questions.

Keywords: Generalized Additive Models; water quality; Chesapeake Bay; trend analysis

1. Software and data availability

The methods described here are incorporated into an R package, 'baytrends' which is available on CRAN (https://CRAN.R-project.org/package=baytrends). All water quality data used are incorporated into the example data set provided with 'baytrends' and are also available from the Chesapeake Bay Program's data hub (Chesapeake Bay Program, 2017a). River flow data are available from the U.S. Geological Survey (https://waterdata.usgs.gov/nwis).

2. Introduction

Major efforts have been underway in coastal and estuarine systems to reduce nutrient inputs in order to minimize eutrophication and improve water quality, aquatic habitats, and hence living resource populations. Examples include multiple Danish Action Plans for the Aquatic Environment which have led to more than 50% nutrient reductions into Danish coastal waters since 1990 (Riemann et al., 2016), dramatic reductions in nutrient loads to Tampa Bay, USA, after citizen demands and legislation in the 1970s to reduce waste water loads kick-started multiple nutrient reduction efforts (Greening et al., 2014), and several European nutrient reduction directives that have led to reduced nutrient loads into, among other places, coastal France lagoons (Le Fur et al., 2019). In addition, in the United States, efforts to improve water quality in the Chesapeake Bay, a large estuary in the mid-Atlantic, have been underway for more than three decades. In 2010, the U.S. EPA implemented a Total Maximum Daily Load (TMDL) for the entire Chesapeake Bay watershed to limit nutrients and sediment flowing into the tidal waters (USEPA, 2010). In systems such as these, where resources are being devoted to reducing nutrients from the watershed, it is critical to be able to explain the water quality response to many diverse stakeholders. At the same time, the interaction of physical, biogeochemical, and anthropogenic forces, further complicated by strong seasonal and interannual patterns, makes deciphering long-term trends in coastal and estuarine systems challenging (e.g., Cloern, 2001; Duarte et al., 2009). Some examples of these interactions include the correlation and relative impacts of freshwater flow and nutrient loads (e.g., Hagy et al., 2004; Paerl et al., 1998), the likely nonlinear impacts of climate change on biological processing and many other factors (Najjar et al., 2010), how sediment-nutrient dynamics can change with oxygen conditions (Testa and Kemp, 2012), and top down controls on phytoplankton populations (e.g., Jassby, 2008).

In the Chesapeake Bay, the states Maryland (MD) and Virginia (VA) have coordinated and conducted water quality monitoring at about 150 locations monthly or semi-monthly since 1984. Initially, a Seasonal Kendall-based approach (Hirsch et al., 1982) was used to evaluate these data for trends. This approach quantifies monotonic change over time and different trends by season and does not assume any specific distribution to the data. Many of the time series, however, follow non-monotonic patterns (e.g., Beck and Murphy, 2017; Harding et al., 2016) that the Seasonal Kendall test cannot identify. In addition, with a desire to see progress from the TMDL-driven actions, a method was needed that could separate the impacts of large interannual variations in freshwater flow from other drivers of variability such as nutrient load reductions. Finally, the flexibility to include other explanatory variables (e.g., climate, nutrients, biology) to help unravel the multiple forces at play could be important in future work.

Based on these current and future assessment needs, a Generalized Additive Model (GAM) approach was proposed. The mathematical formulation of a GAM is typically presented in a very general form (eq 1.)

$$g(E[Y]) = \mu + f_1(x_1) + f_2(x_2) + \ldots + f_m(x_m) \qquad \text{eq 1}$$

In this general format, a function *g()* is applied to the expected value of the dependent variable *Y* (frequently either the natural logarithm or identity). On the right-hand-side, μ is the modeled intercept followed by a sum of smooth functions of the independent variables $x_1,x_2 \ldots x_m$. The mathematical form of each smooth function is typically left unspecified and, in application, an interpolating function such as a spline function is used to approximate the unspecified function (Hastie and Tibshirani, 2004). The number of research applications using GAMs to evaluate water quality change is growing (e.g., Haraguchi et al., 2015; Riemann et al., 2016), and some recent work in our case study location has included evaluation of long-term chlorophyll-*a* in the Chesapeake Bay mainstem and Patuxent River (Beck and Murphy, 2017; Harding et al., 2016), changing seasonal cycles of nitrogen and chlorophyll-*a* in the Bay mainstem (Testa et al., 2018), and factors affecting submerged aquatic vegetation in the southern Bay waters (Lefcheck et al., 2017). These applications demonstrate the usefulness of this technique for evaluating long-term change, but are all relatively local and targeted to one or two parameters. This study focuses on a GAM implementation that can be used to evaluate progress annually at a large suite of stations, at multiple depths, for very different water quality constituents, in order to regularly update the public, policy-makers, and scientists engaged in tracking the water quality of an estuary or coastal system.

We apply this approach to our case study system, while noting findings and features that would need further evaluation in other locations with similar water quality concerns and long-term data records. We include research on model development, incorporation of the impact of freshwater flow variations, an intervention approach to account for laboratory or method changes, and an example demonstrating application of the approach to generate summaries of change over different time periods while adjusting for the effects of season, flow, and methods changes. Although we have built these features into an R package, this manuscript focuses on documenting our approach and is not a user guidance for the 'baytrends' package. Users' guidance can be found within the 'baytrends' package download.

2. Study Area and Data

The Chesapeake Bay and its estuarine tributaries extend through parts of Maryland, Virginia, and Washington D.C. The tidally-influenced waters (Figure 1) have 147 tidal monitoring stations that were used for the examples provided here. The long-term monitoring and laboratory analysis program has existed since 1984 among teams in Virginia (VA Department of Environmental Quality, Old Dominion University, Division of Consolidated Laboratory Services), Maryland (MD Department of Natural Resources and Chesapeake Biological Laboratory), and the Chesapeake Bay Program (CBP). The same field and laboratory methods are followed by the two states within a given year (Chesapeake Bay Program, 2017b). Each of the stations is visited by boat twice a month for most of the year, and monthly during winter. After analysis and quality assurance, all data values are submitted to the Chesapeake Environmental Data Repository and are available online through its DataHub (Chesapeake Bay Program, 2017a).

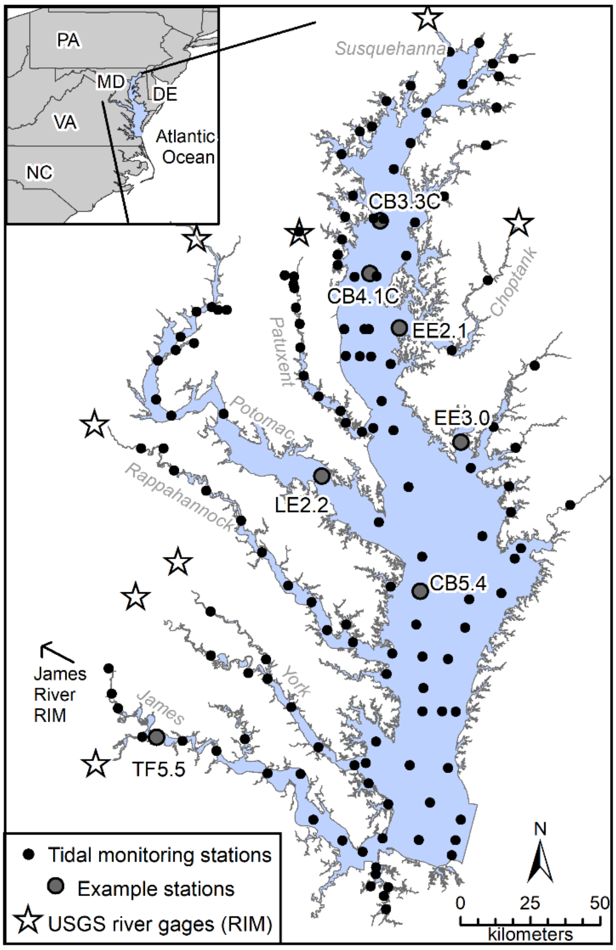


Figure 1. Chesapeake Bay location with 147 tidal monitoring stations used in GAM analyses indicated with circles. Larger labeled circles are station locations used for examples, and stars indicate locations of USGS River Input Monitoring (RIM) stations.

The annual trend assessments using the GAM method described here focus on a subset of the water quality data collected at these stations, and data for all examples shown here are available with the R 'baytrends' package. Parameters and depths typically analyzed are:

- Surface-mixed and bottom-mixed: total nitrogen (TN), dissolved inorganic nitrogen (DIN), total phosphorus (TP), dissolved inorganic phosphorus (orthophosphates, PO4), chlorophyll-*a*, total suspended solids (TSS), water temperature, and salinity.
- Secchi depth, and bottom dissolved oxygen (DO).

The surface-mixed layer is computed by averaging the samples identified as surface (top 0.5 or 1 m) and above-pycnocline. The bottom-mixed layer is the average of the just-above bottom sample and the below-pycnocline sample. If no pycnocline was detected, only the surface and bottom measurements are used. All nutrients, chlorophyll-*a*, and TSS concentrations are log-transformed before analysis. We made the transformation decisions based on experience with these data and typical data distributions observed for these parameters. It is reasonable to consider whether other transformations would be more appropriate, and future work may include building an optimal power transformation into the analysis (Box and Cox, 1964).

Several changes have occurred over the more than 30-year monitoring program, including laboratories used for analysis, how certain parameters are calculated, and method detection limits (i.e., MDLs). Changes like these can impact long-term trend analysis. Some examples of the changes we have had to consider are: a laboratory change in both MD and VA on different dates for different groups of stations; the computation of TN from a sum of nitrate+nitrite and total Kjeldahl nitrogen to a sum of particulate nitrogen plus total dissolved nitrogen; and TP changing from a direct measure to a sum of particulate and dissolved phosphorus.

3. Methods and Implementation

We selected the 'mgcv' package in the R statistical software (Wood, 2018) to fit GAMs. This is in part because of the usefulness of penalized thin plate regression splines (Wood, 2003) for our large-scale analyses, as described below. In addition we make use of the 'mgcv' functionality to fit smooths with more than one variable (Wood, 2006), and uncertainty estimates that are generated using a Bayesian approach (Giampiero and Wood, 2012; Wood, 2013).

3.1 Temporal models

A critical first step in most environmental data analyses is to examine the patterns over time. To do this with the Chesapeake Bay data, we tested a series of increasingly complex GAM model structures. All of these models are built from the basic GAM structure in eq 1, but they are presented below in the syntax used by 'mgcv' which allows us to be more specific about the construction of each approximating function. We specified the models as follows for a response variable *y* at an individual station location:

```
gam0: gam(y ~ cyear +  s(doy,bs='cc'), knots = list(doy = c(1,366)), select=TRUE)

gam1: gam(y ~ cyear + s(cyear, k=gamK1) + s(doy,bs='cc'), knots = list(doy = c(1,366)),
select=TRUE)

gam2: gam(y ~ 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

Note that random error and an intercept are included when these models are fit with 'mgcv.' The variable *cyear* is zero-centered date in decimal form, *doy* is the day of year (accounting for leap years by setting every year to have 366 days), *s()* is a spline function on a variable with *bs='tp'* specifying a penalized thin plate regression spline and *bs='cc'* specifying a cyclic penalized cubic regression spline. If the spline is not specified, the default '*tp*' is used. The *ti()* format specifies a tensor product of two smooths to account for the interacting effects of these two variables (Wood, 2006).

The *tp* spline is a computationally efficient way to approximate a smoother called a thin plate spline (Duchon, 1977), which avoids the need to manually place knots. An upper limit on the number of knots is required (*k*), or else a default of 10 is used. Testing was conducted on the *k* values and the use of *gamK1* for *s(cyear)* is described in Section 4.1 of the Results. The *cc* option is specified to represent the seasonal cycle as a function of *doy*. This spline has the usual property of being continuous to the second derivative at each knot location and, in addition, is constrained to take the same value and have continuous derivatives at the boundaries of the domain to insure a smooth transition across years (Wood, 2006). Unlike the thin plate regression spline, the number of knots is explicitly set for this spline. We use the default of 10 with the knots evenly spread throughout the parameter space, except requiring knots on Dec. 31 and Jan. 1 via the specification shown in the equations. In the above models, setting *select=TRUE* allows for terms to be removed from the GAM during model fitting if they provided no benefit (Wood, 2018). The *select* term is adjustable and depending on the analysis may be set to *FALSE*.

From gam0 to gam2, incrementally more complex temporal structures are fit to each data set. The first model allows for a linear trend that has a seasonal cycle. Adding *s(cyear)* into gam1 allows for a nonlinear pattern over time. The tensor product in gam2, *ti(cyear,doy)*, adds the ability for the seasonal cycle to change over time, or, in other words, for there to be different long-term trends in different seasons of the year. The results of these models can be compared graphically as well as with goodness of fit statistics including adjusted R^2, AIC, and a root mean squared error in prediction (RMSE).

3.2 Flow or salinity models

Accounting for the impact of inter-annual variations in freshwater flow on tidal water quality is key to distinguishing the variability associated with hydrology from changes associated with management actions to reduce nutrient pollution. This task is challenging because estuaries receive and retain the dissolved and particulate materials transported from rivers at varying time scales, meaning that there is not necessarily a single lag or averaging period that results in the best relationship between river flow and estuarine water quality. Moreover, as in many estuaries, there are multiple major rivers flowing into Chesapeake Bay as well as a long and complex shoreline with numerous ungaged smaller rivers and streams. The freshwater residence time from these inputs varies greatly, with modeling estimates of the "age" of the water exiting the Bay into the Ocean ranging from 140 to 310 days (Shen and Wang, 2007). Seasonal variability in the residence time is also a factor (Hagy et al., 2000).

Potential solutions to this problem include using salinity as a tracer of integrated freshwater influence (e.g., Beck and Hagy, 2015; Beck and Murphy, 2017), or providing flexibility in the river or rivers used to address hydrologic forcing. We developed our approach to allow for both of these options. River flow into the Chesapeake Bay is dominated by nine rivers (Figure 1), each of which is monitored as part of the USGS River Input Monitoring (RIM) program. For one option, we identified a set of possible river flow averaging periods for each salinity zone based on published hydrodynamic model evaluations (Shen and Haas, 2004; Shen and Lin, 2006; Shen and Wang, 2007). This allows for different averaging periods to be compared to the water quality data at each station using Spearman rank correlation coefficients to help select a good averaging period for a particular parameter-station-gage match. For the other option, we use *in-situ* salinity observations as a surrogate "flow" variable to account for variations in freshwater input. In saline regions of estuaries, salinity generally varies inversely with freshwater river flow. Using salinity instead of flow avoids the need to find a good spatial and temporal match between the RIM stations and a water quality station, because salinity is measured at the same place and time as the water quality samples.

To accommodate these multiple options, we incorporated a flexible variable that could represent either river flow or salinity (*flw_sal*) in the model structure. In gam2_flwsal, *flw_sal* can be either the salinity measured at the same place, time, and depth as the water quality variable, or the upstream river flow averaged over a period of *n* preceding days (where *n* is determined by picking the best period based on correlation coefficients or by setting *a priori* based on the needs of the study).

gam2_flwsal: gam(y ~ 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

gamK2=c(10,2/3) means that the maximum of 10 or (2/3*number of years) is selected

We preprocess the *flw_sal* variable to remove the seasonal cycle using a simple GAM (i.e., *y ~ s(doy,bs='cc')*). This process allows for the *flw_sal* term to account for any long-term impacts of river flow, rather than just the average seasonal impacts of river flow that can be accounted for with *s(doy).* This treatment of *flw_sal* and its relationship to the seasonal term in the model is something to consider in other applications, depending upon the importance of modeling the seasonal impacts of hydrology explicitly. If river flow is used, we transform the daily values with a log transformation before any other steps. The four *flw_sal* smooth terms allow for: the hydrologic effect to vary with the magnitude of the water quality variable (*s(flw_sal)*); the hydrologic effect to vary at different times of the year (*ti(flw_sal,doy)*); the hydrologic effect to change over time (*ti(flw_sal,cyear)*); and the seasonal impact of hydrology to change over time (*ti(flw_sal,doy,cyear)*). The type of smooth is specified as *tp* or *cc*, as described in Section 3.1. As with the previous gam0, gam1, and gam2 models, the k-value—or upper limit on the basis dimension—needs to be set. This is particularly important with the addition of flow or salinity into the models because of observed concurvity (Buja et al., 1989), which can be described as collinearity between the nonlinear spline functions (Peng et al., 2006) of *cyear* and *flw_sal*. Testing results that led us to our choices of *gamK1* and *gamK2* are presented in Section 4.4.

In summary, gam2_flwsal incorporates the impact of varying watershed hydrology into the modeling. For investigation of the causes of long-term change, it is useful to evaluate results that remove the impact of fluctuations in freshwater input. We conduct the "flow or salinity adjustment" by first fitting the model (gam2_flwsal) to the data, then generating predictions to estimate the mean of the water quality constituent over the normal range of flow. Because the relationship of the constituent to *flw_sal* can be nonlinear, it is not sufficient to simply set *flw_sal* to the average for each day of the year. Instead, we select five values of *flw_sal* representing the (5^{th}, 25^{th},50^{th}, 75^{th}, and 95^{th}) percentiles of the flow distribution for each day, compute the estimate of the constituent for each value, and compute a density-weighted average of these five estimates to get the "flow or salinity adjusted" estimate. These five percentiles are based on the mean of *flw_sal,* which is actually zero because the seasonal cycle has been removed from the *flw_sal* variable and the standard deviation for each *doy.* We found that the standard deviation is not usually constant throughout the year (e.g., Supplemental materials Fig. SM1). Therefore, we retain a data set that contains the average standard deviation of the seasonally-adjusted *flw_sal* variable for each day of the year based on a smooth of observed standard deviations.

3.3 Intervention approach

In either the temporal (Section 3.1) or flow or salinity adjusted models (Section 3.2), there is a need to consider the impact of changes in laboratory analysis or sampling procedure that could create a discontinuity (i.e. a disconnected shift) in the time series. We propose an intervention analysis (e.g., Box and Tiao, 1975) that involves adding a binary (0 or 1) term to gam2 to create gam3, or to gam2_flwsal to create gam3_flwsal:

gam3: gam(y ~ 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, and *intervention* is a binary variable taking the value 0 prior to the discontinuity and 1 after.

gam3_flwsal: gam(y ~ 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, *gamK2=c(10,2/3)* means that the maximum of 10 or (2/3*num years) is selected, and *intervention* is a binary variable taking the value 0 prior to the discontinuity and 1 after.

More than one change can be evaluated by changing *intervention* into a factor variable (e.g, 0, 1, 2..) as well. The p-value and estimated step size of the intervention is output from the gam3 and gam3_flwsal models to determine if the intervention has impacted the data values. This intervention model can also be used to evaluate management actions that are implemented at a point in time such as dam removal, implementation of advanced nutrient removal at a waste water treatment plant, or other changes in the data.

3.4 Percent change

By fitting the five GAM structures above, we can evaluate the shape of a water quality data time series, its relation to hydrology, and the impact of any possible interventions on trend estimates. Frequently we need to simplify these results and identify whether there has been an increase or decrease over time, and how confident we are in that finding. This type of information, however, is not an explicit output from the GAMs that have been presented, so an additional computation is needed.

To address this need, we compute a linear combination of the GAM coefficient vector that can be conceptualized as a percent change between the model estimates at the beginning and end of a period of interest. The period of interest could be the entire record or a subset of it. Once this time period is selected, we must decide how many years at the beginning and end of the period should be used in the computation. One simple option for percent change is to compute the percent difference between the predicted values for the first and last observations in the time series. However, experience has shown that the flexibility of GAMs allows model estimates to become unstable at the very edge of the independent variable space. To remedy this issue, we average predicted values over a short interval of time at each end of the time series. For the purpose of illustration here, we are presenting how we compute the percent change from the first two years to the last two years of the record, but we evaluate the choice of two years by testing different numbers of years at the beginning and end of the record in Section 4.4.

For a fitted GAM from 'mgcv', we can output the vector $\hat{\boldsymbol{\beta}}$ which is the estimated parameter vector, and $\hat{\sum}_{\beta}$ which is the estimated variance-covariance matrix of the parameter vector. These are computed in 'mgcv' following equations 2 and 3 (Wood, 2006):

$$\hat{\boldsymbol{\beta}} = \left(Z^{\mathrm{T}}Z\right)^{-1} Z^{\mathrm{T}}Y \quad \text{(eq 2)}$$

$$\hat{\sum}_{\beta} = s^{2}\left(Z^{\mathrm{T}}Z\right)^{-1} \quad \text{(eq 3)}$$

$$Z = S(X) \quad \text{(eq 4)}$$

where $\boldsymbol{X}$ is the matrix of independent variables (e.g., *cyear, doy, flow_sal*), $\boldsymbol{Z}$ is the matrix of linear predictors in the spline bases for the GAM model computed by the smooth function from $\boldsymbol{X}$, and $\boldsymbol{Y}$ is the dependent variable parameter vector. To compute a percent change over time and its standard error, we define a dependent variable matrix, $\boldsymbol{X_p}$, that represents the 15th day of the month for each month throughout the two years at the beginning and end of the period of interest. So for a difference between 1985-1986 to 2015-2016, $\boldsymbol{X_p}$ for gam2 would be a two column matrix with the first column containing the decimal date for:

Jan 15, 1985; Feb 15, 1985;… Dec 15, 1986; Jan 15, 2015; Feb 15, 2015; … Dec 15, 2016.

The second column would be the day of year corresponding to the 15th of each month repeated four times, once for each of the four years considered. $\boldsymbol{X_p}$ is linearly converted in 'mgcv' via the fitted GAM to the prediction vector ($\boldsymbol{Z_p}$). From $\boldsymbol{Z_p}$, matrix $\boldsymbol{Z_d}$ is computed such that $\boldsymbol{Z_d}\hat{\boldsymbol{\beta}}$ is an estimate of the difference between the first two years and last two years of predictions:

$$\mathbf{Z_d} = \mathbf{dAZ_p} \quad \text{(eq 5)}$$

where $\boldsymbol{d}$ is a simple row vector [-1 1]. $\boldsymbol{A}$ is a two row averaging matrix with the following structure where 1/24 is repeated 24 times followed by 24 zeros on the first row, with the opposite on the second row.

A = [1/24 1/24 1/24 ….. 0 0 0 ….
0 0 0 ….. 1/24 1/24 1/24 ….]

Pre-multiplying $\hat{\boldsymbol{\beta}}$ by $\boldsymbol{Z_d}$ provides a point estimate of the difference of the last n years minus the first n years as a simple linear transformation of the estimated parameter vector.

$$\text{Difference} = Z_d \hat{\boldsymbol{\beta}} \qquad \text{(eq 6)}$$

The standard error of this difference is from the following (as in Rao, 2001):

$$se_{diff} = \sqrt{Z_d \hat{\sum}_\beta Z_d^{\mathrm{T}}} \qquad \text{(eq 7)}$$

4. Results and Discussion

4.1 Temporal model selection and smoothing

Fitting an example dataset with each of the first three GAM models (gam0, gam1, and gam2) demonstrates the impact of each additional smooth term on model shape and fit to the data (Figure 2; Table 1). In this example, there is a clear seasonal cycle, and an apparent decrease over time shown via gam0 (Figure 2a). The R^2 and residual diagnostics (Table 1; Figure SM2) suggest there could be room for improvement with this model structure, and indeed adding the *s(cyear)* term with gam1 demonstrates the nonlinear pattern over time and improves the adjusted R^2 (Figure 2b; Table 1). Inclusion of *s(cyear)* in gam1 makes the p-value on the linear *cyear* term >0.10, essentially making it possible to drop that term from the model (although we retain it for simplicity since it has no impact on the results). With gam2, the addition of the tensor product that allows for a changing seasonal cycle is shown with the shape of the black line as well as with four colored lines that represent mean predictions at four different days of the year (Figure 2c). The gam2 model (Figure 2c) indicates that seasonal cycling of TN is greater in the 1993-1998 period than in the 2008-2013 period, whereas gam1 (Figure 2b) estimates constant seasonal cycling for the period of record. The most complicated model has the best model fit using AIC, RMSE, and adjusted R^2 (Table 1).

Autocorrelation of the residuals is something to consider with these results (Supplemental Materials, Figure SM2). An option we have not implemented, but that we are considering, is a mixed model approach that incorporates autocorrelation into the error structure. For the evaluation of the models presented here, autocorrelated errors may mean underestimated p-values on the model terms. Because our purpose is mostly descriptive, this uncertainty is acceptable for the evaluations described here. Notably, significant autocorrelation is almost always removed with models incorporating hydrology (Section 4.2). If data points were even closer together temporally, or if there was not another explanatory variable such as hydrology to incorporate, users might want to carefully evaluate the use of a mixed model. Residual diagnostics of all following example GAM graphics are presented in the Supplemental Materials with labels corresponding to the figures in the main text.

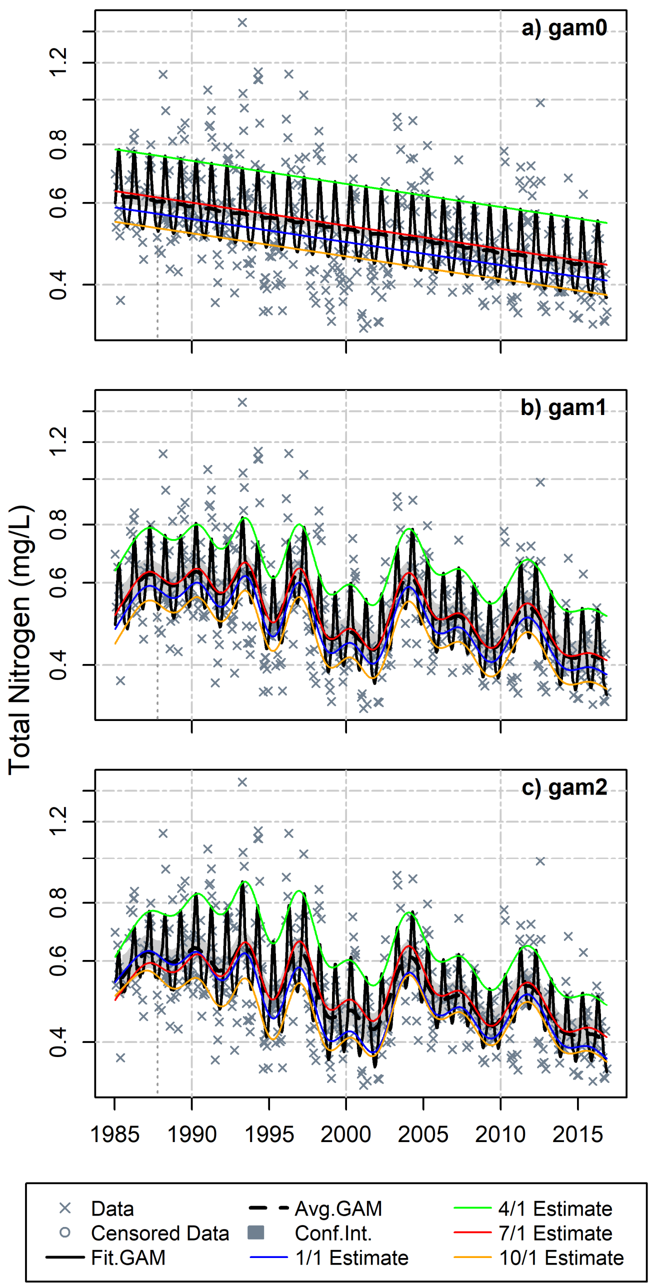


Figure 2. Example output from (a) gam0, (b) gam1 and(c) gam2 for surface TN at a mid-bay station. In each plot, the x marks are the data, circles are a small number of censored data (i.e., reported at the detection limit), the solid black line is the full GAM fit (Fit. GAM), the dotted black line with gray bounds is the average prediction (Avg. GAM) with a 95% confidence bound (Conf. Int.), and four colored lines represent estimates for Jan 1 (blue), Apr. 1 (green), July 1 (red) and Oct. 1 (orange). Results are shown on a log-scale.

Table 1. Details on gam0, 1 and 2 example fits for CB5.4 surface TN (Figure 2)

Model and diagnostics	Term	Linear parameter coefficients			GAM analysis of variance		
		estimate	Std. error	p.value	df	F	p.value
gam0, AIC = -100.9; RMSE=0.213; Adj R^2 = 0.373	intercept	-0.647	0.0102	<0.0001			
	cyear	-0.0114	0.00113	<0.0001	1.00	101	<0.0001
	s(doy)				6.19	17.7	<0.0001
gam1, AIC = -178.2; RMSE=0.191; Adj R^2 = 0.496	intercept	-0.597	0.0577	<0.0001			
	cyear	-0.0472	0.0668	0.48	1.00	0.498	0.48
	s(cyear)				18.9	5.68	<0.0001
	s(doy)				6.83	22.3	<0.0001
gam2, AIC = -195.2; RMSE=0.186; Adj R^2 = 0.522	intercept	-0.614	0.572	<0.0001			
	cyear	0.0260	0.0664	0.70	1.00	0.154	0.70
	s(cyear)				19.0	5.88	<0.0001
	s(doy)				6.85	23.6	<0.0001
	ti(cyear,doy)				6.42	1.84	0.0004

For the purposes of evaluating water quality annually for many parameters, stations, and depths, model selection must be simplified and ideally defined in advance. Therefore, we conducted a comparison between the performances of these three temporal GAMs (gam0,1 and 2) for TN, TP, chlorophyll-*a*, DO, and Secchi disk depth, for most mainstem and tributary stations from 1999-2016. The comparison of model fit is based on AIC, and for 99% of these data sets, AIC values for gam2 are lower than or equal to those for gam1 and gam0 (Table 2). A negative difference means that the gam2 AIC is lower, therefore suggesting a “better” model. It is clear that the gam2 models frequently have much lower AICs than the other models, and when gam0 or gam1 performed better, the difference is very small. Based on these findings, gam2 was selected as the primary temporal-only model for Chesapeake Bay annual tidal trend assessments that are graphically presented to the public, and the model to build upon for incorporating hydrology (i.e., gam2_flwsal). Based on this experience, it is difficult to think of an estuarine water quality scenario where gam2 would not perform as well if not better than gam0 or 1. For instance, a linearly changing data set could be modeled accurately with gam2; there would simply be little variability attributed to the nonlinear terms. In periods of missing data, any of these models will tend to the average with large uncertainty during the data gap.

Table 2. Statistics on AIC ranges ($AIC_{gam2} - AIC_{gam0or1}$) for 1999-2016 data at 137 stations.

parameter	layer	min diff	Quantiles of differences			max diff
			25	50	75	

Chlorophyll-a	B	-34.0	-11.9	-3.71	-0.88	0.02
Chlorophyll-a	S	-43.9	-8.24	-3.5	-0.84	0.01
DO	B	-22.3	-5.53	-2.09	-0.01	0
Secchi depth	NA	-27.5	-6.80	-2.4	-0.18	0.03
TN	B	-31.3	-9.97	-4.55	-1.07	0.01
TN	S	-79.5	-12.3	-5.71	-1.04	5.06
TP	B	-17.3	-4.63	-1.38	-0.16	0.92
TP	S	-23.1	-7.73	-2.78	-0.43	0.01

Another component to GAM fitting involves confirming that the upper level on the flexibility of the spline function is high enough for the thin plate spline-based smooths (Wood, 2006; also Section 3.1). The recommended procedure for selecting the k-value is a manual process to ensure that the fit is not reaching the upper limit (Wood, 2018). The k-value selection is very application-specific and should be re-evaluated with any new location. In this case study, with a large number of analyses to conduct every year, we needed to select a reasonable default k-value. After some testing, it was clear that many of the nonlinear patterns over time are due to multi-year wet and dry periods. To capture these cycles, we set the k-value to 2/3 times the number of years. This allows for a change in direction roughly every 3 years, if needed (e.g., Figure 3b). Less flexibility in the model (Figure 3a, Table SM1) is over-smoothed and poorly fits the data. More flexibility in the model (e.g., Figure 3c, Table SM1) sometimes improves model fit, and allows for a degree of "bumpiness" that essentially captures year-to-year variation. In this Chesapeake Bay study, because our purpose is to examine long-term changes that persist beyond a year, we found that this high amount of model flexibility adds unnecessary computation and interpretation time. In other applications, however, a higher amount of variability may be necessary, perhaps where climatic patterns that vary dramatically year-to-year are the driver of the change being evaluated. In addition, if any other explanatory variables are added to the model, this also requires a re-evaluation of the k-values, as we found when incorporating hydrology effects into the models (next section).

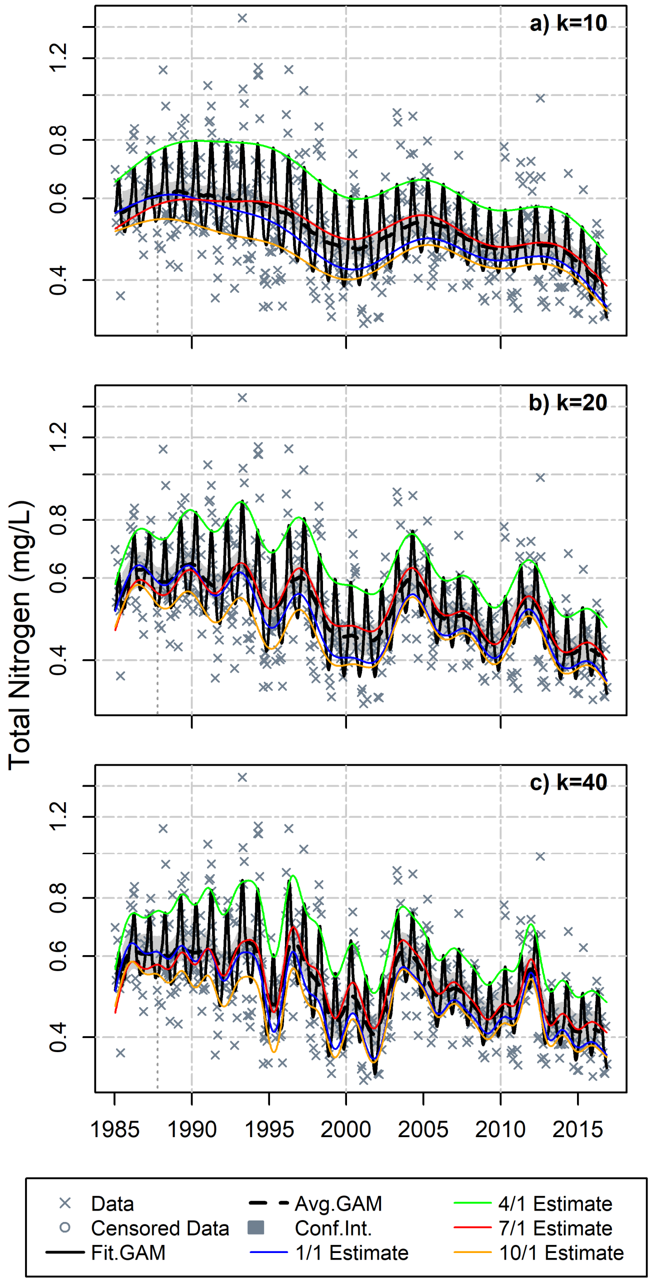


Figure 3. Example showing the effect of adjusting the k-value for the *s(cyear)* smooth with progressively increasing k-value from (a) to (c). All graphs are for surface TN at station CB5.4.

4.2 Incorporating flow or salinity

Like many estuaries, the range and variability in river flow from the major freshwater sources into Chesapeake Bay is quite large (Figure 4). Annual average river flow through the Susquehanna ranged from 670 to 2,040 m^3/s from 1985 to 2016, and for the Potomac River from 180 to 790 m^3/s (Figures 4a, b). An inverse relationship exists between downstream salinity and freshwater river flow with higher flow years (e.g., 1996, 2003-2004, 2011) showing lower annual average salinity in both systems (Figures 4c, d). The impact on water quality of large multi-year variations in freshwater input can be seen in many of the temporal (gam2) graphics. See, for example, the multi-year dip in TN concentrations mid-bay during a drought period extending from 1999-2002 (Figure 3b). This corresponds to a 4-year period of below average river flows (Figure 4a, b) and high salinities (Figure 4 c,d).

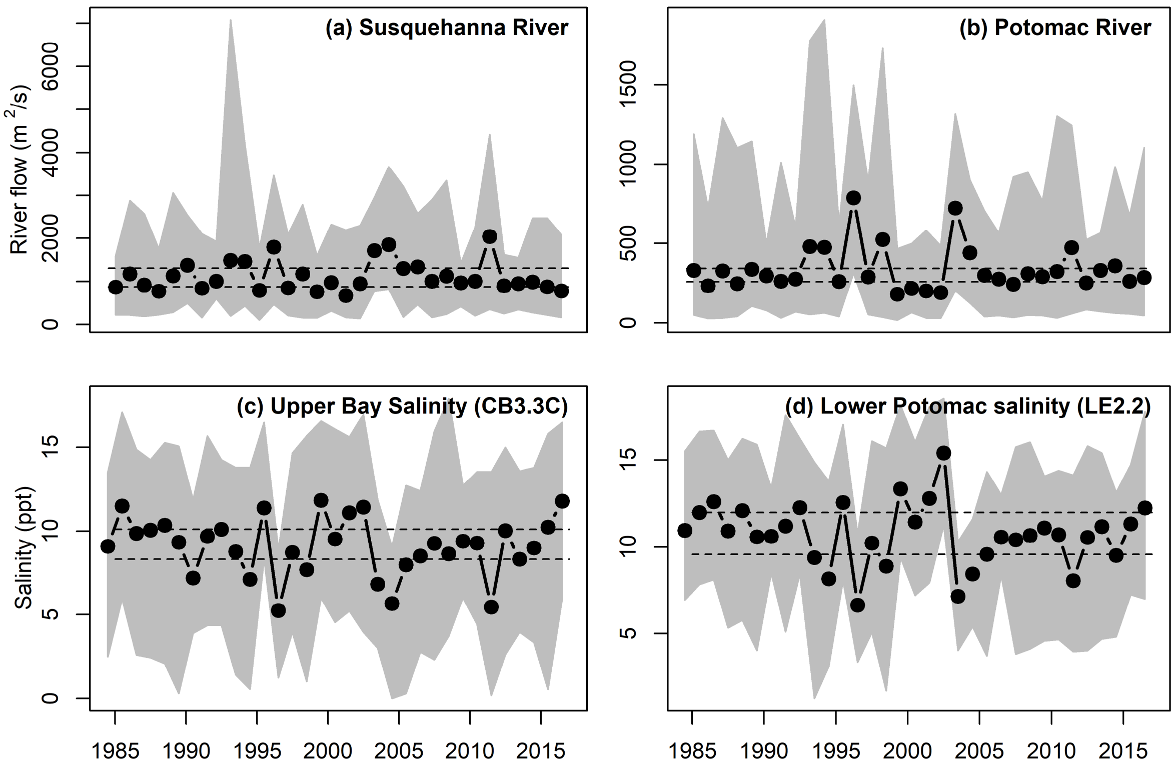


Figure 4. Mean annual river flow from two largest freshwater sources to Chesapeake Bay (Susquehanna River (a) and Potomac River (b)) and mean annual salinity at a station in in mesohaline mainstem Chesapeake Bay (c) and mesohaline Potomac (d). Grey regions show range of monthly average flow or salinity observations. Dotted lines show 25th and 75th quantiles around the annual means.

We fit two gam2_flwsal models to the same example used above, CB5.4 surface TN, with the first using the preceding 90 days of Susquehanna River flow for *flw_sal* (Figure 5a) and the second using salinity for *flw_sal* (Figure 5b). Ninety days was selected as an optimal averaging window by computing the Spearman correlation between the data set and a range of flow options between 15 and 210 days. The magenta dashed lines in these graphs show the full gam2_flwsal estimates, which incorporate seasonal patterns and variability in flow or salinity. The solid black lines show the flow-adjusted results, i.e. estimated conditions if flow or salinity had been average throughout the record. The heavy black dashed lines show the seasonally-adjusted and flow-adjusted long-term trend. The impact of the 1999-2002 drought on the long-term pattern is removed with both the flow and salinity adjustments. A summary of the GAM output (Table 3) shows that including either flow or salinity improves the model fit, and that salinity is slightly more explanatory than flow in this case. At this mid-bay location, salinity may be a better explanatory variable than river flow because salinity integrates the influence of multiple freshwater sources. Due to the nature of estuaries being a mixture of fresh and salt water, in tidal fresh areas of estuaries, there will not be enough salinity to use it as a predictor. In this case study for those fresh regions, nearby gaged river flow is the current choice. To illustrate this example, we analyzed a chlorophyll-*a* time series from the tidal fresh James River. We first fit it with a temporal gam2 (Figure 5c, Table 3) and then with gam2_flwsal using the preceding 10 days of river flow from the James River RIM station (Figure 5d, Table 3). In the future, we will explore using average salinity from downstream locations, which may also be a viable option in an estuary without nearby gaged river flow.

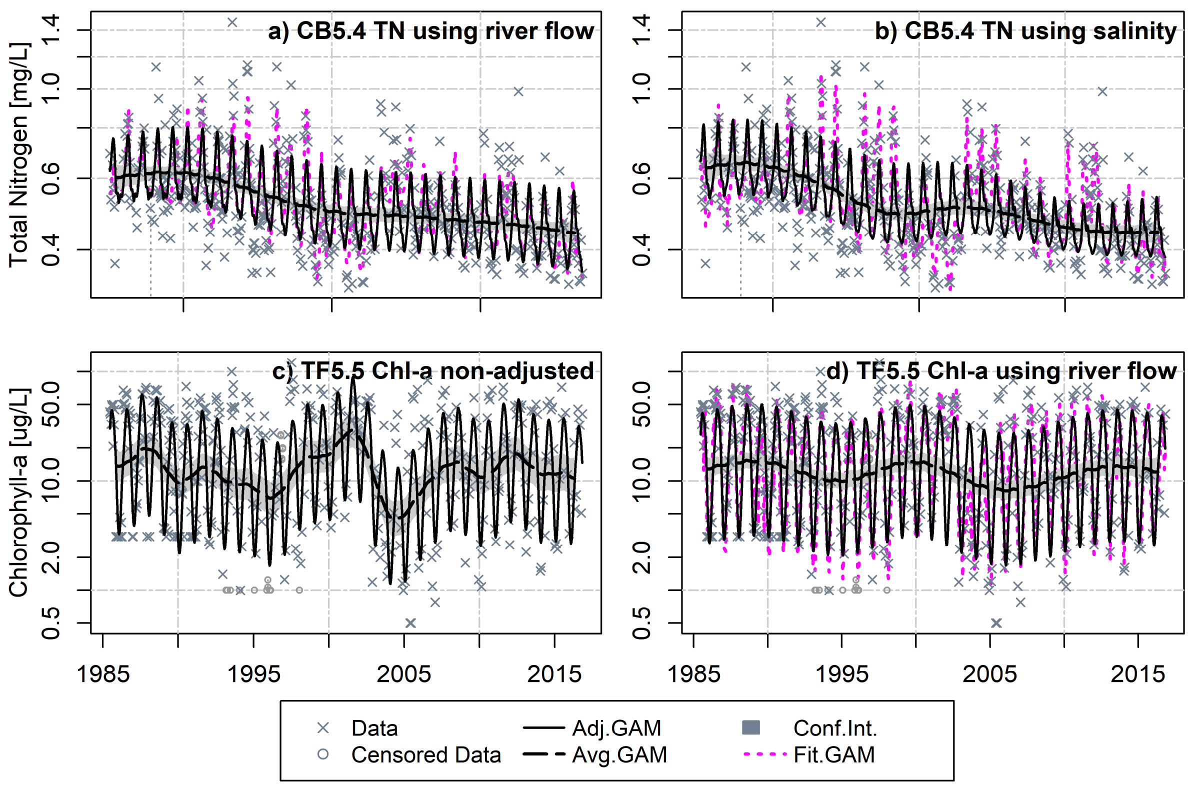


Figure 5. Examples of flow- and salinity-adjustment results for (a) CB5.4 TN flow-adjusted model using previous 90 days of flow from Susquehanna; (b) CB5.4 TN using salinity-adjustment; (c) James River TF5.5 chlorophyll-*a* non-adjusted using gam2; (d) James River TF5.5 chlorophyll-*a* flow-adjusted using preceding 10 days of flow averaged. Adj.GAM refers to estimates after flow or salinity adjustment, Avg.GAM is the seasonally-averaged adjusted model, and Fit.GAM is the full fitted model before adjustment

Table 3. Details on GAM shown in Figure 5.

Station	Model and diagnostics	Term	Linear parameter coefficients			GAM analysis of variance		
			estimate	Std. error	p.value	df	F	p.value
CB5.4	gam2_flwsal with flow, AIC = -324.9; RMSE=0.161; Adj R^2 = 0.641	intercept	-0.635	0.0119	<0.0001			
		cyear	0.0058	0.0101	0.57	1.0	0.33	0.57
		s(cyear)				4.26	3.25	0.006
		s(doy)				7.15	27.0	<0.0001
		ti(cyear,doy)				4.05	1.10	0.001
		s(flw_sal)				1.84	72.0	<0.0001
		ti(flw_sal,doy)				5.12	6.40	<0.0001
		ti(flw_sal,cyear)				4.12	1.27	0.28
		ti(flw_sal,doy,cyear)				2.30	0.13	0.03
CB5.4	gam2_flwsal with salinity, AIC = -443.0; RMSE=0.140; Adj R^2 = 0.729	intercept	-0.621	0.0121	<0.0001			
		cyear	-0.0141	0.0126	0.26	1.00	1.26	0.26
		s(cyear)				5.98	6.25	<0.0001
		s(doy)				7.06	36.0	<0.0001
		ti(cyear,doy)				5.64	2.79	<0.0001
		s(flw_sal)				2.40	81.8	<0.0001
		ti(flw_sal,doy)				5.14	5.54	<0.0001
		ti(flw_sal,cyear)				5.85	3.33	0.002
		ti(flw_sal,doy,cyear)				1.88	0.111	0.02
TF5.5	gam2, AIC = 1098; RMSE=0.865; Adj R^2 = 0.543	intercept	2.23	0.448	<0.0001			
		cyear	-0.216	0.264	0.413	1.00	0.673	0.41
		s(cyear)				16.8	4.43	<0.0001
		s(doy)				5.09	50.9	<0.0001
		ti(cyear,doy)				0.79	0.086	0.25
TF5.5	gam2_flwsal with flow, AIC = 893; RMSE=0.677; Adj R^2 = 0.720	intercept	2.52	0.103	<0.0001			
		cyear	-0.0394	0.0579	0.50	1.0	0.463	0.50
		s(cyear)				6.47	3.90	0.0003
		s(doy)				5.83	78.1	<0.0001
		ti(cyear,doy)				--[a]	--	--
		s(flw_sal)				1.00	244	<0.001
		ti(flw_sal,doy)				7.10	1.75	0.002
		ti(flw_sal,cyear)				1.00	0.279	0.60
		ti(flw_sal,doy,cyear)				2.64	0.097	0.12

[a]Selected out of the model by *select=TRUE*

In initial testing with default k-values on the *cyear* and *flw_sal* smooths, the flow or salinity adjustment did not always remove the impact of the year-to-year variability of freshwater input. After some research, we learned that it was necessary to adjust the relative flexibility of these spline bases in order to account for concurvity (Buja et al., 1989). Concurvity, in this case, means that the smooth term for *flw_sal* can be approximated by the smooth term for *cyear* (Wood, 2018). This occurs when there are long-term non-linear patterns in river flow. One useful approach (based on Peng et al., 2006) to correct for this is to limit the flexibility of the smooth on *cyear* so that more of the variability can be modeled with the smooth on *flw_sal*. After running a comprehensive suite of tests across Chesapeake Bay stations, we found that setting the k-value on *s(cyear)* to one-third times the number of years of the analysis and the k-value on *s(flw_sal)* to two-thirds times the number of years effectively limited the flexibility of the *cyear* smooth function in most cases. This is shown in the equations in Section 3.2 for gam2_flwsal and Section 3.5 for gam3_flwsal, and for an example Secchi data set with flow-adjustment in the Supplemental Materials (Fig. SM5).

These k-values allow for double the amount of variability to be attributed to hydrology than to the s(*cyear)* term. Our selected k-values perform well for automated analysis in Chesapeake Bay, but in any focused individual case (in the Chesapeake Bay or elsewhere), improvements could likely be made to this approach with more testing, particularly on the effect of the k-values in the tensor product interactions. In other systems, we recommend site-specific testing of the k-values because the temporal scale of hydrologic effects could be very different, and our experience showed that possible concurvity cannot be ignored when additional terms are added to an explanatory GAM model.

4.3 Lab or method change interventions

We used the intervention approach described in Section 3.3 to evaluate suspected impacts of laboratory or method changes, which are common in long-term monitoring records. For example, a change in the laboratory analyzing Maryland tributary data resulted in a suspected change to the TSS values in 1998 (Karrh, 2017). We fit the TSS data from station EE3.0 in Fishing Bay with a gam3 model with an intervention on May 1, 1998 (Figure 6). Results strongly indicate a significant change on this date ($p=0.001$; Supplemental Materials Table SM3). From these results, the model predictions were adjusted by the estimated size of the change to show what the long-term pattern would have looked like if the laboratory had been the same throughout the record (Figure 6b). Interviews with laboratory personnel revealed that samples collected prior to the lab change were not rinsed as thoroughly as samples after the change. The prior samples retained some salt, which explains the step down.

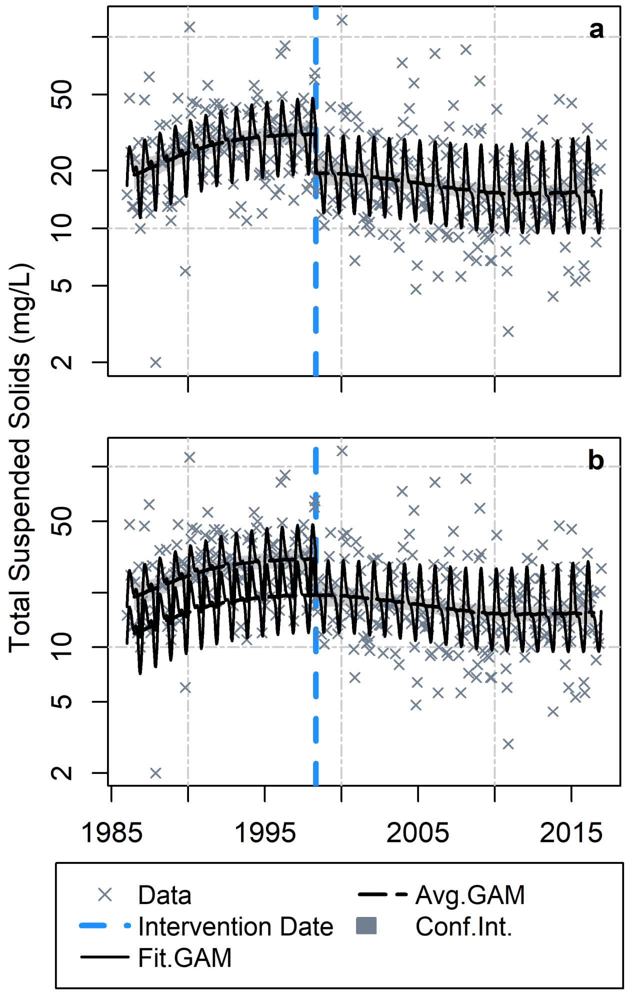


Figure 6. Modeling an intervention for surface TSS at EE3.0. The lab change is indicated with a blue dotted line and (a) modeled response with a change, (b) adjusted model results also shown.

Combining this intervention approach with the flow or salinity adjustment approach via gam3_flwsal helps to distinguish the effect of a method change versus a hydrologic change. For example, a laboratory and method change happened concurrently for total phosphorus evaluations on May 1, 1998 in the smaller MD tributaries (Karrh, 2017). By coincidence, this fell between a higher flow period in 1996 and a lower flow period starting in 1999, both of which appear to influence the TP dataset. When gam3—which includes the intervention but no river flow effect—is tested, the intervention is not significant (intervention p=0.28, AIC=357, Adjusted R^2 = 0.44, Fig7a). But when salinity is included as an explanatory variable with gam3_flwsal, the intervention is significant (intervention p=0.02, AIC=340, Adjusted R^2 = 0.47, Figure 7b), and the model fit improves (Supplemental Materials Table SM4). With gam3_flwsal, a few changes occurred which could have caused this result: the inclusion of salinity could be accounting for some of the shape to the long-term pattern, and/or the lower k-value on the *s(cyear)* term in gam3_flwsal makes the cyear spline stiffer and less likely to capture the possible intervention as part of the *cyear* spline.

This model structure is successful for evaluating known interventions, but identifying possible interventions can be a time consuming process that requires knowledge of the data collection and processing history. Preliminary results from our case studies indicate that when more than one intervention is suspected in a short period of time, distinguishing their effects is challenging (results not shown). More testing will be needed to evaluate such situations.

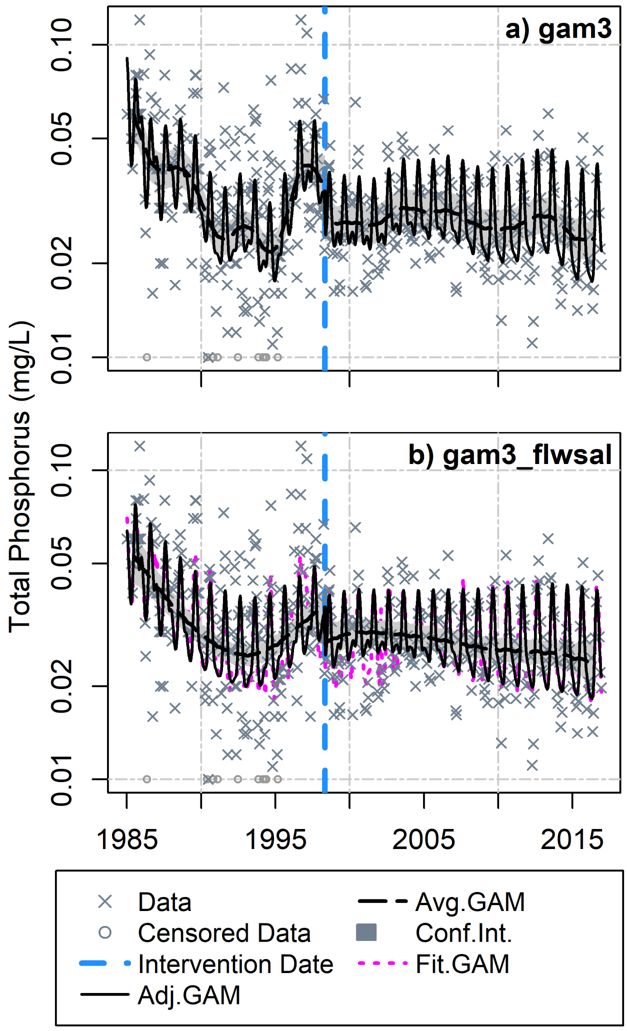
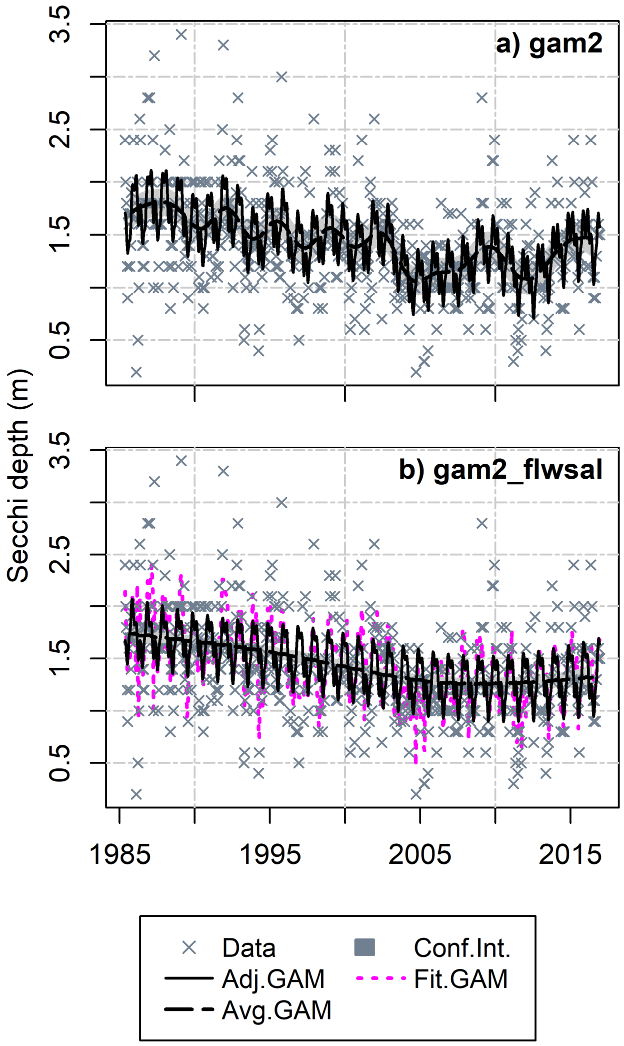


Figure 7. Intervention example at EE2.1 for surface TP without accounting for flow or salinity variation (a), and with accounting for salinity (b).

4.4 Percent change results Bay-wide

After fitting a GAM model using one of the six approaches described above (gam0-gam3_flwsal), an overall measure of change over time can be computed as described in Section 3.4. For example, Secchi depth appears to be decreasing at a station in the mid-bay using both a gam2 (Figure 8a) and salinity-adjusted gam2_flwsal (Figure 8b). The percent change computations on these models (Table 4) support this visual conclusion with an average decrease of 0.26 meters from the gam2 (p-value 0.01) and 0.41 meters in the salinity-adjusted model (p-value <0.0001). Presenting the findings from both GAMs gives us confidence that a decrease occurred independent of the impact of hydrologic variability.



564

565 **Figure 8**. Example of model results where percent change over time is computed for both a (a)

566 temporal model gam2, and (b) salinity-adjusted model gam2_flwsal. Example is for Secchi at CB4.1C.

567

Table 4. Model diagnostics and percent change computation results for CB4.1C Secchi (Figure 8).

Value	**Temporal model (gam2)**	**Model with salinity (gam2_flwsal)**
Model Diagnostics		
AIC	563.6	415.7
RMSE	0.399	0.349
Adj R^2	0.328	0.489
Estimates of change over full period for regular (gam2) or salinity-adjusted (gam2_flwsal) results		
Baseline mean	1.72	1.73
Current mean	1.46	1.32
Estimated difference	-0.259	-0.412
Std. Err. difference	0.102	0.071
95% Confidence interval for difference	(-0.459 , -0.0595)	(-0.551 , -0.273)
Difference p-value	0.011	<0.0001
Period of Record Percent Change Estimate (%)	-15.1%	-23.8%

These computations were repeated for every station for Secchi depth, and summarized with maps for both the long term (1985-2016, Figure 9a,c) and short-term (2007-2016, Figures 9b,d). In the maps of Chesapeake Bay tidal trends, we have decided to show the percent change results using a range of symbols. Specifically, we use an alpha of 0.05 as a cutoff for a “significant” trend, but we also relax that criterion to 0.25 to show a second level of “possible” trends. This is a higher alpha level than normally considered, but this approach allows us to present a richer suite of results that can help managers identify locations were changes might be starting to occur, and hence dig deeper immediately into those results rather than wait until a trend is “significant.” This is similar to the approach used for presenting trends in nutrient and sediments loads into the Chesapeake Bay at the RIM stations (Figure 1) that are generated by the USGS (Hirsch et al., 2015). The Secchi maps show clear spatial and temporal patterns. At most stations, Secchi depth degraded over the longer record, with a change in direction in the last 10 years in the middle parts of Chesapeake Bay. A more thorough investigation of these Secchi patterns is part of a study on long-term water clarity in Chesapeake Bay (Keisman et al., 2018).

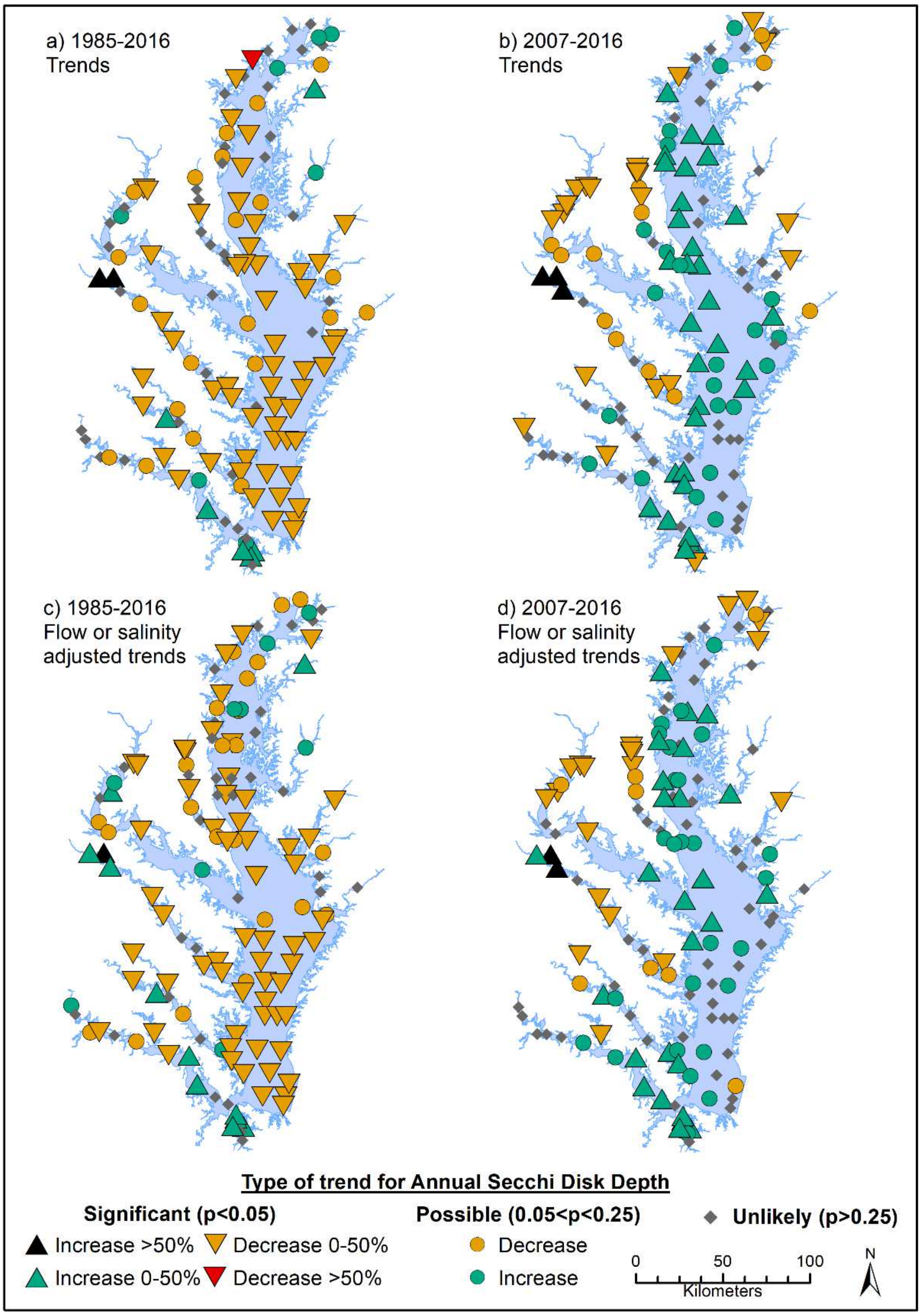
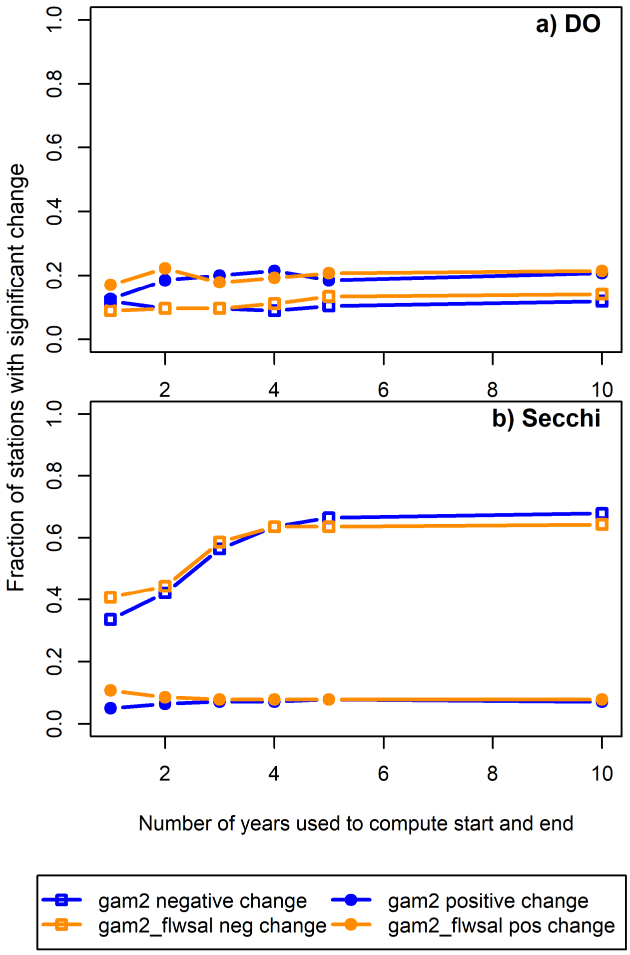


Figure 9. Annual bay-wide percent change results based on gam2 (a and b) and gam2_flwsal (c and 6) for Secchi Depth.

As discussed in Section 3.4, the selection of two years for the start and end of the period of interest in percent change computations is adjustable. We tested how the trend summaries might change if a different number of years was used. Specifically, we computed percent change using either 1, 2, 3, 4, 5 or 10 years as the beginning and ending windows from GAM results for Secchi and bottom DO from 1985-2016 and surface TN and TP from 1999-2016. The results were evaluated by counting how many stations had a significant percent change (i.e., $p<0.05$) for each option. This evaluation is summarized graphically (Figure 10, Figure SM9) and shows that for DO there is almost no change in the number of stations with a significant percent change across the options, but for Secchi there is a noticeable increase in the number of stations with significant percent change when three years is considered instead of two. TN and TP fall between these ranges (Supplemental Materials, Figure SM9). Similarity between flow-adjusted and non-flow adjusted results indicates that differences between the options are not due to hydrologic effects.

The choice of how many years to use at the start and end of the percent change computation must be made based on the purpose of the assessment. One option is to make a selection based on where the graphs are plateauing (perhaps 3 or 4 years in this case). Another is to use fewer years in order to identify a trend change soon after it occurred. We chose to use two years for our Chesapeake Bay assessments to prioritize detection of major shifts soon after they occur. The resulting greater sensitivity to changes in trend direction increases the potential for a reverse finding given a few years of additional data. This choice should be based on the needs of the assessment for other studies or locations.



0

Figure 10. Fraction of total stations with significant (p<0.05) negative (square) or positive (circle) trends when a different number of years is used to compute the beginning and ending periods (x-axis). Blue lines are gam2 results and orange lines are flow-adjusted results for (a) bottom summer DO, and (b) secchi depth.

5. Conclusions

The approach described here to apply GAMs to a diverse set of water quality parameters across a large spatial scale in an estuary can help describe both large scale (e.g., Figure 9) and local (e.g., Figure 2, Figure 5) water quality dynamics. The results from our Chesapeake Bay case study are serving as a starting point for further research evaluations (e.g., Keisman et al., 2018; Testa et al., 2018) as well as for local-scale planning efforts required by the Chesapeake Bay watershed TMDL. Detailed GAM analyses have added to the understanding of water quality conditions in multiple coastal and estuarine regions, including, for example, turbidity off the Gold Coast, Queensland, Australia (Richards et al., 2013) and chlorophyll-*a* and nutrients in Bohai Bay, China (Qiao et al., 2017). The uniqueness of this study, however, stems from our need to implement an approach that retains statistical rigor while being applied operationally every year to a large set of stations and parameters. We accomplished this by conducting testing across a range of scenarios and allowing flexibility in our model specifications. Many of our choices along the way for model complexity, k-values, concurvity adjustment with additional variables, method change adjustments, and visualizing the results over large spatial scales can provide a starting point for application in other systems with long-term water quality monitoring. However, to ensure rigorous results, it would be necessary to work with both local scientists who can test and adjust these choices by, for example, selecting appropriate hydrology adjustment (Section 3.2) and local data providers who are aware of potential data issues (Section 3.3). Certainly, there are analysis questions that are not best addressed with this GAM approach. We generally do not use these model structures for data sets with fewer than eight years of data, because even if there were enough data for a valid model fit, our experience shows that a period of less than 8 years is too short to decipher long-term change from short-term fluctuations. In addition, we are not proposing these techniques for extrapolation, or prediction, beyond the years for which data have been collected. A dynamic modeling approach would be more applicable for generating predictions.

In this study, we focused on model selection, incorporating hydrology, accounting for method changes, and extracting long-term change conclusions from GAM results. Other topics under current investigation by our team include how to appropriately model censored data with an Expectation Maximization approach (Liu et al., 1997), how to systematically incorporate other explanatory variables such as watershed nutrient inputs into the models, and how to identify appropriate scales for spatially aggregating the water quality stations in the GAM approach. Additional refinements that could improve this approach include accounting for residual autocorrelation, alternative data transformations, a more thorough investigation of concurvity, and adjusting model smoothing with k-values.

6. Acknowledgements

We would like to acknowledge Renee Karrh (MDDNR), Mike Lane (ODU), and Monika Arora (ODU), for valuable work and insights testing and implementing this approach to water quality data in MD and VA; the VA and MD tidal monitoring programs and individuals who have worked to collect, measure and quality assure the monitoring data over the last 33 years; Erik Leppo (Tetra Tech) for his work on the R package development; and members of a STAC review committee (Hugh Ellis, JHU; Pan Du, VA Tech; Carl Friedrichs, VIMS; and Vyacheslav Lyubchich, UMCES) who provided helpful comments on this approach. This work was supported in part by the U.S. Environmental Protection Agency [grant EPA/CBP Technical Support 2017 No. 07-5-230480]. This is contribution no. xxx of the University of Maryland Center for Environmental Science.

7. References

Beck, M.W., Hagy, J.D., 2015. Adaptation of a Weighted Regression Approach to Evaluate Water Quality Trends in an Estuary. Environmental Modeling & Assessment 20(6) 637-655. 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 Journal of the American Water Resources Association 53(1) 197-219. 10.1111/1752-1688.12489.

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B (Methodological) 26(2) 211-252.

Box, G.E.P., Tiao, G.C., 1975. Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association 70(349) 70-79. 10.1080/01621459.1975.10480264.

Buja, A., Hastie, T., Tibshirani, R., 1989. Linear Smoothers and Additive Models. The Annals of Statistics 17(2) 453-510. 10.1214/aos/1176347115.

Chesapeake Bay Program, 2017a. CBP Water Quality Database (1984-present). https://www.chesapeakebay.net/what/downloads/cbp_water_quality_database_1984_present.

Chesapeake Bay Program, 2017b. Methods and Quality Assurance for Chesapeake Bay Water Quality Monitoring Programs. CBP/TRS-319-17.

Cloern, J.E., 2001. Our evolving conceptual model of the coastal eutrophication problem. Marine Ecology Progress Series 210 223-253.

Duarte, C.M., Conley, D.J., Carstensen, J., Sánchez-Camacho, M., 2009. Return to Neverland: Shifting Baselines Affect Eutrophication Restoration Targets. Estuaries and Coasts 32(1) 29-36. 10.1007/s12237-008-9111-2.

Duchon, J., 1977. Splines minimizing rotation-invariant semi-norms in Sobolev spaces, Constructive Theory of Functions of Several Variables. Springer, Berlin, Heidelberg, pp. 85-100.

Giampiero, M., Wood, S., 2012. Coverage Properties of Confidence Intervals for Generalized Additive Model Components. Scandinavian Journal of Statistics 39(1) 53-74. 10.1111/j.1467-9469.2011.00760.x.

Greening, H., Janicki, A., Sherwood, E.T., Pribble, R., Johansson, J.O.R., 2014. Ecosystem responses to long-term nutrient management in an urban estuary: Tampa Bay, Florida, USA. Estuarine, Coastal and Shelf Science 151 A1-A16. https://doi.org/10.1016/j.ecss.2014.10.003.

Hagy, J.D., Boynton, W.R., Keefe, C.W., Wood, K.V., 2004. Hypoxia in Chesapeake Bay, 1950–2001: Long-term change in relation to nutrient loading and river flow. Estuaries 27(4) 634-658. 10.1007/BF02907650.

Hagy, J.D., Boynton, W.R., Sanford, L.P., 2000. Estimation of net physical transport and hydraulic residence times for a coastal plain estuary using box models. Estuaries 23(3) 328-340. 10.2307/1353325.

Haraguchi, L., Carstensen, J., Abreu, P.C., Odebrecht, C., 2015. Long-term changes of the phytoplankton community and biomass in the subtropical shallow Patos Lagoon Estuary, Brazil. Estuarine, Coastal and Shelf Science 162 76-87. 10.1016/j.ecss.2015.03.007.

Harding, L.W., Gallegos, C.L., Perry, E.S., Miller, W.D., Adolf, J.E., Mallonee, M.E., Paerl, H.W., 2016. Long-Term Trends of Nutrients and Phytoplankton in Chesapeake Bay. Estuaries and Coasts 39(3) 664-681. 10.1007/s12237-015-0023-7.

Hastie, T., Tibshirani, R., 2004. Generalized Additive Models, Encyclopedia of Statistical Sciences. John Wiley & Sons, Inc.

Hirsch, R.M., Archfield, S.A., De Cicco, L.A., 2015. A bootstrap method for estimating uncertainty of water quality trends. Environmental Modelling & Software 73 148-166. https://doi.org/10.1016/j.envsoft.2015.07.017.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18(1) 107-121. 10.1029/WR018i001p00107.

Jassby, A., 2008. Phytoplankton in the Upper San Francisco Estuary: Recent Biomass Trends, Their Causes, and Their Trophic Significance. San Francisco Estuary and Watershed Science 6(1)https://escholarship.org/uc/item/71h077r1.

Karrh, R., 2017. Personal Communication.

Keisman, J., Friedrichs, C., Buchanan, C., Cornwell, J., Lane, M., Porter, E., Testa, J., Trice, M., Zhang, Q., Zimmerman, R., Batiuk, R., Blomquist, J., Lyubchich, S., Moore, K., Murphy, R., Noe, G., Orth, R.J., Sanford, L., 2018. Understanding and explaining over 30 years of water-clarity trends in Chesapeake Bay: Previous work and new insights. Edgewater, MD. STAC Publication Number 18-XXX.

Le Fur, I., De Wit, R., Plus, M., Oheix, J., Derolez, V., Simier, M., Malet, N., Ouisse, V., 2019. Re-oligotrophication trajectories of macrophyte assemblages in Mediterranean coastal lagoons based on 17-year time-series. Marine Ecology Progress Series 608 13-32.

Lefcheck, J.S., Wilcox, D.J., Murphy, R.R., Marion, S.R., Orth, R.J., 2017. Multiple stressors threaten the imperiled coastal foundation species eelgrass (Zostera marina) in Chesapeake Bay, USA. Global Change Biology 23(9) 3474-3483. 10.1111/gcb.13623.

Liu, S., Lu, J.-C., Kolpin, D.W., Meeker, W.Q., 1997. Analysis of Environmental Data with Censored Observations. Environmental Science & Technology 31(12) 3358-3362. 10.1021/es960695x.

Najjar, R.G., Pyke, C.R., Adams, M.B., Breitburg, D., Hershner, C., Kemp, M., Howarth, R., Mulholland, M.R., Paolisso, M., Secor, D., Sellner, K., Wardrop, D., Wood, R., 2010. Potential climate-change impacts on the Chesapeake Bay. Estuarine, Coastal and Shelf Science 86(1) 1-20. 10.1016/j.ecss.2009.09.026.

Paerl, H.W., Pinckney, J.L., Fear, J.M., Peierls, B.L., 1998. Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. Marine Ecology Progress Series 166 17-25.

Peng, R.D., Dominici, F., Louis, T.A., 2006. Model choice in time series studies of air pollution and mortality. Journal of the Royal Statistical Society: Series A (Statistics in Society) 169(2) 179-203.

Qiao, Y., Feng, J., Cui, S., Zhu, L., 2017. Long-term changes in nutrients, chlorophyll a and their relationships in a semi-enclosed eutrophic ecosystem, Bohai Bay, China. Marine Pollution Bulletin 117(1) 222-228. https://doi.org/10.1016/j.marpolbul.2017.02.002.

Rao, C.R., 2001. Linear Statistical Inference and its Applications, 2nd ed. Wiley.

Richards, R., Hughes, L., Gee, D., Tomlinson, R., 2013. Using generalized additive models for water quality assessments: A case study example from Australia. Journal of Coastal Research 111-116. 10.2112/SI65-020.1.

Riemann, B., Carstensen, J., Dahl, K., Fossing, H., Hansen, J.W., Jakobsen, H.H., Josefson, A.B., Krause-Jensen, D., Markager, S., Stæhr, P.A., Timmermann, K., Windolf, J., Andersen, J.H., 2016. Recovery of Danish Coastal Ecosystems After Reductions in Nutrient Loading: A Holistic Ecosystem Approach. Estuaries and Coasts 39(1) 82-97. 10.1007/s12237-015-9980-0.

Shen, J., Haas, L., 2004. Calculating age and residence time in the tidal York River using three-dimensional model experiments. Estuarine, Coastal and Shelf Science 61(3) 449-461. 10.1016/j.ecss.2004.06.010.

Shen, J., Lin, J., 2006. Modeling study of the influences of tide and stratification on age of water in the tidal James River. Estuarine, Coastal and Shelf Science 68(1) 101-112. 10.1016/j.ecss.2006.01.014.

Shen, J., Wang, H.V., 2007. Determining the age of water and long-term transport timescale of the Chesapeake Bay. Estuarine, Coastal and Shelf Science 74(4) 585-598. 10.1016/j.ecss.2007.05.017.

Testa, J.M., Kemp, W.M., 2012. Hypoxia-induced shifts in nitrogen and phosphorus cycling in Chesapeake Bay. Limnology and Oceanography 57(3) 835-850. 10.4319/lo.2012.57.3.0835.

Testa, J.M., Murphy, R.R., Brady, D.C., Kemp, W.M., 2018. Nutrient- and Climate-Induced Shifts in the Phenology of Linked Biogeochemical Cycles in a Temperate Estuary. Frontiers in Marine Science 510.3389/fmars.2018.00114.

USEPA, 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment. Annapolis. Annapolis, MD. https://www.epa.gov/chesapeake-bay-tmdl/chesapeake-bay-tmdl-document.

Wood, S., 2003. Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65(1) 95-114. 10.1111/1467-9868.00374.

Wood, S., 2006. Generalized Additive Models: An Introduction with R, 1 ed. Chapman and Hall/CRC, Boca Raton, FL.

Wood, S., 2013. On p-values for smooth components of an extended generalized additive model. Biometrika 100(1) 221-228. 10.1093/biomet/ass048.

Wood, S., 2018. mgcv (https://CRAN.R-project.org/package=mgcv), 1.8-23 ed.

A Generalized Additive Model approach to evaluating water quality: Chesapeake Bay Case Study

Rebecca R. Murphy[a], Elgin Perry[b], Jon Harcum[c], and Jennifer Keisman[d]

Highlights:

- A GAM approach was developed for annual estuarine water quality analysis.
- Fresh water flow impacts are incorporated and adjusted for in the model structure.
- Method changes are accounted for using an intervention parameter.
- Insights generated from this case study are informing management efforts for Chesapeake Bay.