# Comparison of weighted regression and additive models for trend evaluation of water quality in tidal waters*

**Marcus W. Beck, Rebecca R. Murphy[1]**

[1]Post-doctorate researcher (Beck), USEPA National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561, Estuarine Data Analyst (Murphy), UMCES at Chesapeake Bay Program, 410 Severn Avenue, Suite 112, Annapolis, MD 21403 (Email/Beck: beck.marcus@epa.gov)

*Version Date: Fri Mar 25 18:07:44 2016 -0500

# ABSTRACT

Two similar statistical approaches, weighted regression on time, discharge, and season (WRTDS) and generalized additive models (GAMs), have recently been used to evaluate long-term trends in chlorophyll *a* (chl-*a*) in estuarine systems. We evaluated WRTDS and GAMs using thirty years of data for a discrete time series of chl-*a* in the Patuxent River Estuary, a well-studied tributary to Chesapeake Bay. Each model was evaluated based on predictive performance against the observed data and ability to reproduce flow-normalized trends with simulated data. For all examples, prediction errors and average between-model differences were small despite differences in computational requirements for each approach. Flow-normalized trends from each model revealed distinct differences in temporal variation in chl-*a* from the upper to lower estuary. Mainstem influences of the Chesapeake Bay were apparent with both models predicting a roughly 65% increase in chl-*a* over time in the lower estuary, whereas flow-normalized predictions for the upper estuary showed a more dynamic pattern, with a nearly 100% increase in chl-*a* in the last 10 years. Comparison of flow-normalized trends estimated from observed data suggested that GAMs were less sensitive to periods with sparse observations, although both models had comparable abilities to remove flow effects from simulated time series of chl-*a*. This study provides valuable guidance for using statistical models in trend analysis, with particular relevance for computational requirements, desired products, and future data needs.


KEY TERMS: chlorophyll, estuaries, additive models, nutrients, Patuxent River Estuary, statistics, time series analysis, weighted regression

# INTRODUCTION

The interpretation of enviromental trends can have widespread implications for the management of natural resources and can facilitate an understanding of ecological factors that mediate system dynamics. An accurate interpretation of trends can depend on the chosen method of analysis, and more importantly, its ability to consider effects of multiple drivers on response endpoints that may be particular to the system of interest. The need to interpret potential impacts of nutrient pollution has been a priority issue for managing aquatic resources (Nixon 1995), particularly for estuaries that serve as focal points of human activities and receiving bodies for upstream hydrologic networks (Paerl *et al.* 2014). Common assessment endpoints for eutrophication in estuaries have included seagrass growth patterns (Steward and Green 2007), frequency and magnitude of oxygen depletion in bottom waters (Paerl 2006), and trophic network connectivity (Powers *et al.* 2005). Additionally, chlorophyll *a* (chl-*a*) concentration provides a measure of the release of phytoplankon communities from nutrient limitation with increasing eutrophication. Chlorophyll time series have been collected for decades in tidal systems (e.g., Tampa Bay, TBEP (Tampa Bay Estuary Program) (2011); Chesapeake Bay, Harding (1994); datasets cited in Monbet (1992), Cloern and Jassby (2010)), although the interpration of trends in observed data has been problematic given the inherent variability of time series data. Identifying the response of chl-*a* to different drivers, such as management actions or increased pollutant loads, can be confounded by natural variation from freshwater inflows (Borsuk *et al.* 2004) or tidal exchange with oceanic outflows (Monbet 1992). Seasonal and spatial variability of chl-*a* dynamics (see Cloern (1996)) can further complicate trend evaluation, such that relatively simple analysis methods may insufficiently describe variation in long-term datasets (Hirsch 2014). More rigorous quantitative tools are needed to create an unambiguous characterization of chl-*a* response independent of variation from confounding variables.

Recent applications of statistical methods to describe water quality dynamics have shown promise in estuaries, specifically weighted regression on time, discharge, and season (WRTDS) and generalized additive modelss (GAMs). The WRTDS method was initially developed to describe water quality trends in rivers (Hirsch *et al.* 2010, Hirsch and De Cicco 2014) and has recently been adapted to describe chl-*a* trends in tidal waters (Beck and Hagy III 2015). A defining characteristic of WRTDS is a weighting scheme that fits a continuous set of parameters to the time series by considering the influence of location in the record and flow values relative to the period of interest. The WRTDS model has been used to model pollutant delivery from tributary sources to Chesapeake Bay (Hirsch *et al.* 2010, Moyer *et al.* 2012, Zhang *et al.* 2013), Lake Champlain (Medalie *et al.* 2012), the Mississippi River (Sprague *et al.* 2011), and is now being used operationally at the US Geological Survey (USGS) to produce nutrient load and concentration trend results annually for tributaries of the Chesapeake Bay (USGS (US Geological Survey) 2015). A comparison to an alternative regression-based model for evaluating nutrient flux, ESTIMATOR, suggested that WRTDS can produce more accurate trend estimates (Moyer *et al.* 2012). As opposed to WRTDS, GAMs were initially developed in a more general context as a modification to generalized linear models to model a response variable as the sum of smoothing functions of different predictors (Hastie and Tibshirani 1990, Wood 2006a). GAMs have recently been used to describe eutrophication endpoints in tidal waters (Haraguchi *et al.* 2015, Harding *et al.* 2015), and exploratory analyses are underway to use GAMs for long-term trend analysis in Chesapeake Bay tidal waters at the Chesapeake Bay Program. Although the approach was not developed specifically for application to water quality problems, GAMs are particularly appealing because they are less computationally intense and provide more accessible estimates of model uncertainty than WRTDS. Both approaches appear to have similar potential to characterize system dynamics, but the relative merits of each have not been evaluated. Quantitative comparisons that describe the accuracy of empirical descriptions and the desired products could inform the use of each model to describe long-term changes in ecosystem characteristics.

The goal of this study is to provide an empirical description of the relative abilities of WRTDS and GAMs to describe long-term changes in time series of eutrophication response endpoints in tidal waters. Two discrete time series covering 1986-2014 from two stations in the Patuxent River estuary are used as a common dataset for evaluating each model. The Patuxent Estuary is a well-studied tributary of the Chesapeake Bay system that has been monitored for several decades with fixed stations along the longitudinal axis. Two stations were chosen as representative time series that differed in the relative contributions of watershed inputs and influences from the mainstem of the Chesapeake, in addition to known historical events that have impacted water quality in the estuary. The specific objectives of the analysis were to 1) provide a narrative comparison of the statistical foundation of each model, both as a general description and as a means to evaluate water quality time series, 2) use each model to develop an empirical description of water quality changes at each monitoring station given known historical changes in water quality drivers, 3) evaluate each models's ability to reproduce flow-normalized trends as known components of simulated time series, and 4) compare each technique's ability to describe changes, as well as the differences in the information provided by each. We conclude with recommendations on the most appropriate use of each method, with particular attention given to computational requirements, uncertainty assessment, and potential needs for additional monitoring data.

## METHODS

### *Study site and water quality data*

The Patuxent River estuary, Maryland, is a tributary to Chesapeake Bay on the Atlantic coast of the United States (Figure 1). The longitudinal axis extends 65 km landward from the confluence with the mesohaline portion of Chesapeake Bay. Estimated total volume at mean low water is 577 x $10^6$ $m^3$ and a surface area of 126 x $10^6$ $m^2$. The lower estuary (below 45 km from the confluence) has a mean width of 2.2 km and depth of 6 m (Cronin and Pritchard 1975), whereas the upper estuary has a a mean width of 0.4 km and mean depth of 2.5 m (Hagy 1996). The lower estuary is seasonally stratified and the upper estuary is vertically mixed. A two-layer circulation pattern occurs in the lower estuary characterized by an upper seaward-flowing layer and a lower landward-flowing layer. A mixed diurnal tide dominates with mean range varying from 0.8 m in the upper estuary to 0.4 m near the mouth (Boicourt and Sanford 1998). The estuary drains a 2300 $km^2$ watershed that is 49% forest, 28% grassland, 12% developed, and 10% cropland (Jordan *et al.* 2003). The USGS stream gage on the Patuxent River at Bowie, Maryland measures discharge from 39% of the watershed. Daily mean discharge from 1985 to 2014 was 11.0 $m^3$ $s^{-1}$, with abnormally high years occuring in 1996 (annual mean 20.0 $m^3$ $s^{-1}$) and 2003 (annual mean 22.5 $m^3$ $s^{-1}$).

The Chesapeake Bay Program and Maryland Department of Natural Resources (MDDNR) maintain a continuous monitoring network for the Patuxent at multiple fixed stations that cover the salinity gradient from estuarine to tidal fresh (http://www.chesapeakebay.net/, Figure 1 and Table 1). Water quality samples have been collected by MDDNR since 1985 at monthly or bimonthly intervals and include salinity, temperature, chl-*a*, dissolved oxygen, and additional dissolved or particulate nutrients and organic carbon. Seasonal variation in chl-*a* is observed across the stations with spring and summer blooms occuring in the upper, oligohaline section, whereas chl-*a* is generally higher in the lower estuary during winter months (Figure 2). Chlorophyll concentrations are generally lowest for all stations in late fall and early winter. Periods of low flow are associated with higher chl-*a* concentrations in the upper estuary, whereas the opposite is observed for high flow. Stations TF1.6 and LE1.2 were chosen as representative time series from different salinity regions to evaluate the water quality models. Observations at each station capture a longitudinal gradient of watershed influences at TF1.6 to mainstem influences from the Chesapeake Bay at LE1.2. Long-term changes in chl-*a* have also been related to historical reductions in nutrient inputs following a statewide ban on phosphorus-based detergents in 1984 and wastewater treatment improvements in the early 1990s that reduced point sources of nitrogen (Lung and Bai 2003, Testa *et al.* 2008). Therefore, the chosen stations provide unique datasets to evaluate the predictive and flow-normalization abilities of each model given the differing contributions of landward and seaward influences on water quality.

Thirty years of chl-*a* and salinity data from 1986 to 2014 were obtained for stations TF1.6 ($n = 522$) and LE1.2 ($n = 530$, Chesapeake Bay Program data hub, Accessed March 23, 2015, http://www.chesapeakebay.net/data). All data were vertically integrated throughout the water column for each date to create a representative sample of water quality. The integration averaged all values after interpolating from the surface to the maximum depth. Observations at the most shallow and deepest sampling depth were repeated for zero depth and maximum depths, respectively, to bound the interpolations within the range of the data. Daily flow data were also obtained from the USGS stream gage station at Bowie, Maryland and merged with the nearest date in the chorophyll and salinity time series. Initial analyses suggested that a moving-window average of discharge for the preceding five days provided a better fit to the chl-*a* data at TF1.6, whereas the salinity record was used as a tracer of discharge at LE1.2. Both chl-*a* and discharge data were log-transformed. Censored data were not present in any of the datasets. Initial quality assurance checks for all monitoring data were conducted following standard protocols adopted by the Chesapeake Bay Program.

*Model descriptions*

**Weighted Regression on Time, Discharge, and Season.**

The WRTDS method relates a response variable, typically a nutrient concentration, to discharge and time to evaluate long-term trends (Hirsch *et al.* 2010, Hirsch and De Cicco 2014). Recent adaptation of WRTDS to tidal waters relates chl-*a* concentration to salinity and time (Beck and Hagy III 2015), where salinity is a tracer of freshwater inputs or tidal changes (R package link in **Appendex S1**). The functional form of the model is a simple regression that relates the natural log of chl-*a* ($Chl$) to decimal time ($T$) and salinity ($Sal$) on a sinuisoidal annual time scale (i.e., cyclical variation by year).

$$\ln{(Chl)} = \beta_0 + \beta_1 T + \beta_2 Sal + \beta_3 \sin{(2\pi T)} + \beta_4 \cos{(2\pi t)} + \epsilon \tag{1}$$

The tidal adaptation of WRTDS uses quantile regression models (Cade and Noon 2003) to characterize trends in different conditional distributions of chl-*a*, e.g., the median or 90th percentile. For comparison to GAMs, a version of WRTDS created by the authors similar the original model in Hirsch *et al.* (2010) was used to characterize the conditional mean of the response (see **Appendix S1**). Mean models require an estimation of the back-transformation bias parameter for response variables in log-space (Hirsch *et al.* 2010). Although back-transformation is developed for WRTDS, a similar approach has not yet been implemented for GAMs. For simplicity and ease of comparison, all units for chl-*a* are reported in log-space unless otherwise noted.

The WRTDS approach obtains fitted values of the response variable by estimating a unique regression model at each point in the time series. Each model is weighted with a three-dimensional window by month, year, and salinity (or flow) such that a unique set of regression parameters for each observation is obtained. For example, a weighted regression centered on a single observation weights other observations in the same year, month, and similar salinity with higher importance, whereas observations for different months, years, or salinities receive lower importance. This weighting approach allows estimation of regression parameters that vary in relation to observed conditions throughout the period of record (Hirsch *et al.* 2010). Optimal window widths can be identified using cross-validation, described below, that evaluates the ability of the model to generalize results with novel datasets.

Predicted values are based on an interpolation matrix from the unique regressions at each time step. A sequence of salinity or flow values based on the minimum and maximum values for the data are used to predict chl-*a* using the observed month and year based on the parameters fit to the observation. Model predictions are based on a bilinear interpolation from the grid using the salinity (flow) and date values closest to observed. Salinity- or flow-normalized values are also obtained from the prediction grid that allow an interpretation of chl-*a* trend that is independent of variation related to freshwater inputs. Normalized predictions are obtained for each observation by collecting the sample of observed salinity or flow values that occur for the same month throughout all years in the dataset. These values are assumed to be equally likely to occur across the time series at that particular month. A normalized value for each point in the time series is the average of the predicted values from each specific model based on the salinity or flow values that are expected to occur for each month.

**Generalized Additive Models.**

A GAM is a statistical model that allows for a linear predictor to be represented as the sum of multiple smooth functions of covariates (Hastie and Tibshirani 1990). In this application, GAMs were constructed with the same explanatory variables as the WRTDS approach: log of chl-*a* was modeled as a function of decimal time, salinity or flow, and day of year (i.e., to capture the annual cycle). The relationships between log-chl-*a* and the covariates were modeled with thin plate regression splines (Wood 2006a) as the smooth functions using the 'mgcv' package in R. To allow for interaction between the model covariates (e.g., seasonal differences in the long-term chl-*a* pattern), a tensor product basis between all three covariates was constructed. The tensor product basis allows for the smooth construct to be a function of any number of covariates, without an isotropy constraint (Wood 2006b). The GAM implementation in 'mgcv' does not require the selection of knots for a spline basis, but instead a reasonable upper limit on the flexibility of the function is set, and a 'wiggliness' penalty is added to create a penalized regression spline structure. The balance between model fit and smoothness is achieved by selecting a smoothness parameter that minimizes the generalized cross-validation score (Wood 2006a).

Predictions with GAMs are straightforward to obtain after the model parameters are selected, and can be obtained along with standard errors which are based on the Bayesian posterior covariance matrix (Wood 2006a). For this comparison, salinity- or flow-normalized GAM predictions were obtained in a manner consistent with that used for WRTDS. The observed salinity or flow values were compiled that occurred in the same month throughout all years in the dataset. These values were assumed to be equally likely to occur at that particular month. A normalized GAM estimate at each date in the record was computed as the average of the predictions obtained using all of the flow or salinity values for that month of the year throughout the record.

**Selection of model parameters.**

The selection of optimal model parameters is a challenge that represents a tradeoff between model precision and ability to generalize to novel datasets. Weighted regression requires identifying optimal half-window widths, whereas the GAM approach used here requires identifying an optimal value for a smoothing parameter that weights the wiggliness of the function against model fit (Wood 2006a). Overfitting a model with excessively small window widths or smoothing parameters will minimize prediction error but prevent extrapolation of results to different datasets. Similarly, underfitting a model with large window widths or smoothing parameters will reduce precision but will improve the ability to generalize results to different datasets. From a statistical perspective, the optimal model parameters provide a balance between over- and under-fitting. Both models use a form of cross-validation to identify model parameters that maximize the precision of model predictions with novel data.

The basic premise of cross-validation is to identify the optimal set of model parameters that minimize prediction error on data not used to develop the model. For the GAM approach, generalized cross-validation is used to obtain the optimal smoothing parameter in an iterative process with penalized likelihood maximization to solve for model coefficients. The effective degrees of freedom of the resulting model vary with the smoothing parameter (Wood 2006a). Similarly, the tidal adaptation of WRTDS used k-fold cross-validation to identify the optimal half-window widths (Efron and Tibshirani 1993, Arlot and Celisse 2010). For a given set of half-window widths, the dataset was separated into ten disjoint sets (k = 10), such that ten models were evaluated for every combination of k - 1 training and remaining test datasets. That is, the training dataset for each fold was all k - 1 folds and the test dataset was the remaining fold, repeated k times. The average prediction error of the test datasets across k folds provided an indication of model performance for the given combination of half-window widths. The optimum window widths were those that provided minimum errors on the test data. Evaluating multiple combinations of window-widths can be computationally intensive. An optimization function was implemented in R (RDCT (R Development Core Team) 2015) to more efficiently evaluate model parameters using a search algorithm. Window widths were searched using the limited-memory modification of the BFGS quasi-Newton method that imposes upper and lower bounds for each parameter (Byrd *et al.* 1995, Nocedal and Wright 2006). The chosen parameters were based on a selected convergence tolerance for the error minimization of the search algorithm that balanced computation time with precision. Specifically, the algorithm converged when the reduction in the minimization function for a given change in parameters was within an acceptable tolerance without excessive search time.

*Comparison of modelled trends*

Separate WRTDS and GAMs were created using the above methods for the chl-*a* time series at TF1.6 and LE1.2. For each model and station, a predicted and flow-normalized (hereafter flow-normalized refers to flow at TF1.6 and salinity at LE1.2) time series was obtained for comparison. The results were compared with summary statistics that evaluated both the predictive performance to describe observed chl-*a* and direct comparisons between the models. Emphasis was on agreement between observed and predicted values, rather than uncertainty associated with parameter estimates or model results. As of writing, methods for estimating confidence intervals of WRTDS have been developed for the original model (Hirsch *et al.* 2015), but have not been fully developed for application to WRTDS in tidal waters. In addition to simple visual evaluation of trends over time, summary statistics used to compare model predictions to observed chl-*a* included root mean square error (RMSE) and average differences. For all comparisons, RMSE comparing each model's predictions to observed chl-*a* (fit) was defined as:

$$RMSE_{fit} = \sqrt{\frac{\sum_{i=1}^{n}\left(Chl_i - \widehat{Chl}_i\right)^2}{n}} \tag{2}$$

where $n$ is the number of observations for a given evaluation, $Chl_i$ is the observed value of chl-*a* for observation $i$, and $\widehat{Chl}_i$ is the predicted value of chl-*a* for observation $i$. RMSE values closer to zero represent model predictions closer to observed. Comparisons between models using

RMSE were performed similarly, using the equation:

$$RMSE_{btw} = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{Chl}_{WRTDS,i} - \widehat{Chl}_{GAM,i}\right)^2}{n}} \tag{3}$$

where the estimated chl-*a* values for each model, $\widehat{Chl}_{i,WRTDS}$ and $\widehat{Chl}_{i,GAM}$, are compared directly. Similarly, average differences (or bias) of predictions between models as a percentage was defined as:

$$\text{Average difference} = \left(\frac{\sum_{i=1}^{n}\widehat{Chl}_{WRTDS,i} - \sum_{i=1}^{n}\widehat{Chl}_{GAM,i}}{\sum_{i=1}^{n}\widehat{Chl}_{GAM,i}}\right) * 100 \tag{4}$$

Positive values indicate that WRTDS provided higher predictions than GAMs on average, whereas the opposite is true for negative values (Moyer *et al.* 2012). Results between models were also evaluated using regressions comparingWRTDS (as the response) and GAM (as the predictor). The regressions were compared to a null model having an intercept of zero and slope of one. Deviation of either the intercept or slope of the regressions from the null model provided evidence of systematic differences between the models. An intercept significantly different from zero was interpreted as an overall difference between the predictions, whereas a slope different from one was interpreted as a difference that varies with relative magnitude of the predictions. Although the signs of the slope and intercept estimates for the comparisons depended on which model was used as the predictor, we were primarily concerned with magnitude of the parameter estimates in the regression comparisons as evidence of systematic differences between each model. The statistical comparisons described above were conducted for the entire time series at each station to evaluate overall performance. Different time periods were also evaluated to identify potential temporal variation in results, which included a comparison of results by annual and seasonal aggregations and periods with different levels of flow using the discharge record at Bowie, Maryland. Annual and seasonal aggregations shown in Figure 2 were evaluated between the models, in addition to evaluating the models at different levels of flow defined by the quartile distributions (min–25%, 25%–median, median–75%, and 75%–max).

Flow-normalized time series were compared similarly but only between models because the true flow-independent component of the observed data is not known and can only be empirically estimated. As described below, an evaluation of flow-normalized data for each model was accomplished using simulated datasets with known components that were independent of discharge. However, a simple comparison of flow-normalized trends by different time periods summarized long-term patterns in the Patuxent River estuary. These comparisons evaluated percent changes of flow-normalized estimates at the beginning and end of each time period. Percent changes within each period were based on annual mean estimates for the first and last three years of flow-normalized chl-*a* estimates, excluding the annual aggregations that had limited annual mean data (i.e., seven years per period). For example, percent change for the January-February-March (JFM) seasonal period compared an average of JFM annual means for 1986 through 1988 to an average of JFM annual means for 2012 through 2014. This approach was used to reduce the influence of abnormal years or missing data on trend estimates.

*Comparison of flow-normalized trends*

The relative abilities of each model to characterize flow-normalized trends in chl-*a* were evaluated using simulated datasets with known components. This approach was used because the flow-independent component of chl-*a* can only be empirically estimated from raw data. Accordingly, the ability of each model to isolate the flow-normalized trend cannot be evaluated with reasonable certainty unless the true signal is known. Following similar concepts in Beck *et al.* (2015), Simulated time series of observed chlrophyll ($Chl_{obs}$) were created as additive components related to flow ($Chl_{flo}$) and a flow-independent biological component of chl-*a* ($Chl_{bio}$):

$$Chl_{obs} = Chl_{flo} + Chl_{bio} \tag{5}$$

A distinction between $Chl_{flo}$ and $Chl_{bio}$ is that the former describes variation in the observed time series with changes in discharge (e.g., concentration dilution with increased flow) and the latter describes a true, desired measure of chl-*a* in the water column that is directly linked to productivity. The biological component of chl-*a* is comparable to an observation in a system that is not affected by flow and is the time series that is estimated by flow-normalization with WRTDS and GAMs.

The simulated time series was created using methods similar to those in Hirsch *et al.* (2015)

and was based on a stochastic model derived from actual flow and water quality measurements to ensure the statistical properties were comparable to existing datasets. This approach allowed us to evaluate GAMs and WRTDS under different sampling regimes (e.g., monthly rather than daily), while ensuring the simulated datasets had statistical properties that were consistent with known time series. Daily flow observations from the USGS stream gage station 01594440 near Bowie, Maryland (38°57′21.3″N, 76°41′37.3″W) were obtained from 1985 to 2014. Daily chl-$a$ records were estimated from fluorescence values from the Jug Bay station (38°46′50.6″N, 76°42′29.1″W) of the Chesapeake Bay Maryland National Estuarine Research Reserve in the upper Patuxent.

Four time series were estimated or simulated from the actual datasets to create the complete, simulated time series: 1) estimated discharge as a stationary seasonal component ($\widehat{Q}_{seas}$), 2) simulated error structure from the residuals of the seasonal discharge model ($\varepsilon_{Q,\,sim}$), 3) estimated chl-$a$ independent of discharge as a stationary seasonal component ($\widehat{Chl}_{seas}$), and 4) simulated error structure from the residuals of the seasonal chl-$a$ model ($\varepsilon_{Chl,\,sim}$). Unless otherwise noted, chl-$a$ and discharge are in ln-transformed units. Each of the four components was used to simulate the components in eq. (5):

$$Chl_{flo} = I\left(\widehat{Q}_{seas} + \sigma_{\varepsilon_{Q,\,sim}} \cdot \varepsilon_{Q,\,sim}\right) \tag{6}$$

$$Chl_{bio} = \widehat{Chl}_{seas} + \sigma_{\widehat{Chl}_{seas}} \cdot \varepsilon_{Chl,\,sim} \tag{7}$$

Standard errors for the residuals of the discharge time series, $\sigma_{\varepsilon_{Q,\,sim}}$, and the seasonal chl-$a$ component, $\sigma_{\widehat{Chl}_{seas}}$, are estimated empirically from the simulated data. The estimated flow time series within the parentheses, $\widehat{Q}_{seas} + \sigma_{\varepsilon_{Q,\,sim}} \cdot \varepsilon_{Q,\,sim}$, is floored at zero to simulate an additive effect of increasing flow on $Chl_{obs}$. Although the actual relationship of water quality measurements with flow is more complex, we assumed that a simple addition was sufficient for the simulations where the primary objective was to create an empirical and linear link between flow and chl-$a$. The vector $I$ (where $0 < I < 1$) is a weighting and unit-conversion vector that translates the terms enclosed in parentheses from flow to chl-$a$ concentration units and allows for the effect of flow to be defined as time-varying. For example, a flow effect that changes from non-existent to positive throughout the period of observation can be simulated by creating a vector ranging from zero to one. For the simulated $Chl_{bio}$ time series, the seasonal and error components were characterized using the daily time series at Jug Bay that likely included an effect of flow in the observed data. For the simulated models, we assumed that the actual flow effect was part of the seasonal component to obtain an accurate estimate of the error component that was independent of both flow and season. In other words, the seasonal component of chl-*a* was modelled with a discharge component to remove any variability related to flow in the residuals. Methods for estimating each of the components in eqs. (6) and (7) are described in detail in **Appendix S2**.

Three time series with monthly sampling frequencies and varying contributions of the flow component ($Chl_{flo}$ in eqs. (5) and (6)) were created from daily simulated time series of $Chl_{obs}$ (**Appendix S2**). One day in each month for each year (e.g., January 5, 2010, February 19, 2010, ..., January 28, 2011, Febuary 1, 2011, etc.) was randomly sampled and used as an approximate monthly time step for each time series. Varying effects of the flow component on observed chl-*a* were creating by multiplying $Chl_{flo}$ by different indicator vectors ($I$ in eq. (6)). The contribution of the flow component varied from non-existent (vector of zeroes), constant (vector of ones), and steadily increasing (continuous vector from zero to one). This created three monthly time series that were used to evaluate each model that were analogous to no influence, constant, and changing influence of the flow component over time (**Appendix S2**). Results were evaluated by first comparing the predicted ($\widehat{Chl}_{obs}$) and observed ($Chl_{obs}$) chloropyll values for each simulation, followed by comparing the flow-normalized results ($\widehat{Chl}_{bio}$) from each model to the original biological chl-*a* ($Chl_{bio}$) component of each simulated time series (eqs. (5) and (7)). The former comparison provided information on relative fit to validate the simulated data, whereas the latter comparison to evaluate flow-normalization was the primary focus of the analysis.

## RESULTS

### *Observed trends and relative fit*

The optimal half-window widths and degrees of freedom for smoothing varied at each station for WRTDS and GAMs, respectively. For WRTDS, optimal half-window widths identified by generalized cross-validation were 0.25 as a proportion of each year, 13.59 years, and 0.25 as a proportion of the total range of salinity for LE1.2, and 0.25 of each year, 6.28 years, and 0.50 of flow at TF1.6. For both stations, the optimization method selected relatively wide windows for the year weights while minimizing the seasonal (annual proportion) and flow component. For

GAMs, the optimal smoothing procedure resulted in a smoother model at LE1.2 than TF1.6 with effective degrees of freedom of 35.5 and 71.4, respectively. The tensor product smooth contruct does not split apart the effective degrees of freedom among the three interacting parameters.

The predicted chl-*a* from each model generally followed patterns in observed chl-*a* from 1986 to 2014 (Figure 3). At LE1.2, each model showed seasonal minima typically in November, whereas maximum chl-*a* was observed in a spring bloom, typically March or April (Figure 4). A secondary, smaller seasonal peak was also observed in late summer from bottom-layer regeneration and upward nutrient transport (Testa *et al.* 2008). Seasonal variation at TF1.6 was noticeably different with an initial peak typically observed in May and a larger dominant bloom occuring in September or October (Figure 4). Elevated chl-*a* concentrations were also more prolonged than those at LE1.2 with only a slight decrease between the two seasonal blooms. A seasonal minimum was typically observed in December or January, followed by a rapid increase in the following months. Differences in magnitude of the seasonal range were also less prononced at LE1.2 compared to TF1.6, with differences throughout the year approximately 3 $\mu$g/L of chl-*a* (arithmetic) at LE1.2 and 7 $\mu$g/L of chl-*a* at TF1.6. Visual evaluation of seasonal trends suggested each model provided similar results, although WRTDS predictions had slightly better fits at the extreme ends of the distribution of chl-*a* (Figure 3a). Normalized predictions for both models were visually distinct from the non-flow-normalized predictions such that seasonal minima and maxima and extreme predictions were not observed with the normalized values. Overall, both models had predictions that provided a more adequate visual description of the range of chl-*a* at TF1.6 as compared to LE1.2 where observed values lower or higher than the predictions were more common.

Quantitative summaries of model fit by site indicated that performance between sites and models was similar, with RMSE ranging from a minimum of 0.50 at TF1.6 for GAM predictions and a maximum of 0.52 at TF1.6 for WRTDS predictions (Table 2). Overall, both models performed similarly, although WRTDS had slightly better performance at LE1.2 and GAMs had slightly better performance at TF1.6 (Table 2). Fit by different time periods generally showed agreement between methods during periods when performance was relatively high or low. For LE1.2, both models had the worst fit during the 2001-2007 annual period (RMSE 0.61 for GAMs, RMSE 0.60 for WRTDS), the April-May-June (AMJ) seasonal periods (0.64 for GAMs, 0.64 for

WRTDS), and periods of high flow (0.64 for GAMs, 0.63 for WRTDS). For TF1.6, models had the worst fit during the 1994-2000 annual period (0.55 for GAMs, 0.58 for WRTDS) and the AMJ seasonal period (0.54 for GAMs, 0.58 for WRTDS). Errors between models were comparable for all flow periods at TF1.6, with the exception of lower errors during low flow (0.45 for GAMs, 0.46 for WRTDS). In general, model performance was partially linked to flow such that fit was improved during periods of low flow, including seasonal or annual periods of low flow. For example, both models at both sites had the best fit during the July-August-September (JAS) period when seasonal flow was minimized (Table 2 and Figure 2).

Results as annual aggregations suggested that chl-*a* patterns between years have not been constant and are considerably different between sites (Figure 3b). Both models showed a gradual and consistent increase in chl-*a* at LE1.2, with values increasing by approximately 1.5 $\mu$g/L (arithmetic) from 1986 to 2014. Predictions at TF1.6 did not show a similar increase from the beginning to the end of the time series, although a dramatic decrease from approximately 12 $\mu$g/L to 6 $\mu$g/L from 2000 to 2006 was observed. By 2014, chl-*a* returned to values similar to those prior to the initial decrease. Flow-normalized predictions that were annually averaged at each site allowed an interpretation of trends that were independent of variation in discharge or salinity (Tables 3 and 4). Overall percent change of ln-transformed chl-*a* concentration from the beginning to the end of the time series at LE1.2 was approximately 20% (Table 3). A slight decrease in chl-*a* at TF1.6 was observed from 1986 to 2014 (Table 4). Changes by annual, seasonal, and flow time periods at LE1.2 were comparable for each time period and model type, although some differences were observed. For example, both models had maximum increases in chl-*a* for the different flow periods for high levels of flow at LE1.2 (25.1% for GAMs, 22.3% for WRTDS). Trends by different time periods were more apparent for TF1.6, particularly as an overall decrease in chl-*a* for both models during the 2001–2007 period and an overall increase during the 2008–2014 period (Table 4). Seasonal changes were especially pronounced during the January-February-March (JFM) and October-November-December (OND) periods where both models showed an increase and decrease, respectively, with differences between the two (JFM period, 9% for GAMs, 32.7% for WRTDS; OND period, $-18.2\%$ for GAMs, $-17.5\%$ for WRTDS). Percent changes by flow quantile were also observed at TF1.6, with the most noticeable difference from LE1.2 being a decrease in chl-*a* during both high and low flow (both models) and relatively larger increases in chl-*a* during moderate flow.

*Comparison of model predictions*

The following describes direct comparisons of model results, whereas the previous section emphasized results relative to trends over time and fit to the observed data. Accordingly, direct comparisons were meant to identify instances when models results were systematically different from each other. Table 5 compares average differences and RMSE of results between each model for the complete time series and different subsets by annual, seasonal, and flow periods. Overall, differences between the models were minor with most percent differences not exceeding 1% and no RMSE values exceeding 0.15. Model differences between different time periods were not apparent for either station, although the largest average difference was observed at TF1.6 for the 2008–2014 time period (3.1%, WRTDS greater than GAMs).

Regressions comparing model results (WRTDS as response, GAMs as predictor) provided additional information about overall differences (significantly different intercept) and differences between the models that varied for different values (significantly different slope) (Table 6, Figure 5). Significant differences were observed for the entire time series such that estimated intercepts and slopes were different from zero and one, respectively, for both stations and model predictions (observed and flow-normalized), excluding intercepts and slopes for the flow-normalized predictions at TF1.6 ($\beta_{0,\,norm}$ and $\beta_{1,\,norm}$). Differences were also observed for the time period subsets, with the most obvious differences occuring for the seasonal aggregations. For example, all comparisons between the models for both sites and model predictions had intercept estimates significantly greater than zero and slope estimates significantly less than one for the AMJ period (Table 6). For almost all significant differences, intercept estimates were greater than zero and slope estimates were less than one. Visual comparisons of results in Figure 5 confirm those in Table 6, particularly differences in the seasonal aggregations.

*Changes in chl-a response to flow over time*

Both models described chl-*a* response with sufficient parameterization of input variables to evaluate variation with flow changes over time. As in Beck and Hagy III (2015), changes in the relationship of chl-*a* to flow can be evaluated by predicting observed chl-*a* across the range of observed flow (or salinity) values for each year in the time series. Visual information obtained from these plots are useful to identify periods of time when chl-*a* was or was not related to changes in flow and may also lead to the development of hypotheses regarding changes in drivers of water quality, e.g., temporal shifts in point-sources to non-point sources of pollution (Hirsch *et al.* 2010, Beck and Hagy III 2015). The only difference between the models in creating the plots is that the three-dimensional prediction grid of chl-*a*, flow, and time created during model fitting is used for WRTDS, whereas the plots for GAMs are based on post-hoc model predictions with covariates defined to vary over a regular grid.

Figure 6 shows the estimated changes from each model in predicted chl-*a* for salinity (LE1.2) or flow (TF1.6) across all years in the study period. The plots are also separated by months of interest to isolate effects of seasonal variation. Visual assessment of the plots suggests that the relationships were dynamic across the study years and varied considerably between LE1.2 and TF1.6. For example, the October plots show decreasing sensitivity of chl-*a* with increasing flow (decreasing salinity) at LE1.2 from early to late in the time series (i.e., a strong, positive relationship changing to a weak relationship over time). Conversely, the opposite trend is observed at TF1.6 in October such that a weak relationship with flow is observed early in the time series and a strong, negative relationship is observed later in the time series, although overall chl-*a* has decreased over time. Additionally, both models provided similar indications of the changes over time, regardless of site or time of year. However, some differences between the models were observed. For example, WRTDS results for January at LE1.2 provided a wider range, or potentially less stable fit of chl-*a* to salinity changes in the earlier years.

*Flow-normalization with simulated data*

WRTDS and GAMs were fit to each of the three simulated datasets, creating six models to evaluate the general fit of observed to predicted ($Chl_{obs} \sim \widehat{Chl}_{obs}$) and biological to flow-normalized chl-*a* ($Chl_{bio} \sim \widehat{Chl}_{bio}$). Models were fit using identical methods as those for the Patuxent time series such that an optimal window width combination for WRTDS and optimal degrees of freedom for smoothing parameters with GAMs were identified. Figure 7 is similar to Figure 6 and shows an example of the changing relationships between chl-*a* and flow across the simulated time series using the results from three optimal WRTDS models. The plots show the varying effects of flow in each simulated dataset over time (no effect, constant, increasing) and that the models appropriately characterized the relationships. For example, a changing response of chl-*a* to salinity is apparent in the third panel of Figure 7 such that no response is observed early in the time series followed by an increase in the response of chl-$a$ to flow later in the time series. Similar patterns were observed for the GAMs.

Comparisons of fit to the simulated time series showed no systematic differences between the models. Overall, WRTDS results had lower RMSE than GAMs for all comparisons except one ($Chl_{obs} \sim \widehat{Chl}_{obs}$, constant flow simulation), although differences in performance were minor (Table 7). Although both models provided similar performance for individual simulations, differences between the simulations were observed. The different effects of flow had a negative effect on the ability of each model to remove the flow component. Comparisons of $Chl_{bio}$ with $\widehat{Chl}_{bio}$ showed the lowest RMSE with no flow effect and the highest with a constant flow effect (Table 7). Different flow effects did not have an influence on the relationship between predicted ($\widehat{Chl}_{obs}$) and observed ($Chl_{obs}$) chl-$a$ such that RMSE for all models and simulations were similar and lower than those comparing the flow-normalized results. Overall, changing the flow component primarily affected the ability of each model to reproduce the flow-normalized component ($\widehat{Chl}_{bio}$) with relatively minor differences between the models.

## DISCUSSION

### *Model comparisons and considerations*

Both WRTDS and GAMs have similar objectives of describing trends from long-term monitoring datasets, whereas more specific applications for each model (e.g., hypothesis testing, assessment of management actions, etc.) will be defined by future needs or research goals. Accordingly, our comparison methods were chosen based on the exploratory needs of the analysis and by considering that each technique provides a potentially novel approach to trend assessment in future applications. The variety of methods for comparing models can provide different information depending on the desired application. An improvement in predictive performance using RMSE, for example, may suggest one model is more advantageous over another if the goal is to reproduce trends, whereas this information may be much less relevant for hypothesis testing. Inferior performance for one metric does not necessarily invalidate an analysis method for all potential applications. An interpretation of the results should consider that the analysis provides an overview with several techiques, given that the purpose of each model will be better defined by future applications.

A general conclusion from our results is that both models provide similar information, both in predictive performance and trends over time in the Patuxent. Comparisons using RMSE provided strikingly similar indications of performance for each model, although some instances were observed where one model had lower errors. Large differences were not observed and we emphasize that any potential improvement in performance at the scale shown in Table 2 is trivial. Prediction errors for either model could easily be improved by slight adjustments of the model parameters. This highlights a potential risk of using prediction error as a performance metric because the values are sensitive to tuning parameters and the statistical characteristics of training datasets. To address this issue, comparable methods for model development were implemented to ensure valid comparisons. Both WRTDS and GAMs used a form of cross-validation that was meant to identify an optimal parameter space that balances over- and under-fitting by using separate training and test datasets. A more generic benefit of cross-validation is that model development is not biased by analyst intervention as the parameters are chosen with predefined heuristics. This paper presents the first application of a statistical method of selecting optimal window widths for WRTDS. Further work should explore use of these methods to develop robust and unbiased parameters for WRTDS.

The comparisons of predictive performance should also be interpreted relative to the statistical foundations of each model. The smoothing process in GAMs, although mathematically involved, rapidly converges to a solution, whereas the fitting process for WRTDS is much longer because a unique regression is estimated for every point in the time series. From a practical perspective, the comparable error estimates for each model's predictions suggests that GAMs are advantageous because there is no apparent benefit of the added computational time of WRTDS. Temporal changes in the relationship between chl-*a* and flow were also comparable. For example, Figure 6 shows similar information for each model, although different methods were used to characterize chl-*a* variation from salinity or flow over time. Additional insight into trends might a logical expectation with the added computational time required to estimate WRTDS interpolation grids. Conventional modelling techniques that have a predefined parameterization and limited parameter space have been described as 'statistical straightjackets' that mold the data to the model (Hirsch 2014). WRTDS is meant to provide a contrasting approach where the data mold the results. GAMs could be overconstrained by following a less flexible model. However, the results do not provide a compelling contrast between GAMs and WRTDS, despite the alternative statistical foundations.

Similarity in results for WRTDS and GAMs may suggest that relationships between time, season, and flow in the Patuxent were adequately described by the statistical theories of each approach, but generalizations of the merits of each model should be made sparingly until additional assessments with alternative datasets. Site selection of TF1.6 and LE1.2 was meant to capture a gradient of watershed to mainstem influences at each location. The known historical changes from management practices (e.g, wastewater treatment, banning of phosphorus-based detergents) and natural events (e.g., storm events, seagrass recovery) that have affected the Patuxent have also provided a unique context for the time series. Additionally, a natural conclusion from this study is that both models were equally 'good' at trend evaluation, although the possibility that both were equally inadequate should also be considered as a potential explanation. Alternative drivers of chl-*a* response that were not explicitly included in each model could limit explanatory power if time, season, and discharge were not the dominant predictors of production. The observation that models capture more of the extreme values at TF1.6 than at LE1.2 (Figure 3a) suggests this may be the case at LE1.2. For example, Beck and Hagy III (2015) evaluated residual variation of WRTDS models in Tampa Bay, Florida in relation to seagrass growth, El Niño effects, and nitrogen inputs. A similar analysis of additional variables at LE1.2 could reveal insight into factors other than time, season, or flow that influence chl-*a* in the lower estuary. Evaluation of alternative sites with different historical contexts could provide further information to support our general conclusion of comparability between methods.

Although our results generally indicated that comparable information was provided by both models, some instances were observed when different information was provided. For example, significant differences in the regression comparisons between the models (Table 6 and Figure 5) typically had intercept estimates greater than zero and slope estimates less than one. This suggests that WRTDS estimates were, on average, larger than GAMs (intercept $> 0$), whereas GAMs fit a wider range of values compared to WRTDS (slope $< 1$). However, these conclusions should be interpreted with caution given the certainty of the results in the context of the analysis method. More robust approaches to evaluate systematic biases, in addition to alternative datasets, should be used to validate these general conclusions. Generally, important differences between the models would be those that would result in a different conclusion if one model was used instead of the other. Although none of the model differences were large, several differences were observed in the patterns of the flow normalized results (Tables 3 and 4). Most notably, the LE1.2 annual percent change results from GAMs suggested that the increase in ln-chl-*a* has become less steep over time (9.6 to 3.2%), whereas the WRTDS results suggested the increase has become steeper over time (1.75 to 6.07%) (Table 3). The seasonal slopes in Table 3 for LE1.2 also suggested different patterns from the two models. The increase in chl-*a* was the smallest in the summer (JAS) from the GAM results, whereas the WRTDS results suggested that the smallest increase over time was in the winter months (JFM). For TF1.6 (Table 4), differences in the percent changes were also observed, with the JFM change from WRTDS more than three times that suggested by GAMs. These slight differences in patterns showed that the models were not identical on the fine-scale. Although we cannot know which model was more accurate in depicting flow-normalized trends in Patuxent chl-*a*, these differences reveal that, in fact, a multiple models approach could be beneficial when making conclusions on a fine temporal scale.

Finally, initial assessment of Figure 6 suggested that WRTDS provided a more dynamic description of chl-*a* response to changes in flow or salinity for specific locations in the record. For example, chl-*a* response over time to salinity changes during January at LE1.2 shows WRTDS describing greater variation than GAMs, particularly for lower salinity values. Additional investigation suggested that these 'novel' descriptions were related to low sample size for the specific location in the record causing instability in the model predictions. WRTDS descriptions may be unstable at extreme or uncommon locations in the data domain where the number of observations with non-zero weights may be limited. Methods for WRTDS have been developed to address this issue (i.e., automated window width increases with low sample sizes, Hirsch *et al.* 2010), although they were not implemented for the current analysis to simplify direct comparisons between models. Similar problems may be avoided with datasets at smaller time steps (e.g., daily), whereas the nutrient time series represent a more coarse resolution at the bimonthly scale.

*Patuxent trends*

Both models provided a detailed description of water quality changes in the Patuxent River estuary. Several trends were described that deserve additional discussion independent of the model comparisons. Annual trends at TF1.6 showed a substantial decrease in chl-*a* that lasted several years, followed by a gradual increase to concentrations similar to those earlier in the time series. By comparison, annual trends in the lower estuary at LE1.2 showed a consistent, linear increase over time. Seasonal patterns and trends related to different flow periods were also described by the models. Spring blooms were commonly observed in the lower estuary, whereas late summer blooms were observed in the upper estuary. Trends related to different flow periods were less obvious, although large increases in chl-*a* were observed for moderate flow levels. Trends in Figure 6 can facilitate an interpretation of changes at each station related to flow effects over time. For example, annual trends in October suggested that the association between flow (decreasing salinity) and chl-*a* have weakened over time at LE1.2. By contrast, trends at TF1.6 showed an increasingly negative relationship between flow and chl-*a* over time. Both models also showed changes in the shape of the relationship between chl-*a* and discharge. For example, a distinct non-linear relationship between chl-*a* and increasing discharge (decreasing salinity) was observed for January predictions at LE1.2 earlier in the record, whereas the trend became more linear near the end of the record. Identifying differences in chl-*a* response at both different flow levels and different seasons could be a first step to identifying influencing factors. The increase over time at LE1.2 is fairly consistent, except for patterns in October at high salinities. Further investigation to reveal what sources are actually being reduced during that period would be insightful.

The results from either model can be used to hypothesize causal links between water quality changes, flow variation, or additional ecosystem characteristics. Previous studies have linked chl-*a* changes and flow relationships to shifts in sources of nutrient pollution (Hirsch *et al.* 2010, Beck and Hagy III 2015). Similarly, historical changes in the Patuxent are likely related to the banning of phosphorus-based detergents in the mid 1980s and wastewater treatment plant upgrades in the early 1990s (Lung and Bai 2003, Testa *et al.* 2008). An investigation of chl-*a* response to both flow changes and ratios of point-source to non-point sources of nutrients could provide valuable information on system dynamics. Historical changes in flow have also affected water quality in the Patuxent. Flow records for the Patuxent show a drought period from 1999 to 2002 that likely contributed to increases in chl-*a* in the upper estuary and decreases in the lower estuary. By contrast, storm events could be linked to lower chl-*a* from estuarine flushing or shifts in concentration along the longitudinal axis (Hagy *et al.* 2006, Murrell *et al.* 2007). The substantial decline in chl-*a* in the upper estuary in the early 2000s coincides with storm events, including the passage of Hurricane Isabel in 2003. However, low concentrations persisted for several years suggesting additional factors may have had separate or additive effects on chl-*a* response. For example, seagrass growth patterns in the upper estuary have followed a similar but inverse pattern as chl-*a*, with an increase in growth in the late 1990s and early 2000s, followed by a decline in recent years after a peak in coverage in 2005 (J. M. Testa, personal communication). This correlation suggests nutrient sequestration by seagrasses, although definitive links have yet to be shown. Comparison to baywide changes for the larger Chesapeake Bay could provide additional explanations, such as the relationship to long-term trends in seagrass growth patterns, additional nutrients, or phytoplankton (Orth *et al.* 2010, Harding *et al.* 2015).

## CONCLUSIONS

The use of data-driven statistical techniques that leverage the descriptive potential of long-term monitoring datasets continues to be a relevant research focus in aquatic systems. Both WRTDS and GAMs are actively being developed for application to monitoring time series and our analysis represents the first quantitative comparison of WRTDS and GAMs to evaluate trends in tidal waters. For the Patuxent River estuary, both models had surprisingly similar abilties to describe observed and flow-normalized trends in chl-*a*. The relative differences between the models were trivial considering computational requirements of each. Some differences in the descriptive capabilities were observed, such as specific periods of the time series where data limitations may have caused instability in model predictions for WRTDS. Our application to simulated datasets with known flow-independent components of chl-*a* provided further indications of similarities between the two approaches. This analysis was the first to rigrously compare both WRTDS and GAMs and further evaluations with alternative datasets should be made to verify our results herein. Although both models provided similar information, the results from either reveal interesing relationships (e.g., flow, nutrient response over time, Figure 6) that can lead to additional hypotheses or analysis to investigate ecosystem dynamics.

Practical applications of each model should consider alternative characteristics of each technique, in addition to the simple quantitative comparisons described above. The use of WRTDS to describe water quality trends in tidal waters, particularly with monthly or bimonthly time series, is a novel application for which the model was never intended. Hirsch *et al.* (2010) developed the original model for streams and rivers using high-resolution, daily time series where time, discharge, and season are dominant characteristics that influence water quality. Although seasonal and flow effects are important drivers of change in estuaries, other physical or biological characteristics may be equally or more important. For example, the extreme ends of the chl-*a* distribution at LE1.2 were not fit well by either model as compared to TF1.6, which suggests additional predictors besides time, discharge, and season may better describe variation in the lower estuary. As such, recent use of GAMs in tidal waters has followed an alternative paradigm where drivers of change are not necessarily known and the time series may have a larger time step with occasional discontinuous intervals (E. S. Perry, personal communication, Harding *et al.* 2015). Although we have quantitatively compared each method to inform decision-making, choosing a technique should also consider characteristics of the dataset, questions of interest, or specifics of the study system. Each model can also provide different products, which we have not specifically addressed above given constraints on similarly comparing each model. For example, confidence intervals that can facilitate hypothesis-testing are readily available GAMs, whereas similar products are not yet available for tidal adaptation of WRTDS (but see Hirsch *et al.* 2015). Likewise, WRTDS has been applied using a quantile regression approach to characterize trends at the extreme concentration distributions of the data that could have important ecological implications (Beck and Hagy III 2015), but similar functionality has not been implemented with GAMs. Accordingly, the results herein provide a partial description of WRTDS and GAMs that should be considered in a broader context for water quality assessment.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article:

**Appendix S1**: The WRTDStidal R package for implementing the tidal adaptation of WRTDS is available for download at https://github.com/fawda123/WRTDStidal

**Appendix S2**: Additional material describing the simulation of daily discharge and chlorophyll time series is included. First, a model for simulating flow-related chl-*a* (eq. (6)) was estimated from the stream gage data as the additive combination of a stationary seasonal component and serially-correlated errors:

$$Q_{seas} = \beta_0 + \beta_1 \sin(2\pi T) + \beta_2 \cos(2\pi T) \tag{8}$$

$$\varepsilon_Q = Q_{seas} - \widehat{Q}_{seas} \tag{9}$$

A seasonal model of flow was estimated using linear regression for time, $T$, on an annual sinusoidal period (eq. (8)). The residuals from this regression, $\varepsilon_Q$ (eq. (9)), were used to estimate the structure of the error distribution for simulating the stochastic component of flow. The error distribution was characterized using an Autoregressive Moving Average (ARMA) model to identify appropriate $p$ and $q$ coefficients (Hyndman and Khandakar 2008). The parameters were chosen using stepwise estimation for nonseasonal univariate time series that minimized Akaike Information Criterion (AIC). The resulting coefficients were used to generate random errors from a standard normal distribution for the length of the original time series, $\varepsilon_{Q,\,sim}$. These stochastic errors were multiplied by the standard deviation of the residuals in eq. (9) (i.e., $\sigma_{\varepsilon_{Q,\,sim}}$ in eq. (6)) and added to the seasonal component in eq. (8) to create a simulated, daily time series of the flow-component for chl-*a*, $Chl_{flo}$ (eq. (6)).

The chl-*a* time series was created using a similar approach. The first step estimated the stationary seasonal component of the chl-*a* time series by fitting a WRTDS model (Hirsch *et al.* 2010) that explicitly included discharge from the gaged station using one year of data from the whole time series:

$$Chl_{seas} = \beta_0 + \beta_1 T + \beta_2 Q + \beta_3 \sin(2\pi T) + \beta_4 \cos(2\pi T) \tag{10}$$

$$\varepsilon_{Chl} = Chl_{seas} - \widehat{Chl}_{seas} \tag{11}$$

This approach was used to isolate an error structure for simulation that was independent of flow and biology, where the seasonal component (as time $T$ on a sinusoidal annual period) was assumed to be related to biological processes. The error distribution was then estimated from the residuals (eq. (11)) as before using an ARMA estimate of the residual parameters, $p$ and $q$. Standard error estimates from the regression used at each point in the one-year time series were also retained for each residual. Errors were simulated ($\varepsilon_{Chl,\,sim}$, eq. (7)) for the entire time series using the estimated auto-regressive structure and multiplied by the corresponding standard error estimate from the regression ($\sigma_{\widehat{Chl}_{seas}}$, eq. (7)). The single year estimate for $Chl_{seas}$ was repeated for every year and added to the error component that covered the entire time series. All simulated errors were rescaled to the range of the original residuals that were used to estimate the distribution. Finally, the simulated flow-component, $Chl_{flo}$, was added to the simulated biological model, $Chl_{bio}$, to create the daily time series, $Chl_{obs}$, in eq. (5). The final time series was then used to compare the relative abilities of WRTDS and GAMs to characterize flow-normalized trends.

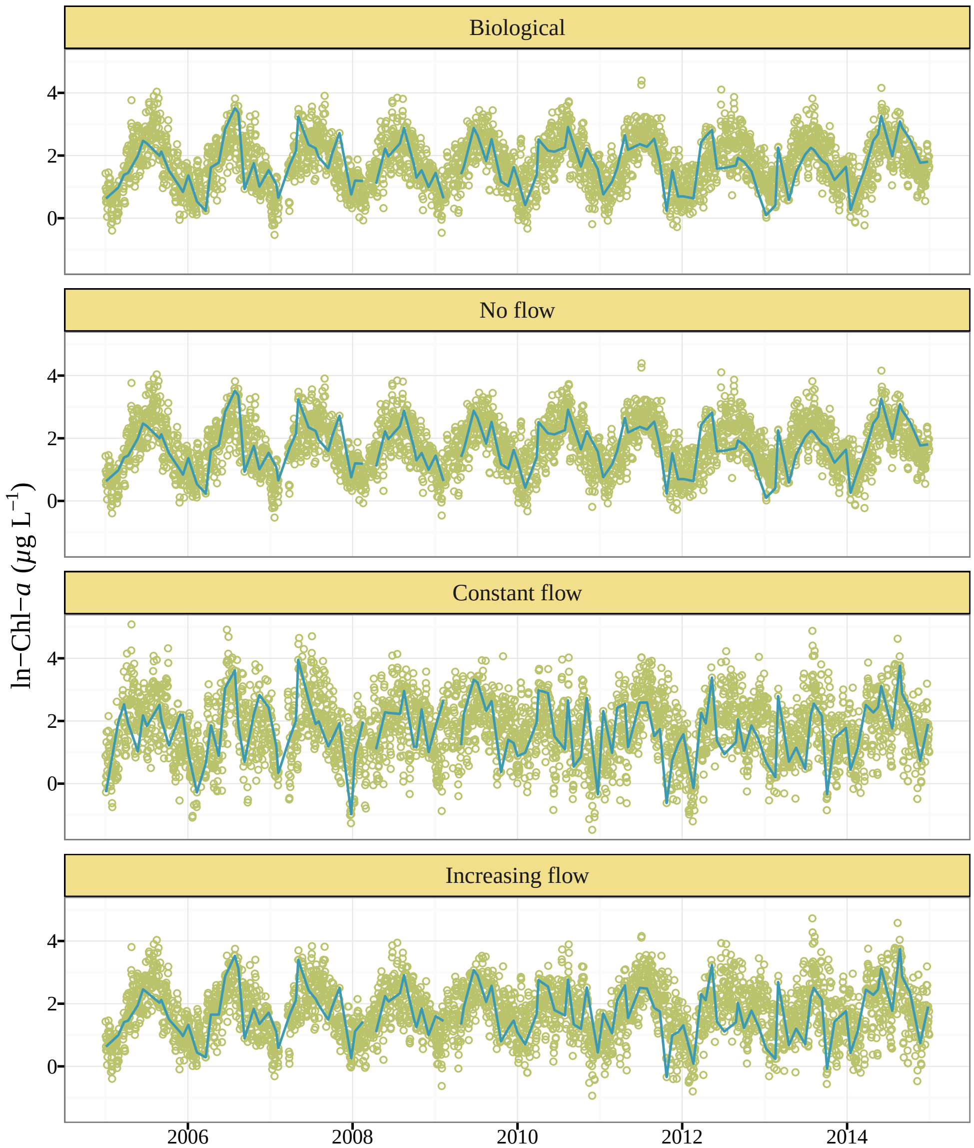


**FIGURE**: Examples of simulated time series for evaluating flow-normalized results from WRTDS and GAMs. The plots show the simulated daily time series (points) and monthly samples (lines) from the daily time series used to evaluate the flow-normalized predictions from WRTDS and GAMs. From top to bottom, the time series show the biological chl-*a* independent of flow and the three simulated datasets that represent different effects of flow: none, constant, and increasing effect. The flow-normalized results for the simulated monthly time series from each model were compared to the first time series (biological chl-*a*) that was independent of flow.

ACKNOWLEDGMENTS

We thank Jim Hagy, Bob Hirsch, Jennifer Keisman, and Elgin Perry for valuable discussions that improved the analysis. Thanks to Jeremy Testa and Jeffrey Chanat for providing valuable comments on an earlier draft. We thank the Chesapeake Bay Program and Maryland Department of Natural Resources for providing data.

# LITERATURE CITED


Arlot, S. and A. Celisse, 2010. A survey of cross-validation procedures for model selection. Statistics Surveys 4:40–79.

Beck, M. W. and J. D. Hagy III, 2015. Adaptation of a weighted regression approach to evaluate water quality trends in an estuary. Environmental Modelling and Assessment 20(6):637–655.

Beck, M. W., J. D. Hagy III, and M. C. Murrell, 2015. Improving estimates of ecosystem metabolism by reducing effects of tidal advection on dissolved oxygen time series. Limnology and Oceanography: Methods .

Boicourt, W. C. and L. P. Sanford, 1998. A hydrodynamic study of the Patuxent River estuary, Technical report, Maryland Department of the Environment, Baltimore, Maryland.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow, 2004. Confounding effect of flow on estuarine response to nitrogen loading. Journal of Environmental Engineering-ASCE 130(6):605–614.

Byrd, R. H., P. Lu, and J. N. adn C. Zhu, 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16(5):1190–1208.

Cade, B. S. and B. R. Noon, 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1(8):412–420.

Cloern, J. E., 1996. Phytoplankton bloom dynamics in coastal ecosystems: A review with some general lessons from sustained investigation of San Francisco Bay, California. Review of Geophysics 34(2):127–168.

Cloern, J. E. and A. D. Jassby, 2010. Patterns and scales of phytoplankton variability in estuarine-coastal ecosystems. Estuaries and Coasts 33(2):230–241.

Cronin, W. B. and D. W. Pritchard, 1975. Additional statistics on the dimensions of the Chesapeake Bay and its tributaries: Cross-section widths and segment volumes per meter depth, Technical Report 42, Reference 75-3, Chesapeak Bay Institute, The Johns Hopkins University, Baltimore, Maryland.

Efron, B. and R. Tibshirani, 1993. An Introduction to the Bootstrap, Chapman and Hall, New York, New York, first edition.

Hagy, J. D., 1996. Residence times and net ecosystem processes in Patuxent River estuary, Master's thesis, University of Maryland, College Park, Maryland.

Hagy, J. D., J. C. Lehrter, and M. C. Murrell, 2006. Effects of hurricanee Ivan on water quality in Pensacola Bay, Florida. Estuaries and Coasts 29(6A):919–925.

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., 1994. Long-term trends in the distribution of phytoplankton in Chesapeake Bay - roles of light, nutrients, and streamflow. Marine Ecology Progress Series 104:267–291.

Harding, L. W., C. L. Gallegos, E. S. Perry, W. D. Miller, J. E. Adolf, M. E. Mallonee, and H. W. Paerl, 2015. Long-term trends of nutrients and phytoplankton in Chesapeake Bay. Estuaries and Coasts .

Hastie, T. and R. Tibshirani, 1990. Generalized Additive Models, Chapman and Hall, London, New York.

Hirsch, R. M., 2014. Large biases in regression-based constituent flux estimates: causes and diagnostic tools. Journal of the American Water Resources Association 50(6):1401–1424.

Hirsch, R. M., S. A. Archfield, and L. A. De Cicco, 2015. A bootstrap method for estimating uncertainty of water quality trends. Environmental Modelling and Software 73:148–166.

Hirsch, R. M. and L. De Cicco, 2014. User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data, Technical Report Techniques and Methods book 4, ch. A10, US Geological Survey, Reston, Virginia. http://pubs.usgs.gov/tm/04/a10/.

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.

Hyndman, R. J. and Y. Khandakar, 2008. Automatic time series forecasting: The forecast package for R. JOurnal of Statistical Software 26(3):1–22.

Jordan, T. E., D. E. Weller, and D. L. Correll, 2003. Sources of nutrient inputs to the Patuxent River estuary. Estuaries and Coasts 26(2A):226–243.

Lung, W. and S. Bai, 2003. A water quality model for the Patuxent estuary: Current conditions and predictions under changing land-use scenarios. Estuaries 26(2A):267–279.

Medalie, L., R. M. Hirsch, and S. A. Archfield, 2012. Use of flow-normalization to evaluate nutrient concentration and flux changes in Lake Champlain tributaries, 1990-2009. Journal of Great Lakes Research 38(SI):58–67.

Monbet, Y., 1992. Control of phytoplankton biomass in estuaries: A comparative analysis of microtidal and macrotidal estuaries. Estuaries 15(4):563–571.

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.

Murrell, M. C., J. D. Hagy, E. M. Lores, and R. M. Greene, 2007. Phytoplankton production and nutrient distributions in a subtropical esuary: Importance of freshwater flow. Estuaries and Coasts 30(3):390–402.

Nixon, S. W., 1995. Coastal marine eutrophication: A definition, social causes, and future concerns. Ophelia 41:199–219.

Nocedal, J. and S. J. Wright, 2006. Numerical Optimization, Springer-Verlag, New York, New York, 2nd edition.

Orth, R. J., M. R. Williams, S. R. Marion, D. J. wilcox, T. J. B. Carruthers, K. A. Moore, W. M. Kemp, W. C. Dennison, N. Rybicki, P. Bergstrom, and R. A. Batiuk, 2010. Long-term trends in submersed aquatic vegetation (SAV) in Chesapeake Bay, USA, related to water quality. Estuaries and Coasts 33(5):1144–1163.

Paerl, H. W., 2006. Assessing and managing nutrient-enhanced eutrophication in estuarine and coastal waters: Interacive effects of human and climatic perturbations. Ecological Engineering 26(1):40–54.

Paerl, H. W., N. S. Hall, B. L. Peierls, and K. L. Rossignol, 2014. Evolving paradigms and challenges in estuarine and coastal eutrophication dynamics in a culturally and climatically stressed world. Estuaries and Coasts 37(2):243–258.

Powers, S. P., C. H. Peterson, R. R. Christian, E. Sullivan, M. J. Powers, M. J. Bishop, and C. P. Buzzelli, 2005. Effects of eutrophication on bottom habitat and prey resources of demersal fishes. Marine Ecology Progress Series 302:233–243.

RDCT (R Development Core Team), 2015. R: A language and environment for statistical computing, v3.2.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Sprague, L. A., R. M. Hirsch, and B. T. Aulenbach, 2011. Nitrate in the Mississippi River and its tributaries, 1980 to 2008: Are we making progress? Environmental Science and Technology 45(17):7209–7216.

Steward, J. S. and W. C. Green, 2007. Setting load limits for nutrients and suspended solids based upon seagrass depth-limit targets. Estuaries and Coasts 30(4):657–670.

TBEP (Tampa Bay Estuary Program), 2011. Tampa Bay Water Atlas. http://www.tampabay.wateratlas.usf.edu/. (Accessed October, 2013).

Testa, J. M., W. M. Kemp, W. R. Boynton, and J. D. Hagy, 2008. Long-term changes in water quality and productivity in the Patuxent River Estuary: 1985 to 2003. Estuaries and Coasts 31(6):1021–1037.

USGS (US Geological Survey), 2015. Water Quality Loads and Trends at Nontidal Monitoring Stations in the Chesapeake Bay Watershed. http://cbrim.er.usgs.gov/. (Accessed November, 2015).

Wood, S. N., 2006a. Generalized Additive Models: An Introduction with R, Chapman and Hall, CRC Press, London, United Kingdom.

—, 2006b. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics 62(4):1025–1036.

Zhang, Q., D. C. Brady, and W. P. Ball, 2013. Long-term seasonal trends of nitrogen, phosphorus, and suspended sediment load from the non-tidal Susquehanna River Basin to Chesapeake Bay. Science of the Total Environment 452-453:208–221.

TABLE 1: Summary characteristics of monitoring stations on the Patuxent River estuary. Chlorophyll and salinity values are based on averages from 1986 to 2014. Stations used for the analysis are in bold. Segments are salinity regions in the Patuxent for the larger Chesapeake Bay area (TF = tidal fresh, OH = oligohaline, MH = mesohaline). See Figure 1 for site locations.

| Station | Lat | Long | Segment | Distance (km) | Depth (m) | ln-Chl ($\mu$g/L ) | Sal (ppt) |
|---|---|---|---|---|---|---|---|
| TF1.3 | 38.81 | -76.71 | TF | 74.90 | 2.9 | 1.52 | 0.00 |
| TF1.4 | 38.77 | -76.71 | TF | 69.50 | 2.0 | 2.31 | 0.02 |
| TF1.5 | 38.71 | -76.70 | TF | 60.30 | 10.6 | 2.88 | 0.27 |
| **TF1.6** | 38.66 | -76.68 | OH | 52.20 | 6.2 | 2.44 | 0.90 |
| TF1.7 | 38.58 | -76.68 | OH | 42.50 | 3.0 | 2.09 | 4.09 |
| RET1.1 | 38.49 | -76.66 | MH | 32.20 | 11.2 | 2.47 | 10.25 |
| LE1.1 | 38.43 | -76.60 | MH | 22.90 | 12.1 | 2.31 | 12.04 |
| **LE1.2** | 38.38 | -76.51 | MH | 13.40 | 17.1 | 2.16 | 12.73 |
| LE1.3 | 38.34 | -76.48 | MH | 8.30 | 23.4 | 2.12 | 12.89 |
| LE1.4 | 38.31 | -76.42 | MH | 0.00 | 15.4 | 2.21 | 13.46 |

TABLE 2: Summaries of model performance using RMSE of observed to predicted ln-chl-*a* for each station (LE1.2 and TF1.6). Deviance for each model as the sum of squared residuals is shown in parentheses. Overall performance for the entire time series is shown at the top with groupings by different time periods below. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge.

| **Period** | **LE1.2** | | **TF1.6** | |
|---|---|---|---|---|
| | GAM | WRTDS | GAM | WRTDS |
| **All** | | | | |
| | 0.51 (139.5) | 0.51 (135.1) | 0.50 (128.4) | 0.52 (138.6) |
| **Annual** | | | | |
| 1986-1993 | 0.50 (41.1) | 0.50 (40.9) | 0.48 (37.2) | 0.49 (39.1) |
| 1994-2000 | 0.51 (34.7) | 0.50 (33.2) | 0.55 (39.3) | 0.58 (44.9) |
| 2001-2007 | 0.61 (51.5) | 0.60 (49.6) | 0.50 (33.7) | 0.53 (37.5) |
| 2008-2014 | 0.37 (12.1) | 0.36 (11.4) | 0.45 (18.2) | 0.44 (17.1) |
| **Seasonal** | | | | |
| JFM | 0.60 (38.1) | 0.58 (35.3) | 0.49 (24.4) | 0.49 (23.8) |
| AMJ | 0.64 (65.2) | 0.64 (65.3) | 0.54 (45.7) | 0.58 (51.9) |
| JAS | 0.35 (19.3) | 0.35 (18.6) | 0.45 (30.4) | 0.46 (32.2) |
| OND | 0.39 (16.8) | 0.38 (15.9) | 0.52 (27.9) | 0.54 (30.7) |
| **Flow** | | | | |
| 1 (Low) | 0.36 (17.4) | 0.36 (16.7) | 0.45 (26.5) | 0.46 (27.7) |
| 2 | 0.43 (24.4) | 0.42 (23.5) | 0.53 (36.6) | 0.54 (37.8) |
| 3 | 0.58 (43.8) | 0.57 (42.9) | 0.49 (31.3) | 0.52 (35.4) |
| 4 (High) | 0.64 (53.9) | 0.63 (52.0) | 0.51 (34.0) | 0.54 (37.7) |

TABLE 3: Summaries of flow-normalized trends from each model at LE1.2 for different time periods. Summaries are averages and percentage changes of ln-chl-*a* ($\mu$g/L ) based on annual means within each category. For example, summary values for high flow for a given model are based on instances of high flow across years. Percentage changes are the differences between the last and first years in the periods. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge.

| **Period** | **GAM** | | **WRTDS** | |
|---|---|---|---|---|
| | Ave. | % Change | Ave. | % Change |
| **All** | | | | |
| | 2.17 | 24.28 | 2.18 | 18.85 |
| **Annual** | | | | |
| 1986-1993 | 1.99 | 9.60 | 2.03 | 1.75 |
| 1994-2000 | 2.12 | 5.49 | 2.12 | 5.50 |
| 2001-2007 | 2.24 | 5.50 | 2.24 | 5.35 |
| 2008-2014 | 2.37 | 3.20 | 2.37 | 6.07 |
| **Seasonal** | | | | |
| JFM | 2.57 | 20.06 | 2.58 | 14.04 |
| AMJ | 2.32 | 31.20 | 2.33 | 22.47 |
| JAS | 2.01 | 18.48 | 2.01 | 19.91 |
| OND | 1.82 | 25.29 | 1.83 | 15.14 |
| **Flow** | | | | |
| Flow 1 (Low) | 1.90 | 20.86 | 1.93 | 16.77 |
| Flow 2 | 2.10 | 13.71 | 2.11 | 7.73 |
| Flow 3 | 2.28 | 15.66 | 2.29 | 9.24 |
| Flow 4 (High) | 2.34 | 25.09 | 2.33 | 22.29 |

TABLE 4: Summaries of flow-normalized trends from each model at TF1.6 for different time periods. Summaries are averages and percentage changes of ln-chl-*a* ($\mu$g/L ) based on annual means within each category. For example, summary values for high flow for a given model are based on instances of high flow across years. Percentage changes are the differences between the last and first years in the periods. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge.

| **Period** | **GAM** | | **WRTDS** | |
|---|---|---|---|---|
| | Ave. | % Change | Ave. | % Change |
| **All** | | | | |
| | 2.43 | -4.81 | 2.44 | -2.28 |
| **Annual** | | | | |
| 1986-1993 | 2.62 | -4.93 | 2.60 | -3.06 |
| 1994-2000 | 2.69 | -5.05 | 2.65 | -3.55 |
| 2001-2007 | 2.15 | -22.42 | 2.19 | -21.51 |
| 2008-2014 | 2.24 | 47.10 | 2.30 | 38.35 |
| **Seasonal** | | | | |
| JFM | 1.52 | 9.03 | 1.48 | 32.72 |
| AMJ | 2.63 | 5.47 | 2.62 | 5.14 |
| JAS | 3.06 | 0.04 | 3.08 | 0.79 |
| OND | 2.17 | -18.16 | 2.20 | -17.55 |
| **Flow** | | | | |
| Flow 1 (Low) | 2.89 | -4.78 | 2.93 | -0.42 |
| Flow 2 | 2.41 | 16.71 | 2.43 | 20.31 |
| Flow 3 | 2.28 | 6.53 | 2.27 | 15.20 |
| Flow 4 (High) | 2.22 | -11.58 | 2.21 | -11.27 |

TABLE 5: Comparison of predicted results between WRTDS and GAMs using average differences (%) and RMSE values at each station. Overall comparisons for the entire time series are shown at the top with groupings by different time periods below. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge. Negative percentages indicate WRTDS predictions were lower than GAM predictions (eq. (4)).

| **Period** | **LE1.2** | | **TF1.6** | |
|---|---|---|---|---|
| | Ave. diff. | RMSE | Ave. diff. | RMSE |
| **All** | | | | |
| | -0.11 | 0.09 | 0.01 | 0.13 |
| **Annual** | | | | |
| 1986-1993 | 0.20 | 0.10 | -0.74 | 0.11 |
| 1994-2000 | 0.34 | 0.09 | -1.29 | 0.15 |
| 2001-2007 | -0.55 | 0.07 | 0.68 | 0.13 |
| 2008-2014 | -0.53 | 0.08 | 3.10 | 0.14 |
| **Seasonal** | | | | |
| JFM | 0.39 | 0.12 | -2.00 | 0.14 |
| AMJ | 0.22 | 0.10 | -0.66 | 0.14 |
| JAS | -0.71 | 0.06 | 0.76 | 0.10 |
| OND | -0.46 | 0.05 | 1.04 | 0.15 |
| **Flow** | | | | |
| Flow 1 (Low) | -0.27 | 0.07 | -0.15 | 0.10 |
| Flow 2 | -0.14 | 0.09 | 0.70 | 0.13 |
| Flow 3 | 0.49 | 0.11 | 1.07 | 0.14 |
| Flow 4 (High) | -0.53 | 0.09 | -1.75 | 0.15 |

TABLE 6: Regression fits comparing predicted (*pred*) and flow-normalized (*norm*) results for WRTDS and GAMs at each station. Values in bold-italic are those where the intercept ($\beta_0$) estimate was significantly different from zero or the slope ($\beta_1$) estimate was significantly different from one. Fits for the entire time series are shown at the top. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge. See Figure 5 for a graphical summary.

| **Period** | **LE1.2** | | **TF1.6** | | **LE1.2** | | **TF1.6** | |
|---|---|---|---|---|---|---|---|---|
| | $\beta_{0,pred}$ | $\beta_{1,pred}$ | $\beta_{0,pred}$ | $\beta_{1,pred}$ | $\beta_{0,norm}$ | $\beta_{1,norm}$ | $\beta_{0,norm}$ | $\beta_{1,norm}$ |
| **All** | | | | | | | | |
| | ***0.05*** | ***0.97*** | ***0.08*** | ***0.97*** | ***0.15*** | ***0.94*** | 0.02 | 0.99 |
| **Annual** | | | | | | | | |
| 1986-1993 | 0.02 | 0.99 | -0.02 | 1.00 | ***0.20*** | ***0.92*** | ***-0.12*** | ***1.03*** |
| 1994-2000 | ***0.16*** | ***0.93*** | -0.03 | 0.99 | ***0.17*** | ***0.92*** | ***-0.12*** | ***1.02*** |
| 2001-2007 | 0.02 | 0.99 | ***0.13*** | ***0.95*** | ***0.06*** | ***0.98*** | ***0.11*** | ***0.97*** |
| 2008-2014 | 0.00 | 1.00 | ***0.12*** | 0.97 | 0.01 | 0.99 | ***0.08*** | 0.99 |
| **Seasonal** | | | | | | | | |
| JFM | -0.01 | 1.01 | 0.09 | ***0.92*** | 0.01 | 1.00 | ***0.20*** | ***0.84*** |
| AMJ | ***0.28*** | ***0.88*** | ***0.27*** | ***0.89*** | ***0.38*** | ***0.84*** | ***0.34*** | ***0.87*** |
| JAS | -0.08 | 1.03 | ***0.34*** | ***0.89*** | ***0.30*** | ***0.85*** | ***0.39*** | ***0.88*** |
| OND | 0.02 | 0.98 | ***0.13*** | ***0.95*** | ***0.38*** | ***0.80*** | 0.03 | 1.00 |
| **Flow** | | | | | | | | |
| Flow 1 (Low) | ***0.14*** | ***0.92*** | -0.03 | 1.01 | ***0.46*** | ***0.77*** | ***0.16*** | ***0.95*** |
| Flow 2 | 0.00 | 1.00 | ***0.12*** | ***0.96*** | ***0.14*** | ***0.94*** | 0.01 | 1.00 |
| Flow 3 | 0.09 | 0.96 | ***0.21*** | ***0.91*** | ***0.12*** | ***0.96*** | -0.02 | 1.00 |
| Flow 4 (High) | 0.09 | ***0.96*** | 0.03 | ***0.97*** | ***0.09*** | ***0.96*** | ***0.09*** | ***0.95*** |

TABLE 7: Summaries of model performance comparing observed chl-$a$ with predicted values ($Chl_{obs} \sim \widehat{Chl}_{obs}$) and biological chl-$a$ with flow-normalized values ($Chl_{bio} \sim \widehat{Chl}_{bio}$) for the three simulated time series (no flow, constant flow, and increasing flow effect). Summaries are RMSE values comparing results from each model (GAM, WRTDS). Deviance for each model as the sum of squared residuals is shown in parentheses.

| Simulations | $Chl_{obs} \sim \widehat{Chl}_{obs}$ | $Chl_{bio} \sim \widehat{Chl}_{bio}$ |
|---|---|---|
| **No flow** | | |
| GAM | 0.51 (31.2) | 0.53 (33.2) |
| WRTDS | 0.50 (29.4) | 0.52 (31.7) |
| **Constant flow** | | |
| GAM | 0.51 (31.2) | 0.58 (39.8) |
| WRTDS | 0.53 (32.8) | 0.57 (38.9) |
| **Increasing flow** | | |
| GAM | 0.51 (31.2) | 0.54 (35.0) |
| WRTDS | 0.50 (29.7) | 0.52 (31.9) |

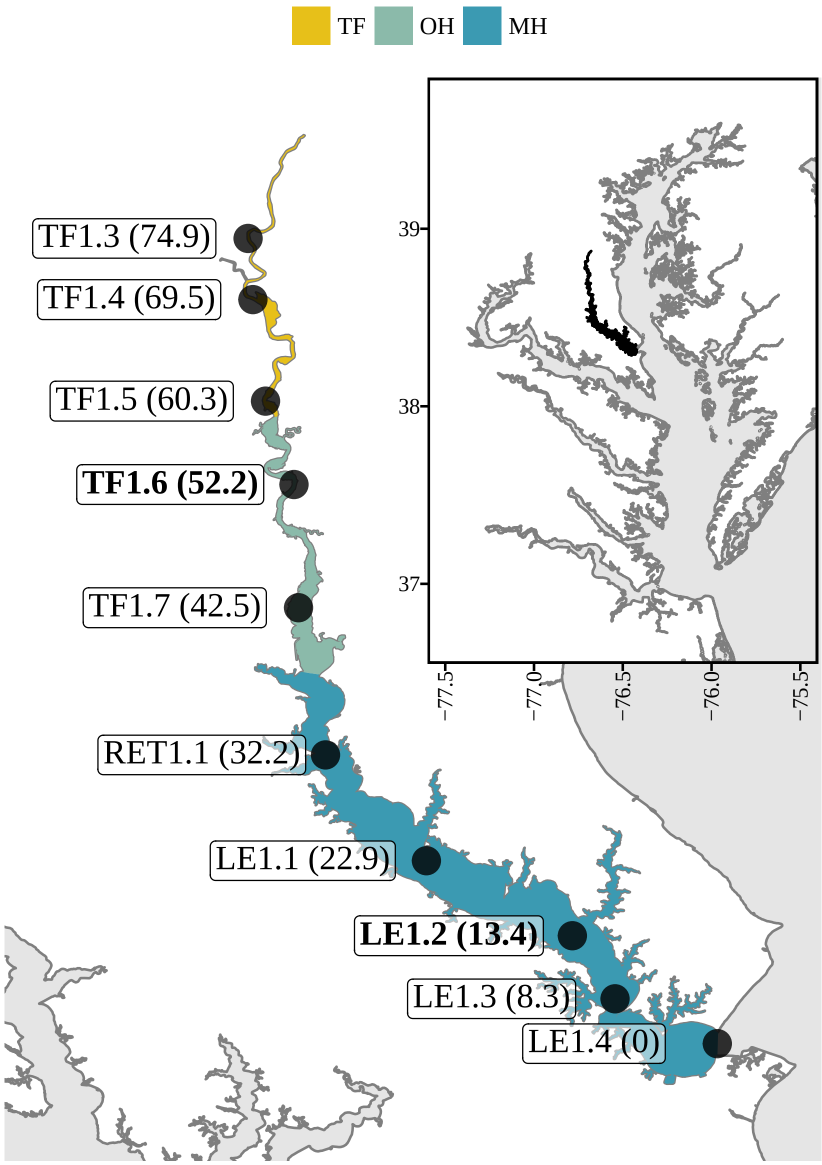


FIGURE 1: Patuxent River estuary with Chesapeake Bay inset. Fixed locations monitored by the Maryland Department of Natural Resources at monthly frequencies are shown along the longitudinal axis with distance from the mouth (km). Study sites are in bold. Salinity regions in the Patuxent for the larger Chesapeake Bay area are also shown (TF = tidal fresh, OH = oligohaline, MH = mesohaline). See Table 1 for a numeric summary of station characteristics.

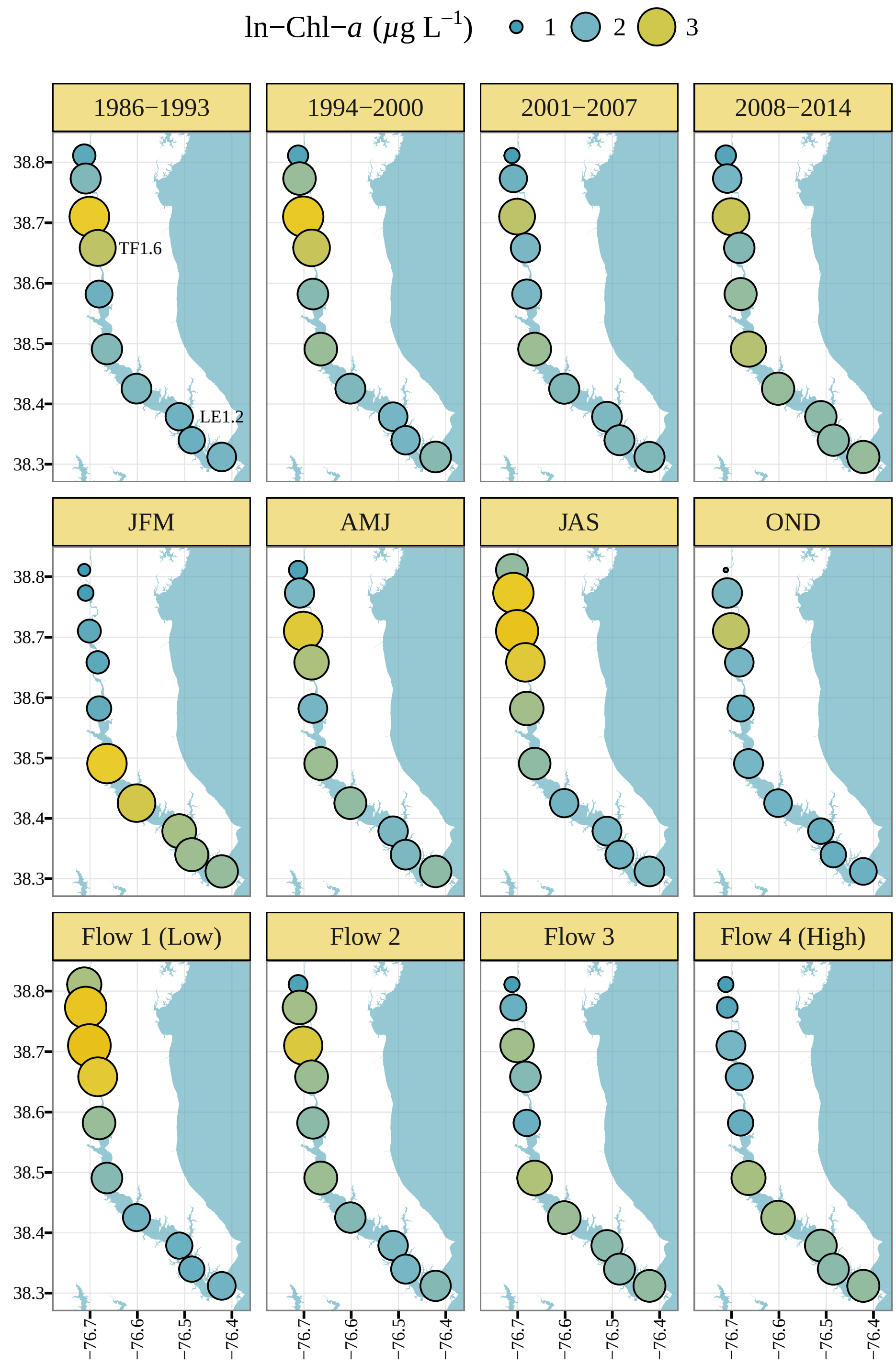


FIGURE 2: Annual, seasonal, and flow differences in chl-*a* trends at each monitoring station in the Patuxent River Estuary. Size and color are proportional medians of ln-chl-*a* by year, season, and flow categories. See Figure 1 for station numbers.

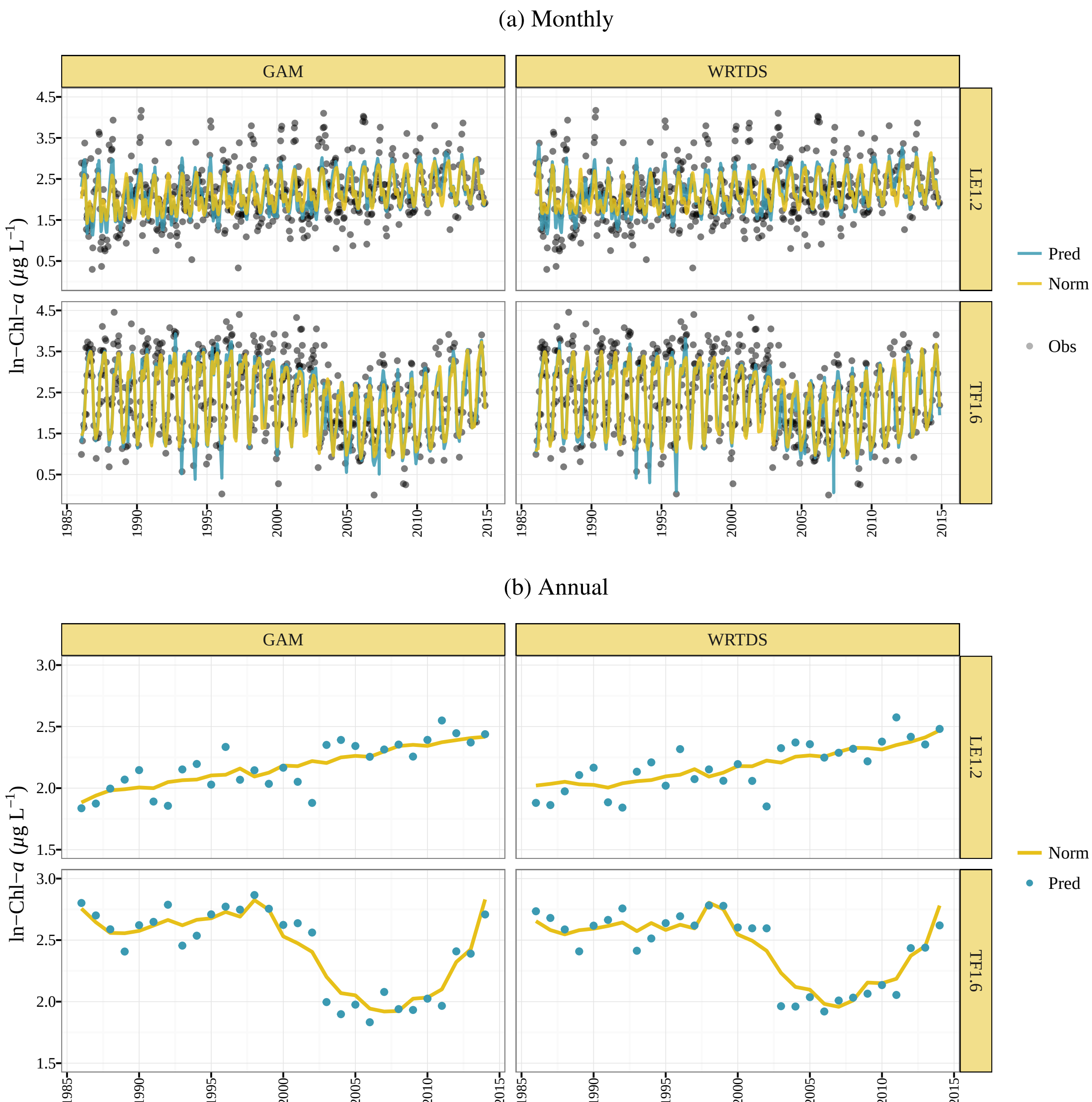


FIGURE 3: Predicted chl-*a* from generalized additive models (GAM) and weighted regression (WRTDS) for LE1.2 and TF1.6 stations on the Patuxent River estuary. Figure 3a shows results at monthly time steps and Figure 3b shows results averaged by year. Values in blue are model predictions and values in yellow are flow-normalized predictions.

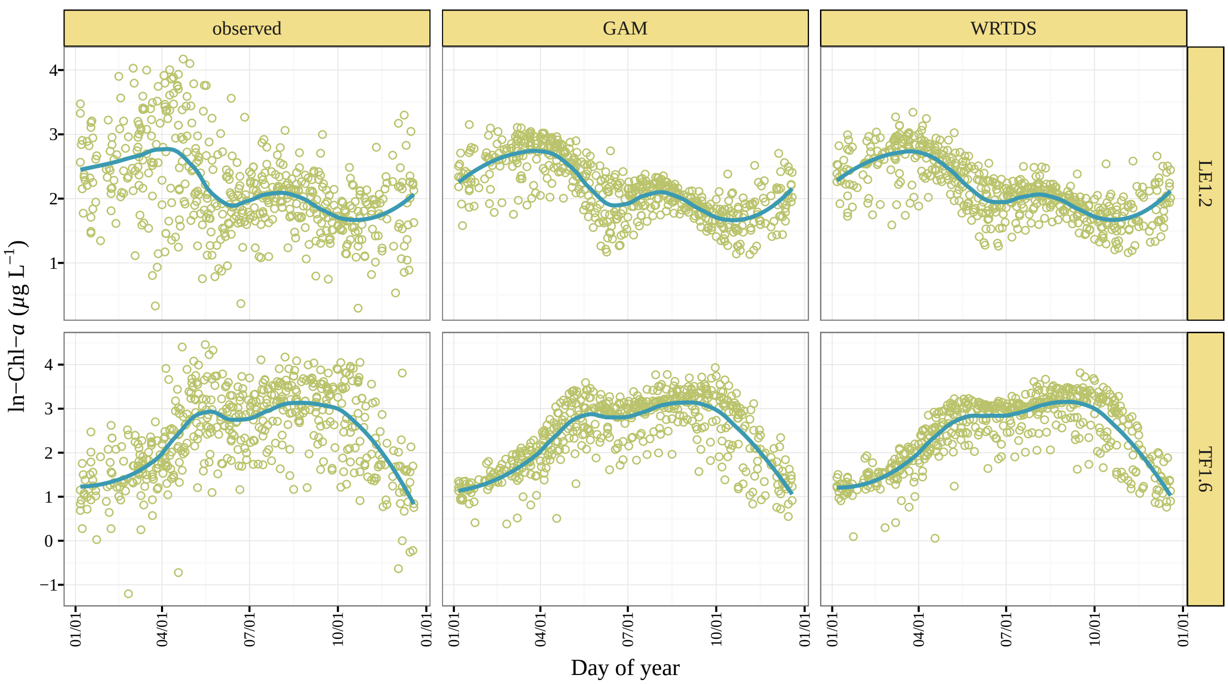


FIGURE 4: Seasonal variation from observed and model predictions of chl-*a* by station. Predictions are points by day of year from 1986 to 2014. The blue line is a loess (locally estimated) polynomial smooth to characterize the seasonal components.

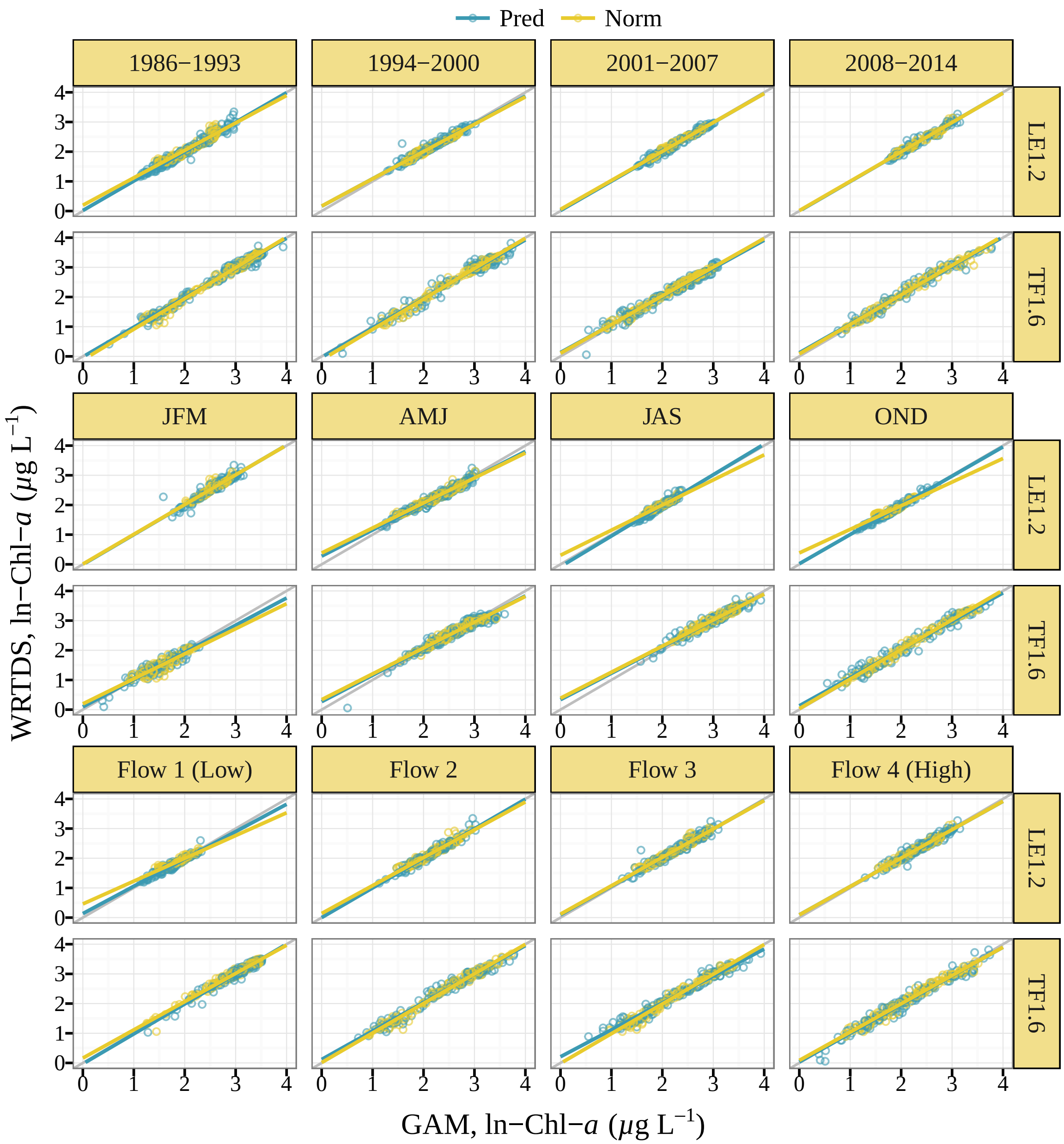


FIGURE 5: Comparison of WRTDS and GAMs results at each station (LE1.2, TF1.6) and different time periods. Predicted and flow-normalized results are shown. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions from the discharge record (low). Regression lines for each model result and 1:1 replacement lines (thin grey) are also shown. See Table 6 for parameter estimates of regression comparisons.

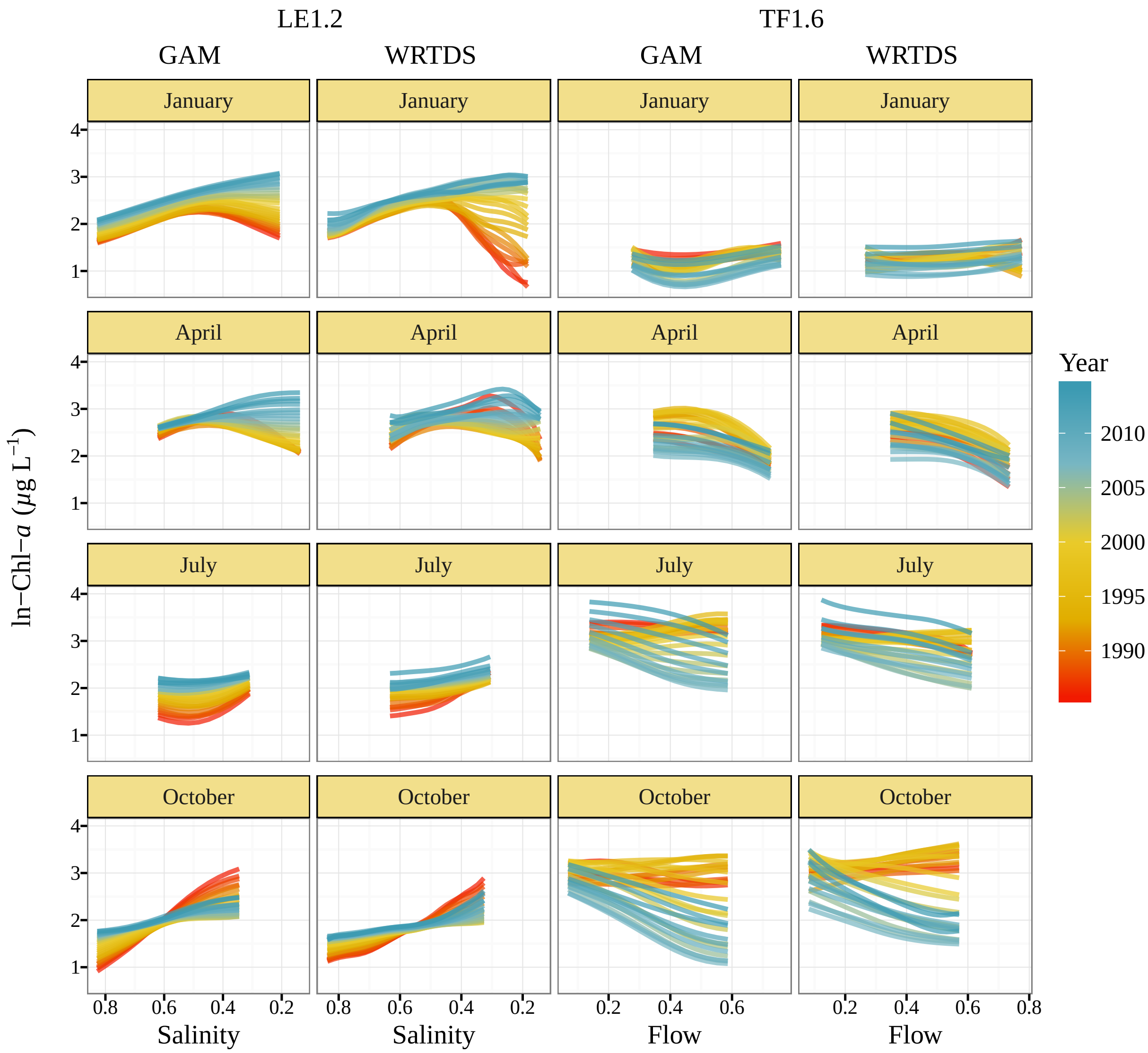


FIGURE 6: Changes in the relationship between chl-*a* and freshwater inputs (salinity decrease, flow increase) across the time series. Separate panels are shown for each station (LE1.2, TF1.6), model type (GAM, WRTDS), and chosen months. Changes over time are shown as different predictions for each year in the time series (1986 to 2014). Salinity was used as a tracer of freshwater inputs at LE1.2, whereas the flow record at Bowie, Maryland was used at TF1.6. The scales of salinity and flow are reversed for comparison of trends. Units are proportions of the total range in the observed data with values in each plot truncated by the monthly 5$^{th}$ and 95$^{th}$ percentiles.

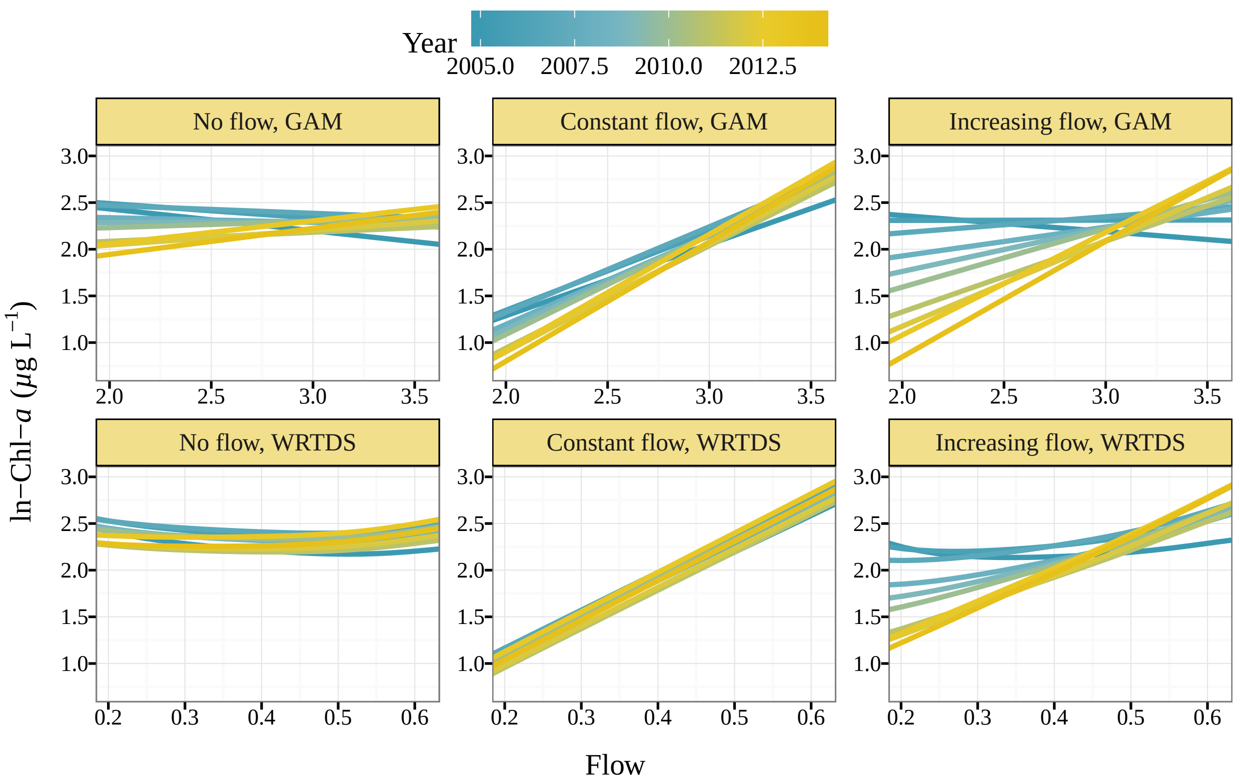


FIGURE 7: Examples of changing relationships between chl-*a* ($\mu$g/L ) and flow (as proportion of the total range) over time (2005–2015) for each simulated time series (**Appendix S2**). The plots are based on August predictions from three WRTDS models for each time series to illustrate the simulated relationships between flow and chl-*a*.