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

Marcus W. Beck[1], Rebecca Murphy[2]

[1]*ORISE Research Participation Program*
USEPA National Health and Environmental Effects Research Laboratory
Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561
Phone: 850-934-2480, Fax: 850-934-2401, Email: beck.marcus@epa.gov

[2]*UMCES at Chesapeake Bay Program*
410 Severn Avenue, Suite 112, Annapolis, MD 21403
Phone: 410-267-9837, Fax: 410-267-5777, Email: rmurphy@chesapeakebay.net

Version Date: Fri Aug 7 19:26:32 2015 -0500

Abstract

Long-term monitoring datasets can provide information to interpret the effects of environmental changes or management actions on ecosystem condition. The ability to link causal effects with potential changes can depend on the chosen method of interpreting trends in the observed data. Two statistical approaches, weighted regression on time, discharge, and season (WRTDS) and {acro:GRM} generalized additive modelss (GAMs), have recently been used to evaluate long-term trends in chlorophyll time series in estuarine systems. Both models provide a similar approach to trend analysis by using context-dependent parameters or smoothing functions that vary continuously and have the potential to identify multiple drivers of change. However, the quantitative capabilities of each model, including descriptions of observed and flow-normalized trends, have not been rigorously compared to determine most appropriate use of each model. We evaluated WRTDS and GAMs using thirty years of data for a monthly time series of chlorophyll in the Patuxent River Estuary, a well-studied tributary to Chesapeake Bay. Each model was evaluated based on goodness of fit to the observed data and ability to reproduce flow-normalized trends with simulated data that had statistical properties comparable to the original dataset. Each model was also evaluated based on concordance of conclusions of water quality changes, and causes thereof, in different time periods. This information will provide researchers with valuable guidance for using statistical models in trend evaluation, with particular relevance for computational requirements, desired products, and future data needs.

Key words: chlorophyll, estuary, generalized additive models, Patuxent River Estuary, trend analysis, weighted regression

1 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 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 chlorophyll 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 chlorophyll 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 chlorophyll 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 chlorophyll trends in tidal waters (Beck and Hagy III 2015). A

{acro:WRTI

{acro:GAM]

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 contextual flow inputs 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), and the Mississippi River (Sprague et al. 2011). 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). Similarly, GAMs can be used to describe variation in a response variable as a sum of smoothing functions for different predictors (Hastie and Tibshirani 1990, Wood 2006). In applications to water quality time series, GAMs are similar to WRTDS in that variable effects through time can be described in relation to seasonal or annual changes. Application of GAMs to describe eutrophication endpoints in tidal waters have not been as extensive as WRTDS, although exploratory analyses have suggested that results are comparable. Moreoever, GAMs are particularly appealing because they are less computationally intense and provide more accessible estimates of model uncertainty than WRTDS. Despite the potential for both approaches to characterize system dynamics, 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. A thirty year time series of monthly chlorophyll observations from the Patuxent River Estuary is 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

4

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.

2   Methods

2.1   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 (Fig. 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 vertically-mixed in the upper estuary. 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 US Geological Survey (USGS) stream gage on the Patuxent {acro:USGS 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 maintains a continuous monitoring network for the Patuxent at ten fixed stations that cover the salinity gradient from estuarine to tidal fresh (http://www.chesapeakebay.net/, Fig. 1). Water quality samples have been collected since 1985 at monthly or bimonthly intervals and include salinity, temperature, chlorophyll *a* (chl-*a*), dissolved {acro:chla 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 primary production is generally higher in the lower estuary during winter months (Fig. 2). Chlorophyll concentrations are generally lowest for all stations in late fall and early winter. Periods of low flow are associated with higher chlorophyll 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 chlorophyll 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 monthly chlorophyll and salinity data from 1986 to 2014 were obtained for stations TF1.6 and LE1.2 from the Chesapeake Bay Program data hub. 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 of each observation provided a better fit to the chlorophyll data. Both chlorophyll and discharge data were log-transformed. Censored data were not present in any of the data sets. Initial quality assurance checks for all monitoring data were conducted following standard protocols adopted by the Chesapeake Bay Program.

2.2 Model descriptions

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

The WRTDS method relates a response endpoint, typically a nutrient concentration, to flow and time to evaluate trends in long-term water quality (Hirsch et al. 2010, Hirsch and De

Cicco 2014). Recent adaptation of WRTDS to tidal waters relates chlorophyll concentration to salinity and time (Beck and Hagy III 2015). The functional form of the model is a simple regression that relates the natural log of chlorophyll (Chl) to decimal time (T) and salinity (Sal) on a sinuisoidal annual time scale (i.e., cyclical variation by year).

$$\ln\left(Chl\right) = \beta_0 + \beta_1 T + \beta_2 Sal + \beta_3 \sin\left(2\pi T\right) + \beta_4 \cos\left(2\pi t\right) + \epsilon \qquad (1) \quad \{\texttt{eqn:funf}}$$

The tidal adaptation of WRTDS uses quantile regression models (Cade and Noon 2003) to characterize trends in different conditional distributions of chlorophyll, e.g., the median or 90th percentile. For comparison to GAMs, the standard WRTDS model that characterizes the conditional mean of the response was used. Mean models require an estimation of the back-transformation bias parameter. This is achieved by accounting for the standard error of residuals for each observation along the time series during back-transformation (Hirsch et al. 2010). Additionally, the WRTDS model uses survival regression as a variation of the weighted Tobit model (Tobin 1958) to account for censored observations beyond the detection limit (Hirsch and De Cicco 2014).

The WRTDS approach obtains fitted values of the response variable by estimating regression parameters for each unique observation. Specifically, a unique regression model is estimated for each point in the period of observation. Each model is weighted by month, year, and salinity 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 (Hirsch et al. 2010). Optimal window widths can be identified using cross-validation, described below, thats 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 values based on the minimum and maximum values for the data are used to predict chlorophyll 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 and date values closest to observed. Salinity-normalized values are also obtained from the prediction grid that allow an interpretation of chlorophyll trend that is independent of variation related to salinity changes. Normalized predictions are obtained for each observation by collecting the sample of observed salinity values that for the same month throughout all years in the dataset. These salinity 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 values that are expected to occur for each month.

2.2.2   Generalized Additive Models

2.2.3   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 GAMs requires identifying the optimal degrees of freedom for the smoothing parameter. Overfitting a model with excessively small window widths or excessive degrees of freedom will minimize prediction error but prevent extrapolation of results to different datasets. Similarly, underfitting a model with large window widths or very few degrees of freedom 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 a novel dataset.

The basic premise of cross-validation is to identify the optimal set of model parameters that minimize prediction error on a dataset that was not used to develop the model. For GAMs (Hastie and Tibshirani 1990, Zuur 2012)...[insert GAMs methods]. Similarly, the tidal adaptation of WRTDS used k-fold cross-validation to identify the optimal half-window widths. For a given set of half-window widhts, the dataset was separated into ten disjoint sets, 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 (Byrd et al. 1995, 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. The chosen parameters were based on a selected convergence tolerance for the error minimization of the search algorithm.

2.3   Comparison of modelled trends

Separate WRTDS and GAMs were created using the above methods for the chlorophyll time series at TF1.6 and LE1.2. Initial analyses indicated that model performance could be improved using the flow record to model chl-*a* at TF1.6 and the salinity record to model chl-*a* at LE1.2. For each model and station, a predicted and flow-normalized (hereafter flow-normalized refers to both flow and salinity) time series was obtained for comparison. The results were compared using several summary statistics that evaluated both the predictive performance to describe observed chlorophyll and direct comparisons between the models. Emphasis was placed on agreement between predictions and observed values, rather than uncertainty associated with parameter estimates. As of writing, methods for estimating confidence intervals of WRTDS have not been fully developed for application to WRTDS in tidal waters. In addition to simple visual evaluation of trends over time and bivariate scatterplot comparisons, summary statistics used to compare model predictions to observed chl-*a* included explained variance (R^2) and root mean square error (RMSE). Model results were compared directly using average differences and evaluation of a regression fit for the results comparing each model to a line with intercept of zero and slope of one. Deviation of the slope of the regression either greater or less than one indicated potential bias for a particular model as under- or over-predicting relative to the other. For all comparisons, RMSE comparing each model's predictions to observed chl-*a* was defined as:

$$RMSE = \sqrt{\frac{\sum\limits_{i=1}^{n} \left(Chl_i - \hat{Chl_i}\right)^2}{n}} \tag{2}$$

where n is the number of observations for a given evaluation, Chl_i is the observed value of chl-*a* for observation i, and $\hat{Chl_i}$ is the predicted value of chl-*a* for observation i. RMSE values closer

to zero represent model predictions closer to observed. Similarly, average difference between models as a percentage was defined as:

$$\text{Average difference} = \left(\frac{\sum\limits_{i=1}^{n} \hat{Chl}_{WRTDS,i} - \sum\limits_{i=1}^{n} \hat{Chl}_{GAM,i}}{\sum\limits_{i=1}^{n} \hat{Chl}_{GAM,i}} \right) * 100 \tag{3}$$

where $\hat{Chl}_{WRTDS,i}$ and $\hat{Chl}_{GAM,i}$ are the predicted values of chl-a for each model at observation i. Positive values indicate that WRTDS provided larger predictions than GAMs on average, whereas the opposite is true for negative values (Moyer et al. 2012).

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. Annual and seasonal aggregations shown in Fig. 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) relevant to each station. Flow-normalized time series were compared similarly but only between the 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 the flow-normalized data was primarily accomplished using simulated datasets with known components that were independent of discharge.

2.4 Comparison of flow-normalized trends

The relative abilities of each model to characterize flow or salinity-normalized trends in chlorophyll were evaluated using simulated datasets with known components. This approach was used because the flow-independent component of chlorophyll is typically not observed in raw data such that the true signal must be empirically estimated. Accordingly, the ability of each model to isolate the flow-normalized trend cannot be evaluated with reasonable certainty unless the true signal is known. Simulated time series of observed chlrophyll (Chl_{obs}) were created as additive components related to flow (Chl_{flo}, analogous to salinity) and a flow-independent

biological component of chlorophyll (Chl_{bio}):

$$Chl_{obs} = Chl_{flo} + Chl_{bio} \qquad (4) \quad \texttt{\{chlobs\}}$$

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 chlorophyll in the water column that is directly linked to primary production. The biological component of chlorophyll is comparable to an observation in a closed 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 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 were obtained from the USGS stream gage station 01594440 near Bowie, Maryland (38°57′21.3″N, 76°41′37.3″W) from 1985 to 2014. Daily chlorophyll records were obtained from the Jug Bay station (38°46′50.6″N, 76°42′29.1″W) of the Chesapeake Bay Maryland National Estuarine Research Reserve. Daily chlorophyll concentrations were estimated from fluorescence values that did not include blue-green algae blooms. Our primary concern was simulating chlorophyll concentrations at monthly or bimonthly timesteps such that taxa-specific concentrations on a daily time step were not relevant.

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 ($\hat{Q}_{seas}$), 2) simulated error structure from the residuals of the seasonal discharge model ($\varepsilon_{Q,sim}$), 3) estimated chlorophyll independent of discharge as a stationary seasonal component ($\hat{Chl}_{seas}$), and 4) simulated error structure from the residuals of the seasonal chlorophyll model ($\varepsilon_{Chl,sim}$). Unless otherwise noted, chlorophyll and discharge are in ln-transformed units. Each of the four

11

components was used to simulate the components in eq. (4):

$$Chl_{flo} = I\left(\hat{Q}_{seas} + \sigma \cdot \varepsilon_{Q,sim}\right) \tag{5} \quad \text{\{chlflo\}}$$

$$Chl_{bio} = \hat{Chl}_{seas} + \sigma \cdot \varepsilon_{Chl,sim} \tag{6} \quad \text{\{chlbio\}}$$

The estimated flow time series within the parentheses, $\hat{Q}_{seas} + \sigma \cdot \varepsilon_{Q,sim}$, 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 chlorophyll. Moreover, the vector I (where $0 \leq I \leq 1$) can be manually changed to represent an independent effect of flow based on the desired simulation. 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. Methods for estimating each of the components in eqs. (5) and (6) are described in detail below.

First, a model for simulating flow-related chlorophyll (eq. (5)) was estimated from the stream gage data as the additive combination of a stationary seasonal component and serially-correlated errors:

$$Q_{seas} = \beta_0 + \beta_1 \sin\left(2\pi T\right) + \beta_2 \cos\left(2\pi T\right) \tag{7} \quad \text{\{qseas\}}$$

$$\varepsilon_Q = Q_{seas} - \hat{Q}_{seas} \tag{8} \quad \text{\{qerr\}}$$

A seasonal model of flow was estimated using linear regression for time, T, on an annual sinusoidal period (eq. (7)). The residuals from this regression, ε_Q (eq. (8)), 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 $\quad \text{\{acro:ARMA}}$ identify appropriate p and q coefficients (Hyndman and Khandakar 2008). The parameters were

12

chosen using stepwise estimation for nonseasonal univariate time series that minimized Akaike {acro:AIC}
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. (8) and added to the
seasonal component in eq. (7) to create a simulated, daily time series of the flow-component for
chlorophyll, Chl_{flo} (eq. (5)).

The chlorophyll time series was created using a similar approach. The first step estimated
the stationary seasonal component of the chlorophyll 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:

$$Chl_{seas} = \beta_0 + \beta_1 T + \beta_2 Q + \beta_3 \sin\left(2\pi T\right) + \beta_4 \cos\left(2\pi T\right) \qquad (9) \quad \{\text{chlseas}\}$$

$$\varepsilon_{Chl} = Chl_{seas} - \hat{Chl}_{seas} \qquad (10) \quad \{\text{chlerr}\}$$

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. (10)) 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. (6)) for the entire year using
the estimated auto-regressive structure and multiplied by the corresponding standard error
estimate from the regression. The entire year was repeated for every year in the observed 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 bilogical model, Chl_{bio}, to create the final chlorophyll-flow time series, Chl_{obs}, in
eq. (4).

A daily time series for the entire period of record was simulated using the above methods
and then used to compare the relative abilities of WRTDS and GAMs to characterize
flow-normalized trends. Multiple time series with a monthly sampling frequencies and varying
contributions of the flow component, (Chl_{flo} in eq. (4)) were created from the daily time series.

One day in each month for each year was randomly sampled to create a monthly time series. Varying contributions of the flow component on observed chlorophyll were creating by multiplying Chl_{flo} by different indicator vectors (I in eq. (5)). The contribution of the flow component varied from constant, non-existent, steadily increasing, and steadily decreasing. Respectively, the vector of coefficients applied to each flow component was a constant vector of ones, a constant vector of zeroes, a linear increase starting at zero and ending at one, and a linear decrease starting at one and ending at zero. This created four monthly time series that were used to evaluate each model. The flow-normalized results of WRTDS and GAMs for each simulated time series were compared to each other and to the original biological chlorophyll component of each time series (Chl_{bio}, eqs. (4) and (6)).

3   Results

Predictions with actual data

Simulations

4   Discussion

Qualitative comparison

- Computational requirements and potential limitations

- Data needs or transferability of each technique to novel datasets

- Products, e.g., conditional quantiles of WRTDS, confidence intervals for GAMs, handling censored data, hypothesis testing vs description

- Appropriate context for using each approach

4.1   Conclusions

References

Beck MW, Hagy III JD. 2015. Adaptation of a weighted regression approach to evaluate water quality trends in an estuary. Environmental Modelling and Assessment, pages 1–19.

Boicourt WC, Sanford LP. 1998. A hydrodynamic study of the Patuxent River estuary. Technical report, Maryland Department of the Environment, Baltimore, Maryland.

Borsuk ME, Stow CA, Reckhow KH. 2004. Confounding effect of flow on estuarine response to nitrogen loading. Journal of Environmental Engineering-ASCE, 130(6):605–614.

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

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

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

Cronin WB, Pritchard DW. 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.

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

Harding LW. 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.

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

Hirsch RM. 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 RM, De Cicco L. 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 RM, Moyer DL, Archfield SA. 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 RJ, Khandakar Y. 2008. Automatic time series forecasting: The forecast package for r. JOurnal of Statistical Software, 26(3):1–22.

Jordan TE, Weller DE, Correll DL. 2003. Sources of nutrient inputs to the Patuxent River estuary. Estuaries and Coasts, 26(2A):226–243.

Lung W, Bai S. 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, Hirsch RM, Archfield SA. 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 DL, Hirsch RM, Hyer KE. 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.

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

Paerl HW. 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 HW, Hall NS, Peierls BL, Rossignol KL. 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 SP, Peterson CH, Christian RR, Sullivan E, Powers MJ, Bishop MJ, Buzzelli CP. 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 LA, Hirsch RM, Aulenbach BT. 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 JS, Green WC. 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 JM, Kemp WM, Boynton WR, Hagy JD. 2008. Long-term changes in water quality and productivity in the Patuxent River Estuary: 1985 to 2003. Estuaries and Coasts, 31(6):1021–1037.

Tobin J. 1958. Estimation of relationships for limited dependent variables. Econometrica, 26(1):24–36.

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

Zhang Q, Brady DC, Ball WP. 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.

Zuur AF. 2012. A Beginner's Guide to Generalized Additive Models in R. Highland Statistics Ltd., Newburgh, United Kingdom.

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 Fig. 1 for a map.

tab:statsum

Station	Lat	Long	Segment	Distance (km)	Depth (m)	Chl (μg L^{-1})	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: Summary of model performances at LE1.2 for WRTDS and GAMs during different time periods. Overall performance for the entire time series is 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.

tab:perfLE12

Period	**RMSE**		R^2	
	WRTDS	GAM	WRTDS	GAM
All				
	0.51	0.55	0.48	0.39
Annual				
1986-1993	0.50	0.56	0.53	0.42
1994-2000	0.50	0.56	0.42	0.26
2001-2007	0.60	0.62	0.34	0.30
2008-2014	0.35	0.38	0.60	0.55
Seasonal				
JFM	0.58	0.65	0.33	0.16
AMJ	0.64	0.69	0.41	0.33
JAS	0.35	0.38	0.31	0.18
OND	0.38	0.41	0.44	0.37
Flow				
1 (Low)	0.36	0.40	0.33	0.16
2	0.42	0.47	0.40	0.26
3	0.58	0.59	0.47	0.44
4 (High)	0.63	0.69	0.34	0.19

Table 3: Summary of model performances at TF1.6 for WRTDS and GAMs during different time periods. Overall performance for the entire time series is 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. tab:perfTF16

Period	RMSE		R^2	
	WRTDS	GAM	WRTDS	GAM
All				
	0.52	0.56	0.71	0.66
Annual				
1986-1993	0.49	0.52	0.71	0.67
1994-2000	0.58	0.65	0.67	0.59
2001-2007	0.53	0.54	0.65	0.64
2008-2014	0.44	0.53	0.77	0.67
Seasonal				
JFM	0.48	0.54	0.38	0.22
AMJ	0.58	0.58	0.48	0.47
JAS	0.46	0.49	0.49	0.42
OND	0.54	0.64	0.75	0.64
Flow				
1 (Low)	0.46	0.48	0.43	0.38
2	0.54	0.56	0.67	0.65
3	0.52	0.56	0.69	0.64
4 (High)	0.54	0.64	0.72	0.60

Table 4: Average differences (%) between WRTDS and GAMs at each station during different time periods. Differences for predicted and flow-normalized results from each model are provided. Overall differences for the entire time series is 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.
tab:avediff

Period	LE1.2		TF1.6	
	Pred	Norm	Pred	Norm
All				
	-0.12	-0.05	0.03	0.16
Annual				
1986-1993	0.59	-1.07	0.45	-0.45
1994-2000	0.36	-1.43	-0.14	-2.39
2001-2007	-0.61	2.37	0.50	3.61
2008-2014	-1.13	1.18	-1.05	1.19
Seasonal				
JFM	-0.15	-1.94	-1.56	-2.08
AMJ	-0.11	-0.82	0.39	-1.67
JAS	-0.20	0.79	0.43	0.84
OND	0.01	0.87	0.97	3.66
Flow				
Flow 1 (Low)	-0.20	-0.85	0.86	2.20
Flow 2	-0.60	2.58	-0.69	-0.81
Flow 3	1.35	0.81	-0.14	-0.30
Flow 4 (High)	-1.03	-2.92	0.20	-0.93

Table 5: Regression fits comparing WRTDS and GAMs predictions at each station during different time periods. Fits for the entire time series are shown at the top. The second slope column (Slope$'$) describes a regression that was forced through the origin. Values in bold-italic are those where the intercept estimate was significantly different from zero or the slope estimate was significantly different from one. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge.

tab:regpred

Period	Intercept		Slope		Slope$'$	
	LE1.2	TF1.6	LE1.2	TF1.6	LE1.2	TF1.6
All						
	0.08	*0.15*	*0.96*	*0.93*	1.00	*0.99*
Annual						
1986-1993	*-0.24*	*-0.29*	*1.13*	*1.10*	1.01	1.00
1994-2000	*0.42*	0.03	*0.81*	0.97	1.00	*0.98*
2001-2007	0.17	*0.44*	*0.92*	*0.81*	0.99	1.00
2008-2014	0.03	*0.25*	0.97	*0.90*	*0.99*	1.00
Seasonal						
JFM	0.06	*0.50*	0.98	*0.65*	1.00	*0.96*
AMJ	0.13	*0.47*	0.94	*0.82*	1.00	*0.99*
JAS	*0.40*	*0.57*	*0.79*	*0.81*	1.00	1.00
OND	0.04	0.14	0.98	0.94	1.00	1.00
Flow						
Flow 1 (Low)	*0.20*	*0.40*	*0.88*	*0.86*	1.00	*0.99*
Flow 2	-0.01	*0.21*	1.00	*0.94*	0.99	*1.02*
Flow 3	0.09	0.10	0.98	0.95	1.01	0.99
Flow 4 (High)	*0.25*	*0.21*	*0.89*	*0.87*	0.99	*0.96*

Table 6: Regression fits comparing WRTDS and GAMs flow-normalized predictions at each station during different time periods. Fits for the entire time series are shown at the top. The second slope column (Slope$'$) describes a regression that was forced through the origin. Values in bold-italic are those where the intercept estimate was significantly different from zero or the slope estimate was significantly different from one. Time periods are annual groupings every seven years (top), seasonal groupings by monthly quarters (middle), and flow periods based on quantile distributions of discharge. tab:regnorm

| Period	Intercept		Slope		Slope$'$	
	LE1.2	TF1.6	LE1.2	TF1.6	LE1.2	TF1.6
All						
	0.28	*0.21*	*0.87*	*0.91*	1.00	0.99
Annual						
1986-1993	*0.29*	*-0.28*	*0.86*	*1.10*	1.00	1.00
1994-2000	*0.35*	-0.10	*0.83*	1.00	0.99	*0.97*
2001-2007	*0.43*	*0.66*	*0.81*	*0.73*	1.00	1.00
2008-2014	0.00	*0.20*	0.99	*0.92*	*0.99*	1.00
Seasonal						
JFM	0.02	*0.76*	0.98	*0.47*	*0.98*	*0.95*
AMJ	*0.36*	*0.79*	*0.85*	*0.69*	1.00	*0.98*
JAS	*0.81*	*0.68*	*0.60*	*0.78*	1.00	1.00
OND	*0.72*	*0.21*	*0.61*	0.94	1.00	*1.03*
Flow						
Flow 1 (Low)	*0.73*	*0.31*	*0.62*	*0.91*	1.00	*1.02*
Flow 2	*0.35*	*0.28*	*0.83*	*0.88*	*0.99*	*0.98*
Flow 3	*0.25*	*0.19*	*0.89*	*0.90*	1.00	*0.98*
Flow 4 (High)	*0.30*	*0.29*	*0.88*	*0.86*	1.00	*0.98*
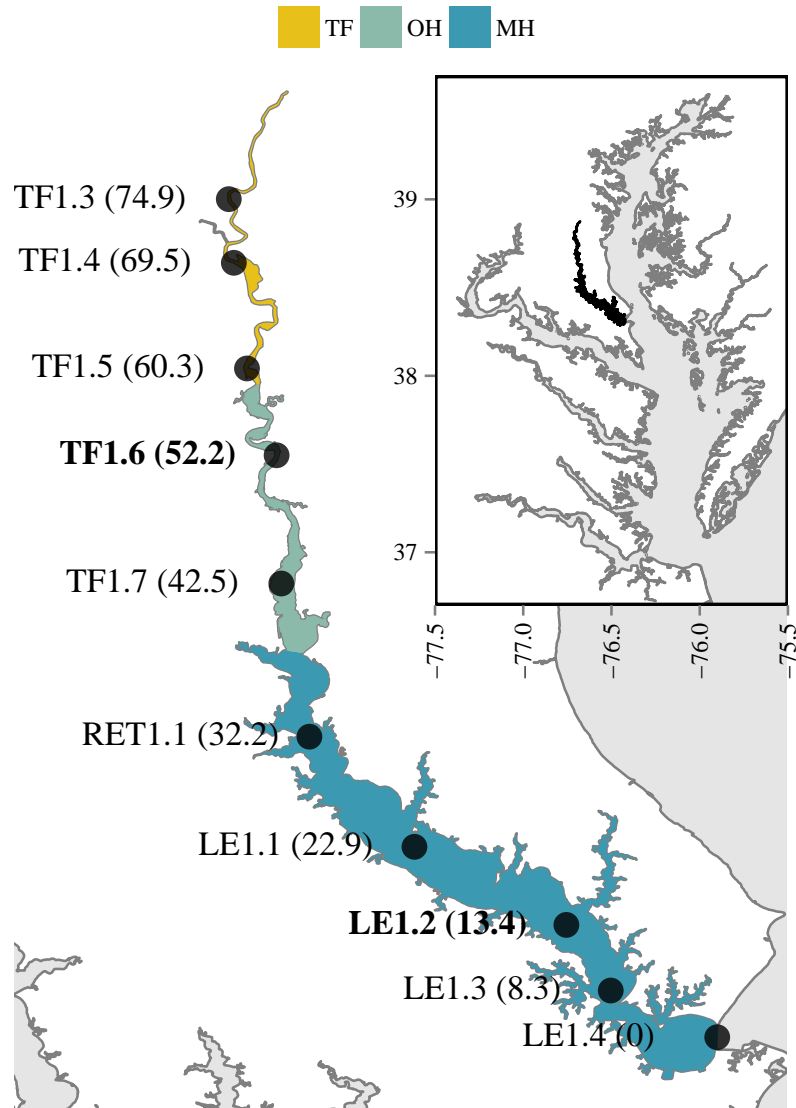
Fig. 1: Patuxent River estuary with Chesapeake Bay inset. Fixed locations monitored by the Chesapeake Bay Program at monthly frequencies are shown along the longitudinal axis with distance from the mouth (km). Study sites are in bold. Salinity regions in the Patuxent for the larger Chesapeake Bay area are also shown (TF = tidal fresh, OH = oligohaline, MH = mesohaline). See Table 1 for a numeric summary of station characteristics.
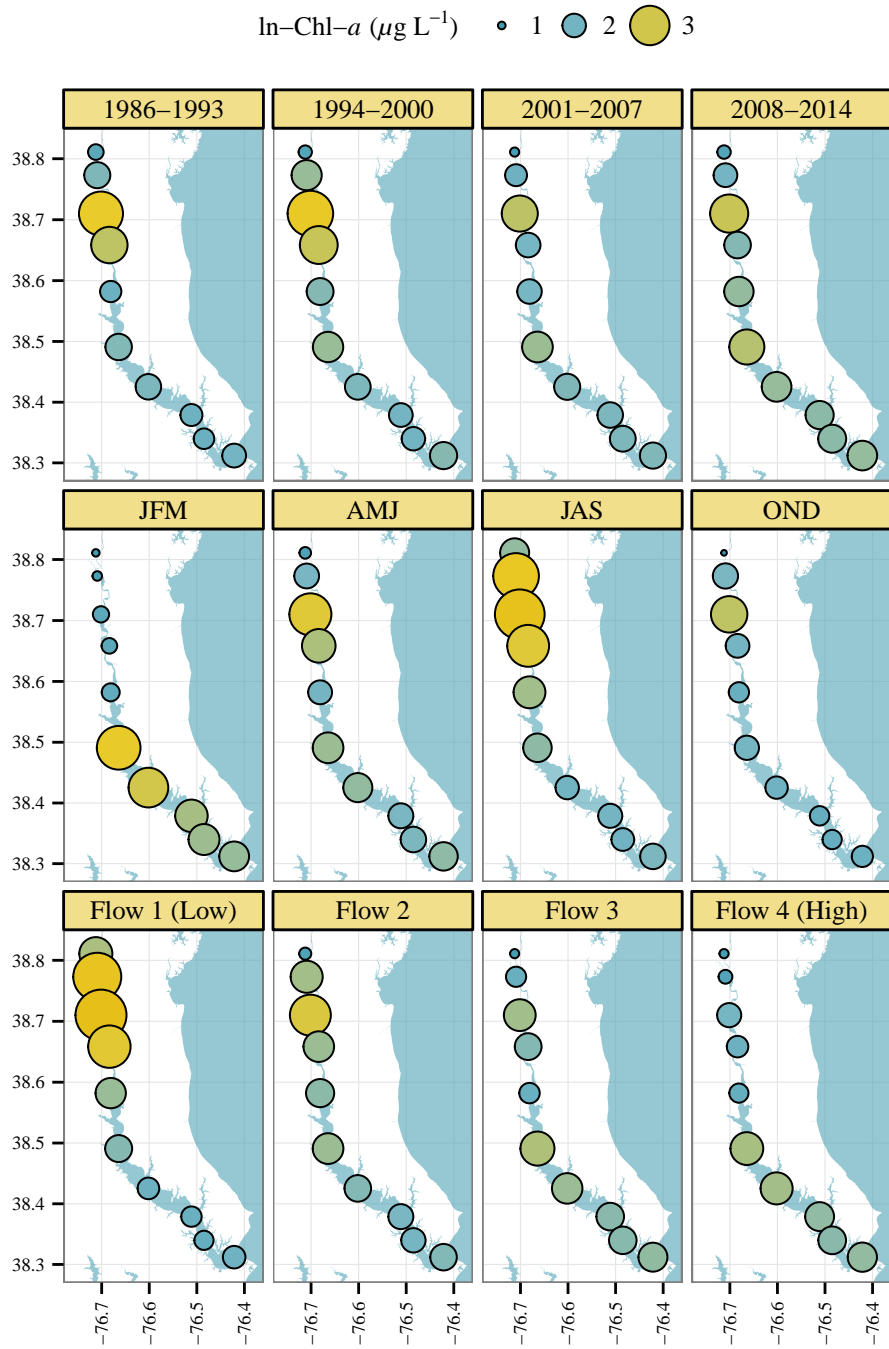
{fig:map}

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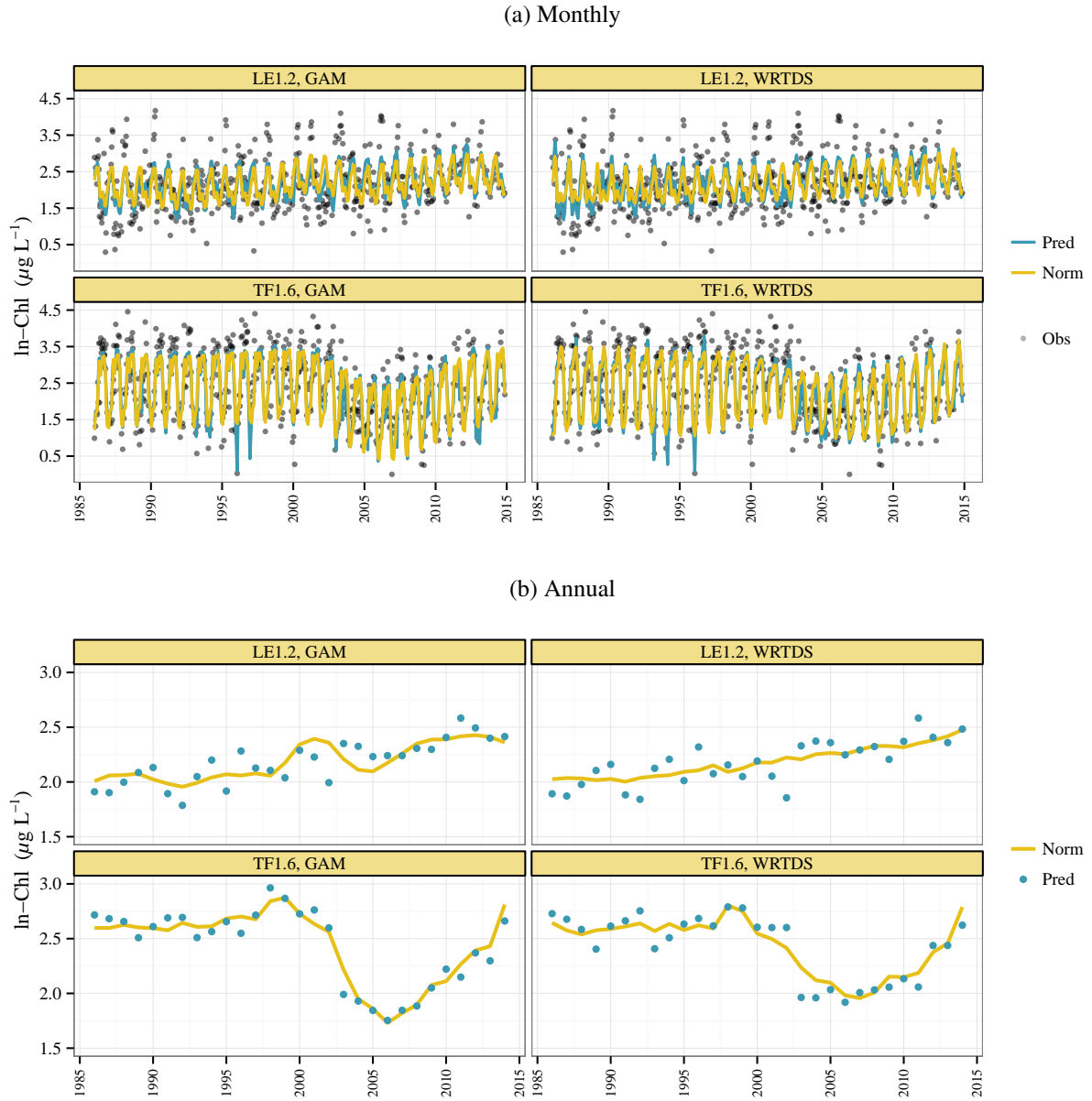
{fig:chly

(a) Monthly

(b) Annual

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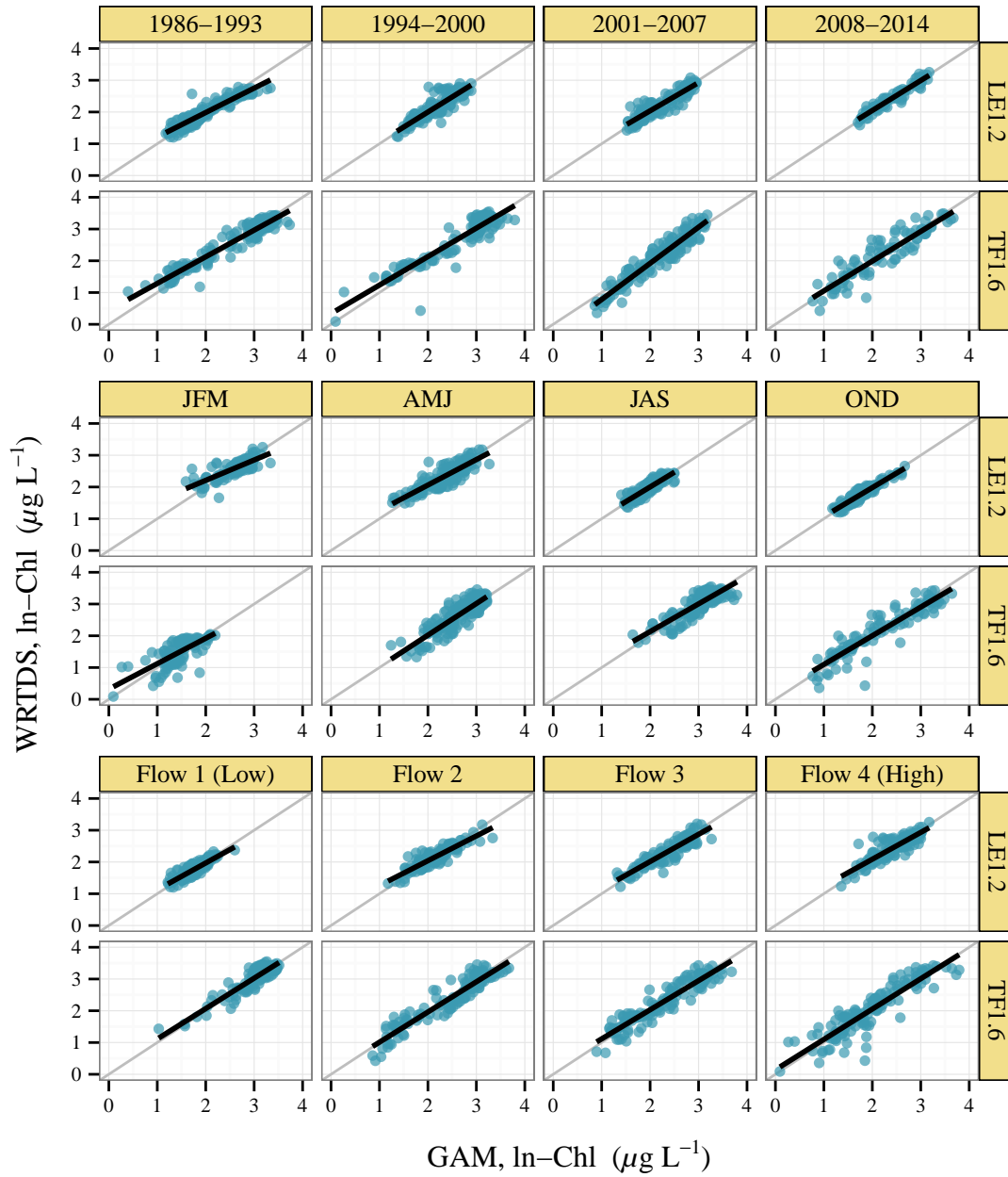
{fig:pred}

Fig. 4: Comparison of WRTDS and GAMs predictions of chl-*a* at each station (LE1.2, TF1.6) and different time periods. 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 (thick black) and 1:1 replacement lines (thin grey) are also shown.

{fig:wrtds