SUPPORTING INFORMATION

Numerical and qualitative contrasts of two statistical models for water quality change in tidal waters

Marcus W. Beck, Rebecca R. Murphy

Research Ecologist (Beck), US Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561; Estuarine Data Analyst (Murphy), University of Maryland Center for Environmental Science at Chesapeake Bay Program, 410 Severn Avenue, Suite 112, Annapolis, MD 21403 (Email/Beck: beck.marcus@epa.gov)

Appendix B: Additional material describing the simulation of daily discharge and chl-a time series.

Appendix C: Supplementary figure of regression comparisons between WRTDS and GAMs.

Appendix B: Additional material describing the simulation of daily discharge and chl-*a* time series.

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:

$$Q_{seas} = \beta_0 + \beta_1 \sin(2\pi T) + \beta_2 \cos(2\pi T) \quad \text{(B1)}$$

$$\varepsilon_Q = Q_{seas} - \hat{Q}_{seas} \quad \text{(B2)}$$

A seasonal model of flow was estimated using linear regression for time, *T*, on an annual sinusoidal period (eq. (B1)). The residuals from this regression, ε_Q (eq. (B2)), 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. (B2) (i.e., σ_ε in eq. (6)) and added to the seasonal component in eq. (B1) 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:

$$Chl_{seas} = \beta_0 + \beta_1 T + \beta_2 Q + \beta_3 \sin(2\pi T) + \beta_4 \cos(2\pi T) \quad \text{(B3)}$$

$$\varepsilon_{Chl} = Chl_{seas} - \hat{Chl}_{seas} \quad \text{(B4)}$$

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. (B4)) 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 and used as a measure of the scale parameter. 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 ($\sigma_{\hat{Chl}_{seas}}$, eq. (7)). The Chl_{seas} component was estimated again for the whole time series using a WRTDS model with no weights, which was then added to the product of the error component simulated for the whole time series and repeated one-year scale parameter. 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 daily time series (Figure B1) was then sampled monthly to compare the relative abilities of WRTDS and GAMs to characterize flow-normalized trends (Figure B2).

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.

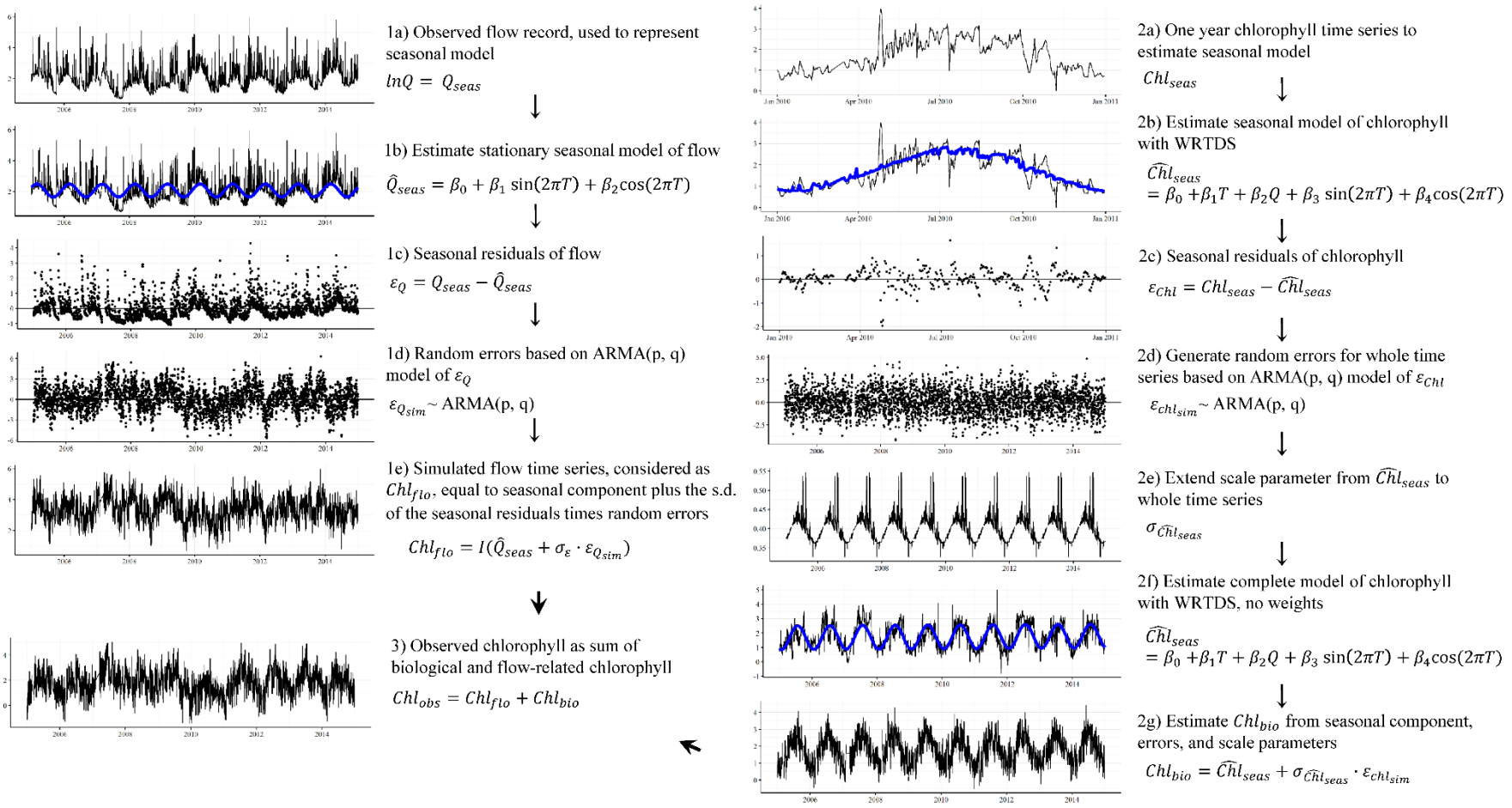

FIGURE B1: Workflow for creating simulated time series of observed chlorophyll described by eqs. (5-7, B1-B4) and supporting text. Steps 1a-1e describe simulating the flow component of chlorophyll (Chl_{flo}, eq. (6)), steps 2a-2g describe simulating the biological component of chlorophyll (Chl_{bio}, eq. (7)), and step three combines the two to create observed chlorophyll (Chl_{obs}, eq. (5)). The flow component is assumed constant throughout the time series (I = 1 in step 1e, eq. (6)). Functions for estimating these time series are available in the WRTDStidal R package, link in Appendix A.

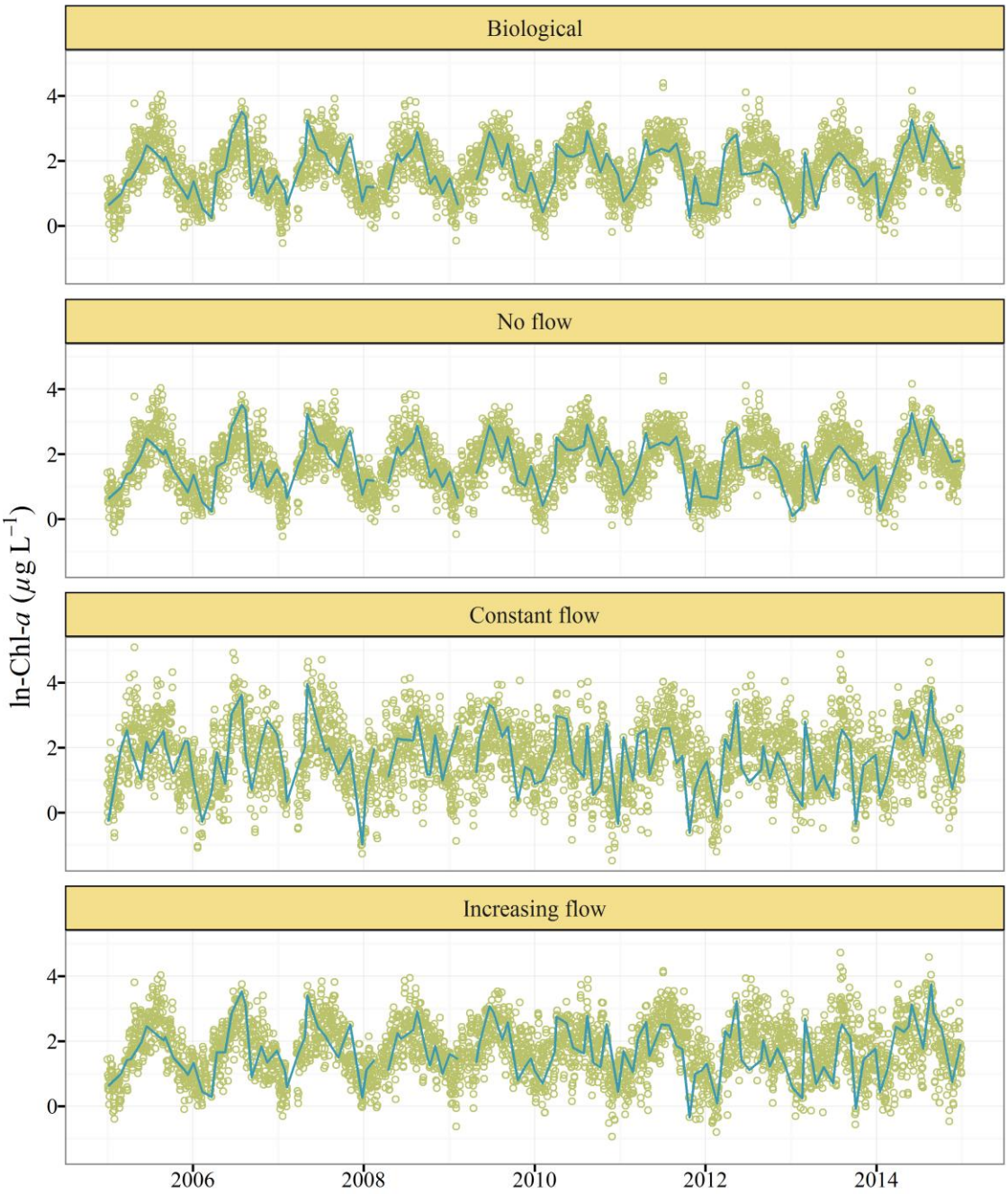

FIGURE B2: Examples of simulated time series for evaluating flow-normalized results from WRTDS and GAMs. The plots show the simulated daily time series (points) and monthly samples (lines) from the daily time series used to evaluate the flow-normalized predictions. From top to bottom, the time series 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.

Appendix C: Supplementary figure of regression comparisons between WRTDS and GAMs.

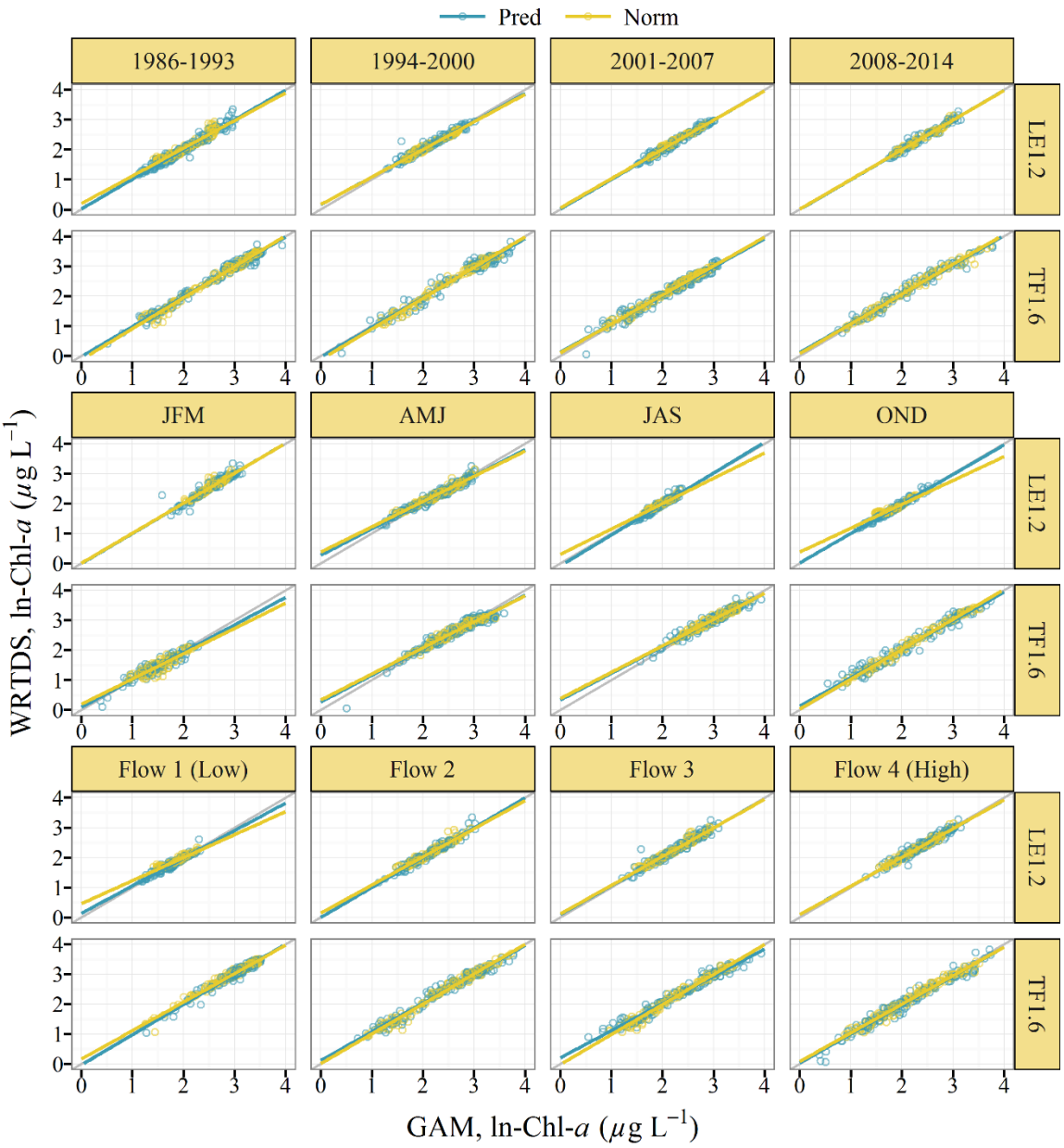

FIGURE C1: Comparison of WRTDS and GAM results at each station (LE1.2, TF1.6) and different time periods. Predicted and flow-normalized results are shown. Time periods are annual groupings every seven years (top), seasonal groupings (middle), and flow periods based on quantile distributions from the discharge record (bottom). Regression lines for each model result and 1:1 replacement lines (thin grey) are also shown.