# Appendix A. Supplementary materials

## S1 Dimensions of the model segments

Table S1 Dimensions used in the mechanistic model formulation.

|  |  |  |
| --- | --- | --- |
| Dimension | Segment | Units |
| Upper | Middle | Bend |
| Average depth, *H* | 2.51 | 2.83 | 3.67 | m |
| Volume, *V* | 47.94 | 134.54 | 170.35 | ∙106 m3 |
| Length  | 10.36 | 9.57 | 10.20 | ∙103 m |

## S2 Lower model boundary conditions (chl-*a*, DIN, OP)

Daily values of lower boundary conditions for chl-*a* (*alb*), DIN (*nlb*), OP (*plb*) for Models 2 and 3 were estimated via linear regressions. Models were built assuming water quality observations at Middle and Bend segments (indexed 2 and 3, respectively in eqs. S2.1-S2.3) as predictors and downstream observations at sampling location 140 (Fig. 1) as responses. The resulting regressions (eqs. S2.1-S2.3) explained 25%, 82%, and 67% of chl-*a*, DIN, and OP variability, respectively. Lower boundary conditions were dynamically simulated based on these regression coefficients and the model predictions for Middle and Bend segments.

|  |  |
| --- | --- |
| *alb* =9.56 + 0.36∙*a*3 | (S2.1) |
| *nlb =* 21.25 + 0.67∙*n*3 − 0.16∙*n*2 | (S2.2) |
| *plb =* 1.93 + 0.72∙*p*3 | (S2.3) |

## S3 Prior and calibrated posterior parameter distributions summary

Table S3.1 Prior parameters distributions for statistical Model 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Description** | **Units** | **Prior** |
| β0 | intercept | ln(μg/L) | *N*(μβ0*,*σβ0) |
| β*Q,l* | ln(*Q*2) before breakpoint | ln(μg/L)/ln(m3/d) | *N*(μβ*Q*,*l*,σβ*Q*,*l*) |
| β*Q*,h | ln(*Q*2) after breakpoint | ln(μg/L)/ln(m3/d) | *N*(μβ*Q*,*h*,σβ*Q*,*h*) |
| β*TN* | *TN* | ln(μg/L)∙L/μg | *N*(μβ*TN*,σβ*TN*) |
| *bp* | breakpoint | ln(m3/d) | *N*(16,2) |
| β*T* | *T* | ln(μg/L)/oC | *N*(0,10) |
| μβ0 | mean for β0 | ln(μg/L) | *N*(0,10) |
| μβ*Q*,*l* | mean for β*Q*,*l* | ln(μg/L)/ln(m3/d) | *N*(0,10) |
| μβ*Q*,*h* | mean for β*Q*,*h* | ln(μg/L)/ln(m3/d) | *N*(0,10) |
| μβ*TN* | mean for β*TN* | ln(μg/L)∙L/μg | *N*(0,10) |
| σβ0 | SD for β0 | ln(μg/L) | *tN*(0,10) |
| σβ*Q*,*l* | SD for β*Q*,*l* | ln(μg/L)/ln(m3/d) | *tN*(0,10) |
| σβ*Q*,*h* | SD for β*Q*,*h* | ln(μg/L)/ln(m3/d) | *tN*(0,10) |
| σβ*TN* | SD for β*TN* | ln(μg/L)∙L/μg | *tN*(0,10) |
| σ*a* | residual SD for ln(*a*) | ln(μg/L) | *tN*(0,1) |

Table S3.2 Summary of posterior parameter distributions for Models 1, 2, 3 with calibrated mean values and standard errors in parenthesis. Parameters which include zero in the 95% credible interval are underlined.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Description** | **Units** | **Posterior** |
| *Statistical model* | *Segment* |
| *Upper* | *Middle* | *Bend* |
| β0 | intercept | ln(μg/L) | 4.95 (2.13) | 1.02 (0.93) | -0.94 (0.73) |
| β*Q,l* | ln(*Q*2) before breakpoint | ln(μg/L)/ln(m3/d) | -0.18 (0.15) | 0.04 (0.07) | 0.19 (0.05) |
| β*Q*,h | ln(*Q*2) after breakpoint | ln(μg/L)/ln(m3/d) | -0.75 (0.16) | -1.36 (0.15) | -1.65 (0.24) |
| β*TN* | *TN* | ln(μg/L)∙L/μg | 0.00068 (0.00031) | 0.00130 (0.00038) | 0.00082 (0.00034) |
| *bp* | breakpoint | ln(m3/d) | 15.55 (0.21) | 16.47 (0.13) | 16.99 (0.11) |
| β*T* | *T* | ln(μg/L)/oC | 0.00067 (0.00301) |
| μβ0 | mean for β0 | ln(μg/L) | 1.86 (0.94) |
| μβ*Q*,*l* | mean for β*Q*,*l* | ln(μg/L)/ln(m3/d) | -0.0021 (0.78) |
| μβ*Q*,*h* | mean for β*Q*,*h* | ln(μg/L)/ln(m3/d) | -1.23 (1.32) |
| μβ*TN* | mean for β*TN* | ln(μg/L)∙L/μg | 0.0009 (0.0033) |
| σβ0 | SD for β0 | ln(μg/L) | 4.32 (2.80) |
| σβ*Q*,*l* | SD for β*Q*,*l* | ln(μg/L)/ln(m3/d) | 0.79 (1.28) |
| σβ*Q*,*h* | SD for β*Q*,*h* | ln(μg/L)/ln(m3/d) | 1.59 (2.01) |
| σβ*TN* | SD for β*TN* | ln(μg/L)∙L/μg | 0.0025 (0.0091) |
| σ*a* | residual SD for ln(*a*) | ln(μg/L) | 0.66 (0.01) |
| *Mechanistic* | *Model 2* | *Model 3* |
| *kg* | growth rate | d-1 | 0.81 (0.05) | 0.63 (0.04) |
| θ*g* | temperature correction | *—* | 1.00 (0.01) | 1.01 (0.01) |
| *rna* | ratio of *n* to *a* | μg*n*/μg*a* | 7.76 (0.43) | 13.11 (0.77) |
| *ksn* | half-sat constant, *n* | μg/L | 26.23 (1.38) | 26.70 (1.29) |
| *kr* | loss rate | d-1 | 0.09 (0.01) | 0.06 (0.01) |
| θ*r* | temperature correction | *—* | 1.02 (0.01) | 1.03 (0.01) |
| *km* | recycling rate | d-1 | 0.04 (0.01) | 0.05 (0.01) |
| θ*m* | temperature correction | *—* | 1.06 (0.02) | 1.06 (0.01) |
| ν*s* | settling rate | m/d | 0.25 (0.06) | 0.16 (0.07) |
| *Is* | optimal light level | W/m2 | 40.43 (4.03) | 37.39 (2.59) |
| *kd* | *n* removal rate | d-1 | 0.08 (0.01) | 0.07 (0.01) |
| θ*d* | temperature correction | *—* | 1.03 (0.01) | 0.99 (0.01) |
| σ*n* | residual SD for ln(*n*) | ln(μg/L) | 0.88 (0.02) | 0.89 (0.02) |
| *ksp* | half-sat constant, *p* | μg/L | — | 1.49 (0.09) |
| *rpa* | ratio of *p* to *a* | g*p*/g*a* | 0.53 (0.04) |
| *Pf* | *p* flux from the sediment | μg/m2/d | 357 (62) |
| θ*p* | temperature correction for *Pf* | *—* | 1.37 (0.03) |
| σ*p* | residual SD for ln(*p*) | ln(μg/L) | 0.75 (0.01) |
| σ*a* | residual SD for ln(*a*) | ln(μg/L) | 0.67 (0.01) | 0.68 (0.01) |

## S4 Light limitation of phytoplankton growth and extinction coefficient

Light limitation was estimated using the following formulation for integrated water depth (Chapra, 2008):

|  |  |
| --- | --- |
|  | (S4.1) |

where *f* is a photoperiod (fraction of day) calculated based on day of the year and geographical location (Agafonkin and Thieurmel, 2018), *Is* (W/m2) is optimal light level for phytoplankton growth, *Ia* (W/m2) is average daily photosynthetically available light intensity, *ke* (m-1) is light extinction coefficient.

Daily values of light extinction coefficient, *ke* (m-1) were estimated from Bayesian regression model with varying intercept using chl*-a* concentration and hydrometeorological variables as predictors. As meteorological and flow conditions might have a lagged effect on light extinction coefficient, 2-, 10-, and 60-day averages of wind speed, air temperature and river discharge were included as candidate predictors (other intermediate averaging periods were also considered, but were found to be highly correlated with those listed above). Both untransformed and natural log-transformed discharges were also considered as potential predictors. First, simple linear regressions were built for each of three segments, selecting the predictive variables through a process of an exhaustive search (Lumley, 2017). Best fit models were chosen based on Bayesian information criterion (BIC), which prioritizes predictive performance while penalizing for overfitting (Faraway, 2015). Multiple linear regression models explained 41%, 45% and 49% of variability in light extinction coefficient *ke* in Upper, Middle and Bend segments, respectively and are shown in eqs. S4.2-S4.4.

|  |  |
| --- | --- |
|  | (S4.2) |
|  | (S4.3) |
|  | (S4.4) |

Optimal hydro-meteorological variables and averaging periods were selected based on their frequency of selection in these segment-specific regressions. Then, these variables were used to construct a Bayesian model to predict *ke*, allowing only the intercept term to vary by segment. This approach was employed assuming that light attenuation is dependent on the same variables throughout the estuary, given clustering effects of the different segments (Gelman and Hill, 2007). The resulting hierarchical model allowed for distinguishing between algal and non-algal sources of light attenuation and provided means of *ke* dynamical estimation within the mechanistic models.

The light extinction coefficient is estimated via Bayesian linear regression with varying intercept and chl*-a* and hydrometeorological inputs as covariates. The BIC-selected predictors include chl*-a*, 10-day natural log-transformed flow (ln(*Q*10)), 60-day temperature (*T*60), and 2-day wind speed (*WS*2), and the associated significantly positive coefficients (β*kea*, β*keQ*, β*keTa*, and β*keWS*, respectively) with the mean values showed in eq. S4.5. The segment-specific intercepts decrease from upstream to downstream (mean β*ke*0 are −6.45, −6.65 and −6.96 for Upper, Middle and Bend segments, respectively (Fig. S4.1)). A self-shading effect of phytoplankton increases with higher chl-*a* concentration, reflected via coefficient β*kea* (Fig. S4.1), which magnitude is within the range of previously reported values for other systems (Obrador and Pretus, 2008; Thomann and Fitzpatrick, 1982). The resulting regression was visually checked for normality of residuals, and it explains 50% of *ke* variability in all segments (Fig. S4.2).

|  |  |
| --- | --- |
| . | (S4.5) |



Fig. S4.1 Estimated parameter distributions of the linear regression model with varying intercept for light extinction coefficient *ke*. Y-axis represents relative probability density.



Fig. S4.2 Observed light extinction coefficient *ke* (m-1) vs predicted via Bayesian linear regression model for Upper (red), Middle (orange) and Bend (light-blue) model segments.

## S5 Model performance

Table S5 Full and cross-validated Model 1, 2, 3 summary statistics for ln(chl-*a*).

|  |  |
| --- | --- |
|  | Segment |
| Upper | Middle | Bend | All |
| Model | Full | CV | Full | CV | Full | CV | Full | CV |
| *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) | *R*2 | RMSE,ln(µg/L) |
| 1 | 0.46 | 0.74 | 0.42 | 0.76 | 0.37 | 0.62 | 0.29 | 0.66 | 0.23 | 0.59 | 0.15 | 0.63 | 0.43 | 0.65 | 0.37 | 0.68 |
| 2 | 0.38 | 0.79 | 0.35 | 0.81 | 0.40 | 0.60 | 0.38 | 0.61 | 0.14 | 0.62 | 0.12 | 0.63 | 0.38 | 0.68 | 0.36 | 0.69 |
| 3 | 0.42 | 0.76 | 0.43 | 0.76 | 0.38 | 0.61 | 0.31 | 0.64 | 0.10 | 0.63 | -0.02 | 0.67 | 0.38 | 0.67 | 0.35 | 0.70 |



Fig. S5 chl-*a* simulation from Model 3 for 2003, including median predictions (black line), and 90% predictive interval associated with parameter and residual uncertainty (light grey). Blue x’s and error bars represent the mean and range of observations.

## S6 Probability of exceeding chl-*a* criterion of 40 µg/L



Fig. S6 Median probability of chl-*a* exceeding 40 µg/L in any day of the year predicted via Model 3 as a function of variation in loading (both TP and TN), shown for three segments, four seasons, and three hydrologic conditions.

# References

Agafonkin, V., Thieurmel, B., 2018. suncalc: Compute Sun Position, Sunlight Phases, Moon Position and Lunar Phase.

Chapra, S., 2008. Surface Water-Quality Modeling. Waveland Press, Inc.

Faraway, J.J., 2015. Linear Models with R, 2nd ed. Taylor & Francis Group.

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models, Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.

Lumley, T., 2017. leaps: Regression Subset Selection. Based on Fortran code by Alan Miller.

Obrador, B., Pretus, J.L., 2008. Light regime and components of turbidity in a Mediterranean coastal lagoon. Estuar. Coast. Shelf Sci. 77, 123–133. https://doi.org/10.1016/j.ecss.2007.09.008

Thomann, R., Fitzpatrick, J., 1982. Calibration and verification of a mathematical model of the eutrophication of the Potomac Estuary. New Jersey. https://doi.org/10.13140/2.1.2533.1526