University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Publications, Agencies and Staff of the U.S. Department of Commerce

US Department of Commerce

1-1-2011

PREDICTING THE HYPOXIC-VOLUME IN CHESAPEAKE BAY WITH THE STREETER–PHELPS MODEL: A BAYESIAN APPROACH

Yong Liu
Peking University, yongliu@pku.edu.cn

George B. Arhonditsis
University of Toronto

Craig A. Stow
NOAA Great Lakes Environmental Research Laboratory

Donald Scavia
University of Michigan - Ann Arbor

Follow this and additional works at: http://digitalcommons.unl.edu/usdeptcommercepub

Part of the Environmental Sciences Commons

Liu, Yong; Arhonditsis, George B.; Stow, Craig A.; and Scavia, Donald, "PREDICTING THE HYPOXIC-VOLUME IN CHESAPEAKE BAY WITH THE STREETER–PHELPS MODEL: A BAYESIAN APPROACH" (2011). *Publications, Agencies and Staff of the U.S. Department of Commerce.* Paper 288.
http://digitalcommons.unl.edu/usdeptcommercepub/288

This Article is brought to you for free and open access by the US Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

AMERICAN WATER RESOURCES ASSOCIATION

PREDICTING THE HYPOXIC-VOLUME IN CHESAPEAKE BAY WITH THE STREETER–PHELPS MODEL: A BAYESIAN APPROACH[1]

Yong Liu, George B. Arhonditsis, Craig A. Stow, and Donald Scavia[2]

ABSTRACT: Hypoxia is a long-standing threat to the integrity of the Chesapeake Bay ecosystem. In this study, we introduce a Bayesian framework that aims to guide the parameter estimation of a Streeter–Phelps model when only hypoxic volume data are available. We present a modeling exercise that addresses a hypothetical scenario under which the only data available are hypoxic volume estimates. To address the identification problem of the model, we formulated informative priors based on available literature information and previous knowledge from the system. Our analysis shows that the use of hypoxic volume data results in reasonable predictive uncertainty, although the variances of the marginal posterior parameter distributions are usually greater than those obtained from fitting the model to dissolved oxygen (DO) profiles. Numerical experiments of joint parameter estimation were also used to facilitate the selection of more parsimonious models that effectively balance between complexity and performance. Parameters with relatively stable posterior means over time and narrow uncertainty bounds were considered as temporally constant, while those with time varying posterior patterns were used to accommodate the interannual variability by assigning year-specific values. Finally, our study offers prescriptive guidelines on how this model can be used to address the hypoxia forecasting in the Chesapeake Bay area.

(KEY TERMS: hypoxia; Chesapeake Bay; Bayesian inference; Markov chain Monte Carlo; Streeter–Phelps model; uncertainty analysis; eutrophication.)

Liu, Yong, George B. Arhonditsis, Craig A. Stow, and Donald Scavia, 2011. Predicting the Hypoxic-Volume in Chesapeake Bay with the Streeter–Phelps Model: A Bayesian Approach. *Journal of the American Water Resources Association* (JAWRA) 1-15. DOI: 10.1111/j.1752-1688.2011.00588.x

INTRODUCTION

Hypoxia (dissolved oxygen [DO] ≤2 mg/l) is a serious problem for estuaries in the United States (Bricker *et al.*, 2007) and has received considerable scientific and policy attention (e.g., NRC, 2000; Diaz and Rosenberg, 2008). Low oxygen concentrations in the Chesapeake Bay were first reported in the 1930s, with a significant increase in severity and spatial extent observed since the 1950s (Newcombe and Horne, 1938; Officer *et al.*, 1984; Cooper and Brush, 1991; Boesch *et al.*, 2001). The causes and consequences of Chesapeake Bay hypoxia have been the focus of research and policy action, including a series of intergovernmental agreements and nutrient

[1]Paper No. JAWRA-10-0127-P of the *Journal of the American Water Resources Association* (JAWRA). Received August 16, 2010; accepted June 7, 2011. © 2011 American Water Resources Association. **Discussions are open until six months from print publication**.

[2]Respectively, Research Professor, College of Environmental Science and Engineering, The Key Laboratory of Water and Sediment Sciences Ministry of Education, Peking University, Beijing 100871, China; Associate Professor, Department of Physical and Environmental Sciences, University of Toronto, Toronto, Canada M1C 1A4; Senior Research Scientist, NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan 48105-2945; and Professor, School of Natural Resources & Environment, University of Michigan, Ann Arbor, Michigan 48109 (E-Mail/Liu: yongliu@pku.edu.cn).

This article is a U.S. government work, and is not subject to copyright in the United States.

reduction efforts (Boesch *et al.*, 2007). An analysis of the relationship between hypoxia and nitrate loading for two periods, 1950-1979 and 1980-2001, demonstrated that the Bay's ability to assimilate nitrogen inputs was reduced in the latter period (Boesch *et al.*, 2001; Kemp *et al.*, 2005). This loss of assimilative capacity may indicate a state change, implying that disproportionately large nitrogen load reductions may be required to bring hypoxia under control (Scheffer and Carpenter, 2003).

For the Chesapeake Bay, and elsewhere, models have become important tools for guiding decisions on nutrient load reduction (NRC, 2000; Arhonditsis and Brett, 2004; Arhonditsis *et al.*, 2006). In developing and assessing such models, there are three key issues that need to be addressed before the models can be used for decision making: (1) establishing a reliable and cost-effective model; (2) estimating model parameter values; and (3) assessing model uncertainties. Several types of models have been developed for the Chesapeake Bay, ranging from complex, mechanistic models (e.g., Cerco and Cole, 1993; Cerco and Noel, 2004) to simple statistical models (Hagy *et al.*, 2004; Kemp *et al.*, 2005); each with features that make them differentially amenable to effectively addressing these three issues. Complex models can be used to organize explicit, detailed assumptions about the relationships among important ecosystem processes and to test the logical consequences of those assumptions under different circumstances. However, complex models are usually over-parameterized, requiring modelers to adopt a combination of literature values and expert judgment to select key parameter values. Additionally, complex models usually carry a high computational demand, which inevitably makes it difficult to conduct thorough sensitivity or uncertainty analyses. Alternatively, statistical models are computationally more efficient, but generally do not resolve physical, ecological, and biogeochemical processes. By acknowledging that models at both ends of the complexity spectrum have different strengths and weaknesses, simple biophysically based water quality models, tested against observations, were recommended by the National Research Council (NRC, 2001) as a cost-effective way to inform decision making.

Consistent with that recommendation, Scavia *et al.* (2006) used a simple process-based approach, the Streeter–Phelps equation (Streeter and Phelps, 1925), to model Chesapeake Bay DO conditions. Scavia *et al.* (2006) used fixed *a priori* parameter values based on a combination of available information and expert judgment, and the uncertainty associated with two parameters (initial DO deficit and downstream advection) was estimated using Monte Carlo simulations. Stow and Scavia (2009) introduced a Bayesian configuration of the Chesapeake Bay model to estimate DO profiles. Their work demonstrated that the Bayesian approach is useful in a decision context because it can combine both expert judgment and rigorous parameter estimation to yield model forecasts along with the underlying uncertainty.

While the Streeter–Phelps model has been shown useful to predict longitudinal DO profiles in Chesapeake Bay, goals for hypoxia management are defined in terms of reducing the volume of hypoxic water. Thus, it may not be necessary to predict the full DO profile if the hypoxic volume can be accurately estimated. Additionally, in some systems such as the northern Gulf of Mexico, although hypoxic area is estimated annually, detailed DO profile measurements are not generally available for calibrating the Streeter–Phelps model. Even without this information the Streeter–Phelps model has been successfully used to predict Gulf of Mexico hypoxia, but with calibration directly to hypoxic area rather than to full DO profiles (Scavia *et al.*, 2003, 2004; Scavia and Donnelly, 2007).

In this study, we introduce a Bayesian framework to guide the parameter estimation of a Streeter–Phelps model when only hypoxic volume data are available. In particular, we use the Chesapeake Bay as a case study to present a modeling exercise that addresses a hypothetical scenario under which the only information available is hypoxic volume estimates. Our analysis evaluates the credibility of the proposed statistical approach by undertaking a series of modeling experiments of individual and joint parameter estimation, whereby the model parameterization and the associated predictive uncertainty are compared to those derived from conventional Streeter–Phelps model fitted to DO profiles. The purpose of this study is to examine to what extent the proposed Bayesian framework is suitable (1) to achieve realistic parameter values and acceptable predictive capacity; (2) to gain insights into possible regime shifts or other structural changes that may have occurred during the study period; and (3) to dictate areas where future model augmentations should focus on. Finally, we offer prescriptive guidelines on how this model can be used to address the hypoxia forecasting in the Chesapeake Bay.

MATERIALS AND METHODS

Data Source

Data used in this study are the same as reported in Scavia *et al.* (2006) for oxygen concentration profiles and hypoxic volumes along the main stem of the

Chesapeake Bay for 36 years from 1950 to 2003. There were 137 values reported for the DO profiles for each of the 36 years, which were computed by interpolating observations to populate a regular grid with dimensions, first at 1-m resolution in the vertical and then at 1 km in the horizontal direction across constant depths (Hagy *et al.*, 2004). The January-May average total nitrogen loads in 1945-1978 from the Susquehanna River were computed using empirical relationships from corresponding nitrate loads at Harrisburg, Pennsylvania (Hagy *et al.*, 2004). The 1979-2004 loads were from USGS, based directly on frequent measurements of Susquehanna River total nitrogen concentrations at Conowingo and were derived using the seven-parameter log-linear regression model described by Cohn *et al.* (1989).

Model Equations

The model is an adaptation of the Streeter–Phelps river model which predicts downstream oxygen concentrations from point sources of organic matter loads (Chapra, 1997). Chesapeake Bay is stratified vertically in summer, with surface waters flowing seaward and bottom waters flowing landward. As such, a key assumption for this application is that watershed nutrient loads can be used to approximate a point source of organic matter deposited to sub-pycnocline depths at the southern end of the mid-Bay region (37°48′N, 76°W; c. 220 km from the Susquehanna River mouth), and that the organic matter decays as it is transported up estuary below the pycnocline. The nutrient nitrogen load was converted to organic carbon via the Redfield C:N ratio (106:16 or 5.67 g C/g N), and then converted to a biologic oxygen demand via O_2:C (0.9, or 2.4 g O_2/g C). The validity of this assumption was first examined in Scavia *et al.* (2006) and was found to result in realistic representation of the ecological processes underlying the system dynamics. The steady-state solution used in Scavia *et al.* (2006) and Stow and Scavia (2009) for Chesapeake Bay DO profiles along downstream distance can be written as:

$$\mathrm{DO} = \mathrm{DO_s} - \frac{k_\mathrm{d}\mathrm{BOD_u}(F)}{k_\mathrm{r} - k_\mathrm{d}}\left(e^{-k_\mathrm{d}\frac{x}{v}} - e^{-k_\mathrm{r}\frac{x}{v}}\right) - D_\mathrm{i}e^{-k_\mathrm{r}\frac{x}{v}} + \varepsilon \tag{1}$$

where DO is the measured DO concentration (mg/l), $\mathrm{DO_s}$ is the saturation oxygen concentration (mg/l), k_d is the BOD decay coefficient (1/day), k_r is the reaeration coefficient (1/day), $\mathrm{BOD_u}$ is the ultimate BOD (mg/l), F is the fraction of surface organic carbon production that settles below the pycnocline, x is the downstream distance (km), D_i is the initial DO deficit at x = 0 (mg/l), and ε is model error. In the original Streeter–Phelps formulation, v represents net downstream advection. However, earlier predictions of the Chesapeake Bay DO profiles were based on variant v across years, while other parameters were retained constant. The selection of the parameter v, as opposed to the cross-pycnocline flux or the sub-pycnocline decay rate parameters, was justified by three basic reasons: (1) it was assumed that the other two parameters were much less likely to vary across time, (2) there were good *in situ* measurements to compare those modeled flux rates with observations, and (3) the original physical meaning of v (i.e., downstream advection) in the steady-state application of this river model has little meaning (Scavia *et al.*, 2006). In this regard, the role of v is conceptually similar to a time-variant structural error, and therefore can potentially account for the impact of un-modeled ecological processes and may accommodate structural shifts in the ecosystem functioning that have occurred during the study period. Aggregating those uncertainties into one coefficient allowed their characterization through Monte Carlo simulations in previous applications (Scavia *et al.*, 2003; Scavia and Donnelly, 2007). In the present study though, we will revisit the assumption of a temporally varying v. The impact of the alternative approach (i.e., temporally constant v) on the model parameterization will be examined, along with the predictive uncertainty. It should also be noted that in this application k_r reflects oxygen flux across the pycnocline, not with the atmosphere. A multiplier (K) was added to equation (1), replacing k_r with $k_\mathrm{r}{*}K$, to scale the previously calculated cross-pycnocline flux coefficient k_r which varies spatially along the system (Hagy *et al.*, 2004), but not across years.

The observed annual hypoxic volumes V (km^3) in the system were linked with the predicted hypoxic length L_m (km), using the regression equation originally derived from existing measurements in the system (Scavia *et al.*, 2006), i.e.,

$$\begin{aligned} &\text{If} \quad \mathrm{DO_{mh}} < 2\ mg/L \quad \text{then } \mathrm{w_h} = 1, \quad \text{else } \mathrm{w_h} = 0, \\ &L_m = \sum_{h=1}^{137} l_h w_h \\ &\mathrm{V} = 0.00391\ L_m^2 + \varepsilon_v \end{aligned} \tag{2}$$

where l_h is length of the segment h in which the predicted oxygen concentration (DO_{mh}) is below 2 mg/l; ε_V is the error term that reflects the discrepancy between the "observed" hypoxic volume estimates and those predicted from our model. In this study, we will examine the differences in the parameter

estimation and the predictive uncertainty that may arise when the calibration exercise is founded upon the hypoxic volume data and the DO concentrations are not part of the fitting process. The hypoxic volume fitting is essentially an *ad hoc* feature to the DO profile model, and thus the predictive uncertainty associated with the latter model is propagated through the former one. Yet, our analysis addresses the hypothetical scenario under which DO data do not exist and the corresponding error cannot be quantified, but rather is treated as an implicit component of the fitting exercise.

Bayesian Framework for Hypoxia Modeling

Bayesian statistics provide rigorous methods for uncertainty analysis and key information for decision making in the context of environmental management (Reckhow, 1994; Stow *et al.*, 2006). All unknown parameters θ are treated as random variables and their distributions are derived from past information and current observations from the system (Borsuk *et al.*, 2001). Bayesian inference is based on Bayes' Theorem (Gill, 2002):

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int_\theta p(\theta)p(y|\theta)d\theta} \propto p(\theta)p(y|\theta) \tag{3}$$

where $p(\theta|y)$ is the posterior probability of θ, which is the conditional distribution of the parameters after observation of the data; θ is the estimated parameter; $p(\theta)$ is the prior probability of θ; $p\ (y|\theta)$ is the likelihood function, which represents the probability for the occurrence of the observations y given different realizations of the postulated mechanistic relationship between the response and predictor variables, i.e., Equation (1).

Model Fitting to DO Profiles. Transferring equation (1) into Bayes' Theorem, the error term ε was assumed to be normally distributed with zero mean and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$. The likelihood function of equation (1) can be written as:

$$\begin{aligned} &\mathrm{DO_{gh}} \sim N(\mu_{\mathrm{gh}}, \sigma^2) \\ &\mu_{gh} = \mathrm{DO}_S - \frac{k_{\mathrm{d}}\mathrm{BOD_u}(F_g)}{k_{\mathrm{r}} - k_{\mathrm{d}}}\left(e^{-k_{\mathrm{d}}\frac{xg}{v}} - e^{-k_{\mathrm{r}}\frac{xg}{v}}\right) - D_i e^{-k_{\mathrm{r}}\frac{xg}{v}} \end{aligned} \tag{4}$$

where h represents the 137 observed DO estimates along the transect; and g represents the 36 years

Model Fitting to Hypoxic Volume Data. The likelihood function for hypoxic volume, with the error term assumed to be normally distributed with zero mean and variance of σ_V^2 is:

$$\begin{aligned} &\text{If } \mu_{\mathrm{gh}} < 2 \text{ mg/L} \quad \text{then } \mathrm{w_{gh}} = 1, \quad \text{else } \mathrm{w_{gh}} = 0 \\ &L_{mg} = \sum_{h=1}^{137} l_{gh} w_{gh} \\ &V_g \sim N(0.0039\, L_{mg}^2, \sigma_V^2) \end{aligned} \tag{5}$$

In this analysis, model parameters were scrutinized in two stages: (1) we started with values used by Scavia *et al.* (2006) and systematically relaxed the fixed prior values for individual parameters; and then (2) we formed parameter subsets (refer to the Results section for details) to examine if the joint estimation of some parameters can further improve model performance while keeping the uncertainty of the underlying model predictions at realistic levels. Selected parameters were allowed to differ by year using a Bayesian structure, founded upon the assumption that the year-specific estimates arise from a common normal distribution. For example, for a selected parameter θ, the year-specific values θ_i were all derived from a normal distribution of the form: $\theta_i \sim N(\theta_\mathrm{mean}, \theta.\mathrm{sd})$, in which θ_mean and $\theta.\mathrm{sd}$ represent the parameters of the common normal prior. The θ_mean values were set equal to the averages of the estimated values from Stow and Scavia (2009), while the ranges between the corresponding maximum and minimum values were assumed to contain 95% of the prior probability mass. Using this approach, we obtained the following informative normal priors: D_i ~N(0.30,1.20); F ~N(0.75,0.13)I(0,1); K ~N(0.6, 0.20)I(0,1); k_d ~N(0.11,0.05)I(0,); and v ~N(2.5, 0.77) I(0,), in which the numbers in the brackets represent the mean and the standard deviation of the corresponding normal distributions; and I(,) denotes the censoring imposed to eliminate negative values (and values greater than one) during the Bayesian updating of the model. With the parameters K and F, we alternatively used a number of beta distributions that allocated a greater probability mass to parameter values higher than 0.5, and the results remained practically unaltered. It should also be clarified that the fact that these informative priors are loosely based on the parameter estimates presented by Stow and Scavia (2009) may seem that we are "double dipping" into the same dataset. Yet, the actual practice followed merely aimed to avoid unnecessary divergence of the Markov chain Monte Carlo (MCMC) sampling into ecologically unrealistic regions of the parameter space that were unlikely to hold true in the real world. Namely, the informative priors for

each parameter were based on the corresponding values across all the years of the study period, which represented a quite broad range (see Figures 2-4 of Stow and Scavia, 2009). From our experience, relatively similar priors would have been derived, even if we had reviewed the pertinent literature and subsequently used the framework presented by Arhonditsis *et al.* (2007). Thus, what may be criticized as "double counting" of the same piece of information is actually a formal framework of using existing knowledge to achieve an ecologically meaningful foundation of our model.

Markov chain Monte Carlo is a method to draw samples from multidimensional distributions for numerical integration (Qian *et al.*, 2003). The idea underlying the MCMC implementation in Bayesian inference is to construct a Markov process whose stationary distribution is the model posterior distribution, and then run the process long enough to produce an accurate approximation of this distribution (Malve and Qian, 2006). There are many methods (e.g., Gibbs sampler) for obtaining sequence of realizations from the posterior model distributions (Gelfand and Smith, 1990), but all of them are special cases of the general Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970). We used WinBUGS (version 1.4.3; Lunn *et al.*, 2000), called from R (version 2.6.0; R2WinBUGS version 2.1-8; Gelman and Hill, 2007). The MCMC sampling was carried out using four chains, each with 20,000 iterations (first 10,000 discarded after model convergence); and samples for each unknown quantity were taken from the next 10,000 iterations using a thin equal to 40 to reduce serial correlation. Statistical inference was based on the resulting 1,000 MCMC samples. A potential scale reduction factor, *Rhat*, was produced in package R2WinBUGS to determine the model convergence (at convergence, Rhat = 1.0). *Rhat* is approximately equal to the square root of the variance if the mixture of all the chains divided by the average within-chain variance; if it is greater than 1.0, the chains have not mixed well (Gelman and Hill, 2007).

Three measures of fit were combined to test and compare model results: (1) the deviance information criterion (DIC), a Bayesian measure of model fit that penalizes the complexity of the model structure, and therefore a lower DIC value suggests a more favorable model (Spiegelhalter *et al.*, 2002); (2) model residual standard error; and (3) the coefficient of determination, R^2, between the original data and predicted mean values. The former measure enables the selection of the most parsimonious models that effectively balance between model-fit and complexity, while the R^2 was mainly selected because of its common use in the modeling practice for assessing goodness-of-fit.

For a Bayesian model with data y, unknown parameters θ, and the likelihood function $p(y \mid \theta)$, the deviance is defined as (Spiegelhalter *et al.*, 2002):

$$D(\theta) = -2\log[p(y|\theta)] + c \tag{6}$$

where c is a constant. The effective number of parameters in the model is:

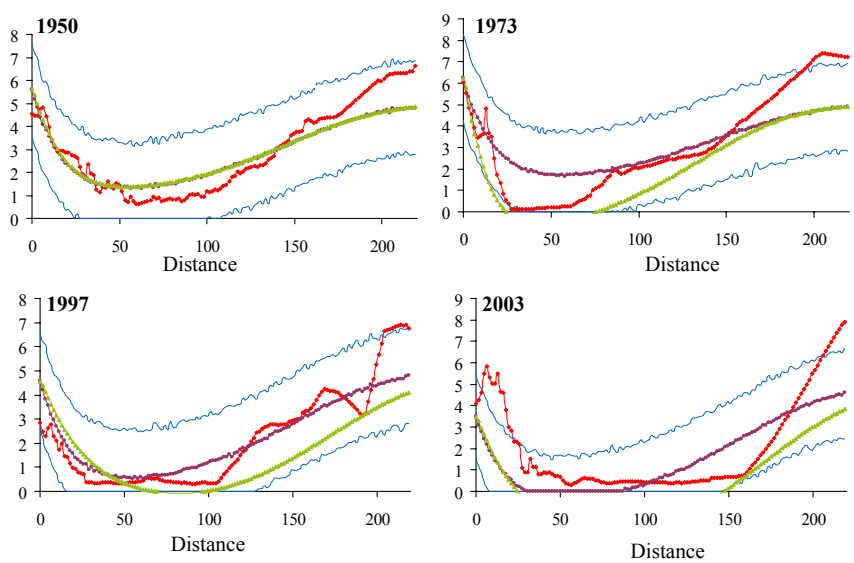

FIGURE 1. Comparison Between Observed (red line) and Predicted Mean (purple line) Dissolved Oxygen (DO) Profiles in Chesapeake Bay Using Year-Specific *K* Estimates. The blue lines represent the 95% credible intervals. The green line corresponds to the DO profile predicted from Scavia *et al.* (2006). Distance is kilometers downstream from the Susquehanna River.

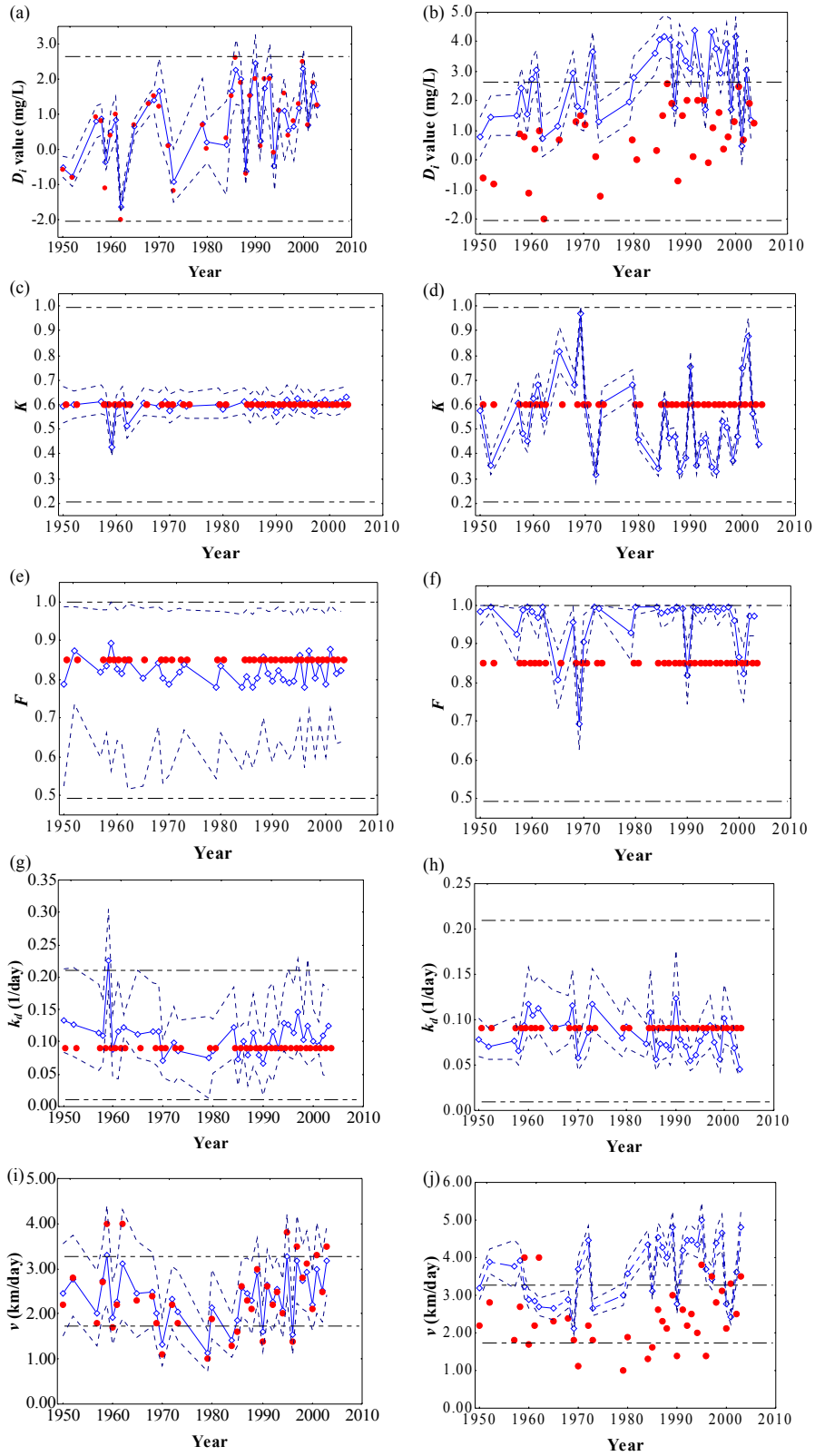

FIGURE 2. Comparison of the Year-Specific Estimates for the Parameters D_i, K, F, k_d, and v Fitted to the Hypoxic Volume (left panels) and the Dissolved Oxygen (DO) Profile Data (right panels). The red circles correspond to the parameter values reported in Scavia *et al.* (2006). The blue diamonds and dashed lines correspond to the posterior mean values along with the 95% credible intervals. The two dashed horizontal lines represent the 95% credible intervals of the prior distributions.

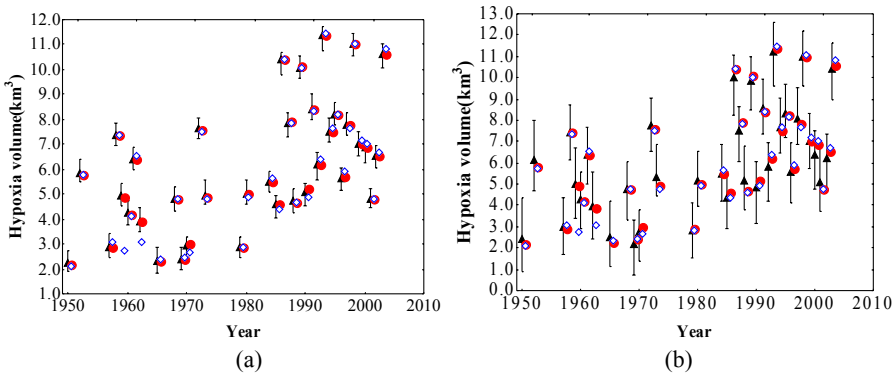

FIGURE 3. Comparison Between Observed (red circles) and Predicted Mean Hypoxic Volume (black triangles) Values in Chesapeake Bay Using Year-Specific Estimates for the Parameters (A) D_i, F and (B) D_i, k_d, and F. The lower and upper bars represent the 2.5 and 97.5% credible bounds. The blue diamonds correspond to the predicted values in Scavia *et al.* (2006).

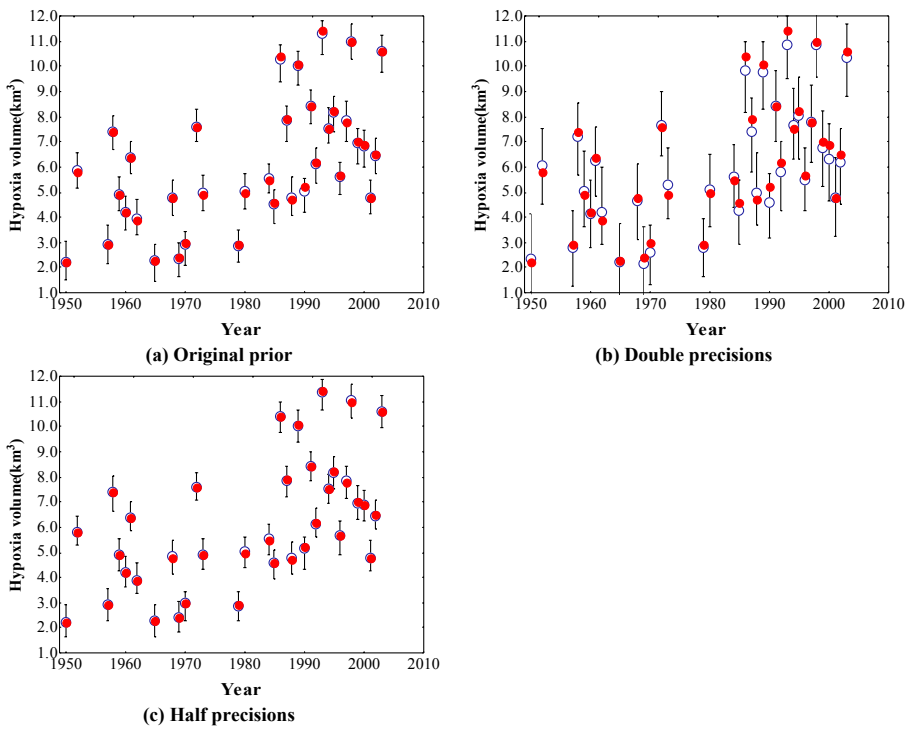

FIGURE 4. Sensitivity of the Hypoxic Volume (km^3) Predictions to the Prior Distributions Assigned to the Parameter D_i. the Original Prior Based on the Informative Normal Distribution of D_i ~N(0.30,0.70) Was Examined With Double (1.40) and Half Precision (0.35).

$$pD = \overline{D(\theta)} - D(\overline{\theta}) \tag{7}$$

where $\overline{\theta}$ is the expectation of θ. $\overline{D(\theta)}$ is the expectation of $D(\theta)$:

$$\overline{D(\theta)} = E^{\theta}[D(\theta)] \tag{8}$$

Moreover, DIC is calculated as:

$$DIC = pD + \overline{D(\theta)} \tag{9}$$

Thus, with this Bayesian model comparison, we first assessed model fit or model "adequacy" (sensu Spiegelhalter *et al.*, 2002), $\overline{D(\theta)}$, and then we penalized complexity, pD.

R^2 is defined as (Gelman and Hill, 2007):

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{10}$$

where, SS_E and SS_T are the sum of squared errors and total sum of squares, respectively; y_i and y'_i are the original data and predicted mean values, respectively; $\overline{y_i}$ is the mean of the observations y_i; and n is the number of observations. It should be noted though that the Bayesian approach generates a predictive distribution and not a single value for each variable (y'_i), and thus the use of R^2 is essentially a non-Bayesian ("point") assessment of the model performance. For the same reason, the posterior mean values of the model error terms σ and σ_V should not be confused with the point estimate of the error derived when using the mean model outputs (e.g., an $R^2 = 1$ does not imply that the mean value of σ or σ_V is zero).

TABLE 1. Goodness-of-Fit of the Hypoxic Volume and Dissolved Oxygen (DO) Profile Models Using Year-Specific Estimates for the Parameters D_i, v, K, k_d, and F.

Estimated Parameter	R^2	Model Residual Standard Error		
		Mean	2.5%	97.5%
Hypoxic Volume Model				
D_i	1.00	0.20	0.06	0.57
v	0.97	1.27	0.79	1.94
K	1.00	0.58	0.11	0.97
k_d	0.99	0.13	0.07	0.22
F	0.98	1.71	1.22	2.36
DO Profile Model				
D_i	0.56	1.50	1.47	1.53
v	0.76	1.20	1.17	1.22
K	0.78	1.03	1.01	1.05
k_d	0.64	1.69	1.65	1.72
F	0.75	1.21	1.19	1.24

RESULTS

Individual Parameter Estimation

In the first phase of our analysis, we selected five parameters to estimate them individually with the DO profile and the hypoxic volume datasets. In the former case, we fitted the model to the DO observations along the system, while the latter inverse solution exercise was only based on the observed annual hypoxic volume estimates. For these modeling experiments, one parameter was estimated and the other four were assigned the values of an earlier calibration of the model presented by Scavia *et al.* (2006). We then evaluated the relative influence of each parameter on model performance to determine the optimal parameter combinations that will be examined in the second phase of our analysis. D_i, v, K, k_d, and F are the model parameters being estimated, while DO_s and BOD_u represent error-free boundary conditions and inputs of the model, respectively.

For the hypoxic volume data, the use of year-specific D_i, k_d, and K estimates provided the lowest model residual standard error and highest R^2 values (Figure 1 and Table 1). Further, the derivation of year-specific posteriors for v and F was also associated with broader prediction uncertainty bands of the hypoxic volume levels, while k_d resulted in the narrowest (Supporting Information). A different way of viewing this result is that there is a wider range of predicted outputs that have been assigned "some" likelihood during the updating of the v and F distributions. The fact that the MCMC search resulted in a "flatter" predictive posterior can conceivably be explained by – at least – two factors: (1) an erroneous prior specification of the relative plausibility of the different parameter values that may lead to an exploration of suboptimal regions of the model posterior; and/or (2) the relationship between the specific parameters and the corresponding model output is steep and approximately linear (or even higher order) within the parameter range examined and as such every incremental change is accompanied by a significant change of the hypoxic volume predictions during the MCMC sampling. The latter finding can be interpreted as evidence that the model is more sensitive to the prior specifications of the parameters v and F, and that the predictive uncertainty can be significantly controlled by formulating more articulate priors for these parameters based on site-specific information.

When the model was fitted to the DO profiles, the year-specific K estimates resulted in the lowest model residual standard error, and highest R^2 values, while the assumption of a year-specific D_i gave the worst fit (Table 1). In all the five cases, most DO observations are included within the 95% credible intervals (e.g., Figure 1 for the case of year-specific K estimation). Posterior predictions and the deterministic solutions presented by Scavia *et al.* (2006) were comparable for some years (e.g., 1950, 1987, 1999, 2002), but the deterministic approach resulted in notably lower DO profiles for others (e.g., 1973, 1997, 2003). By contrast, the Bayesian 2.5 and 97.5% uncertainty bounds include most of the observations as well as Scavia *et al.*'s (2006) predictions.

To gain insight into how the inference changes depending on the data used to calibrate the model, we compared the parameter estimates between the model fit against the DO profile and the hypoxic volume data (Figure 2). Among the five parameters, the year-specific K estimates have the narrowest uncertainty bands when using the hypoxic volume data to

parameterize the model, and the same was true for k_d with both the hypoxic volume and DO profile datasets. Parameter estimates for the hypoxic volume data were comparable to those used in Scavia *et al.* (2006); especially those derived for D_i (Figure 2a), K (Figure 2c), and v (Figure 2i). By contrast, the F and k_d posterior estimates were correspondingly higher and lower compared to the values reported by Scavia *et al.* (2006). We also note the relatively constant F, k_d, and K values over time when fitting the model to the hypoxic volume data, which renders support to the assumption of keeping them constant across the 36 years of the study period.

Figure 2 also illustrates more dramatic differences between the posterior estimates from the DO profile data and those derived from the original work (Scavia *et al.*, 2006). The posterior year-specific D_i estimates have temporal patterns similar to those reported in Scavia *et al.* (2006) but with generally higher values (Figure 2b). Most F posteriors were lying within the 0.90-1.00 range, with main exceptions being the 1965 and 2001 estimates (Figure 2f). The posterior estimates for K vary between 0.30 and 0.70; roughly around the constant 0.60 reported in Scavia *et al.* (2006) (Figure 2d). Similarly, the original k_d value of 0.09 falls within the 0.04–0.13 range derived from the Bayesian updating (Figure 2h). Bayesian-estimated year-specific v values are higher than the 1.0–4.0 values reported in Scavia *et al.* (2006), with a notably consistent temporal pattern that ranges between 2.4 and 4.9, that is, a systematic decrease in the early period was followed by a consistent increase in the most recent years (see also following discussion) (Figure 2j).

The discrepancy between the prior and posterior distributions offers insights into the effects of the Bayesian updating on the parameter estimation process (Arhonditsis *et al.*, 2007). That is, how much can we learn for the different parameters when having the hypoxic volume dataset at hand? We found that the coefficients of the variation (or "relative standard deviation", CV = standard deviation/mean value) of the posterior distributions were significantly reduced for all the parameters considered, but there was substantial difference among the parameters. Notably, the D_i and K posterior CV values were reduced from 3.98 to 0.396 and from 0.33 to 0.042, respectively. The latter result reflects the significant reduction of the D_i (74%) and K (89%) standard deviations. For F, k_d, and v, we found a relatively smaller reduction of the parametric uncertainty after the model updating process. Namely, the F, k_d, and v CV values were reduced from 0.170, 0.455, and 0.307 to 0.118, 0.265, and 0.167, respectively. While the differences in the uncertainty reduction among the different parameters partly reflect the soundness of their prior characterization, these results can also be used to dictate which parameters of our mathematical model can be reasonably delineated through an inverse solution exercise and which ones warrant direct estimation in the field (assuming that they represent real ecological process and they are not "effective" parameters used for modeling purposes).

Joint Parameter Estimation With the Hypoxic Volume Dataset

The lessons learned from the individual parameter estimation were used to optimize the size of the calibration vector examined, thereby avoiding an over-inflation of the predictive uncertainty during the joint parameter estimation exercise. The main criterion used to decide which parameters should be kept constant in the subsequent numerical experiments of joint parameter estimation was primarily the stability of the central tendency of the posterior parameter distributions and secondarily the "narrowness" of the corresponding credible (uncertainty) intervals. That is, we selected parameters with relatively similar most likely values over time and high confidence (or low uncertainty) about these values. Based on the previous results, the parameter K appears to fully satisfy this criterion when using the hypoxic volume data. Relatively stable posterior means were also found for the parameter F, but the corresponding credible intervals were quite wide. The parameter k_d was another candidate for being treated as constant that has the additional advantage of maintaining the trend of temporal stability with both the hypoxic volume and the DO data. Yet, the high model error (and relatively low R^2) resulting from the latter case (see Table 1) suggests that the values assigned to the rest of the parameters restricted this univariate search in a suboptimal region of the model posterior and thus our exercise may not reveal the full range of dynamics associated with the parameter k_d. Because we also assumed that the scaling factor K likely demonstrates less temporal variability, we selected this parameter over the sub-pycnocline decay rate k_d to remain constant for the subsequent numerical experiments. In particular, we assigned the original value derived by Scavia *et al.* (2006) when estimating different combinations of the remaining parameters (D_i, v, k_d, F).

Subsequently, we conducted nine numerical experiments: (1) joint estimation of year-specific D_i and F values; (2) joint estimation of year-specific D_i, k_d, and F values; (3) joint estimation of year-specific D_i, v, and F values; (4) joint estimation of year-specific v and temporally constant D_i, F values; (5) joint estimation of year-specific F and temporally constant D_i, v values; (6) joint estimation of year-specific D_i and

temporally constant F, v values; (7) joint estimation of year-specific k_d and temporally constant F, D_i values; (8) joint estimation of year-specific F and temporally constant k_d, D_i values; and (9) joint estimation of year-specific D_i and temporally constant k_d, F values. In all cases, the parameter not being part of the calibration exercise was set equal to the value used in Scavia *et al.* (2006).

Among the nine experiments, the R^2, DIC, and model residual standard error values indicate that experiments (1), (3), and (9) were associated with the best model performance (Table 2, Figure 3, and Supporting Information). Notably, when using the DIC as the primary criterion, the ninth experiment appears to overwhelmingly outperform the rest of the parameter combinations examined, while the fifth one resulted in the worst fit to the observed data. The latter finding challenges the practice of treating the parameters v and D_i as temporally constant, suggesting that it is certainly more appropriate to base our hypoxic volume predictions on year-specific estimates for these particular parameters. We also compared the posterior estimates from the best performing parameter combinations with those reported by Scavia *et al.* (2006). The purpose was to explore the variability of the parameter estimates and to determine which ones should be retained constant in future modeling exercises. For example, the joint estimation of year-specific D_i and F (first experiment) provided D_i values higher than in Scavia *et al.* (2006), while the F estimates were generally around the 0.85 value reported in Scavia *et al.* (2006). When estimating year-specific D_i, v, and F simultaneously (third experiment), the predicted annual hypoxic volumes are quite close to the observed ones based again on F values that range around the 0.85 level, although the D_i and v values differ significantly from those reported in Scavia *et al.* (2006). Joint estimation of year-specific D_i and temporally constant k_d and F values (ninth experiment) provides k_d and F posterior estimates that are slightly lower and year-specific D_i values that are higher compared to Scavia *et al.* (2006). Other interesting comparisons are provided in the Supporting Information section.

Robustness Analysis to the Specification of the Parameter Priors

We examined the robustness (i.e., sensitivity) of the hypoxic volume predictions to the specification of the parameter priors by doubling and cutting in half the precision of the original priors. For illustration purposes, two cases are presented herein: estimation of D_i alone (Figure 4) and joint estimation of year-specific v with temporally constant D_i and F values (Figure 5). When doubling the precision of the parameter priors (i.e., cutting the variance in half), the parameter space becomes narrower and thus the MCMC sampling focuses on suboptimal regions. Consequently, the overall model performance became relatively worse, as indicated by the R^2 values and the wider credible intervals of the predicted hypoxic volumes. When the precisions were reduced (i.e., increasing the variance), the prior parameter space was expanded, which in principle increases the odds of locating the global optima of the model. The fact that there was no significant improvement of model performance increases our confidence that the original specification of the informative priors did permit sufficient coverage of the model likelihood and that the posterior inferences drawn herein were not biased from the selection of the parameter priors.

DISCUSSION

Given the growing evidence that hypoxia is perhaps one of the greatest threats to the integrity of

TABLE 2. Goodness-of-Fit of the Hypoxic Volume Model Based on Different Parameter Combinations.

			Model Residual Standard Error		
Estimated Parameters	R^2	**DIC**	**Mean**	**2.5%**	**97.5%**
1. Year-specific D_i and F	1.00	388	0.13	0.04	0.44
2. Year-specific D_i, k_d, and F	0.99	585	0.45	0.24	0.91
3. Year-specific D_i, v, and F	1.00	564	0.44	0.04	1.93
4. Year-specific v, temporally constant D_i and F	0.98	374	1.10	0.30	2.15
5. Year-specific F, temporally constant D_i and v	0.25	601	2.31	1.84	2.94
6. Year-specific D_i, temporally constant F and v	0.29	182	2.28	1.76	2.96
7. Year-specific k_d, temporally constant F and D_i	0.95	514	0.74	0.50	1.06
8. Year-specific F, temporally constant k_d and D_i	0.86	557	1.09	0.80	1.48
9. Year-specific D_i, temporally constant k_d and F	1.00	44	0.25	0.05	0.69

Note: DIC, deviance information criterion.

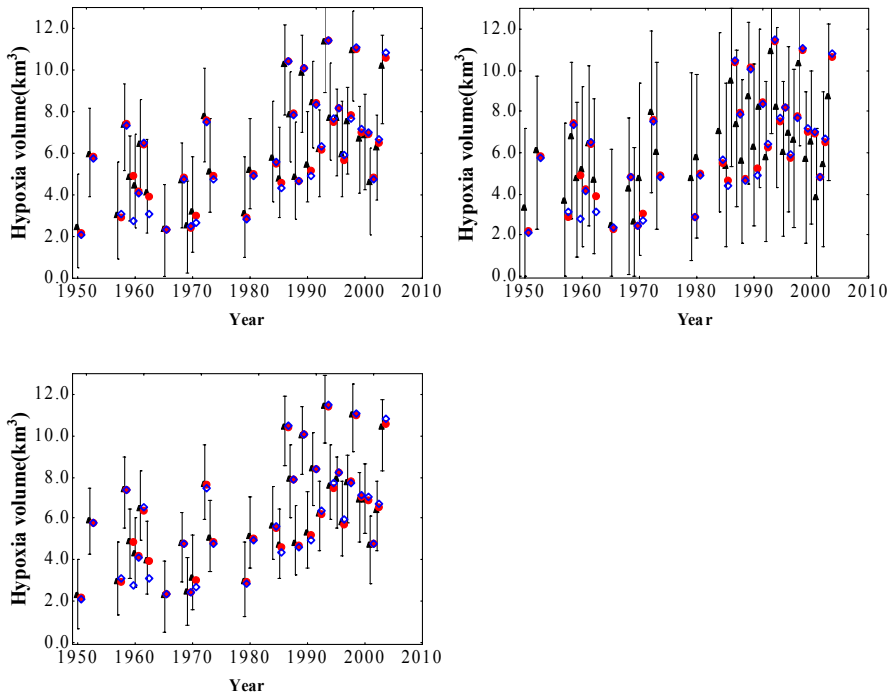

FIGURE 5: Sensitivity of the Hypoxic Volume Predictions to the Prior Distributions Assigned to the Parameters v, D_i, F. (A) The original priors are based on the following normal distributions: $D_i \sim N(0.30,0.70)$, $F \sim N(0.75,60)\mathrm{I}(0,1)$, and $v \sim N(2.5, 1.7)\mathrm{I}(0,)$; the original precisions are replaced by double (B), and half precisions (C).

aquatic ecosystems, the question of model credibility and the development of rigorous methods for assessing the uncertainties associated with parameter values and predictions have recently received considerable attention in hypoxia modeling. Scavia *et al.* (2006) used a deterministic Streeter–Phelps model to predict Chesapeake Bay DO concentrations and hypoxia volume conditions, and assessed the uncertainty associated with the initial DO deficit (D_i) and the calibration term (v) with Monte Carlo simulations. More recently, Stow and Scavia (2009) introduced a Bayesian configuration of the same model to demonstrate a systematic multiple parameter estimation and to assess the uncertainty underlying the DO profile model predictions. The present analysis extended the use of this Bayesian framework to fit the model against hypoxic volume data and examined how parameter estimation and predictive uncertainty can vary depending on the type of data used to calibrate the model.

Examination of the hypoxic volume case stems from the pragmatic need to overcome the lack of DO data (e.g., when only hypoxic volume or area estimates are available), and to improve the predictive power of hypoxia modeling given that the decision-making process is often guided from the hypoxia volume projections rather than the forecasted DO concentrations. The switch from DO concentrations to hypoxic volume data though raises a basic problem in that model fitting is now based on one datum per year, that is, annual hypoxic volume, and therefore the empirical estimation of year-specific parameters becomes problematic. In this regard, the dilemma arising is that the time-varying parameters are more realistic and can offer insights with respect to regime shifts or temporal changes of the ecosystem functioning but, from a statistical inference standpoint, the fitting process lacks replication. To address this problem, we opted for informative priors founded upon literature information and previous knowledge from the system (Arhonditsis *et al.*, 2007). The incorporation of additional information along with the dataset during the model fitting process aimed to reduce the disparity between what we wanted to learn from the model and what was actually observed is the primary reason for the poor model identifiability (Reichert and Omlin, 1997). While the adoption of informative priors is arguably a valuable way to alleviate the problem of model overidentification, the impact of the somewhat controversial inclusion of prior information on the robustness of the model results was further assessed by relaxing the confidence placed on these priors.

Our first set of numerical experiments (i.e., individual parameter estimation) indicated that the calibration of the Streeter–Phelps model against the

hypoxic volume data provides a good fit along with reasonable predictive uncertainty, although the variances of the marginal posterior parameter distributions were greater than those obtained from fitting the model to the observed DO profiles. While the latter result may cast doubt on the ability of this approach to offer insights into the ecosystem functioning, the substantial degree of updating of some of the parameter posteriors (i.e., the initial DO deficit and the scaling factor K) suggests that our knowledge did improve relative to what we knew prior to the calibration. Furthermore, the limited uncertainty reduction of k_d suggests that the relatively narrow prior characterization was sound; and thus is adequate for achieving optimal model fit. Our analysis also indicates that the hypoxic volume predictions are more sensitive to the prior specifications of some parameters (F, v), and therefore the predictive uncertainty can be significantly controlled by formulating, whenever possible, prior distributions with central tendency and dispersion values that more realistically represent the dynamics of the system. For example, it seems feasible to delineate an articulate prior for the fraction of surface organic carbon production that settles below the pycnocline (F) in the Chesapeake Bay, but it may be more difficult with the net downstream advection v. The ambiguity associated with the interpretation of the parameter v can conceivably be addressed by a spatially explicit analytical solution of the original differential equation of the Streeter–Phelps model, which will relax the assumption of constant flow velocity along the system. The existing hydrodynamic work in the area can effectively meet the data requirements of the reformulated model (e.g., flow profiles), whereas an error-in-variables (or measurement error) model may then be used to accommodate the significantly reduced (relative to the present approach) temporal variability of the parameter v.

The results derived from the joint estimation of different parameter combinations can be used to determine which parameters should have year-specific or temporally constant character, and thus to develop more parsimonious models that effectively balance between performance and complexity. Parameters with relatively stable posterior means over time and, if possible, narrow credible intervals can be considered as temporally constant, while those with time-varying posterior patterns can be used to accommodate the interannual variability by assigning year-specific values. For example, similar to the Scavia *et al.* (2006) study, we found that the posteriors of the fraction of the surface organic carbon production that settles below the pycnocline (F), the scaling factor (K) and the BOD decay coefficient (k_d) displayed little variation over the 36 years when fitted against the hypoxia volume data (Figure 2); so we may be able to hold them constant for hypoxia forecasts without missing much of the interannual system variability. By contrast, the posteriors of the initial DO deficit (D_i) and the calibration term (v) varied significantly over time, providing evidence that the assignment of temporally constant values is probably inappropriate. More importantly, the year-to-year variations of their values likely reflect a gradual change in the prevailing conditions in the system and/or a possible regime shift.

In this regard, our approach can conceivably assist with future hypoxia forecasting, as the time-varying parameter estimates can be linked to potentially important ecological/physical process changes. Namely, our model parameterization revealed an increasing trend for the upstream DO deficit, whereas the v values were characterized by a U-shape pattern with a global minimum in the early 1980s. A possible explanation for why v decreased is that during the earlier period, increased surface production was actually moving down estuary past the boundary of the Streeter–Phelps model and outside of the mesohaline region that influences hypoxia (Liu and Scavia, 2010). This is also consistent with Harding and Perry's (1997) assertions that showed Chl *a* concentrations in the mesohaline portion of the Bay (inside the boundary of our model) remained relatively constant after the 1950s while concentrations in the polyhaline region (outside the boundary of the model) increased; presumably in response to increasing loads (see Figures 1 and 2 of Harding and Perry 1997). In contrast to the earlier period, the v increase after the 1980s is consistent with empirical evidence that, for a given freshwater discharge rate, the recent observed hypoxic volume in Chesapeake Bay have been two or more times greater than that of 1949 to the 1980s. The distinctly higher D_i values during the same period are also on par with the latter assertion, suggesting that a threshold of nutrient loading as well as in the ecosystem buffering capacity has been reached. Apparently, the Bay has become more susceptible to the development of hypoxic conditions after the 1980s, and since then the volume of hypoxic waters has remained consistently high regardless of the hydraulic loading (Boesch *et al.*, 2001).

This kind of structural changes and/or regime shift will definitely affect the forecasting statements supported by the model. For example, we can expect higher v and D_i values for forecasting purposes in recent years than in the 1950s-1970s, that is, for a given nutrient load, higher v and D_i values will lead to higher hypoxic volume projections, which implies that a higher loading reduction will be necessary for achieving the targeted hypoxic volume goal (Liu and

Scavia, 2010). The present modeling exercise has, thus, demonstrated the challenges and risks of hypoxia control under the likelihood of ecosystem regime shifts. We also note that our illustration postulates conditional independence among the year-specific values of the different parameters, which may not be the best strategy for forecasting purposes. An alternative approach would have been the sequential updating of the model in time (i.e., the posteriors of one year provide the priors for the next year), whereby the updated information about the parameters will be "discounted" by adding a stochastic disturbance term to represent the aging of information with the passage of time. This type of dynamic modeling framework may support probabilistic risk assessment that more effectively integrates past experience and present information with the future response of the system (Dorazio and Johnson, 2003; Qian and Reckhow, 2007).

Another critical test of the credibility of the hypoxia volume model is its ability to accurately reproduce DO concentrations in the estuary, given that the DO data are not part of the fitting process. To assess the extent to which this premise holds true, we illustrate the comparison between the predicted DO profiles from the two approaches when estimating year-specific v and temporally constant D_i and F (Figure 6). Despite the accurate hypoxia volume predictions presented herein, we noted that there are clear differences between the two predicted DO profiles. Namely, the calibration of the model against the hypoxia volume data tends to result in higher discrepancies from the observed DO levels along the estuary and relatively similar conclusions were drawn from the rest of the numerical experiments (not presented here). That is, for some years, the hypoxia volume model accurately reproduces the length of the estuary below 2 mg/l but displaces the profile up or down estuary. The question this raises is "Does this displacement matter if the prediction goal is for hypoxia volume?" We suggest that, in most cases, it does not. If, however, the spatial distribution of low oxygen is of primary interest, other models with more realistic physics will likely be required.

The Bayesian inference techniques can help with the quantification of the predictive uncertainty which is important information for model-based decision making. Bayes' Theorem provides a rigorous framework to estimate predictive uncertainty based on a given model structure and data. For the purpose of prediction, the Bayesian approach generates a posterior predictive distribution that represents the current estimate of the value of the response variable, taking into account both the uncertainty about the parameters and the uncertainty that remains when the parameters are known (Arhonditsis *et al.*, 2007,

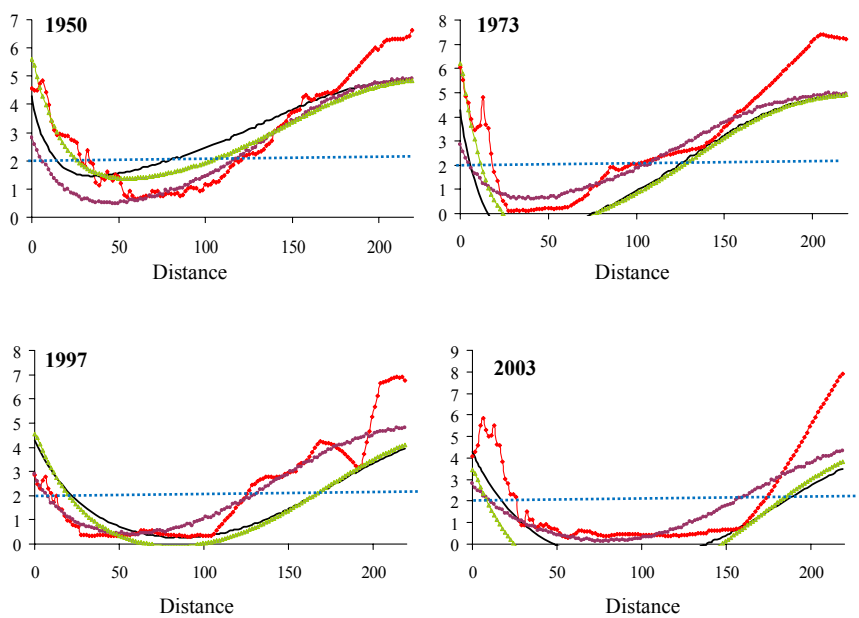

FIGURE 6. Comparison Between Observed (red line) and Predicted Mean (purple and black lines) Dissolved Oxygen (DO) Profiles in Chesapeake Bay Using Year-Specific v Estimates, While Keeping D_i and F Temporally Constant. Purple and black lines correspond to the predicted DO data from DO profile model and hypoxic volume model, respectively. The green line corresponds to the DO profile predicted from Scavia *et al.* (2006), while the horizontal dashed line represents the hypoxia threshold, 2 mg/l. Distance (X-axis) is kilometers downstream from the Susquehanna River; Y-axis is DO (mg/l).

2008a,b). Therefore, estimates of the uncertainty of Bayesian model predictions are more realistic than those based on the classical procedures. Predictions are expressed as probability distributions, thereby conveying significantly more information than point estimates in regards to uncertainty. The presentation of the model outputs as probabilistic assessment of water quality conditions will make the model results more credible and appealing to decision makers and stakeholders. Thus, the often deceptive deterministic statements are avoided and the water quality goals are set by explicitly acknowledging an inevitable risk of non-attainment, the level of which is subject to decisions that reflect different socioeconomic values and environmental concerns (Arhonditsis *et al.*, 2008a,b). Furthermore, the Bayesian (iterative) nature of the proposed approach is conceptually similar to the policy practice of adaptive management, i.e., an iterative implementation strategy that is recommended to address the often-substantial uncertainty associated with water quality model forecasts, avoid the implementation of inefficient and flawed management plans. Adaptive implementation or "learning while doing" augments initial model forecasts of management schemes with post-implementation monitoring, i.e., the initial model forecast serves as the Bayesian prior, the post-implementation monitoring data serve as the sample information (the likelihood), and the resulting posterior probability (the integration of monitoring and modeling) provides the basis for revised (and improved) management actions (Zhang and Arhonditsis, 2008).

In conclusion, we used a Bayesian configuration of the modified Streeter–Phelps equation to predict the DO profiles and the hypoxic volume in the Chesapeake Bay for 36 years between 1950 and 2003. Our analysis suggests that fitting the model against both DO and hypoxic volume data provides realistic predictions along with uncertainty bounds of the spatiotemporal hypoxia patterns, although the inference regarding the parameter posteriors can significantly vary depending on the type of data and calibration parameters used. Although not illustrated here, the present modeling exercise can also provide the foundation of posterior simulations for determining the optimal nutrient loading reductions (with a realistic margin of safety) that can ultimately allow the system to comply with the targeted hypoxia volume goal. We also advocate use of informative priors, whenever available knowledge permits, to alleviate the identification problems that arise when switching from DO profile to hypoxia volume data. Finally, the numerical experiments of joint parameter estimation presented herein can be particularly useful for optimizing the model complexity.

SUPPORTING INFORMATION

Additional results of our modeling analysis can be found in the online version of this article.

Figure S1. Comparison between observed and predicted hypoxic volume (km^3) values using year-specific parameter estimates. The lower and upper bars represent the 2.5 and 97.5% credible bounds. The red circles correspond to the observed values.

Figure S2. Hypoxic volume predictions based on joint estimation of D_i, v, and F.

Figure S3. Hypoxic volume predictions using various combinations of D_i, v, k_d, and F.

Figure S4. Parameter estimation with the Streeter–Phelps model fitted to the hypoxic volume data: (a) joint estimation of year-specific D_i and F values; (b) joint estimation of year-specific D_i and temporally constant k_d and F values. The red dots correspond to the Bayesian estimated values and the lines correspond to the values presented by Scavia *et al.* (2006).

Figure S5. Joint estimation of the parameters D_i, v, and F with the Streeter–Phelps model fitted to the hypoxic volume data.

Figure S6. Hypoxic volume predictions based on various combinations of D_i, k_d, and F. The red circle corresponds to the observed data. The diamond corresponds to the values presented by Scavia *et al.* (2006). The triangle corresponds to the predictions when estimating year-specific k_d values and keeping constant F and D_i. The cross corresponds to the predictions when estimating year-specific F values and keeping constant k_d and D_i. The square corresponds to the predictions when estimating year-specific D_i values and keeping constant k_d and F.

Table S1. Goodness-of-fit to the observed DO profiles when the hypoxic volume dataset is used to calibrate the Streeter–Phelps model.

Please note: Neither AWRA nor Wiley-Blackwell is responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the comments from two anonymous reviewers and Dr. James D. Hagy III for use of the data amassed in his dissertation. This work is contribution number 116 of the Coastal Hypoxia Research Program and was supported in part by grant NA05NOS4781204 from NOAA's Center for Sponsored Coastal Ocean Research, and in part by Scientific Research Foundation for the Returned Overseas Chinese Scholars, SEM and China National Water Pollution Control Program (2008ZX07102-001). This is GLERL contribution number 1591.

LITERATURE CITED

Arhonditsis, G.B., B.A. Adams-VanHarn, L. Nielsen, C.A. Stow, and K.H. Reckhow, 2006. Evaluation of the Current State of Mechanistic Aquatic Biogeochemical Modeling: Citation Analysis and Future Perspectives. Environmental Science & Technology 40:6547-6554.

Arhonditsis, G.B. and M.T. Brett, 2004. Evaluation of the Current State of Mechanistic Aquatic Biogeochemical Modelling. Marine Ecology-Progress Series 271:13-26.

Arhonditsis, G.B., D. Papantou, W. Zhang, G. Perhar, E. Massos, and M. Shi, 2008b. Bayesian Calibration of Mechanistic Aquatic Biogeochemical Models and Benefits for Environmental Management. Journal of Marine Systems 73:8-30.

Arhonditsis, G.B., G. Perhar, W. Zhang, E. Massos, M. Shi, and A. Das, 2008a. Addressing Equifinality and Uncertainty in Eutrophication Models. Water Recourses Research 44:W01420.

Arhonditsis, G.B., S.S. Qian, C.A. Stow, C.E. Lamon, and K.H. Reckhow, 2007. Eutrophication Risk Assessment Using Bayesian Calibration of Process-Based Models: Application to a Mesotrophic Lake. Ecological Modelling 28(2-4):215-229.

Boesch, D.F., R.B. Brinsfield, and R.E. Magnien, 2001. Chesapeake Bay Eutrophication: Scientific Understanding, Ecosystem Restoration and Challenges for Agriculture. Journal of Environment Quality 30:303-320.

Boesch, D.F., V.J. Coles, D.G. Kimmel, and W.D. Miller, 2007. Ramifications of Climate Change for Chesapeake Bay Hypoxia. *In*: Regional Impacts of Climate Change: Four Case Studies in the United States, Kristie L. Ebi, Gerald A. Meehl, Dominique Bachelet, Robert R. Twilley and Donald F. Boesch (Editors). Pew Center on Global Climate Change, Arlington, Virginia, pp. 54-70.

Borsuk, M.E., D. Higdon, C.A. Stow, and K.H. Reckhow, 2001. A Bayesian Hierarchical Model to Predict Benthic Oxygen Demand From Organic Matter Loading in Estuaries and Coastal Zones. Ecological Modelling 143:165-181.

Bricker, S., B. Longstaff, W. Dennison, A. Jones, K. Boicourt, C. Wicks, and J. Woerner, 2007. Effects of Nutrient Enrichment in the Nation's Estuaries: A Decade of Change. NOAA Coastal Ocean Program Decision Analysis Series No. 26. National Centers for Coastal Ocean Science, Silver Spring, Maryland, 328 pp.

Cerco, C.F. and T.M. Cole, 1993. Three-Dimensional Eutrophication Model of Chesapeake Bay. Journal of Environmental Engineering 119:1006-1025.

Cerco, C. and M. Noel, 2004. Process-Based Primary Production Modeling in Chesapeake Bay. Marine Ecology Progress Series 282:45-58.

Chapra, S.C., 1997. Surface Water-Quality Modeling. McGraw-Hill, New York.

Cohn, T.A., L.L. Delong, E.J. Gilroy, R.M. Hirsch, and R.M. Wells, 1989. Estimating Constituent Loads. Water Resources Research 25:937-942.

Cooper, S.R. and G.S. Brush, 1991. Long-Term History of Chesapeake Bay Anoxia. Science 254:992-996.

Diaz, R.J. and R. Rosenberg, 2008. Spreading Dead Zones and Consequences for Marine Ecosystems. Science 321:926-929.

Dorazio, R.M. and F.A. Johnson, 2003. Bayesian Inference and Decision Theory - A Framework for Decision Making in Natural Resource Management. Ecological Applications 13:556-563.

Gelfand, A.E. and A.F.M. Smith, 1990. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85:398-409.

Gelman, A. and J. Hill, 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.

Gill, J., 2002. Bayesian Methods: A Social and Behavioral Sciences Approach. Chapman & Hall/CRC, Boca Raton, Florida.

Hagy, J.D., W.R. Boynton, C.W. Keefe, and K.V. Wood, 2004. Hypoxia in Chesapeake Bay, 1950–2001: Long-Term Change in Relation to Nutrient Loading and River Flow. Estuaries 27:634-658.

Harding, L.W., and E. Perry, Jr., 1997. Long-Term Increase of Phytoplankton Biomass in Chesapeake Bay. Marine Ecology Progress Series 157:39-52.

Hastings, W.K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57(1):97-109.

Kemp, W.M., W.R. Boynton, J.E. Adolf, D.F. Boesch, W.C. Boicourt, G. Brush, J.C. Cornwell, T.R. Fisher, P.M. Glibert, J.D. Hagy, L.W. Harding, E.D. Houde, D.G. Kimmel, W.D. Miller, R.I.E. Newell, M.R. Roman, E.M. Smith, and J.C. Stevenson, 2005. Eutrophication of Chesapeake Bay: Historical Trends and Ecological Interactions. Marine Ecology Progress Series 303: 1-29.

Liu, Y. and D. Scavia, 2010. Analysis of the Chesapeake Bay Hypoxia Regime Shift: Insights from Two Simple Mechanistic Models. Estuaries and Coasts 33(3):629-639.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalter, 2000. WinBUGS - a Bayesian Modelling Framework: Concepts, Structure, and Extensibility. Statistics and Computing 10:325-337.

Malve, O. and S.S. Qian, 2006. Estimating Nutrients and Chlorophyll a Relationships in Finnish Lakes. Environmental Science and Technology 40(24):7848-7853.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, 1953. Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics 21(6):1087-1092.

Newcombe, C.L. and W.A. Horne, 1938. Oxygen-Poor Waters of the Chesapeake Bay. Science 88:80-81.

NRC, 2000. Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution. National Research Council, National Academy Press, Washington, D.C.

NRC, 2001. Assessing the TMDL Approach to Water Quality Management. National Research Council, National Academy Press, Washington, D.C.

Officer, C.B., R.B. Biggs, J.L. Taft, L.E. Cronin, M.A. Tyler, and W.R. Boynton, 1984. Chesapeake Bay Anoxia: Origin, Development, and Significance. Science 223:22-27.

Qian, S.S. and K.H. Reckhow, 2007. Combining Model Results and Monitoring Data for Water Quality Assessment. Environmental Science and Technology 41:5008-5013.

Qian, S.S., C.A. Stow, and M.E. Borsuk, 2003. On Monte Carlo Methods for Bayesian Inference. Ecological Modelling 159(2-3): 269-277.

Reckhow, K.H., 1994. Importance of Scientific Uncertainty in Decision-Making. Environmental Management 18:161-166.

Reichert, P. and M. Omlin, 1997. On the Usefulness of Over Parameterized Ecological Models. Ecological Modelling 95:289-299.

Scavia, D. and K.A. Donnelly, 2007. Reassessing Hypoxia Forecasts for the Gulf of Mexico. Environmental Science and Technology 41:8111-8117.

Scavia, D., D. Justic, and V.J. Bierman, Jr., 2004. Reducing Hypoxia in the Gulf of Mexico: Advice From Three Models. Estuaries 27:419-425.

Scavia, D., E.L.A. Kelly, and J.D. Hagy, 2006. A Simple Model for Forecasting the Effects of Nitrogen Loads on Chesapeake Bay Hypoxia. Estuaries and Coasts 29:674-684.

Scavia, D., N.N. Rabalais, R.E. Turner, D. Justic, and W. Wiseman, Jr., 2003. Predicting the Response of Gulf of Mexico Hypoxia to Variations in Mississippi River Nitrogen Load. Limnology and Oceanography 48:951-956.

Scheffer, M. and S.R. Carpenter, 2003. Catastrophic Regime Shifts in Ecosystems: Linking Theory to Observation. Trends in Ecology & Evolution 18:648-656.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde, 2002. Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society: Series B 64:583-640.

Stow, C.A., K.H. Reckhow, and S.S. Qian, 2006. A Bayesian Approach to Retransformation Bias in Transformed Regression. Ecology 87:1472-1477.

Stow, C.A. and D. Scavia, 2009. Modeling Hypoxia in the Chesapeake Bay: Ensemble Estimation Using a Bayesian Hierarchical Model. Journal of Marine Systems 76:244-250.

Streeter, H.W. and E.B. Phelps, 1925. A Study in the Pollution and Natural Purification of the Ohio River, III Factors Concerning the Phenomena of Oxidation and Reaeration. US Public Health Service, Public Health Bulletin No. 146, Feb 1925 Reprinted by US PHEW, PHA 1958.

Zhang, W. and G.B. Arhonditsis, 2008. Predicting the Frequency of Water Quality Standard Violations Using Bayesian Calibration of Eutrophication Models. Journal of Great Lakes Research 34:698-720.