



TOTAL MAXIMUM DAILY LOAD CRITERIA ASSESSMENT USING MONITORING AND MODELING DATA¹

Jeni Keisman and Gary Shenk²

ABSTRACT: Applications of Total Maximum Daily Load (TMDL) criteria for complex estuarine systems like Chesapeake Bay have been limited by difficulties in estimating precisely how changes in input loads will impact ambient water quality. A method to deal with this limitation combines the strengths of the Chesapeake Bay's Water Quality Sediment Transport Model (WQSTM), which simulates load response, and the Chesapeake Bay Program's robust historical monitoring dataset. The method uses linear regression to apply simulated relative load responses to historical observations of water quality at a given location and time. Steps to optimize the application of regression analysis were to: (1) determine the best temporal and spatial scale for applying the WQSTM scenarios, (2) determine whether the WQSTM method remained valid with significant perturbation from calibration conditions, and (3) evaluate the need for log transformation of both dissolved oxygen (DO) and chlorophyll *a* (CHL) datasets. The final method used simple linear regression at the single month, single WQSTM grid cell scale to quantify changes in DO and CHL resulting from simulated load reduction scenarios. The resulting linear equations were applied to historical monitoring data to produce a set of "scenario-modified" DO or CHL concentration estimates. The utility of the regression method was validated by its ability to estimate progressively increasing attainment in support of the 2010 Chesapeake Bay TMDL.

(KEY TERMS: Chesapeake Bay; Chesapeake TMDL; TMDLs; integrated environmental models; watershed management; water quality standards; dissolved oxygen; chlorophyll; simulations; decision support systems.)

Keisman, Jeni and Gary Shenk, 2013. Total Maximum Daily Load Criteria Assessment Using Monitoring and Modeling Data. *Journal of the American Water Resources Association* (JAWRA) 1-16. DOI: 10.1111/jawr.12111

INTRODUCTION

Under the authority of the 1972 Clean Water Act (CWA), The U.S. Environmental Protection Agency's (USEPA) Total Maximum Daily Load (TMDL) Program requires that states identify, for any impaired water body, the pollutant load reductions required in order for the given water body to attain its stated ambient water quality standards (NRC, 2001). The science of TMDL development for ambient water

quality criteria is relatively new, stemming from the USEPA's initial promulgation of TMDL regulations in 1992. Since that time, the limited number of TMDLs that have been developed for complex estuarine systems have relied primarily on modeled predictions of pollutant reductions to estimate percent reductions from a base load (NYDEC and CT DEEP, 2000; Wool *et al.*, 2003).

For the Chesapeake Bay estuary, early work to elucidate TMDL needs focused on identifying the appropriate water quality criteria and standards (see

¹Paper No. JAWRA-12-0067-P of the *Journal of the American Water Resources Association* (JAWRA). Received March 15, 2012; accepted February 21, 2013. © 2013 American Water Resources Association. **Discussions are open until six months from print publication.**

²Respectively, Assistant Research Scientist (Keisman), Center for Environmental Science, University of Maryland, Cambridge, Maryland at the time this paper was prepared, now Science & Technology Policy Fellow, American Association for the Advancement of Science (AAAS), Washington, D.C. 20005; and Integrated Analysis Coordinator (Shenk), U.S. EPA Chesapeake Bay Program Office, 410 Severn Avenue, Suite 112, Annapolis, Maryland 21403 (E-mail/Shenk: gshenk@chesapeakebay.net).

USEPA, 2003; Tango and Batiuk, this issue) and on developing the Chesapeake Bay Water Quality Sediment Transport Model (WQSTM). The WQSTM is a coupled CH3D finite-difference model and CE-QUAL-ICM finite-volume water quality model designed to simulate the response of key estuarine water quality parameters to pollutant reductions (Cерco and Cole, 1993; Cerco, 1995; Cerco and Noel, 2004; Wang *et al.*, 2006; Cerco and Noel, this issue; Cerco *et al.*, this issue). In this context, simulations are called “scenarios.” Scenarios represent an assumed set of management actions and/or nutrient loadings simulated by the Chesapeake Bay Program’s watershed model (Linker *et al.*, 2008; Shenk and Linker, this issue), as well as the simulated responses of key variables in the WQSTM (see Wang *et al.* 2006 for examples). For the purposes of method development, dissolved oxygen (DO) and chlorophyll *a* (CHL) were the variables of interest. DO serves as the primary indicator of water quality for the Chesapeake Bay TMDL.

Although the WQSTM estimates changes in water quality due to changes in input loads with reasonable accuracy (Cerco *et al.*, 2010, this issue; Cerco and Noel, this issue), a perfect calibration of simulated concentrations to observed data is impossible (NRC, 2007). As mentioned by both National Academies’ reports (NRC, 2001, 2007), integration of modeling efforts with monitoring data can improve analytical outcomes. Producing a TMDL for this large, complex system provides a unique opportunity to develop and test a novel integration of these resources, particularly given the amount of monitoring data available for the Chesapeake and its tributaries.

Environmental observations remain the best tool for estimating water quality conditions at any given point in space and time. Given that discrepancies exist between monitoring data and the WQSTM with regard to the determination of criteria attainment, it is appropriate to use the more accurate monitoring data to assess attainment. However, model simulations still provide one of the best available predictions of the magnitude of improvement that can be expected with implementation of pollutant load control practices. To take advantage of the complementary strengths of the WQSTM (for simulating load response) and the Chesapeake Bay Program’s robust historical monitoring dataset (for determining water quality status), a method was developed that would apply simulated estimates of relative load response to historical observations of water quality. This would allow prediction of the nitrogen and phosphorus load reductions needed to attain water quality standards.

To estimate the Chesapeake’s response to nutrient reductions, numerous progressively stringent nutrient loading scenarios (such as PR2000, Tier1, Tier2, Tier3, and E3; see Figure 8) were developed and run through

the WQSTM. Several iterations of loading scenarios were produced. The above-mentioned scenarios were generated and used during the method development process. After the method development phase was complete, more iterations of loading scenarios were generated as part of the TMDL development process. The latter scenarios are referenced later in this article. The most stringent scenario developed was referred to as “Everyone, Everything, Everywhere” or “E3.” It assumed implementation of all available nutrient management practices (e.g., forest buffers, stream restoration, cover crops, etc.) at all available locations. By comparing the results of a scenario such as the E3 scenario to those of the model’s calibration run (“cal”), one could obtain an estimate of the response of the system to the proposed nutrient reductions.

An important requirement for the method that was developed was that it had to allow assessment of predicted attainment of water quality standards using the same methods that are used to assess actual attainment of water quality standards (methods are documented in USEPA, 2003, 2008a, 2010a). To summarize, methods were developed to assess attainment of water quality standards for DO and CHL parameters through a joint effort using analytical and scientific expertise from academic, state, and federal participants. Every two years, these methods are applied to all available monitoring data to determine attainment of water quality standards. The states then use these results to develop their “303(d) listing” reports, which are submitted to the EPA “as required by section 303(d) of the CWA and its implementing regulations at Title 40 of the Code of Federal Regulations section 130.7” (USEPA, 2010b). Using the same methodology to develop the TMDL that is used to conduct assessments for 303(d) listing reports lends consistency and validity to the TMDL process.

This article describes the development of such a method, and its application to development of the Chesapeake Bay TMDL.

METHODS

To determine the nutrient load reductions necessary to attain water quality standards in the Chesapeake and its tidal tributaries, a method was needed that could combine the comparative accuracy of the monitoring data with the ability of the WQSTM to predict changes in response to nitrogen and phosphorus load reductions. Preliminary studies indicated that regression analysis would provide a sound basis for this application; however, a few key questions needed to be investigated to determine the most

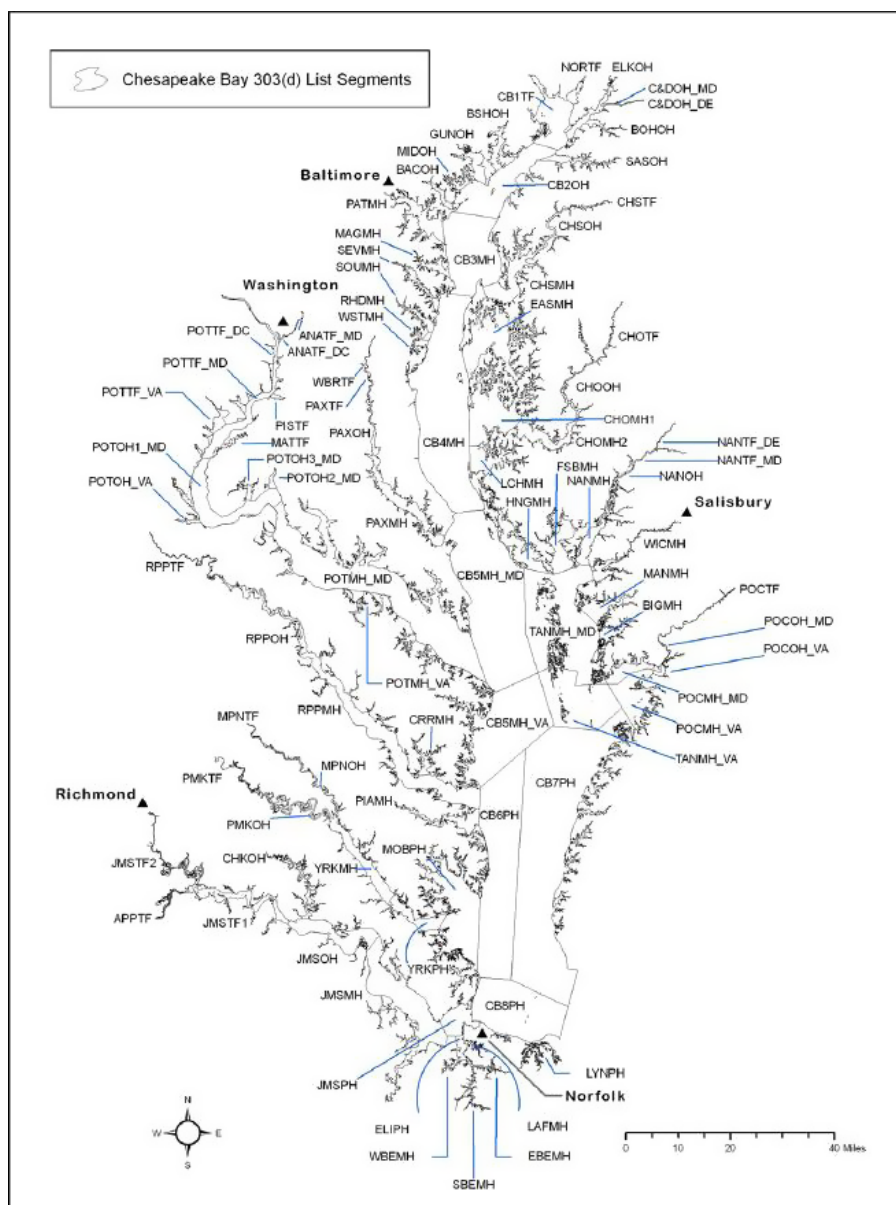


FIGURE 1. The 92 Chesapeake Bay Segments (CBPSEGs) Used for Regression Method Development (USEPA, 2008a).

appropriate application of regression analysis to this problem. Specifically, analysis was needed to: (1) determine the appropriate temporal and spatial scale on which to evaluate the responses of key variables to the management scenarios and apply those responses to monitoring data and (2) determine whether the method remained valid with significant perturbation of the model from calibration conditions. As is customary with analysis of these types of data, the need for log transformations of both DO and CHL datasets was also evaluated.

These questions were explored in the context of the temporal and spatial distinctions made for different habitats within the Chesapeake throughout the

course of any given year. For the purposes of water quality criteria assessment, the Chesapeake and its tidal tributaries were divided into 92 individual segments (CBPSEGs) as shown in Figure 1 (USEPA, 2008a). The volume of water within some segments was further subdivided into Open Water, Deep Water, and Deep Channel designated uses. Deep Water and Deep Channel designated uses apply only to DO. In the Chesapeake, the term “designated use” is used to define areas throughout the Bay and its tidal tributaries that are differentiated by physical characteristics such as salinity and depth and that have particular water quality criteria. Water quality criteria for the different designated uses are

designed to protect the aquatic life that depends on various habitats within the Chesapeake Bay. The Open Water designated use comprises the volume of water occurring above the upper layer of the pycnocline in summer months (June-September), whereas in nonsummer months it encompasses the entire water column. Water quality criteria for the Open Water designated use are designed to protect a variety of sportfish species that inhabit the surface waters of the Chesapeake year-round. The Migratory and Spawning designated use occurs in late winter/early spring, primarily in the upper reaches of tidal rivers and creeks. Its criteria are designed to protect aquatic species that use these areas for spawning. Criteria developed for the Deep Water and Deep Channel designated uses are applied only to the summer months (June-September), and are designed to protect aquatic life that inhabits deeper waters while recognizing that seasonal hypoxia is a natural condition of some regions. The Deep Water designated use comprises the volume of water occurring between the lower and upper boundaries of the pycnocline. The Deep Channel designated use comprises those regions occurring below the lower region of the pycnocline, primarily in the main stem regions of the Chesapeake Bay. See Tango and Batiuk (this issue) for a more detailed explanation of the Bay's designated uses.

A series of analyses was conducted to address the above questions regarding spatial and temporal scales, as well as validity across model simulations. Appropriate temporal and spatial scales for analysis were explored in the context of the different designated uses and segments of the Chesapeake Bay. Finally, the utility of the regression method that was developed was validated through demonstration of its ability to predict continued improvement of key water quality indicators (i.e., DO and CHL) as a consequence of progressively more stringent nutrient reduction scenarios.

Spatial and Temporal Scale of Regressions

It was assumed that to best estimate the response of DO and CHL criteria to load reductions at monitoring station locations, regressions should be applied at the smallest scales possible. Thus, an objective of method development was to minimize the scale of regression application until regression statistics began to deteriorate. For analyses of DO, hourly DO measurements were extracted from the WQSTM output for the calibration run and for each management (nutrient loading) scenario. For analyses of CHL, daily measurements were extracted. The WQSTM operates on a grid of more than 50,000 cells (Figure 2). For the

purposes of this method, model measurements were extracted for only those cells corresponding to sampling locations in the monitoring dataset. Simulated CHL measurements were selected from only those WQSTM grid cells corresponding to surface CHL measurements, whereas for DO, WQSTM grid cells were matched with vertical sampling profiles at each monitoring station. Linear regression equations were calculated at the spatial scale of single WQSTM grid cells representing the locations of individual measurements in the monitoring dataset.

Regressions were generated for all WQSTM cells that matched spatially with the long-term Chesapeake Bay tidal water quality monitoring stations and vertical sampling locations throughout the water column. The regressions were generated using all simulated values (hourly for DO; daily for CHL) for any given month when historical monitoring observations occurred. The result was a unique linear regression equation for each monitoring location, month, and sampling depth (Figure 3).

Performance of the regressions, i.e., the ability to simulate a load response appropriate to the range of empirical observations, was evaluated at both the spatial scale of individual cells, and at the scale of individual segments. Temporally, performance of the regressions was evaluated at the following scales: criteria season (for all years aggregated); monthly for all years (i.e., "all Junes"); and monthly by year (i.e., "June 1985, June 1986, June 1987..."). Both the coefficient of determination (r^2) and jackknife statistics were compared across temporal scales (see below for more discussion of these statistics).

Log Transformation of DO and CHL Measurements

Regressions of log-transformed DO concentrations for the E3 scenario and the calibration run of the WQSTM were compared with those of nontransformed concentrations. Comparisons were made for the Deep Channel, Deep Water, Migratory and Spawning, and Open Water designated uses.

Dissolved oxygen concentrations are bounded by zero and saturation, and are thus relatively well behaved. Therefore, it is unnecessary to log transform them prior to analysis as demonstrated in Clarke and Ainsworth (1993). In contrast, log-normal distribution of CHL concentrations has been observed in numerous analyses of data from the Chesapeake estuary as well as from other systems (Harding and Perry, 1997; USEPA, 2010a). To further confirm the suitability of analysis with untransformed DO data and with log-transformed CHL data, respectively, the effect of log transformation on regression statistics was explored.

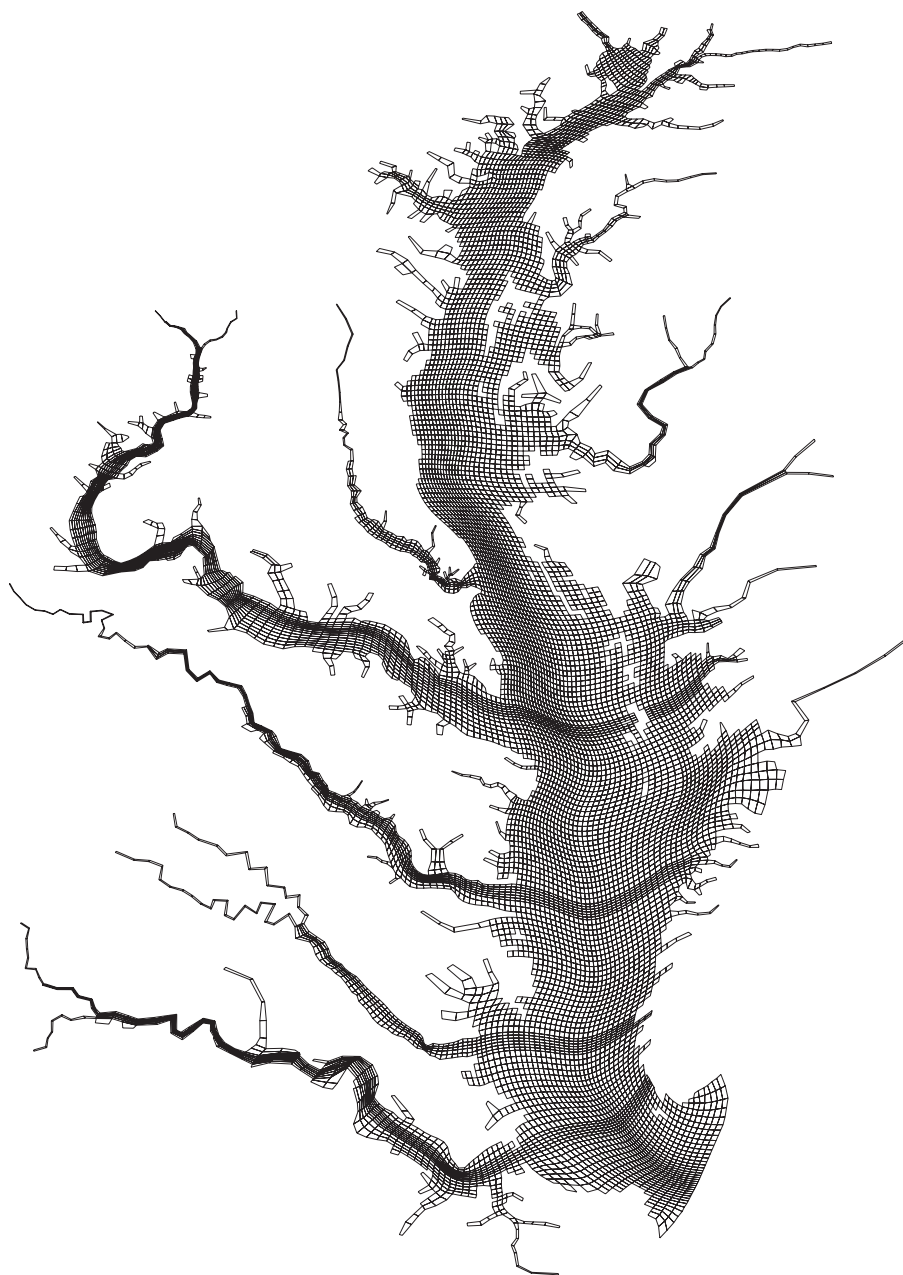


FIGURE 2. The Water Quality Sediment Transport Model Grid, a Three-Dimensional Framework of 57,000 Cells.

Scenario Effect on Regression Method Validation

To evaluate the strength of the regression method as a function of load reduction scenario, regression statistics (r^2 and jackknife) for the E3 scenario (regressed against the calibration run) were compared against statistics generated for a series of progressively less stringent load reduction scenarios (Tier3, Tier2, Tier1, and PR2000). The E3 scenario assumed application of management actions to the fullest possible extent with theoretical maximum levels of managed controls on load sources (Shenk and Linker, this issue). The IMSL Fortran Numeri-

cal Stat Library was used to calculate regression and jackknife statistics. The RLINE function generated regression statistics; it uses least squares to fit a line to a set of data points. Jackknife statistics were generated using the ROTIN function, which “computes diagnostics for detection of outliers and influential data points” (IMSL, 2007). Essentially, it uses resampling to estimate the bias (the tendency of the sample correlation to over- or underestimate the true, known correlation) of a sample statistic. The purpose of the jackknife estimator (a dimensionless value) was to judge the regressions not just by the r^2 , but also to rule out spurious correlations

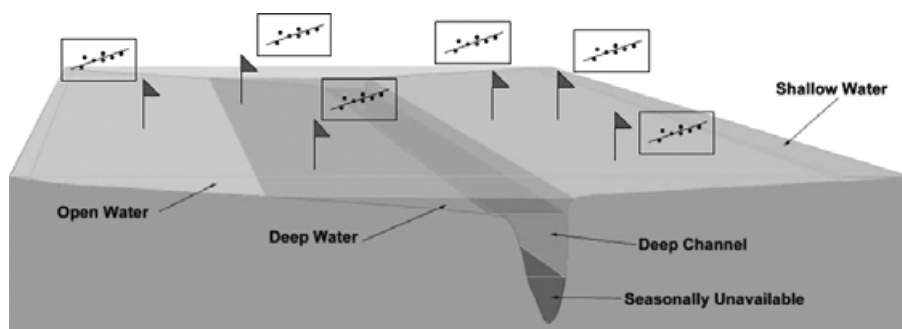


FIGURE 3. An Individual Regression Equation Is Generated for Each Monitoring Station and Month. For dissolved oxygen, a regression equation is generated for each Water Quality Sediment Transport Model cell that is matched to a vertical profile of monitoring observations. For chlorophyll a (CHL), a single equation is generated for the surface cell, which corresponds to a surface CHL monitoring observation.

where one or a few outlying data points might have created the illusion of a relevant model based on an inflated r^2 .

RESULTS

Spatial and Temporal Scale of Regression Analysis

Both CHL and DO from the E3 scenario run were regressed against concentrations from the calibration run on a segment-wide spatial scale and at the scale of individual cells. For both the main stem Bay and the Virginia tributaries, CHL was generally better correlated at the cell-level than at the scale of Chesapeake Bay segments (Figure 4). For DO, there was little or no difference between r^2 at the segment *vs.* the cell scale (data not shown). Results of regressions on different temporal scales — seasons pooled for all years, months pooled for all years, and individual months — were mixed for both CHL and DO, but did not deteriorate appreciably from seasonal and multiyear to monthly and single-year scales (data not shown). Jackknife tests for sample bias consistently showed definitively less bias at the temporal scale of individual months in the main stem of the Bay (Figure 5) as well as in tributary segments (data not shown).

Log Transformation of Data

Regression statistics demonstrated little or no benefit of log-transforming DO concentrations prior to regression (Figure 6). DO concentrations simulated in the E3 load reduction scenario were either equally well or better correlated with the calibration run without log transformation for all designated uses. Similarly, both jackknife estimates and skewness of

the sample dataset showed either no difference, or slightly greater bias when DO concentrations were log transformed (Table 1). Log transformation had little effect on regression and jackknife statistics for CHL concentrations when averaged over seasons and segments; skewness of the log-transformed calibration dataset increased or was unchanged for the CB1TF, CB2OH, and CB3MH Bay segments but decreased for the E3 scenario dataset. Skewness is a measure of asymmetry in the distribution of a dataset. A positive skew indicates that the right tail of the distribution is longer than the left tail; a negative skew indicates that the left tail is longer. For the remainder of Chesapeake Bay segments, skewness improved with log transformation. Similarly, skewness of the calibration dataset in the fresher, upstream James River segments increased slightly while it decreased, or improved, for all James River segments in the E3 scenario dataset (Table 2). Patterns were similar for other tributaries (data not shown).

Scenario Effect

To determine the expected effect of reduced pollutant loads on DO and CHL parameters, the simulated parameter concentrations from the WQSTM's calibration scenario were compared to the parameter concentrations from a given load reduction scenario. This was accomplished by relating each month's worth of values from the calibration scenario for a given location to the same month's worth of values from the load reduction scenario at the same location. The resulting linear regression equation represented the degree of change (in DO or CHL concentration) from the calibration scenario to the load reduction scenario. Figure 7 illustrates an idealized case in which a DO concentration of 2 milligrams per liter (mg/l; x -axis) in the

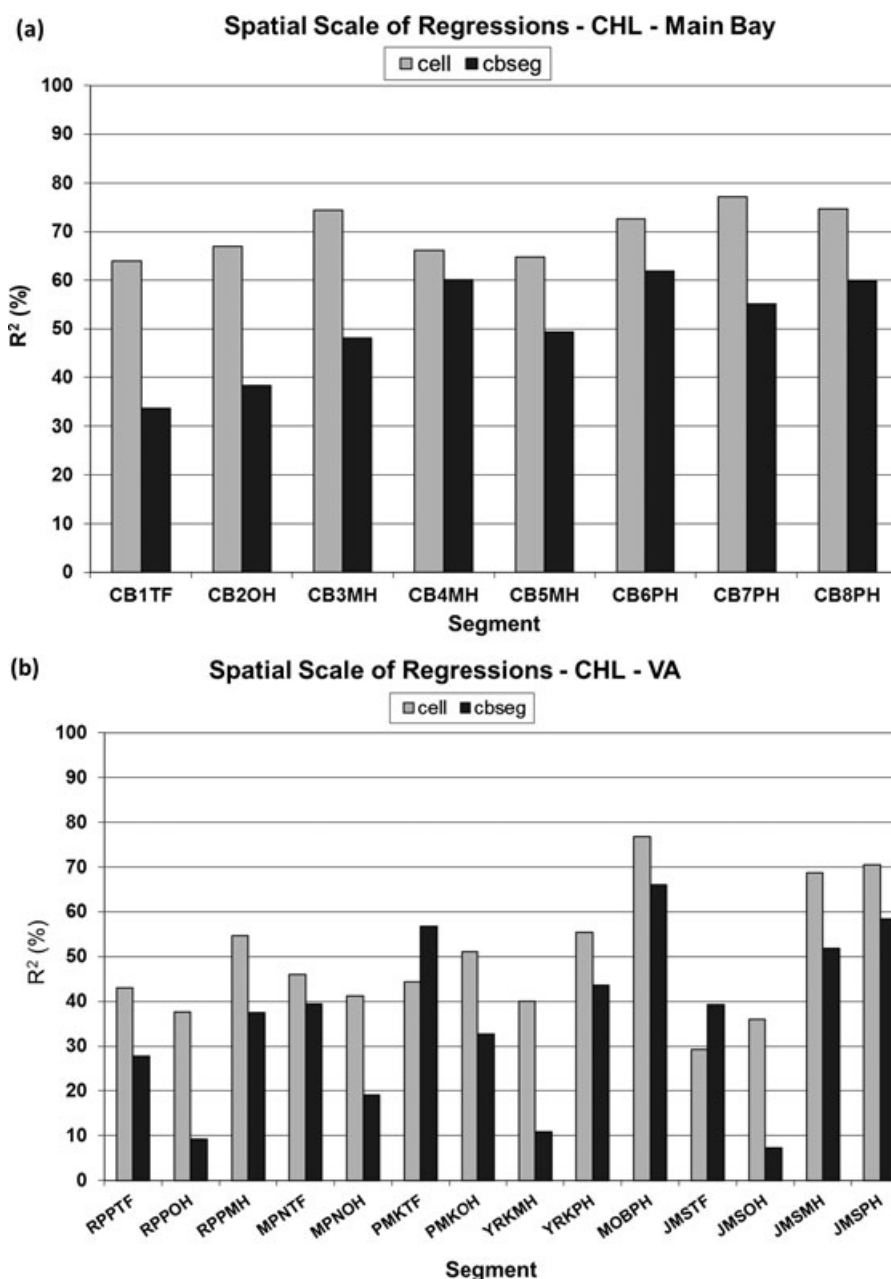


FIGURE 4. Comparison of Chlorophyll *a* (CHL) Regressions (r^2 ; y-axis) by Segment for the Main Stem Bay (a) and the Virginia Tributaries (b). CHL concentrations were evaluated at the scale of individual cells (gray bars), and at the scale of segments (black bars). Segments in (a) are listed in longitudinal order along the x-axis from northernmost (CB1TF) to southernmost (CB8PH) segment. For each tributary in Virginia (b), segments are listed in order from head to mouth.

calibration scenario becomes 3.6 mg/l (y-axis) in the load reduction scenario. Regression equations were generated for all monitoring stations, all months, and all years.

A comparison across main stem segments revealed that the average r^2 of regressions between a given scenario and the calibration run of the WQSTM decreased as scenarios became progressively more stringent (Figure 8). This pattern held for both CHL and DO regres-

sions, and across all main stem segments. Jackknife results were more mixed: the pattern of bias varied longitudinally along the main stem, although it was qualitatively similar for CHL and DO parameters (Table 3). This effect could also be observed when looking at a single location on the WQSTM grid. In one example corresponding to a well-mixed mid-Bay location (monitoring station CB4.3C), r^2 varied with depth as well as across scenarios. In cells located at

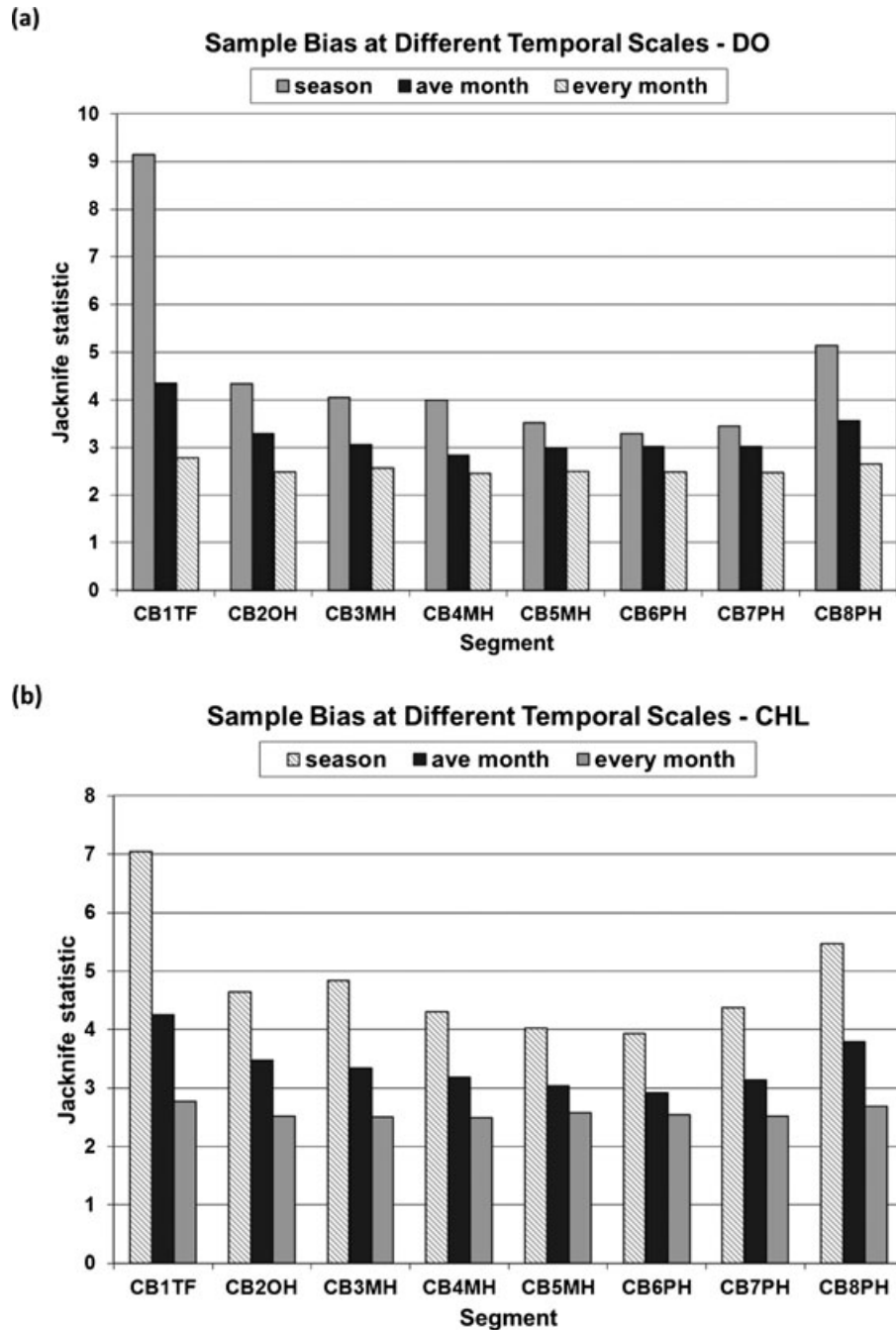


FIGURE 5. Comparison of Sample Bias (jackknife statistic) by Segment for Dissolved Oxygen (DO) (a) and Chlorophyll *a* (CHL) (b) at Different Temporal Scales. Temporal scales tested were seasons pooled (e.g., “all springs”; hatched bars), months pooled (e.g., “all Aprils”; black bars), or individual months (e.g., “April 1990”; gray bars). Lower jackknife estimates represent less sample bias.

mid-depth in the water column, scenarios were more tightly correlated (e.g., $r^2 > 0.95$) and r^2 decreased negligibly (e.g., from 0.99 to 0.97) as loads decreased from the calibration run to the E3 run (Table 4). At grid cell depths near the surface and bottom in the same location, scenarios were less well correlated (e.g., $r^2 = 0.80$) and decreased to a greater degree across loading scenarios (e.g., from 0.80 to 0.73).

DISCUSSION

Regression Method for Scenario Modification of Monitoring Data

Based on the results described above, the most appropriate spatial scale for calculation of linear

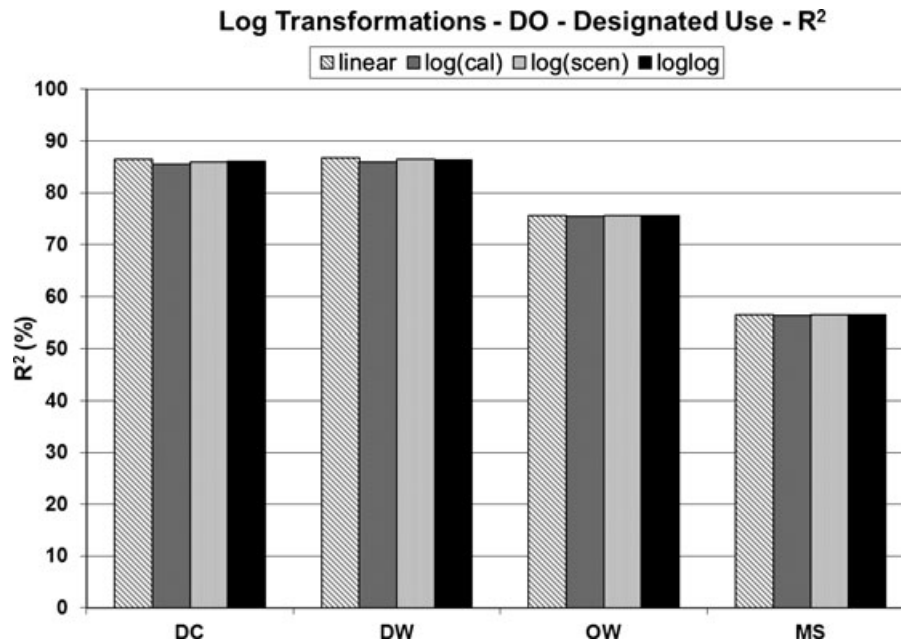


FIGURE 6. Regression Statistics of Nontransformed and Log-Transformed Dissolved Oxygen (DO) Concentrations for the DO Designated Uses. DC, Deep Channel designated use; DW, Deep Water; OW, Open Water; MS, Migratory and Spawning. See Introduction for a description of the Bay's designated uses. Concentrations from the E3 scenario run of the Water Quality Sediment Transport Model (WQSTM) were compared to output from the WQSTM run in calibration mode. "Linear" hatched bars are linear-to-linear regressions. Data from the calibration run were log transformed for the "log(cal)" (dark gray) bars; data from the E3 scenario were log transformed for the "log(scen)", or light gray, bars; data from both datasets were transformed for the "loglog" regression (black bars).

TABLE 1. Estimates of Sample Bias (jackknife) and Skewness of Dissolved Oxygen (DO) Concentrations by Designated Use (DU) for the Calibration (cal) and E3 Scenario (scen) Datasets. For jackknife estimates, "linear" denotes regression of nontransformed calibration and scenario data; "log(cal)" denotes log-transformed calibration data regressed against nontransformed scenario data; "log(scen)" denotes nontransformed calibration data regressed against log-transformed scenario data; "loglog" denotes log-transformed calibration data regressed against log-transformed scenario data. Skewness was evaluated for nontransformed data ("calib lin" and "scen lin"), and for log-transformed calibration (calib log) and scenario (scen log) datasets.

DU	Jackknife Estimate				Skewness of Dataset			
	Linear	log (cal)	log (scen)	loglog	calib lin	calib log	scen lin	scen log
Deep Channel	2.51	2.67	2.56	2.62	0.44	0.52	0.42	0.45
Deep Water	2.45	2.53	2.49	2.49	0.44	0.46	0.41	0.42
Migratory	2.47	2.49	2.48	2.48	0.45	0.49	0.47	0.48
Open Water	2.50	2.52	2.51	2.51	0.43	0.46	0.43	0.44

Note: DO, log transformation, sample bias and skewness by designated use.

regressions was determined to be the model cell corresponding to each 1-m sample depth at each monitoring station for DO, and the cell corresponding to the surface depth of each monitoring station location for CHL. The temporal scale of regressions was established as an individual month's worth of matching data for each cell corresponding to a monitoring station, sample depth, and sample date. A monthly scale uses the data points for all paired hourly (DO) or daily (CHL) values of the calibration and scenario outputs to generate a given regression equation. This has the advantage of providing a

sufficient number of points to establish each regression, while using a time period that integrates other seasonally changing influences such as the effect of temperature on reaction rates and physical processes (i.e., DO saturation in water). Aggregating to a more coarse temporal and/or spatial scale would increase scalar inconsistency between the monitoring and modeled datasets without improving regression statistics. It was determined that DO data did not require transformation before analysis, but that it was appropriate to log-transform CHL data before regression.

TABLE 2. Average Skewness (Skew Stat) of Chlorophyll *a* (CHL) Data When Log Transformed (“log”) Compared to Nontransformed Datasets (“linear”).

All CHL	Ave. of Skew Stat for CHL Datasets			
	Calibration Data		E3 Scenario Data	
	Linear	Log	Linear	Log
CBPSEG				
CB1TF	0.49	0.65	0.66	0.62
CB2OH	0.51	0.50	0.49	0.47
CB3MH	0.51	0.53	0.51	0.47
CB4MH	0.48	0.47	0.46	0.41
CB5MH	0.48	0.47	0.46	0.41
CB6PH	0.46	0.42	0.47	0.43
CB7PH	0.50	0.42	0.53	0.44
CB8PH	0.58	0.40	0.67	0.51
JMSTF	0.63	0.58	0.69	0.56
JMSOH	0.46	0.48	0.45	0.44
JMSMH	0.62	0.57	0.51	0.47
JMSPH	0.60	0.52	0.51	0.44

Note: CBPSEG, Chesapeake Bay Segments.

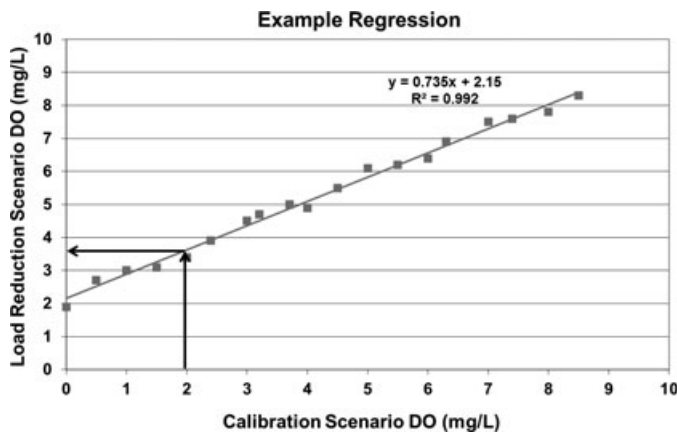


FIGURE 7. Hypothetical Example of a Linear Regression Between Water Quality Sediment Transport Model Calibration (*x*-axis) and Scenario (*y*-axis) Data.

Scenario Effect

While slight degradation of regression statistics was observed as management scenarios became more stringent, correlations remained strong on average ($r^2 > 0.80$) and demonstrated practical utility in simulating improving conditions with progressive load reductions. The decrease in r^2 that was observed with load reductions merits further study. As load reductions became more stringent, greater variability in the response of water quality parameters was observed. Whether the relationship between the calibration run and the management scenario becomes less linear with greater load reductions, or whether there is simply more scatter, is yet to be determined.

Although DO is well understood to be a function of loading (Hagy *et al.*, 2004; Kemp *et al.*, 2005; Murphy *et al.*, 2011; Cerco *et al.*, this issue), this relationship is complex both spatially and temporally. Furthermore, it is mediated through interacting physical and biological factors, such as temperature, algal growth, and stratification. Boundary effects are also likely. For example, DO is bounded by zero and saturation, and may interact with these boundaries differently at different levels of loading. The gridded nature of the WQSTM also generates spatial boundary effects that may contribute additional complexity to these relationships. Future work to better understand these interactions will further improve our understanding of the mechanisms driving the functioning of the WQSTM.

Demonstrated Utility for TMDL Development

The regression method described here was used to produce “scenario-modified” datasets for use in developing the Chesapeake Bay TMDL. Specifically, for any given management scenario that was run on the WQSTM, linear equations were calculated by regressing the scenario’s simulation results against the model calibration results for each grid cell relating to a monitoring station location and month. The linear equations were then applied to the monitoring observations for each relevant location and month. The result of applying a linear equation to an observed data point (*x*) was to produce what can be referred to as a “scenario-modified” data point (*y*). When repeated for all data points in the monitoring dataset, a “scenario-modified” dataset was produced. Because this dataset was identical in format and sampling density to the original monitoring dataset, the same assessment methods used to generate the 303(d) listing reports could be used for predicting attainment of water quality standards under various management scenarios. This is consistent with the principle of planning load reductions with the model, and testing achievement of the criteria with the monitoring data.

Application of the regression method for scenario modification was validated during development of the Chesapeake Bay TMDL. The regression method was applied to historical observations for the 10-year period 1991–2000, for which the final TMDL was developed. The application of linear regressions from individual cells of the WQSTM generated improvements in low DO concentrations. Figure 9 is an example of a single regression used to estimate changes expected in monitoring data for DO in an area of the Bay experiencing frequent low DO concentrations. This particular regression was applied only to observations occurring at this specific depth, in

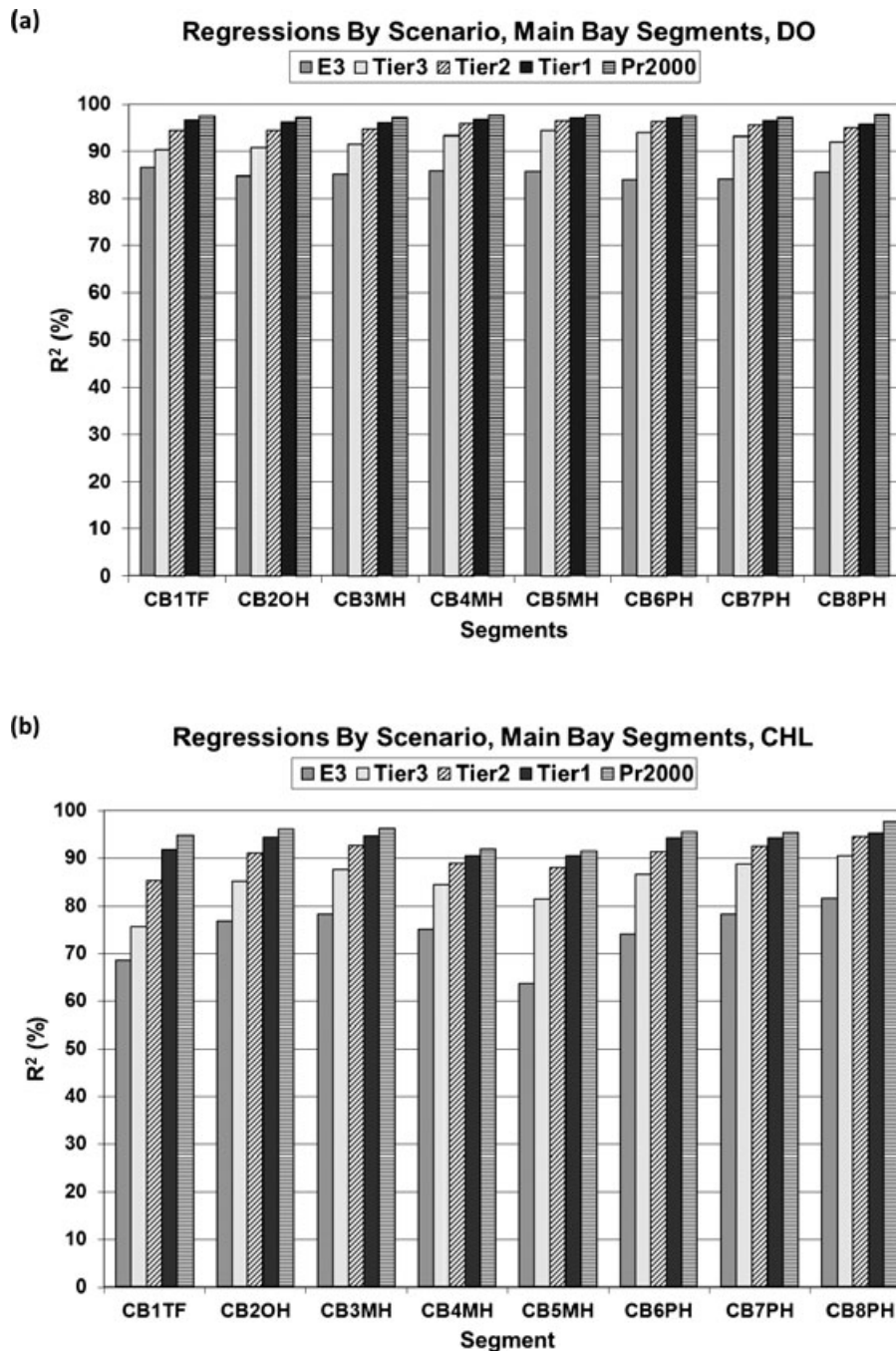


FIGURE 8. Average Regressions of (a) Dissolved Oxygen (DO) and (b) Chlorophyll *a* (CHL) for Each Main Stem Segment, for Five Load Reduction Scenarios. The output from each scenario was regressed against the output from the calibration run of Water Quality Sediment Transport Model. The first bar in each group represents the E3 scenario, which is the most stringent load reduction scenario. Each successive bar in the group represents a scenario with progressively higher (i.e., less stringent) nitrogen and phosphorus loads. For all segments, R^2 decreased as loading scenarios digressed further from the calibration run. This pattern was more pronounced for CHL than for DO.

July 1994, near the CB4.3C monitoring station. Hundreds of such regressions were necessary to generate a particular scenario dataset. This particular regression had a higher R^2 value than the average in main Bay segments for the E3 scenario as shown in Figure 8. Scenarios with a lesser reduction from the calibration generally had an R^2 similar to the rela-

tionship shown in Figure 9. Note that the regression line, denoting the expected change in observed DO under the E3 scenario, was not parallel to the 1:1 line. This indicated that the WQSTM was predicting a higher increase in DO at lower observed levels. This relationship is to be expected since DO changes due to decreased eutrophication would tend to

TABLE 3. Jackknife Estimates of Sample Bias for Dissolved Oxygen (DO) and Chlorophyll *a* (CHL) Across Scenarios, for Each Main Stem Bay Segment.

CBPSEG	DO					CHL				
	E3	Tier3	Tier2	Tier1	Pr2000	E3	Tier3	Tier2	Tier1	Pr2000
CB1TF	2.79	2.82	2.81	2.79	2.82	2.77	2.76	2.75	2.70	2.72
CB2OH	2.48	2.48	2.48	2.50	2.49	2.52	2.52	2.52	2.53	2.53
CB3MH	2.57	2.57	2.57	2.57	2.57	2.51	2.54	2.54	2.55	2.55
CB4MH	2.45	2.48	2.51	2.52	2.51	2.49	2.51	2.53	2.51	2.51
CB5MH	2.49	2.49	2.51	2.51	2.52	2.57	2.60	2.59	2.59	2.57
CB6PH	2.48	2.54	2.57	2.56	2.54	2.54	2.66	2.63	2.62	2.56
CB7PH	2.46	2.50	2.51	2.54	2.56	2.51	2.55	2.56	2.59	2.59
CB8PH	2.66	2.71	2.75	2.79	2.81	2.69	2.72	2.77	2.82	2.85

Note: CBPSEG, Chesapeake Bay Segments.

TABLE 4. Regressions by Depth for the E3 Scenario (a) and Regressions Across Scenarios at About 24 Feet Below the Surface (b).

(a)	
Cell Depth (ft)	r^2
4	0.8
8	0.91
24	0.97
36	0.97
(b)	
Scenario	r^2
Calibration	1
2009N	0.99
P1TMDL1N	0.98
2010_E3	0.97

disappear as the observed DO tended toward saturation. Scenario-modified datasets were assessed using standard 303(d) assessment methods for Chesapeake Bay DO and CHL criteria (USEPA, 2008a, 2010a). For purposes of establishing the Chesapeake Bay TMDL, the three-year assessment period of 1993-1995 was established as the “critical period” for determining attainment, although for informational purposes all three-year periods from 1991 to 2000 were assessed. Progressive improvement in DO and CHL standards attainment was demonstrated with continued nitrogen and phosphorus load reductions (Figure 10).

In Figure 10, the scenarios cover a range of both: (1) scoping scenarios used to examine the degree of nutrient load reduction required for DO criteria attainment and (2) estimates of the response of DO concentrations to key scenarios like the E3 scenario. The scenarios shown in Figure 10 illustrate the progressive improvement of DO concentrations that was simulated as pollutant loads were reduced. Scenarios are organized from highest nitrogen load to lowest,

with the total nitrogen load in millions of pounds (mpN) delivered to the tidal Bay displayed along the *x*-axis. The names of the scenarios are as follows: (1) the 1991-2000 Base Calibration (309 mpN), (2) the 2009 Scenario (248 mpN), (3) the Target Load Option A Scenario (200 mpN), (4) the Tributary Strategy Scenario (192 mpN), (5) the 190 Loading Scenario (190 mpN), (6) the 179 Loading Scenario (179 mpN), (7) the 170 Loading Scenario (170 mpN), and (8) the E3 Scenario (141 mpN). More information on these scenarios and their associated loads of phosphorus and sediment can be found in USEPA (2010c). Often observed in the DO attainment plots is a “stalling” of response of DO attainment to nutrient load reductions at low levels of criteria violation, as can be seen in Figure 10 for TMDL segment CB5MH-MD at the 1% nonattainment (i.e., criteria violation) level. This is an artifact of the DO criteria assessment method that uses time and space attainment curves to assess attainment of the DO criteria in a particulate TMDL segment (Tango and Batiuk, this issue). In particular, the low level of nonattainment often involves DO estimates of a particular bottom cell (or cells) in the model that is in nonattainment and is relatively unresponsive to nutrient reductions, while overall in the segment DO continues to improve with additional nutrient reductions.

For 7 of well over 100 segment-designated use combinations evaluated for the Chesapeake Bay TMDL, the location of the relevant monitoring station and/or boundary effects of the WQSTM prevented the use of this method for determining load response (Table 5; see USEPA, 2010b for details). These segments were almost universally located in small, shallow tributaries, often represented by only one or two WQSTM grid cells directly adjacent to land. Such conditions constrain the WQSTM’s ability to effectively integrate multiple drivers of DO concentrations. As a result, its ability to simulate water quality changes in response to dramatically reduced loads is also limited in these regions. In these cases, water

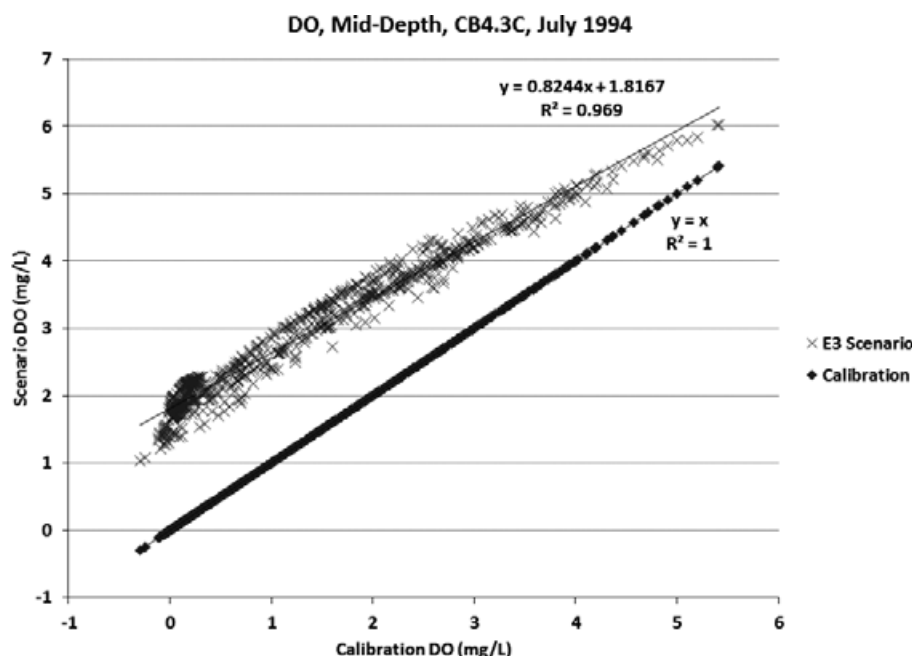


FIGURE 9. Regression of Dissolved Oxygen (DO) Concentrations (mg/l) for a Single Cell Mid-Way Down the Water Column at Station CB4.3C, for the Month of July 1994. DO concentrations from the calibration run are regressed against themselves, resulting in a regression equation of $y = x$. Regression of DO concentrations from the E3 scenario (y-axis) against the calibration run (x-axis) demonstrates that low DO concentrations are projected to improve in this location as pollutant loads are reduced. The linear equation shown can be used to generate a scenario-modified DO value (y) from a given historical observation at this same location from July 1994 (x).

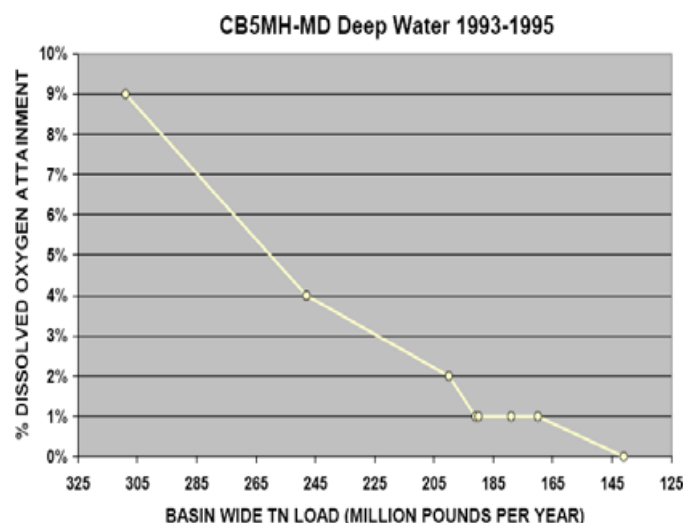


FIGURE 10. Progressive Reduction in Nonattainment of Dissolved Oxygen (DO) Standards (y-axis) Was Observed for DO Datasets that Were Modified in Regressions Generated Using Water Quality Sediment Transport Model Output from Increasingly Stringent Load Reduction Scenarios (x-axis). The example shown here is for the Deep Water designated use in Maryland's portion of mid-Bay segment CB5MH, for the critical assessment period of 1993-1995.

quality criteria violations failed to resolve under increasingly stringent load reduction scenarios. This persistent nonattainment in scenario-modified datasets was generally found to result from two or more of the following factors: (1) less than expected change in DO concentrations from the calibration scenario to

a given reduced nutrient loading scenario; (2) poor agreement between model-simulated and historically observed DO concentrations for a particular location and historical period; and (3) a limited number of unusually and/or very low DO concentrations that would be difficult to “scenario-modify” into attainment under any conditions. In these cases, additional lines of evidence — such as historical incidence of hypoxia and degree of load reduction already required in the relevant watershed — were combined with expert judgment to determine whether further load reductions were necessary. It is important to note that these exceptions accounted for only about one percent of the volume of the Bay's waters. Thus, for the vast majority of the Chesapeake Bay and its tidal tributaries, the regression method described here allowed the identification of the nitrogen and phosphorus load reductions required to attain water quality standards in the Chesapeake Bay and its tidal tributaries.

Comparison with Other TMDL Approaches

More than 40,000 TMDLs, representing impairments caused by a wide variety of pollutants, have been developed and approved by the USEPA since 1995 (USEPA, 2012). Complexity varies from TMDLs for single streams (i.e., “single segment”) impaired by one or two pollutants, to large water bodies receiving

TABLE 5. Chesapeake Bay Designated Use Segments Showing Percent Nonattainment of the Applicable Bay Dissolved Oxygen Water Quality Sediment Under the Basin-Wide Nitrogen and Phosphorus Target Loadings (million pounds per year).

	309 TN 19.5 TP 8,950 TSS 1993-1995	248 TN 16.6 TP 8,110 TSS 1993-1995	200 TN 15 TP 6,390 TSS 1993-1995	191 TN 14.4 TP 6,482 TSS 1993-1995	190 TN 13 TP 6,123 TSS 1993-1995	190 TN 12.7 TP 6,030 TSS 1993-1995	179 TN 12.0 TP 5,510 TSS 1993-1995	170 TN 11.3 TP 5,650 TSS 1993-1995	141 TN 8.5 TP 5,060 TSS 1993-1995	All Forest 1993-1995
Open Water Summer Monthly										
GUNOH	5	5	5	5	5	5	5	5	5	5
MANMH	1	5	5	5	5	5	5	5	5	0
ANATF_MD	39	19	18	12	12	12	11	11	0	0
PMKTF	11	5	5	5	5	5	5	2	1	1
WBEMH	11	15	8	8	8	8	8	8	0	0
WICMH	11	11	15	5	5	5	5	5	5	4
Deep Water										
MAGMH	35	35	16	16	16	3	3	1	1	0

Notes: Values are given in percentages.

CBPSEG, Chesapeake Bay Segments; GUNOH, Gunpowder River; MANMH, Manokin River; ANATF_MD, Anacostia River, Maryland; PMKTF, Upper Pamunkey River; WEBMH, Western Branch Elizabeth River; WICMH, Wicomico River; MAGMH, Magothy River; TN, total nitrogen; TP, total phosphorus; TSS, total suspended solids. Source: Chesapeake Bay TMDL, Appendix M.

inputs from multiple subwatersheds. Watershed TMDLs are differentiated from single-segment TMDLs by the need to assess a system comprising multiple connected water bodies and their associated watersheds (USEPA, 2008b). Thus, it has become common practice to employ modeling approaches. These range from relatively simple statistical models relating load delivery to response of key indicators, to complex process modeling efforts. The latter approach can use a watershed model, a model that simulates conditions in a receiving water body, or a coupled approach that combines the use of watershed and receiving water-body models. Multiwatershed TMDLs often use the coupled approach.

For example, the Hydrological Simulation Program - Fortran (HSPF) dynamical watershed model was coupled with the Environmental Fluid Dynamics Code (EFDC) 3D hydrodynamic receiving water model to develop the Suwannee River Basin TMDL in Georgia (GADNR and GA EPD, 2001). After model calibration, a “pristine” scenario (representing unaltered forestation and wetland conditions) was used to quantify target DO concentrations, and then a series of loading scenarios were used to identify a final acceptable loading capacity. In another example, a coupled hydrodynamic (CH3) and water quality (CE-QUAL-ICM) model was used to develop the TMDL for Delaware’s inland bays (Indian River, Indian River Bay, and Rehoboth Bay) (DE DNREC, 1998). As with other cases, output from a series of progressively restrictive nutrient loading scenarios was directly evaluated to determine necessary pollutant reductions. In yet another example, nutrient dynamics in the Cahaba River and watershed (Alabama) were simulated by coupling the Loading Simulation Program in the C++ watershed model with a one-dimensional receiving water model “EPDRIV1” based on a version of the CE-QUAL model developed by the Army Corps of Engineers (Shoemaker *et al.*, 2005; USEPA, 2005). A mass-balance Excel spreadsheet model then integrated dynamical model results with empirical data representing streamflow and nutrient sources. The results were used to compare observed phosphorus concentrations to target concentrations. While empirical data were used in this case to estimate existing conditions, again model simulations were the tool for estimating nutrient load reductions needed to meet target conditions. Finally, investigators in North Carolina developed a novel approach in which they supplemented the coupled hydrodynamic (EFDC) and water quality model (WASP6) approach with a Bayesian probability network model (“Neu-BERN”) to support development of a TMDL for the Neuse River Estuary (Borsuk *et al.*, 2002; Stow *et al.*, 2003; Wool *et al.* 2003). Supplementing the more traditional approach with a probabilistic model allowed

them to provide policy makers with a range of target reductions defined by explicit decisions regarding the desired margin of safety.

While all of these are critical and informative approaches for predicting water body response to simulated load reductions, it is recognized that models are an incomplete representation of real-world conditions. Thus, subject experts have recommended the development of methods that combine modeling simulations and monitoring datasets to inform TMDL development (NRC, 2001, 2007). The method described here represents a first effort in this direction. To our knowledge, the Chesapeake Bay TMDL represents the only existing case for which process model outputs have been used to modify empirical observations to develop predictive hybrid datasets.

Well-founded modeling applications are a triad of monitoring, research, and modeling. The field has long held that closer integration of the three would improve confidence in model performance and help remove discrepancies between model predictions and observations (NRC, 1994; Boesch, 2000, 2001). The method described in this article is a direct application of the principle that prediction, observation, and understanding are fundamentally interrelated elements of effective water quality management.

ACKNOWLEDGMENTS

The authors acknowledge the Chesapeake Bay Program (CBP) Modeling and Monitoring teams for technical review of the material and the CBP Water Quality Goal Implementation Team for contributions to method development. Thanks also to Carl Cerco for encouragement and editing.

LITERATURE CITED

- Boesch, D.F., 2000. Measuring the Health of the Chesapeake Bay: Toward Integration and Prediction. *Environmental Research Letters* 82:134-142.
- Boesch, D.F., 2001. Science and Integrated Drainage Basin-Coastal Management: The Chesapeake Bay and the Mississippi Delta. *In: Science and Integrated Coastal Management*, B. von Bodungen and K. Turner (Editors). Dahlem University Press, Berlin, Germany, pp. 37-50.
- Borsuk, M.E., C.A. Stow, and K.H. Reckhow, 2002. Predicting the Frequency of Water Quality Standard Violations: A Probabilistic Approach for TMDL Development. *Environmental Science and Technology* 36(10):2109-2115.
- Cerco, C., 1995. Response of Chesapeake Bay to Nutrient Load Reductions. *Journal of Environmental Engineering* 121(8):549-557.
- Cerco, C.F. and T. Cole, 1993. Three-Dimensional Eutrophication Model of Chesapeake Bay. *Journal of Environmental Engineering* 119(6):1006-1025.
- Cerco, C.F., S.C. Kim, and M.R. Noel, 2010. The 2010 Chesapeake Bay Eutrophication Model. A Report to the US Environmental Protection Agency and to the US Army Corps of Engineer Baltimore District, US Army Engineer Research and Development Center, Vicksburg, Maryland. http://www.chesapeakebay.net/publications/title/the_2010_chesapeake_bay_eutrophication_model1, accessed March 2012.
- Cerco, C.F. and M.R. Noel, 2004. The 2002 Chesapeake Bay Eutrophication Model. EPA 903-R-04-004. U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis, Maryland.
- Cerco, C.F. and M.R. Noel, this issue. Twenty-One-Year Simulation of Chesapeake Bay Water Quality Using the CE-QUAL-ICM Eutrophication Model. *Journal of the American Water Resources Association*, doi: 10.1111/jawr.12107.
- Cerco, C.F., M.R. Noel, and P. Wang, this issue. The Shallow-Water Component of the Chesapeake Bay Environmental Model Package. *Journal of the American Water Resources Association*, doi: 10.1111/jawr.12106.
- Clarke, K.R. and M. Ainsworth, 1993. A Method of Linking Multivariate Community Structure to Environmental Variables. *Marine Ecology Progress Series* 92:205-219.
- DE DNREC (Delaware Department of Natural Resources and Environmental Control), 1998. Total Maximum Daily Load (TMDL) Analysis for Indian River, Indian River Bay, and Rehoboth Bay, Delaware. Watershed Assessment Section, Division of Water Resources, Delaware Department of Natural Resources and Environmental Control, Dover, Delaware. <http://www.dnrec.state.de.us/DNREC2000/Library/Misc/Unorg/ibfinaltmdl.pdf>, accessed October 2012.
- GADNR (Georgia Department of Natural Resources) and GA EPD (Georgia Environmental Protection Division), 2001. Suwannee River Basin Dissolved Oxygen TMDL Submittals. Georgia Department of Natural Resources and Environmental Protection Department, Atlanta, Georgia. http://www.gaepd.org/Files_PDF/techguide/wpb/TMDL/Suwannee/FinalSuwanneeDOTMDLs.pdf, accessed October 2012.
- Hagy, J.D., W.R. Boynton, C.W. Wood, and K.V. Wood, 2004. Hypoxia in Chesapeake Bay, 1950-2001: Long-Term Changes in Relation to Nutrient Loading and River Flow. *Estuaries* 27:634-658.
- Harding, Jr., L.W. and E. Perry, 1997. Long-Term Increase of Phytoplankton Biomass in Chesapeake Bay. *Marine Ecology Progress Series* 157:39-52.
- IMSL, 2007. IMSL Fortran Numerical Stat Library, Visual Numerics, Inc. <http://www.roguewave.com/Portals/0/products/imsl-numerical-libraries/fortran-library/docs/6.0/stat/default.htm>, accessed October 2012.
- 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. Gilbert, 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.
- Linker, L.C., G.W. Shenk, P. Wang, and R. Batiuk, 2008. Integration of Modeling, Research, and Monitoring in the Chesapeake Bay Program. *In: Management of Water Quality and Irrigation Techniques*, J. Albiac and A. Dinar (Editors). Earthscan, London, United Kingdom, pp. 41-60.
- Murphy, R.M., W.M. Kemp, and W.P. Ball, 2011. Long-Term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading. *Estuaries and Coasts* 34:1293-1309.
- NRC (National Research Council), 1994. Priorities for Coastal Ecosystem Science. National Academies Press, Washington, D.C.
- NRC (National Research Council), 2001. Assessing the TMDL Approach to Water Quality Management. Committee to Assess the Scientific Basis of the Total Maximum Daily Load Approach to Water Pollution Reduction, Water Science and Technology Board, National Research Council. The National Academies Press, Washington, D.C. <http://www.nap.edu/catalog/10146.html>, accessed March 2012.

- NRC (National Research Council), 2007. Models in Environmental Regulatory Decision Making. Committee on Models in the Regulatory Decision Process, Board on Environmental Studies and Toxicology, National Research Council. The National Academies Press, Washington, D.C. http://www.nap.edu/catalog.php?record_id=11972, accessed September 2012.
- NYDEC (New York State Department of Environmental Conservation) and CT DEEP (Connecticut Department of Energy and Environmental Protection), 2000. A Total Maximum Daily Load Analysis to Achieve Water Quality Standards for Dissolved Oxygen in Long Island Sound: Prepared in Conformance with Section 303(d) of the Clean Water Act and the Long Island Sound Study. New York Department of Environmental Conservation and Connecticut Department of Environmental Protection December 2000, Albany, New York. <http://longislandsoundstudy.net/wp-content/uploads/2010/03/Tmdl.pdf>, accessed March 2012.
- Shenk, G.W. and L.C. Linker, this issue. Development and Application of the 2010 Chesapeake Bay Watershed Total Maximum Daily Load Model. *Journal of the American Water Resources Association*, doi: 10.1111/jawr.12109.
- Shoemaker, L., T. Dai, and J. Koenig, 2005. TMDL Model Evaluation and Research Needs. EPA/600/R-05/149. TetraTech, Inc., Fairfax, Virginia.
- Stow, C.A., C. Roessler, M.E. Borsuk, J.D. Bowen, and K.H. Reckhow, 2003. Comparison of Estuarine Water Quality Models for Total Maximum Daily Load Development in Neuse River Estuary. *Journal of Water Resources Planning and Management* 129(4): 307-314.
- Tango, P. and R. Batiuk, this issue. Deriving Chesapeake Bay Water Quality Standards. *Journal of the American Water Resources Association*, doi: 10.1111/jawr.12108.
- USEPA (U.S. Environmental Protection Agency), 2003. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries. EPA 903-R-03-002. U.S. Environmental Protection Agency, Region 3, Chesapeake Bay Program Office, Annapolis, Maryland.
- USEPA (U.S. Environmental Protection Agency), 2005. A Dynamic One-Dimensional Model of Hydrodynamics and Water Quality (EPDRiv 1): EPD-RIV1 Fact Sheet. <http://www.epa.gov/athens/wwqtsc/html/epd-riv1.html>, accessed October 2012.
- USEPA (U.S. Environmental Protection Agency), 2008a. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries — 2008 Technical Support for Criteria Assessment Protocols Addendum. EPA 903-R-08-001. CBP/TRS 290-08. U.S. Environmental Protection Agency, Region 3, Chesapeake Bay Program Office, Annapolis, Maryland.
- USEPA (U.S. Environmental Protection Agency), 2008b. Handbook for Developing Watershed TMDLs. DRAFT December 15, 2008. U.S. Environmental Protection Agency, Office of Wetlands, Oceans & Watersheds, Washington, D.C.
- USEPA (U.S. Environmental Protection Agency), 2010a. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries: 2010 Technical Support for Criteria Assessment Protocols Addendum. May 2010. EPA 903-R-10-002. CBP/TRS 301-10. U.S. Environmental Protection Agency, Region 3 Chesapeake Bay Program Office, Annapolis, Maryland.
- USEPA (U.S. Environmental Protection Agency), 2010b. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment. U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland. <http://www.epa.gov/reg3wapd/tmdl/ChesapeakeBay/tmdlexec.html>, accessed October 2012.
- USEPA (U.S. Environmental Protection Agency), 2010c. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment Appendix M. Chesapeake Bay Water Quality/Sediment Transport Model Management Scenario Attainment Assessment Results and 2008 303(d) Chesapeake Bay List Assessment Results: Subappendix M1 — Chesapeake Bay Dissolved Oxygen Criteria Assessment Results. U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, Maryland. <http://www.epa.gov/reg3wapd/tmdl/ChesapeakeBay/tmdlexec.html>, accessed October 2012.
- USEPA (U.S. Environmental Protection Agency), 2012. Approved TMDLs by State. http://ofmpub.epa.gov/tmdl_waters10/attains_nation_cy.control?p_report_type=T#APRTMDLS, accessed October 2012.
- Wang, P., L.C. Linker, R. Batiuk, and C. Cerco, 2006. Surface Analysis of Chesapeake Bay Water Quality Response to Different Nutrient and Sediment Loads. *Journal of Environmental Engineering* 132(3):377-383.
- Wool, T.A., S.R. Davie, and H.N. Rodriguez, 2003. Development of Three-Dimensional Hydrodynamic and Water Quality Models to Support Total Maximum Daily Load Decision Process for the Neuse River Estuary, North Carolina. *Journal of Water Resources Planning and Management* 129(4):295-306.