

# Incremental Improvements in Chesapeake Bay Environmental Model Package

Carl F. Cerco, A.M.ASCE,<sup>1</sup> and Mark R. Noel<sup>2</sup>

**Abstract:** The performance of the Chesapeake Bay Environmental Model Package is examined in four steps of model development. The steps include initial application, grid refinements, addition of living resources, and grid refinements with recalibration. Performance statistics are presented for the mainstem bay and for the James River, a major tributary. Computed salinity has the lowest relative error. Computed total phosphorus and surface chlorophyll have the greatest relative error. Errors in the bay are lower than in the James River. The capacity of the model has increased substantially over more than a decade but quantitative performance, measured by the summary statistics, has reached a plateau. Limited spatial sampling, uncertainty in loading, and difficulty in assigning boundary conditions are among the factors that limit the accuracy that can be attained with the model.

**DOI:** 10.1061/(ASCE)0733-9372(2005)131:5(745)

**CE Database subject headings:** Model accuracy; Chesapeake Bay; Eutrophication; Nutrients; Salinity; Chlorophyll; Dissolved oxygen.

## Introduction

The Chesapeake Bay Environmental Model Package (CBEMP) has been undergoing continual improvement since it was first delivered more than a decade ago. We examine here quantitative measures of performance at major steps in model development. Our objective is to identify the developments that contribute most significantly to improved model performance. We examine the mainstem of Chesapeake Bay (Fig. 1) as well as the James River, a major estuarine tributary. The bay is saline throughout and receives the vast majority of loading from distributed sources. Greatest algal biomass is found during the spring bloom although the highest ( $>30 \mu\text{g/L}$ ) chlorophyll concentrations occur in surface waters of the upper bay during the summer. Complete anoxia is a recurrent phenomenon during the summer in deep waters of the bay trench. In contrast, the James River receives a significant loading fraction from urban point sources at the fall line and near the mouth. Half the James River length is tidal fresh water subject to high ( $>50 \mu\text{g/L}$ ) summer chlorophyll concentrations. The lower river exhibits hypoxic (dissolved oxygen  $<3 \text{ mg/L}$ ) but not anoxic conditions. A second objective of our work is to compare model performance in these two systems and relate performance to system characteristics.

<sup>1</sup>Research Hydrologist, U.S. Army Engineer Research and Development Center, Mail Stop EP-W, 3909 Halls Ferry Rd., Vicksburg, MS 39180. E-mail: cercoc@wes.army.mil

<sup>2</sup>Mathematician, U.S. Army Engineer Research and Development Center, Mail Stop EP-W, 3909 Halls Ferry Rd., Vicksburg, MS 39180.

Note. Discussion open until October 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on October 9, 2003; approved on August 16, 2004. This paper is part of the *Journal of Environmental Engineering*, Vol. 131, No. 5, May 1, 2005. ©ASCE, ISSN 0733-9372/2005/5-745-754/\$25.00.

## Four Steps

### Step 1: 4,000-Cell Model

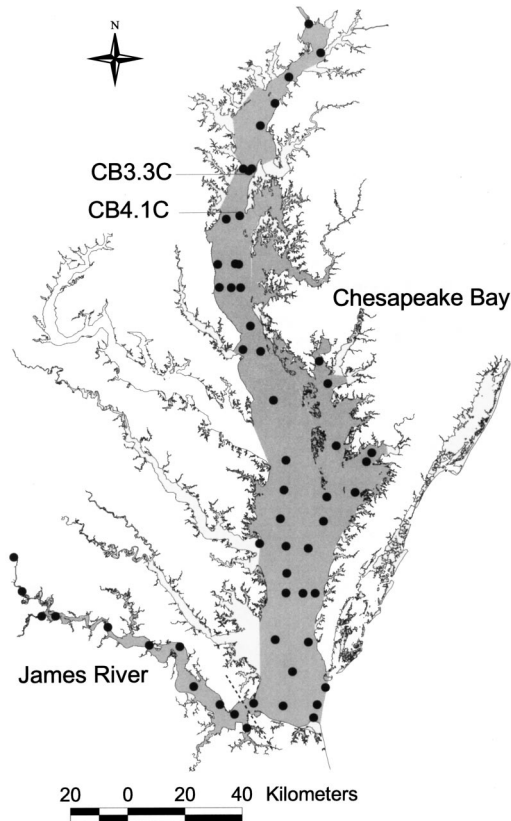
Step 1 is the CBEMP as completed in 1992 (Cerco and Cole 1993). The model simulated the years 1984–1986 on a grid of 4,000 elements (Table 1) and incorporated 22 state variables. The eutrophication model represented physical variables as well as the carbon, nitrogen, phosphorus, oxygen, and silica cycles, but no living resources or trophic levels above phytoplankton. Transport processes were provided by the *CH3D-WES* hydrodynamic model (Johnson et al. 1993). A Lagrangian averaging process (Dortch et al. 1992) was employed to compute tidal-average transport for the eutrophication model from intratidal quantities computed by *CH3D-WES*. Distributed loads were computed by Phase 1 of the U.S. EPA's Chesapeake Bay Watershed Model (WSM), a highly modified version of the HSPF model (Bicknell et al. 1996).

### Step 2: 10,000-Cell Model

In Step 2, the model was transferred to a grid of 10,000 elements (Table 1) which extended outside the bay mouth onto the continental shelf (Cerco and Meyers 2000). The Lagrangian-average hydrodynamics were replaced with intratidal hydrodynamics specified at 2-h intervals. Fall-line and distributed loads were updated to Phase 4.2 of the WSM, point-source loading estimates were revised, and the simulation period was extended to 1985–1994. All model parameters and coefficients were maintained at their Step 1 values.

### Step 3: Virginia Tributaries Model

In Step 3, new kinetics that incorporated living resources were installed. The grid and hydrodynamics from Step 2 were retained. This run introduced two zooplankton groups, deposit- and filter-feeding benthos, and submerged aquatic vegetation to the model (Cerco and Meyers 2000). The addition of living resources necessitated extensive recalibration and revision of parameters



**Fig. 1.** Chesapeake Bay and the James River showing sample stations

and coefficients. Atmospheric loads to the water surface were revised and phosphorus loads from bank erosion were incorporated.

#### Step 4: 12,000-Cell Model

In Step 4 the model was moved to a grid of 12,000 elements (Table 1). Kinetics formulations were the same as Step 3. Distributed loads were updated to Phase 4.3 of the WSM. The objective of this phase was to prepare the model for management use in meeting new dissolved oxygen, water clarity, and chlorophyll goals. Both the hydrodynamic and water quality models were extensively recalibrated to improve computations of these three key parameters on the new grid.

**Table 1.** Characteristics of Three Model Grids

Grid property	Mainstem Bay			James River		
	Step 1	Steps 2, 3	Step 4	Step 1	Steps 2, 3	Step 4
Cells	2,850	5,153	6,546	139	678	700
Surface cells	421	897	1,215	42	258	258
Volume ( $10^9$ m <sup>3</sup> )	77.9	68.7	64.8	3.0	2.9	2.8
Surface area ( $10^9$ m <sup>2</sup> )	7.04	7.00	6.75	0.51	0.59	0.52
Mean depth (m)	11.1	9.8	9.6	5.8	4.8	5.3
Maximum depth (m)	23.5	26.5	26.5	15.9	18.9	18.9
Mean cell volume ( $10^6$ m <sup>3</sup> )	27.3	13.3	9.9	21.4	4.2	3.9
Mean cell surface area ( $10^6$ m <sup>2</sup> )	16.7	7.8	5.6	12.2	2.3	2.0

**Table 2.** Number of Model-Data Comparisons

Observation	Mainstem Bay	James River
Salinity	3,481	820
Total nitrogen	3,487	829
Total phosphorus	3,425	830
Dissolved oxygen	523	115
Light attenuation	1,646	368
Chlorophyll	1,677	339

#### Statistics and Other Computations

Comparisons were based on a continuous simulation of 1985–1986. These years were selected because they were common to all model applications. For salinity, total nitrogen, and total phosphorus, effects of revisions on model performance were examined by computing temporal-volumetric mean concentrations over the simulation period. For chlorophyll and light attenuation, surface computations only were examined. Dissolved oxygen was restricted to computations in the bottom layer during the summer. We formed this restriction for two reasons. First, critical dissolved oxygen concentrations, of interest to management, occur in deep waters during the summer. Second, dissolved oxygen at the surface is near saturation. Including the extensive volume of saturated water into the summary “dilutes” the effect of model revisions on the critical oxygen calculations.

Quantitative statistics were determined through comparison of model computations with observations from 11 (James River) to 46 stations (Chesapeake Bay) sampled once or twice per month (Fig. 1). For salinity, total nitrogen, and total phosphorus, comparisons were made with surface and bottom samples. For chlorophyll and light attenuation, surface samples only were examined while dissolved oxygen comparisons were restricted to bottom samples collected in the summer. The number of model-data comparisons ranged from roughly 100 to 3,400 (Table 2).

Statistics computed were mean error (ME), absolute mean error (AME), and relative error (RE). ME and RE were defined by Cerco and Cole (1993). AME is given by

$$AME = \frac{\sum |O - P|}{N} \quad (1)$$

in which  $O$  = observation;  $P$  = prediction; and  $N$  = number of observations. The ME describes whether the model overestimates or underestimates the observations, on average. The AME is a measure of the distance between individual observations and compu-

**Table 3.** Nutrient Loads at Multiple Model Steps

System	Load	Nitrogen (kg/day)				Phosphorus (kg/day)			
		Step 1	Step 2	Step 3	Step 4	Step 1	Step 2	Step 3	Step 4
James River	Point source	27,270	27,806	27,806	27,864	3,907	4,233	4,233	4,240
	Distributed	20,083	20,319	20,319	24,477	3,113	3,480	6,133	6,703
	Atmospheric	2,126	2,459	1,256	1,254	90	104	104	105
	Total	49,479	50,584	49,381	53,595	7,110	7,817	10,471	11,047
Mainstem Bay	Point source	28,771	28,091	28,091	28,372	2,117	2,143	2,143	2,152
	Distributed	182,892	179,213	179,213	198,430	9,691	7,981	9,384	14,158
	Atmospheric	32,931	31,828	16,322	16,789	1,396	1,368	1,369	1,408
	Total	244,594	239,132	223,626	243,591	13,204	11,492	12,895	17,717

tations. RE provides a statistic suitable for comparison between different variables or systems.

## Results

### Grid Revisions, Step 2

The entire system was regridded for the tributary refinement effort. The surface plane was determined from geographic information system shoreline maps. Over 2.4 million soundings were obtained from the National Ocean Service Hydrographic Data Base. As a result of the regridding, volume and mean depth of both systems decreased (Table 1). These changes originated in improved model representation of shallow, shoreline regions. Maximum depth increased in both systems as a result of improved representation of channel geometry. This step was an intermediate in the development of the complete tributary refinement effort. Before implementing major additions to the eutrophication model, we decided to isolate and quantify the effects of the revised grid.

Alterations in nutrient loads accompanied the change to the new grid (Table 3). These were the result of revisions to the WSM, improved point-source data, and alterations in surface

area. For practical purposes, nitrogen loading was unchanged from the initial 4,000-cell run while changes in phosphorus loading were roughly 10%.

Grid revisions increased salinity in the bay and decreased salinity in the James (Table 4). Changes in computed total nitrogen and phosphorus are linked to changes in computed salinity. Increased salinity in the bay indicates a larger proportion of ocean water relative to fresh water. Diminished salinity in the James indicates a larger proportion of freshwater relative to ocean water. Loads to these systems are nutrient enriched, relative to ocean water. A greater proportion of freshwater should result in higher nutrient concentrations. A greater proportion of ocean water should result in lower nutrient concentrations. The decrease in total nitrogen and phosphorus in the bay and increase in total nitrogen and phosphorus in the James conform to these principles (Table 4). Changes in concentration move ME away from the ideal value of zero in both systems (Table 5). For the most part, AME increases (Table 6). Total nitrogen in the bay presents a dilemma, however. The ME moves away from zero, indicating diminished model performance while the AME decreases indicating improved model performance.

Computed chlorophyll and oxygen result from interactions of multiple physical and biochemical processes. For the grid alter-

**Table 4.** Temporal-Volumetric Mean Computed Concentrations

Substance	System	Step 1	Step 2	Step 3	Step 4
Salinity (ppt)	Mainstem	18.2	19.3	—	16.7
	James River	14.4	11.3	—	11.8
Light attenuation (1/m)	Mainstem	1.01	1.00	1.15	1.03
	James River	2.55	2.63	2.72	2.28
Chlorophyll ( $\mu\text{g/L}$ )	Mainstem	9.77	8.55	8.96	7.98
	James River	9.81	11.51	12.81	14.02
Total nitrogen (mg/L)	Mainstem	0.61	0.58	0.62	0.60
	James River	0.76	0.91	0.93	0.93
Total phosphorus (mg/L)	Mainstem	0.026	0.025	0.031	0.028
	James River	0.076	0.090	0.103	0.110
Dissolved oxygen (mg/L)	Mainstem	6.02	6.29	4.95	5.16
	James River	7.83	7.26	6.5	6.23

**Table 5.** Mean Error Statistics

Substance	System	Step 1	Step 2	Step 3	Step 4
Salinity (ppt)	Mainstem	-0.275	-0.652	—	1.688
	James River	-0.949	0.552	—	0.324
Light attenuation (1/m)	Mainstem	-0.085	0.015	-0.147	0.065
	James River	-0.329	-0.449	-0.425	0.009
Chlorophyll ( $\mu\text{g/L}$ )	Mainstem	-1.549	-0.152	-1.520	-0.325
	James River	0.242	-0.298	-4.840	-2.470
Total nitrogen (mg/L)	Mainstem	0.002	0.059	0.023	0.025
	James River	0.108	-0.182	-0.174	-0.205
Total phosphorus (mg/L)	Mainstem	0.014	0.017	0.011	0.012
	James River	-0.008	-0.033	-0.046	-0.053
Dissolved oxygen (mg/L)	Mainstem	-0.775	-1.109	0.522	-0.032
	James River	-1.880	-0.472	0.258	-0.534

ations, these substances conform to a basic paradigm. Decreased nutrients in the bay induce decreased chlorophyll concentration and higher bottom-water dissolved oxygen. Increased nutrients in the James produce increased chlorophyll concentration and lower bottom-water dissolved oxygen (Table 4). In the bay, the revised chlorophyll concentrations are an improvement, as indicated by the ME statistic. In the James, computed chlorophyll moves from a value which is low, on average, to a value which is high, on average, and slightly larger in magnitude. Changes in AME are consistent with changes in ME. These quantities diminish in the bay and increase in the James. The ME statistic indicates bottom-water dissolved oxygen was initially high in both systems. The grid revisions produce increased ME in the bay and diminished ME in the James. AME in the James diminishes, consistent with the diminished ME. In the bay, a dilemma occurs again as indicated by diminished AME despite increased ME.

### ***Tributary Refinements, Step 3***

The tributary refinements involved the addition of living resources, revisions in model formulations, and revised load estimates, on top of the grid changes (Cercio and Meyers 2000). Light attenuation in the Step 1 model was based on chlorophyll concentration and fall-line flow. Fall-line flow was employed as a surrogate for suspended solids load and location of the turbidity maximum, two factors that influence extinction. Since Step 2 was an interim phase on the way to the complete tributary refinements, the effort to relate attenuation to flow on the new grid was not completed. Instead, attenuation was assigned, on a monthly basis, to large regions of the system. For the completed tributary refinements, Step 3, attenuation was computed as a function of fixed and volatile solids, both of which were model state variables (Cercio and Meyers 2000).

**Table 6.** Absolute Mean Error Statistics

Substance	System	Step 1	Step 2	Step 3	Step 4
Salinity (ppt)	Mainstem	1.768	1.720	—	2.325
	James River	1.600	1.628	—	2.043
Light attenuation (1/m)	Mainstem	0.367	0.344	0.469	0.385
	James River	0.723	0.986	1.324	0.879
Chlorophyll ( $\mu\text{g/L}$ )	Mainstem	4.967	4.401	4.900	4.717
	James River	8.640	11.450	9.170	7.741
Total nitrogen (mg/L)	Mainstem	0.149	0.141	0.143	0.145
	James River	0.307	0.442	0.443	0.453
Total phosphorus (mg/L)	Mainstem	0.018	0.020	0.017	0.017
	James River	0.051	0.064	0.070	0.080
Dissolved oxygen (mg/L)	Mainstem	1.938	1.691	1.362	1.181
	James River	2.010	1.288	2.413	2.702

In the Step 1 model, total phosphorus predictions in the bay were less than observed (Cercio and Cole 1993). Omission of phosphorus loads associated with bank erosion was suggested as a source of the discrepancy. For the Virginia Tributaries application, bank loads were introduced in the form of particulate inorganic phosphorus (PIP). Bank solids loads were taken from Corps of Engineers estimates (U.S. Army 1990). Net solids loads to the water column were taken as 20% of the total. (The balance is sand that forms beaches along erosion banks.) PIP associated with the bank solids, 0.1% by weight, was taken from Keefe (1994). Distributed PIP loads were calculated as 0.1% of the solids load provided by the WSM. To keep total phosphorus loads consistent with the WSM, the new PIP load was subtracted from the WSM "organic" phosphorus load. This strategy was successful in the bay. Increased phosphorus loads (Table 3) consisted of bank loads. In the James, the estimated PIP load exceeded the original organic phosphorus load and a substantial increase in distributed phosphorus load (now all PIP) resulted. The PIP introduced from bank loading ( $\approx 90 \text{ kg day}^{-1}$ ) was trivial by comparison to the increased fall line load. Investigations (Keefe 1994; Conley et al. 1995) suggest little, if any, of the PIP exchanges with the dissolved phosphorus pool. Consequently, PIP was treated as inert within the modeled water column. A fundamental distinction was that PIP was transported and deposited identically to fixed solids while particulate organic phosphorus was transported and deposited as with other particulate organic matter.

Atmospheric loads (Table 3) for the Step 1 model were based on EPA estimates supplemented with data from the National Atmospheric Deposition Program (NADP). Loads were estimated on a seasonal basis and incorporated no spatial or intra-annual variations. For the tributary refinements, nitrogen loads were based exclusively on more extensive NADP data. Loads varied temporally and spatially based on daily rainfall records at stations throughout the basin. The revised atmospheric nitrogen loads were roughly half the previously estimated loads. Atmospheric phosphorus loads in the tributary refinements remained at their original values.

Total nitrogen and phosphorus concentrations in both systems increased over their previously computed, Step 2, values (Table 4). Since total nitrogen loads diminished, the increase in total nitrogen must be attributed to the major recalibration that accompanied the introduction of living resources. For phosphorus, the effects of the recalibration cannot be isolated from the effects of increased loads although it is reasonable to assume the processes that produced high nitrogen, overall, also increased phosphorus. The increase in total nitrogen moved the ME in both systems closer to zero (Table 5). AME (Table 6) showed little change. All statistics for total phosphorus also improved in the bay. In the James, the incorporation of PIP raised total phosphorus concentrations which were already too high. ME moved further away from zero and AME increased.

Relating light attenuation to computed solids produced higher light attenuation in both the bay and James River (Table 4). As a result, ME moved away from the ideal value of zero (Table 5), and AME increased (Table 6).

Surface chlorophyll concentrations increased by small amounts (Table 4) in both the bay and James River. The increased nutrient concentrations, especially nitrogen, likely play a role in the increased chlorophyll but these cannot be isolated from the revised kinetics and new state variables. The higher chlorophyll concentrations increased concentrations that were already high, on average, as indicated by the ME statistic (Table 5). In the bay, model performance diminished, as indicated by the AME

(Table 6). The James repeated a previous dilemma. Performance measured by the AME improved even though performance measured by the ME diminished.

Computed bottom dissolved oxygen diminished in both systems (Table 4). Dissolved oxygen which was high, on average, became low, on average and the computed mean moved closer to the observed mean (Table 5). The origin of the altered dissolved oxygen is difficult to isolate although the basic paradigm illustrated in Step 2 appears to hold. That is, nutrient concentrations increased in Step 3, leading to higher chlorophyll concentrations and lower bottom dissolved oxygen. In the bay, AME diminished, consistent with the improvement in the ME statistic (Table 6). The James produced a now-familiar dilemma. AME increased despite the improvement in the ME statistic.

#### **Grid and Calibration Refinements, Step 4**

The grid was again revised in Step 4. Revisions in shallow, resource-rich areas were intended to improve the living-resource aspects of the model. At the same time, regulatory forces were shaping the direction of management efforts. Portions of Chesapeake Bay were listed as "impaired." Impairments were defined as low dissolved oxygen, excessive chlorophyll, and diminished water clarity. Management emphasis shifted from living resources back to living-resource indicators: dissolved oxygen, chlorophyll, and clarity. A model recalibration was undertaken, with emphasis on improved accuracy in the computation of the three key indicators. The recalibration included a complete reevaluation of the hydrodynamic model. In the water quality model, attention was devoted to reproducing key processes as well as concentrations. In particular, computations of primary production and respiration, which were undercomputed in the earlier steps, were improved.

For Step 4, the Watershed Model was upgraded from Phase 4.2 to Phase 4.3. The revised WSM computed larger distributed nitrogen loads resulting in an  $\approx 8\%$  increase in total load to both systems (Table 3). Phosphorus loading increased substantially in the bay, primarily due to increased estimates of bank solids loads.

The new grid and revised hydrodynamic model diminished computed salinity in the bay by several ppt and increased salinity by a lesser amount in the James River (Table 4). Salinity which was initially high, on average, in the bay became low, on average, and more distant from the observed mean (Table 5). Salinity which was initially low in the James River moved closer to the observed mean. Salinity AME in the bay grew larger, consistent with the larger magnitude ME (Table 6). AME also increased in the James despite the improved computation of mean salinity.

Computed total nitrogen in the bay diminished (Table 4) despite the increased load; computations which were low, on average, became lower still (Table 5). Computed total nitrogen increased in the James; computations that were high grew still higher. Total phosphorus followed the pattern of total nitrogen. Computed concentrations diminished in the bay and increased in the James. AME for nitrogen and phosphorus showed little change in the bay (Table 6). These statistics for both substances increased in the James. The concentration changes in the bay appear due to altered boundary conditions and/or circulation near the mouth. A budget created for the Step 1 model indicated the bay exported 80 t total nitrogen per day and imported 8 t per day total phosphorus across the bay mouth. A similar budget for the Step 4 model indicated nitrogen export increased to 113 t per day while phosphorus import diminished to 2 t per day.

Computed light attenuation diminished in both the bay and

James Rivers (Table 4), accompanied by dramatic improvements in the ME statistic (Table 5). For the most part, corresponding improvements occurred in the AME (Table 6).

The response of surface chlorophyll and bottom dissolved oxygen to the Step 4 changes followed the precedent established in previous runs (Table 4). Decreases in total nutrients in the bay were accompanied by decreased chlorophyll and increased bottom dissolved oxygen. Increases in total nutrients in the James were accompanied by increased chlorophyll and decreased bottom dissolved oxygen. Computed chlorophyll and dissolved oxygen in the bay both moved closer to observed means (Table 5). The AME statistics in the bay showed corresponding improvements (Table 6). The response of ME statistics in the James showed no correspondence with the change in concentrations (Table 5). Although the average chlorophyll in the river increased, the average at observation stations decreased so that ME which initially indicated high chlorophyll computations improved. Dissolved oxygen at sample stations increased despite the riverwide decrease. The ME statistic which originally indicated low dissolved oxygen computations, on average, reversed to indicate high concentrations, on average. The AME statistic indicates improved chlorophyll calculations in the James while this same statistic indicates deterioration in dissolved oxygen calculations (Table 6).

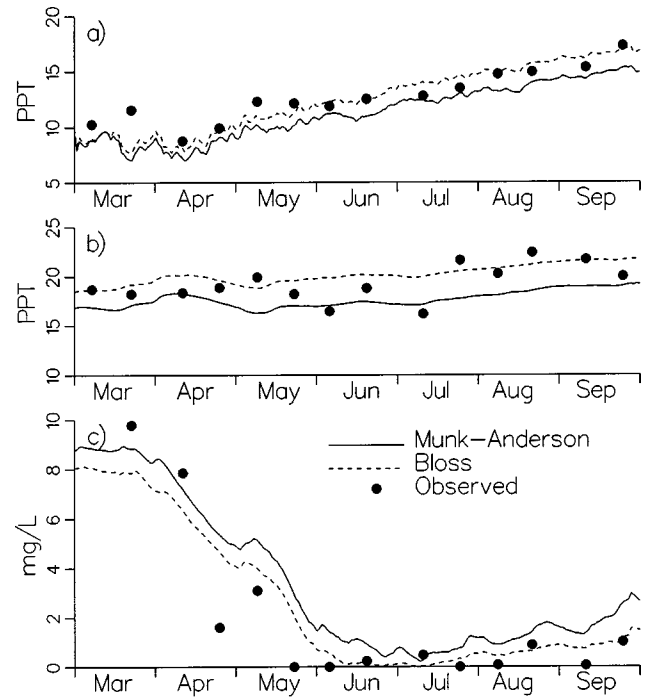
## Discussion

### Grid Revisions

In an ideal world, refinement of the model grid produces improved performance. A larger number of smaller cells improves resolution of geometry and should improve representation of transport processes. As the number of cells increases to a limit in which truncation is the only source of error, then more cells must result in less truncation error and improved model performance. In a less-than-ideal world, grid refinements do not infallibly produce model improvements. Although our system is now gridded on the order of  $10^4$  cells, we are far from the point at which truncation is the only source of error. Circulation and eutrophication processes are influenced by subgrid scale processes that must be parameterized into a feasible computational grid. Grid alterations carry the risk of consciously or unconsciously changing a parameterization that provides successful model performance.

The grid went through three stages of development in this study. The Step 1 grid was developed from a navigation chart. The model realistically represented the longitudinal dissolved oxygen distribution in the bay (Cercio and Cole 1993). For the Virginia tributary refinements, we switched to automated grid generation based on digital data. Several intermediate grids were produced prior to the Step 2 grid. These produced substantially different and completely unacceptable longitudinal dissolved oxygen distributions in the bay. The problem was traced to a narrow channel in the upper bay that was averaged out in the automatic grid generation. The channel was restored and the model produced acceptable results.

Our Step 1 hydrodynamics were developed ahead of, and independently from, the eutrophication model. Our success with this approach led us to declare that, in the future, rigorous hydrodynamics would eliminate tinkering with transport to calibrate water quality (Cercio and Cole 1993). Our enthusiasm for absolute hydrodynamic rigor was premature. Our experience with multiple grids leading to Step 2 showed that hydrodynamics could not be



**Fig. 2.** Effect of Bloss and Munk-Anderson relationships on computed (a) Surface salinity; (b) bottom salinity; and (c) bottom dissolved oxygen at station CB4.1C, March-September 1985.

developed independently of water quality. By the time we arrived at Step 4, we realized that intercalibration of both hydrodynamics and water quality was necessary.

The hydrodynamic model computed vertical viscosity using a two-equation turbulence closure scheme (Johnson et al. 1991). Vertical diffusivity was determined through the Prandtl number (the ratio of turbulent momentum transport to turbulent mass transport). Numerous formulations have been proposed to describe the effect of stratification on Prandtl number. Formulations tested in the hydrodynamic model included the Bloss formulation (Bloss et al. 1988) and the Munk-Anderson relationship (Munk and Anderson 1948). Both formulations related Prandtl number to Richardson number (the ratio of turbulence suppression, via vertical density gradient, to turbulence creation, via velocity shear). For any Richardson number, the Bloss relationship provided lower vertical diffusivity than the Munk-Anderson relationship.

Sensitivity runs were performed with the hydrodynamic and water quality models. The appropriate relationship for Prandtl number was impossible to determine through examination of salinity alone. Visual, and even statistical, comparisons were indeterminate (Fig. 2). The Bloss relationship consistently provided lower bottom dissolved oxygen computations, however. During the summer, in the bay, when dissolved oxygen computations are crucial, the lower concentrations provided by the Bloss relationship were closer to observations so the Bloss relationship was selected. If the two models were not tested together, the Munk-Anderson relationship might have been employed resulting in profound impact on subsequent computed bottom dissolved oxygen.

### Loads and Boundary Conditions

A summary of model behavior can be gained from viewing the systemwide mean computed concentrations (Table 4). The model

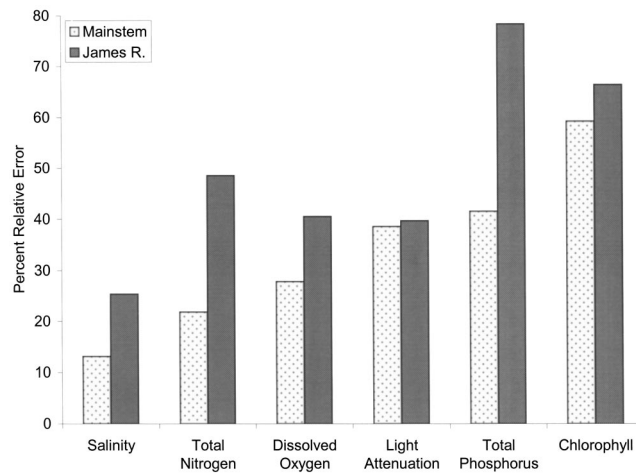


Fig. 3. Step 4 relative errors in the Mainstem Bay and James River

conforms to long-held paradigms that indicate increasing nutrient concentrations produce higher chlorophyll and diminished bottom-water dissolved oxygen concentrations. It is revealing to explore the processes that altered the nutrient concentrations. No kinetics parameters or processes were changed from Step 1 to Step 2. Changes in loadings were small. The changes in total nutrients were attributable to altered circulation that accompanied the expansion of the grid onto the continental shelf. Nutrient concentrations in the bay diminished from Step 3 to Step 4 despite increased loading. The changes in nutrient concentrations were attributed to altered nutrient exchange at the mouth of the bay. More nitrogen was exported and less phosphorus was imported than in the earliest model implementation. These behaviors indicate the fundamental importance of assigning system boundary conditions. These include both concentration boundaries and circulation processes.

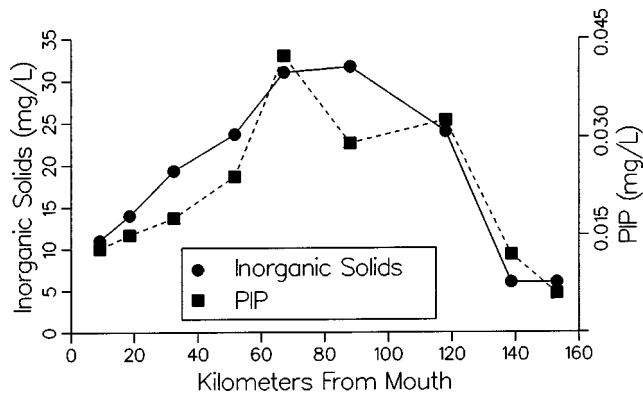
Assigning these boundary conditions in an estuary can be extraordinarily difficult. Our earlier work presented a dilemma (Cercó 1995). The earliest model version used observed concentrations at multiple stations across the mouth of the bay. We were reluctant to use observed concentrations in load-reduction scenarios, however. We feared these concentrations might change as a result of nutrient load reductions in the headwaters of the bay. A major intent of the Step 2 extension of the grid onto the continental shelf was to move the open boundaries beyond the influence of loads to the bay. This extension traded one set of problems for another. Open boundaries were moved from a location with abundant observations to a location with few or no observations. By specifying constant concentrations at the edges of the Step 2 grid we were able to represent mean observed concentrations at the bay mouth but variability, especially of total nitrogen, was not represented (Cercó and Meyers 2000). The Step 4 recalibration was confounded by the inability to specify boundary conditions at the bay mouth. An expedient was adopted in which observed conditions at the bay mouth were projected to the edges of the grid and kinetics processes were disabled outside the bay mouth. This process produced reasonable concentrations at the bay mouth although mass transport across the mouth differed from earlier model versions. The different mass transport affected nutrients, chlorophyll, and dissolved oxygen systemwide.

### Relative Errors

The RE statistic provides a measure for comparing performance across model variables and between systems. Salinity and chlorophyll represent two extremes of RE (Fig. 3). Salinity has the least error; chlorophyll displays some of the greatest RE. Salinity has the least error because the processes that determine salinity are purely physical transport processes. Moreover, salinity is bounded at the upper and lower ends by oceanic salinity and freshwater. The salinity statistics represent a lower bound on model performance. Since salinity is a measure of physical transport processes, other variables cannot perform better than salinity unless compensatory errors are present in loading, boundary conditions, or other factors. Chlorophyll represents the other extreme; chlorophyll is nearly a pure biological variable. The effects of loading and boundary conditions are negligible relative to internal production and losses. No analog to oceanic salinity exists to limit the observed and computed concentrations at the upper end. Errors in transport processes and computed nitrogen, phosphorus, and light attenuation propagate into errors in computed chlorophyll.

One factor that contributes to large chlorophyll errors is the distribution of chlorophyll in patches. Weiss et al. (1997) examined the chlorophyll distribution in lateral transects across the bay. The maximum distance over which two chlorophyll values were correlated was always less than 2 km and often less than 1 km. Average cell dimensions in the bay are  $\approx 2.5 \text{ km} \times 2.5 \text{ km}$  and computed concentrations represent averages over this area. As a consequence of patchiness, point observations of chlorophyll are not necessarily representative of spatial averages over model cell sizes. We expect that the effects of patchiness “average out” when mean statistics are computed over large numbers of observations. When single observations are examined, however, as in the AME and RE statistics, large errors are possible. The patchy distribution of chlorophyll suggests smaller cell sizes are appropriate. Smaller cell sizes will reduce the error introduced by averaging over cell area, but grid refinements alone will not identify nor simulate the complex physical and biological interactions that create chlorophyll patchiness.

We expect RE in total nitrogen and total phosphorus to be similar since they have similar origins and play similar roles in the model ecosystem. A closer look, however, reveals substantial

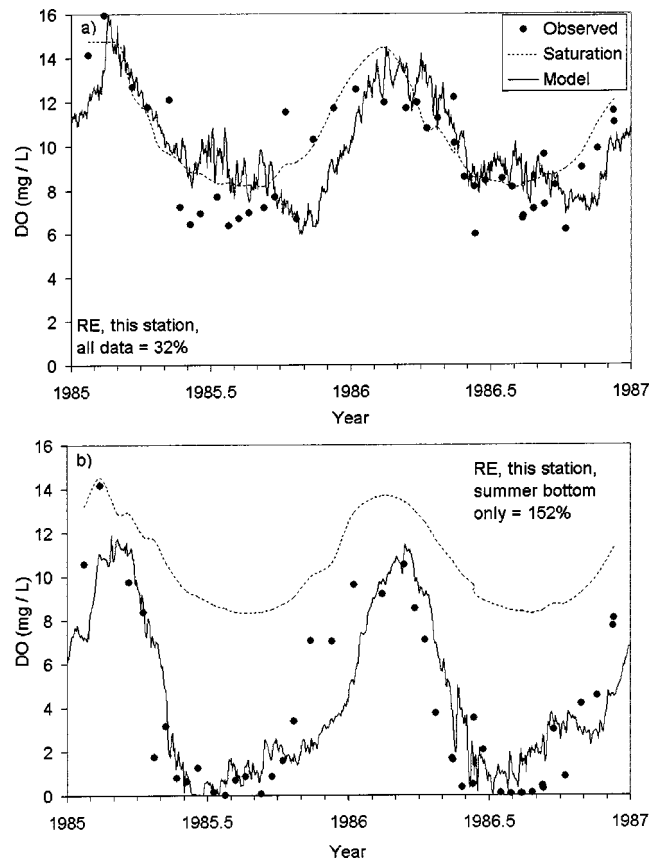


**Fig. 4.** Association of inorganic solids and particulate inorganic phosphorus along the James River axis. Median concentrations from 1994 surveys.

differences between these two nutrients. The most significant difference is the presence of a particulate inorganic phosphorus component. The inorganic fraction of particulate phosphorus ranges from 23 to 86% in upper Chesapeake Bay (Keefe 1994). The inorganic fraction of particulate nitrogen is less than 1%. The processes that affect exchange between the particulate and dissolved nutrient pools differ for the organic and inorganic particulates. We also perceive differences in transport and settling. Transport processes for particulate organic phosphorus are expected to be similar to other organic detritus while transport of PIP is similar to inorganic suspended solids (Fig. 4). We distinguished PIP in our Step 3 model. PIP was considered biologically inert and was transported identically to inorganic suspended solids. This approach was, no doubt, a more realistic representation of the phosphorus cycle although performance statistics did not indicate any particular improvement. Ultimately, the approach was abandoned. Reasons included absence of precise loading data, absence of information on kinetics processes, and limitations to suspended solids transport. We think the distinction between the two phosphorus components can and should be reactivated. Substantial improvement in model performance will require accompanying investigations into PIP kinetics and loads. A rigorous sediment transport component will also be required.

More than 20 years ago, Thomann (1982) reported that the median RE in 15 dissolved oxygen models was 10%. Estuarine components of the New York harbor system exhibited RE of 5–35%. Our results, at the upper end of this range, are attributable to restriction of statistical computations to summer, bottom dissolved oxygen. When we employ year-round, surface and bottom observations, our RE is  $\approx 10\%$ , consistent with the previous summary. We feel, however, that use of all observations does not provide a good measure of model performance. When all observations are considered, the predictive ability of the model is strongly influenced by the ability to calculate saturation dissolved oxygen concentration (Fig. 5). Computation of summer, bottom dissolved oxygen requires that the model correctly portray interactions of physical and biological processes. These include vertical mixing, respiration, and sediment oxygen demand. Although our RE is higher than previously reported, it provides a better measure of the ability of the model to compute dissolved oxygen as a function of multiple modeled processes.

RE for most substances is larger in the James River than in Chesapeake Bay (Fig. 3). Larger errors in the tributaries than in



**Fig. 5.** Observed, saturated, and modeled (a) Surface; and (b) bottom dissolved oxygen at Station CB3.3C

the bay have been a characteristic of the model since its earliest stages. In the past, we attributed the disparity to the coarse grid in the tributaries, relative to the bay. Improving the initial coarse grid was the primary justification for the Virginia Tributary Refinements (Steps 2 and 3). Initial instinct is to claim the James River grid is still not fine enough and undergo another round of grid refinements. In the latest Step, however, James River cells are less than half the size of bay cells (Table 1). We have already demonstrated that finer grids do not necessarily yield improved results. We believe the disparity is related to loading rather than resolution. Head-of-tide loads at the bay and major tributaries are well-monitored. The monitoring program provides observations for calibration of the WSM as well as for independent computation of loads. In contrast, no means exist to verify loads below the fall line. These include loadings from local watersheds, point-source loads, atmospheric loads to the water surface, and loads from bank erosion. In the James, a primary source of uncertainty is in point-source loads. Point sources comprise 70–80% of the below-fall-line nutrient load in the James (Table 3). Information provided by local agencies presently characterizes each point source by an annual mean, however, with no daily, monthly, or seasonal variation.

When normalized by receiving-water volume, below-fall-line nutrient loads to the James River are an order of magnitude larger than loads to the bay (Table 7). Normalized by cross-sectional area, below-fall-line loads to the James River exceed loads to the bay by a factor of 5 (Table 7). Although uncertainty exists in below-fall-line loads to both systems, the effects of uncertainty are damped by the enormous volume of the bay, relative to the

**Table 7.** Normalized below-Fall-Line Nutrient Loads

Normalized load	Nutrient	Mainstem Bay	James River
Load per unit volume (mg/m <sup>3</sup> /day)	N	1.25	12.93
	P	0.18	2.16
Load per unit cross section (kg/m <sup>2</sup> /day)	N	0.409	1.991
	P	0.059	0.333

river. We believe the James River will always demonstrate higher model error, relative to the bay, until a higher degree of accuracy can be assigned to James River loads.

### **What Have We Gained?**

We were surprised when we assembled the statistics from multiple model stages. We expected to see monotonic improvements in performance as the model developed. Although a few parameters, notably summer, bottom dissolved oxygen in the bay, have shown the expected improvement, other substances have shown little improvement while statistics for a few substances have steadily deteriorated. As a whole, the model performance, judged by summary statistics, has been largely stationary despite more than a decade of enhancements.

The modeler's first instinct when a model is not performing as expected is to question the data. We believe a superb monitoring program is in place in the Chesapeake Bay system. Still, characteristics of the monitoring program contribute to the lack of quantitative improvement. Most sample stations are located in the deepest portion of the cross section. The grid improvements were instituted to improve model representation in shallow, resource-rich areas. A specific goal of the first round of grid refinements was to accurately represent surface area and depth within areas that support, or may support, submerged aquatic vegetation. Refinement of the grid in these areas has minimal impact on quantities, such as bottom-water dissolved oxygen, measured in the channel. We believe refinement of the model representation in littoral zones provided a vastly more useful model. Still, the present monitoring program does not allow quantitative assessment of performance in areas that were subject to the greatest change in resolution and in modeled processes.

The primary model improvements have been in capacity rather than in quantitative performance. Light attenuation in the Step 1 model was primarily based on observed relationships to location and flow. The predictive portion of the initial model was restricted to changes in chlorophyll. In Step 3, suspended solids were introduced to the model suite and light attenuation was related to concentration of organic and inorganic solids. This effort to predict light attenuation initially diminished model performance compared to the original model. In Step 4, attention was devoted to refining the suspended solids model. As a consequence, model performance is now comparable to the original model. No doubt exists that the present model, which allows evaluation of solids management actions on attenuation, is superior to the initial model although statistical performance is similar.

The Step 3 model introduced living resources to the model suite. The introduction was management-driven. Managers desired quantitative estimates of improvements in vegetation and shellfish rather than quantification of improvements in living-resource parameters such as dissolved oxygen. We modelers hoped the new living resources would improve model performance. For example, the introduction of zooplankton might im-

prove chlorophyll predictions through dynamic computation of the predation term. Overall, quantitative improvements due to the introduction of living resources are difficult to perceive. Still, a model that can directly quantify the impact of management actions on valuable living resources must be viewed as superior to a model that cannot make these quantifications, even if statistical performance of the two models is similar.

### **Use of Statistics**

Calibration of various steps of the model was accomplished by visual comparison of observations and computations. Statistics, produced at the completion of major model phases, were intended to describe model performance. Statistics were not employed to guide the model calibration. We have demonstrated here that summary statistics can be very useful in evaluating and interpreting model performance. We might have made better use of summary statistics during the various model applications rather than in this retrospective view. Still, we caution against the use of statistics as the sole or even primary guide to model calibration. Our work has revealed shortcomings as well as strengths in statistical applications. One difficulty is in judging the significance of statistical changes. At what point do changes in summary statistics indicate important changes in the state of model calibration? A larger difficulty is in judging conflicting statistics. In multiple cases examined, performance evaluated by ME diminished while performance evaluated by AME improved. Which model application is superior? Another problem lies in the inability to optimize all substances in all regions. The Step 4 model has the best mainstem bay dissolved oxygen statistics of all runs. At the same time, the salinity statistics are the worst. From Step 3 to Step 4, dissolved oxygen statistics improved in the bay but grew worse in the James River. Statistics are valuable as measures of model performance but, inevitably, human judgment must be involved in determining the optimal state of model calibration.

### **Conclusions**

Improvements to the CBEMP over a 10-year period have been primarily in model capacity. Enhancements include more detailed spatial resolution, predictive capability for light attenuation, and direct calculation of living resources. Quantitative performance, measured by summary statistics, has reached a plateau. Limited spatial sampling, uncertainty in loading, and difficulty in assigning boundary conditions are among the factors that limit the accuracy that can be attained with the model.

### **Acknowledgments**

The CBEMP is sponsored by the Chesapeake Bay Program Office, U.S. Environmental Protection Agency, and by the U.S. Army Engineers Baltimore District. Permission was granted by the Chief of Engineers to publish this information.

### **Notation**

*The following symbols are used in this paper:*

$N$  = number of observations;

$O$  = observed concentration; and

$P$  = computed concentration.

## References

- Bicknell, B. et al. (1996). *Hydrologic simulation program—FORTRAN user's manual for release 11*, United States Environmental Protection Agency Environmental Research Laboratory, Athens, Ga.
- Bloss, S., Lehfeldt, R., and Patterson, J. (1988). "Modeling turbulent transport in stratified estuary." *J. Hydraul. Eng.*, 114(9), 1115–1133.
- Cerco, C. (1995). "Response of Chesapeake Bay to nutrient load reductions." *J. Environ. Eng.*, 121(8), 549–557.
- Cerco, C., and Cole, T. (1993). "Three-dimensional eutrophication model of Chesapeake Bay." *J. Environ. Eng.*, 119(6), 1006–1025.
- Cerco, C., and Meyers, M. (2000). "Tributary refinements to Chesapeake Bay model." *J. Environ. Eng.*, 126(2), 164–167.
- Conley, D., Smith, W., Cornwell, J., and Fisher, T. (1995). "Transformations of particle-bound phosphorus at the land-sea interface." *Estuarine Coastal Mar. Sci.*, 40, 161–176.
- Dortch, M., Chapman, R., and Abt, S. (1992). "Application of three-dimensional Lagrangian residual transport." *J. Hydraul. Eng.*, 118(6), 831–848.
- Johnson, B., Kim, K., Heath, R., and Butler, L. (1991). "Development and verification of a three-dimensional numerical hydrodynamic, salinity and temperature model of Chesapeake Bay." *Technical Rep. HL-91-7*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.
- Johnson, B., Kim, K., Heath, R., Hsieh, B., and Butler, L. (1993). "Validation of three-dimensional hydrodynamic model of Chesapeake Bay." *J. Hydraul. Eng.*, 119(1), 2–20.
- Keefe, C. (1994). "The contribution of inorganic compounds to the particulate carbon, nitrogen, and phosphorus in suspended matter and surface sediments of Chesapeake Bay." *Estuaries*, 17(1B), 122–130.
- Munk, W., and Anderson, E. (1948). "Notes on the theory of the thermocline." *J. Mar. Res.*, 3, 276–295.
- Thomann, R. (1982). "Verification of water quality models." *J. Environ. Eng. Div. (Am. Soc. Civ. Eng.)*, 108(5), 923–940.
- United States Army Corps of Engineers Baltimore District. (1990). *Chesapeake Bay shoreline erosion study feasibility report*, Baltimore, Md.
- Weiss, G., Harding, L., Itsweire, E., and Campbell, J. (1997). "Characterizing lateral variability of phytoplankton chlorophyll in Chesapeake Bay with aircraft ocean color data." *Mar. Ecol.: Prog. Ser.*, 149, 183–199.