New York & New Jersey Harbor

Contaminant Assessment and Reduction Project (CARP)

Model Evaluation Group Final Peer-Review Report

Joel Baker
Chair, CARP Model Evaluation Group
Chesapeake Biological Laboratory
University of Maryland

1. Overall scope of review. The overall objective of the Contaminant Assessment and Reduction Project (CARP) is to quantify loadings and movement of chemical contaminants in the New York & New Jersey Harbor with sufficient accuracy to explain current and to predict future concentrations of chemicals in Harbor sediments and organisms. Integral to this effort is the development of a coupled water quality model. Due to the complexity of contaminant transport in the estuary and the range of chemical classes of interest (PCBs, PAHs, pesticides, mercury, and metals), no existing model was suitable to meet the needs of the CARP program. Specifically, it was recognized at the beginning of CARP that a linked model that combines hydrodynamics, sediment transport, organic carbon dynamics, oxygen and sulfur cycling, and contaminant partitioning and bioaccumulation would be needed. These design specifications were detailed in a request for proposals, and subsequent proposals were evaluated by an external peer review group.

Developing models of this nature requires a large number of decisions and interpretations, each of which impacts the final model performance and capability. Given the complexity, it is very difficult to review the structure of the final model without first participating in reviews throughout the course of the model development. For this reason, CARP established a Model Evaluation Group (MEG). The purpose of the MEG was to provide technical review and guidance to the model development group throughout the process to insure that each component of the overall model was thoroughly evaluated and its performance understood before the linked model was assembled. The MEG met regularly with CARP participants throughout the study, including those responsible for the CARP field programs and the model development group. When necessary, targeted meetings or conference calls were arranged to address specific issues that arose during the field program and the model development. The Hudson River Foundation aided the MEG by providing meeting space and logistics. The MEG reviewed each component of the field program (ambient monitoring, loadings calculations) and model as they were completed.

Throughout the CARP program, the MEG continually evaluated the model against the following core questions:

1. Does each component of the CARP model reflect the best available science?

2. Are all available field data appropriately used in model development and calibration?
3. Does the model structure and complexity, especially spatial and temporal resolution, match the intended uses and the available data from the study area?
4. How accurate are the model results, and what tools are available to evaluate and communicate model uncertainty?

2. MEG Members

Joel Baker, University of Maryland	organic contaminant cycling
Frank Bohlen, University of Connecticut	sediment transport, dredging
Richard Bopp, RPI	organic contaminant cycling
Joe DePinto, LimnoTech, Inc.	water quality modeling
Joe DiLorenzo, Najarian and Associates	hydrodynamics
William Fitzgerald, University of Connecticut	mercury and metals cycling
W. Rockwell Geyer Woods Hole Oceanographic Institution	sediment transport/hydrodynamics
Lawrence Sanford[*], University of Maryland	sediment transport
Jay Taft, Harvard University	organic carbon/water quality modeling

*resigned

3. Overall Evaluation of the CARP Model

The design of the CARP model was driven by the original objective to predict contaminant concentrations in water, sediments, and biota throughout the Harbor over time scales of years to decades. These spatial and temporal scales, coupled with the available field observations for calibration, dictated to some extent the model resolution. The model is designed to, and best used for, exploring spatial trends across the entire Harbor, rather than detailed analysis within specific tributaries, berthings, or embayments. For example, the model grid is fairly coarse in many tributaries—this is likely suitable for predicting transport of chemicals on the scale of the entire Harbor, but insufficient to predict detailed behavior within these tributaries. Based on our review, the MEG makes the following conclusions:

1. The CARP model provides an important framework to organize and interpret field data and to identify gaps in data and knowledge.
2. The CARP model is technically sound and represents the best and latest science available at the time of model development.
3. The CARP model development utilizes available data to the greatest extent possible in terms of it complexity (*i.e.*, the spatial and temporal resolution and the process kinetics).
4. The CARP model is suitable for use in predicting relationships between external contaminant loadings and concentrations in water and sediment throughout the NY/NJ Harbor. Predictions of concentrations in biota are less certain due to, among other things, relatively few calibration data.

5. Of the chemical classes modeled, mercury is the most difficult due to its complex geochemistry in the estuary, especially the processes controlling conversion of inorganic mercury to methylmercury.
6. The relationships between current loadings and ambient chemical concentrations are fairly well characterized, but it is more difficult to predict the long-term rate of change in concentrations.
7. The response time of the system is long due to storage in sediments, and the projections are sensitive to the sediment initial conditions and especially to the model parameterization of net sedimentation rates in various parts of the system.
8. Uncertainties increase with time in the projections. This is a characteristic of all models.
9. The model is based on six years of observations. Extreme events for all parts of the system are not included in the projections.
10. The influence of the Hudson River on Newark Bay is uncertain. The CARP model shows a relatively high amount of sediment entering Newark Bay from the Hudson River, which might have important management implications. This behavior needs to be verified by further field studies.
11. The CARP Program has advanced the understanding of contaminant behavior in the Harbor. This increase in understanding should continue to evolve by further field investigations and model refinement.
12. The model should be continually maintained and improved. Findings from the CARP model should drive these studies, i.e., investigate transport of sediments into Newark Bay.
13. A particular research need identified by the CARP model development exercise is to improve the empirical relationships and mechanistic understanding of contaminant levels in biota, sediment, and water. A related need with respect to the bioaccumulation model is the quantitative formulation of food web bioenergetics and trophic transfer processes.

4. Results of Individual Model Component Reviews

Hydrodynamics. Overall, the report demonstrates that the model is a useful tool for simulating the hydrodynamics of the system during the CARP years. The model passes accurate water depths to other sub-models. Generally, the model passes representative water temperatures for computation of temperature-dependent rate constants and kinetics. The model passes representative bottom velocities for the computation of bottom stress, except in upper Newark Bay. The model appears to provide an adequate simulation of constituent (salt) transport, except during the summer of 2002.

Specific recommendations for the hydrodynamic model include:

1. Investigate why the model appears to: (1) underestimate bottom velocity in Upper Newark Bay; (2) underestimate salinity in summer 2002; and (3) reduce semi-diurnal tidal variations in salinity.
2. Characterize the system's salient hydrodynamic features and transport processes and discuss how well the model reproduces such features.

3. Provide values for <u>all</u> adjustable model coefficients and revise and expand error analysis methods and discussion.
4. In the linkage section of the report, describe how diffusivities are computed and aggregated for passage to sub-models

In the future, additional salinity comparisons for the CARP years may be performed using available NOAA data at Sandy Hook and WHOI data at various Hudson Estuary locations. The latter would provide an excellent test of the model's ability to reproduce documented patterns of fortnightly variability in salinity stratification/de-stratification and provide greater model confidence.

Sediment Transport. Strengths: The CARP model is a state-of-the-art model, with prognostic equations for sediment and carbon distributions. The carbon model is well calibrated and shows significant predictive skill. Predicted suspended sediment distributions often in agreement with observations (except near turbidity-maximum region).

Weaknesses: In the lower Hudson River section of the model, sediment trapping distributions are inconsistent with observations of temporally-variable sediment storage, especially along the river flanks (e.g., dredging data, Woodruff et al., 2001). The fine scale temporal and spatial sediment transport processes in this are not captured by the model, but the implications of this on the broader, system-wide scale are not known. Significant adjustment of parameters were required to achieve reasonable fit to field data. In the draft CARP model documentation, justification of the values chosen for various parameters and the detailed description of the model set-up generally were not consistent with peer-reviewed publication-quality.

Hydrophobic Organic Contaminant (HOC) Fate and Transport. Overall, the CARP modeling group has done an excellent job of developing this model. They had to model a very complex system and include a large suite of chemicals, with somewhat marginal data considering the size and complexity of the system and the questions. This made the task quite difficult, and the MEG is overall impressed with their compilation and synthesis of data from all possible sources, including the literature and previous theory, to pull together a well-integrated whole system model. Having said that, the summation of the utility of this model is that it has value in predicting the direction and rate of surface sediment response in various zones of the model (i.e., Newark Bay, lower Hudson River, etc.) to a change in loading from a given source category or from a given source area. It is not clear how much of the uncertainty that will be encountered in trying to predict the absolute value of a chemical's response to a specific load change or the time it will take to achieve that value.

Bioaccumulation Modeling.

Strengths:

1. Linked carbon flow and contaminant models allow investigation of potential impacts of changing nutrient loads and food web structure on HOC accumulation.

2. Ratio approach is straightforward and consistent with regulatory programs; bioaccumulation modeling allows mechanisms underlying trends in ratio-derived values to be explored.
3. Structure allows further integration/expansion with fisheries models (IBMs, bioenergetics).

Weaknesses:

1. Uniform food web structure and ratios (BAF/BSAF) in space and time.
2. Relatively sparse Harbor-specific data describing BAF and BSAF for key species
3. Predicts 'average' concentrations in biota, but risk assessors may be more interested in extremes (especially for human health assessments).

Mercury and Metals Modeling. At the beginning of the CARP program, it was recognized that modeling mercury accumulation in NY & NJ Harbor organisms would be one of the most challenging aspects of model development. Unlike most target organic pollutants, mercury must undergo a key biogeochemical process in the environment (methylation) to be transformed to the species that bioaccumulates in organisms. Modeling the relationship between external loadings (of both inorganic and methyl mercury) and resulting mercury concentrations in biota requires explicit calculation of methylation and demethylation rates within the estuary. These rates depend upon several factors, including sulfate reduction rates (sulfate-reducing microorganisms are thought to be responsible for mercury methylation) and inorganic mercury bioavailability (controlled in turn by speciation with organic matter, sulfides, and other phases). The overall rate of *in situ* mercury methylation depends in the 'controlling' (*e.g.* slowest) process, which is likely either microbial activity (which scales to the sulfate reduction rate) or inorganic mercury bioavailability (which depends on geochemical speciation in the sediments). In addition, the concentration of methyl mercury in the estuary is a balance between production *via* methylation and loss by demethylation reactions. It is likely that the relative importance of these competing processes differs spatially across the NY&NJ Harbor.

The initial CARP modeling strategy sought to take advantage of the existing calibrated sulfur and redox chemistry modules of the System-Wide Eutrophication Model (SWEM) as a chassis for developing the CARP mercury cycling model. The approach was to calculate the mercury methylation rates from modeled sulfate reduction rates. Independent research conducted concurrently with the CARP model development shows that in certain areas of the NY & NJ Harbor model domain that mercury bioavailability may be the controlling process rather than sulfate reduction rates. Overall, our knowledge of mercury dynamics in estuaries was relatively incomplete at the beginning of the CARP program and has rapidly evolved in recent years. Modeling always lags a bit behind the most recent research finding—this is not a criticism but rather reflects the dynamic between model development and fundamental research. The CARP mercury model is an appropriate framework, and as additional research results become available, they should be incorporated into subsequent versions of the CARP model.

5. Appendices - Original Model Evaluation Group Written Reviews

Model Evaluation Group (MEG) Review of: "A Model for the Evaluation and Management of Contaminants of Concern in Water, Sediment, and Biota in the NY/NJ Harbor Estuary Hydrodynamic Sub-Model" by HydroQual, Inc.

This review examines HydroQual's Hydrodynamic Sub-Model Report (the "Report") for the Contaminant Assessment and Reduction Project (CARP). The objective of the Report is to provide a comprehensive description of the CARP hydrodynamic sub-model and its suitability for use in CARP fate and transport calculations. The Report provides a comprehensive description of the application of the CARP hydrodynamic model, but not of the model itself. However, the reader is provided with ample references (e.g., Blumberg et al. 1999) and cross-references (e.g., Blumberg and Mellor, 1987) that describe this well-known and frequently used model. The objective of the CARP hydrodynamic sub-model computations is to provide time series of water depths, current velocities, volume transport rates (in three dimensions), salinities, temperatures and dispersion coefficients. These hourly averaged variables are passed to other CARP sub-models to calculate advective and dispersive fluxes, temperature-and-salinity–dependent rate constants and kinetics coefficients, bottom shear stresses, etc.

Many others have reviewed previous applications of this hydrodynamic model to the NY-NJ Harbor Estuary. These include peer reviews in scientific and engineering journals (Blumberg et al. 1999; Blumberg and Pritchard, 1997), reviews by three previous Model Evaluation Groups (MEGs), and reviews by various government agencies and their consultants. Accordingly, this review focuses primarily on the performance of this model in its present application (CARP).

The model's computational grid is similar to previously reported versions. The specified horizontal grid spacing ranges from a minimum of about 100 m in inland rivers to 50 km in New York Bight. The grid still appears somewhat coarse in the important Newark Bay region, with 1-2 computational cell widths. The model grid includes ten vertical layers to resolve the estuarine vertical structure.

The model is forced primarily with observed physical data (with some empirical adjustments) at the model's boundaries. As in previous applications (Blumberg et al. 1999), the modelers assume that the water surface elevation at the offshore boundaries is a superposition of an astronomical tidal component, a sub-tidal component (assumed to be proportional to observed, low-pass-filtered sea level at Sandy Hook), and a prescribed steady geostrophic component (based on surface slope data collected in August 1978). The model includes a well-documented radiation boundary condition (Blumberg and Kantha, 1985). Also, salinity and temperature are prescribed at the offshore boundary based on mean monthly climatological data reported in the 1998 NOAA World Ocean Atlas (adjusted by 2 ppt or 2°C).

At the water surface, wind stress and heat flux are calculated based on the usual suite of observed meteorological variables (at J.F.K airport, Albany airport, Bridgeport airport and several offshore buoys). Fresh water inflows are input to the model at 25 locations based on flow data recorded at the limited number (34) of available U.S.G.S. gaging station in the tributary rivers. Also, river discharge temperatures are assumed to be the same as daily surface temperatures recorded at the Battery (a first-order approximation). The model benefits from HydroQuals' extensive database of reported wastewater treatment plant discharge data (and flow statistics) and previous model estimates for discharges from CSOs and stormwater runoff.

The Report does not provide the selected values for some of the model calibration coefficients. These include the selected bottom drag coefficients and horizontal diffusion parameter, C_s. However, previously selected values for these adjustable parameters are provided in Blumberg et al. (1999).

The validity of model assumptions and parameter selections may be assessed through the Report's numerous comparisons of model results with observed data. These include some continuous model-data comparisons for four variables (tidal elevation, temperature, bottom salinity and velocity) at the CARP Newark Bay/Kills monitoring stations, and continuous comparisons for one variable (tidal elevation) at various additional NOAA monitoring stations. Continuous comparisons for temperature, bottom salinity and velocity are not provided outside of the Newark Bay/Kills complex. Instead, model results are compared to discrete data collected by several agencies at varying intervals (e.g., weekly, bi-weekly, monthly, seasonal, etc.) throughout the model domain.

Visually, the model appears to reproduce well the observed tidal elevation variations at almost all stations. This includes the timing and magnitude of the observed peaks and troughs, the spring-neap variation, the sub-tidal variability, and the low sub-tidal variability. In particular, there is good agreement displayed (in Appendix 4) for the sub-tidal sea level components at several stations. This agreement suggests that the model adequately simulates the sub-tidal volume transport. Model agreement is less favorable at Montauk Point, though this station is beyond the main area of interest (and located within a small embayment). Unfortunately, the Report provides no quantitative skill assessment for simulated water surface elevations or currents. However, a favorable quantitative assessment was provided in the previous paper (Blumberg et al. 1999; Tables 1-2), and similar results are anticipated here.

Also, the present (CARP) model appears to reproduce well the observed tidal current variations at Newark Bay station 3 (Lower Newark Bay) and Perth Amboy. Also, a good visual comparison is presented in year 2002 for the two Kills stations. At Newark Bay station 1 (upper Newark Bay), the model appears to reproduce the observed velocity pattern at the surface; however, the model appears to underestimate observed bottom currents consistently (by a factor of 2 or more). This disagreement

HOC FATE AND TRANSPORT

REVIEW AND COMMENTS ON CARP HOC FATE AND TRANSPORT MODELING COMPONENT

Joseph V. DePinto
August 30, 2006

General Approach

The CARP modeling framework includes a linked series of fine-scale models (hydrodynamic, sediment transport, organic carbon sorbent, contaminant fate and transport, and bioaccumulation) to allow the prediction of water column, sediment and biota concentrations of a suite of contaminants in the NY-NJ Harbor system as a function of various potential source control measures that might be implemented. The MEG has been conducting an ongoing peer review of this model development effort that includes model construction and configuration, model calibration, model evaluation including diagnostic analysis, and ultimately model evaluation. This document presents a review of the HOC fate and transport part of the overall CARP model. The primary question is our assessment of the strengths and weaknesses of the model and our recommendation as to what it can be used for with confidence and what uses of the model may be problematic or entail excess uncertainty.

In assessing the above primary question, I have reviewed three aspects of the overall model development process: 1) model framework; 2) model calibration; and 3) model diagnostic and evaluation results. Comments on each of these three aspects are presented below followed by a general overview statement on this part of the overall CARP model.

Model Framework

Spatial resolution - fine enough to assess relative contamination in different areas of the system including navigation channels, but not fine enough to determine response of specific port/slip areas to load changes

Processes and process formulations –

three phase, site-specific partitioning on homolog basis is well developed and reasonable. Use a value of A_{DOC} = 0.08; seems reasonable. Also, use three-phase partitioning in bed with a higher value of A_{DOC}.

chemical transformations not used for PCBs or dioxins/furans – seems OK;

Volatilization uses a single, temperature corrected mass transfer rate of 1 m/d based on SF_6 experiments. Model does not include wind effects on this rate, claiming dominance of tidal velocities. **I am not sure this is valid throughout the system; provide evidence of lack of sensitivity to wind**.

sediment bed processes – porewater diffusion uses values from SWEM (2-3 X molecular diffusion) seems OK unless bioturbation is causing this rate to be higher – bed mixing process is based on bioturbation with benthic biomass proportional to deposition of labile organic carbon. This process is crucial to establishing the upper mixed layer depth, which strongly impacts system response time. **Are there dated cores that can provide data-based estimate of mixing depth?**

Model Calibration

Model input data development for the current conditions calibration was a major part of the study. Tributary load estimates were very important. The approach to develop tributary loads was reasonable, given the level of data available.

Configuration of sediment bed for F&T model and establishment of initial conditions is crucial for a short-term calibration. This is a source of uncertainty.

Current calibration using 1998-2002 really only looks at water column relative to current external loads; cannot really calibrate the sediment and sediment-water exchange processes that govern long-term trends. See later where they estimate ~30 years for system to come to steady-state with current loads; this is because the sediment response is much slower than the water column. So, a four year start up for calibration does not really eliminate errors in setting sediment initial conditions.

Calibration is evaluated with three types of plots: plots of model versus data for pairs that match in space and time; cumulative probability plots (these allow one to look for bias at either high or low values of state variables in the system); and traditional time series plots at various locations in the system. My qualitative evaluation of the current conditions calibration for PCBs and dioxins is as follows (see Figures 4.7 and 4.8 for summary; detailed results in Appendices 4B-PCBs and 4C-dioxins/furans):

For mono- and di-PCBs model is biased high throughout the concentration range;

For tri- and tetra-PCBs model is biased high at low concentrations, which occur largely in the bight – perhaps model is not volatilizing these lower chlorinated compounds fast enough and it shows up as the model accumulating more in the bight than is being measured;

For the higher PCB homologs, the calibration is pretty good, within a factor of 10 in water column;

For the dioxins/furans, the model calibration is fairly good – within a factor of 10 throughout. But it tends to be biased a little high for water column concentration (due to dissolved phase) of most congeners.

Model Diagnostic and Evaluation

HQI claimed that model uncertainty analysis was beyond the scope of the project and stated that EPA said it was difficult to do a quantitative uncertainty analysis for contaminant fate and transport models; and, therefore, they did not perform it.

- HQI did a long-term (96 years) run to steady-state with the current loads and zero sediment initial conditions. Found that 32 years is sufficient to reach sediment steady-state for this scenario. In addition to the time to steady-state, the endpoint sediment concentrations indicate where the sediments will end up when the current higher sediment concentrations come into steady-state with the current loads. This is a very useful diagnostic (see figure 5-1 for example and figure on penta-PCB from appendix 10). This and other plots show that the sediments in many areas are declining under the current load conditions (i.e., the sediments are still reflecting historic loads). **I would have liked to have seen a time plot of surface sediment concentrations for a few model cells beginning with initial interpolated conditions and running to steady-state.**
- I would have been very interested in seeing rusults of model runs for site-specific K_{oc}'s very K_{ow}-based K_{oc}'s. Does it make a different in the steady-state surface sediment concentrations?
- **We were told that wind effects were not important for air-water gas phase exchange, because tidal effects were dominant. I am not sure I believe this for New York harbor area and in the Bight. I would like to see a plot of computed mass transfer for this area for a velocity-driven formulation in comparison with a wind-driven formulation for the typical range of velocities and wind.**
- **Section 6 contained results of a 1965 – 2002 hindcast confirmation of the model.**
  - Hydrodynamics, sediment transport, and organic carbon were simulated using surrogate years from 1988-89, 1994-95, and 1998-2002 based on comparison with Hudson River flow for a given year. **It is not clear to me that there was any consideration of a change in nutrient loads that might have affected organic carbon production during a given period in the simulation. A significantly higher carbon production in early years might have meant a larger sedimentation rate and hence contaminant burial rate during this period. Is this possible?**
  - Cs hindcast looks quite good. They have done a good job of capturing the time and space (especially in the HR, upper, and lower bay) gradients in sediments. **This result speaks well of the models ability to capture the sediment response time. This is a very important model evaluation result.**
  - The PCB hindcast was conducted by re-constructing the historical PCB loading in the system in a manner similar to what was done for the Delaware River problem. Appendix 7A (time series) and 7B (spatial transects) display the results. The PCB hindcast is not nearly as good as the Cs hindcast, probably because of larger uncertainty in reproducing historical time series of loads. There are some fairly bad time series for some cells in the upper HR (see plots for cells 10,44 and 14,83). Also, the dry weight normalized sediment concentrations seem to compare better to data than the organic carbon normalized values; this is probably due to changes in f_{oc} over time that are not captured by the model.
- **Section 12: Loading Component Analysis**
  - This section looked at the contribution of various loading categories (atmospheric, ocean boundary, CSO's, STP's, tributaries, head of tide, and in-place sediment) to the sediment and water column concentrations over time. In general, the results indicate that it is almost all

about in-place sediment levels (and the HR load for PCBs) and the rate at which they decline. Some examples (see Figures 12.2 and 12.3) demonstrate this result. This finding places a premium on getting the sediment – water exchange and sediment burial rates correct, and I think for the most part this model does that well in those areas that are important for management questions.

Summary

Overall, HQI has done an excellent job of developing this model. They had to model a very complex system and include a large suite of chemicals, with somewhat marginal data considering the size and complexity of the system and the questions. This made the task quite difficult, and I was overall impressed with their compilation and synthesis of data from all possible sources, including the literature and previous theory, to pull together a well-integrated whole system model. **Having said that, my summation of the utility of this model is that it has value in predicting the direction and rate of surface sediment response in various zones of the model (i.e., Newark Bay, lower Hudson River, etc.) to a change in loading from a given source category or from a given source area. It is not clear of the uncertainty that will be encountered in trying to predict the absolute value of a chemical's response to a specific load change or the time it will take to achieve that value.**

HOC BIOACCUMULATION

Review of Sections 8 - 11 of CARP Report: Food Chain/Bioaccumulation J.E. Baker

Overall, the CARP Food Chain/Bioaccumulation model is based upon a well-established framework that has been successfully used to model PCB accumulation in the Hudson River and elsewhere. The formulations describing accumulation from water, diet, and sediment are consistent with the literature and represent the state-of-the-science for this type of aggregated modeling. The model is constructed to predict the average concentration of each target contaminant based on the properties of the contaminant (PCB homolog, etc.) in the 'average' biota (*e.g.*, an individual fish or worm that exhibits average behavior) throughout the study area. The model parameters are derived from literature values (physical properties of chemicals, bioenergetic rates) and field measurements (concentrations of target contaminants in biota and sediment). While little new conceptual ground was broken during the development of the CARP Food Chain/Bioaccumulation model, Hydroqual did an excellent job of applying their existing model, with minor modifications, to the task at hand.

Strengths of the model: The CARP model is the first comprehensive contaminant model that links hydrodynamics, sediment transport, organic carbon and dissolved oxygen dynamics, and contaminant partitioning and bioaccumulation in one integrated model. This linked approach allows for interesting integrated management questions to be addressed, such as 'how will altering nutrient conditions indirectly affect PCB or mercury bioaccumulation as moderated by changes to the carbon and sulfur cycles in the Harbor?'. The model rests on a sound fundamental footing, and although calibration issues remain (as described below), I have confidence in the overall ability of the model to predict concentrations of target contaminants in the NY/NJ Harbor food web.

Weaknesses of the model: The model calibration is best for PCB homologs, PAHs, and most pesticides (I did not review the cadmium and mercury sections; this is left for other MEG members to complete as part of the metals/mercury review). The model's ability to reproduce observed levels of dioxins and furans in the food web is much poorer. Whether this reflects problems with the field calibration data, errors in parameter estimation, or some missing process in the model will require further investigation. Since the model behaves well for the other organic contaminants, it is difficult to see how problems with the dioxin/furan calibration stem from errors in the model structure.

An inherent characteristic of the model is that it predicts time- and space-averaged concentrations of contaminants in the food web. Numerous studies in the NY/NJ Harbor and elsewhere demonstrate that contaminant levels vary considerably within a species at a site, likely due to differences in energetics, mobility, migration patterns, etc. While the model in most cases captures the average behavior, risk assessors may be more interested in the distribution of contaminant concentrations within a species. A complementary but more probabilistic model would be required.

Specific discussion points.

Section 8

1. What are the consequences of using modeled dissolved phase and sediment concentrations to calculate BAFs and BSAFs? While this is addressed later (Section 11), it would be helpful to have a direct comparison of modeled and field measured concentrations for each bioaccumulation zone.

Section 9

2. Why was the top 10 cm of sediment chosen? What are the consequences? What is the oxygen penetration depth in each zone, and how does this affect the vertical distribution of biota?
3. Lipid content of 5% for phytoplankton seems very high. What is the reference for this?
4. Minor typo of pg 26: type aerial = areal
5. migration of striped bass-how important is this in moderating the PCB levels? What would a 'worse case' scenario look like with adult striped bass remaining in contaminated zones?
6. How important is making the gill uptake rate inversely proportional to the dissolved oxygen concentration? Is this a necessary complication? It is nice to do here since DO is modeled, but is this necessarily a requirement for other models?
7. Chemical assimilation efficiency is not a function of Kow in the model. Comment on the implications on homolog profiles. Also, the chemical assimilation rates seem to be low relative to the literature.

Section 10.

8, BAF for zooplankton = 6xKow with no further increase at higher trophic levels

BAF in fish are lower than zooplankton for lower homologs–dietary exposure is less or that metabolism is occurring–what about kinetics? Not wild about metabolic explanations–need more literature review to support.

9. Dioxin/furan BAFs are way too low–look at chapter 11 for explanation–this is a big deal, may be a problem with the fish data OR indicate the model is over-predicting dissolved dioxin/furan concentrations.
10. Pg 36. Not clear you can argue that PAHs reach equilibrium with surrounding water while PCBs of comparable size/hydrophobicity do not.

Will defer the cadmium BAF review to others on the MEG.

11. Pg 38. How much of this discussion could be influenced by bias in the sediment concentrations–either being off spatially or by assuming 10 cm depth?
12. Pg 40. I agree that the 28 day bioaccumulation test (dredge material bioaccumulation protocols) under-predict BSAF due to slow accumulation rates. We recently found similar results in Baltimore Harbor.

13. Pg 42. Why do metals BSAFs from field sediment measurements agree with those from model but PCBs do not? Cannot argue that the field grab samples do not represent the biota zones for PCBs but do for metals.

14. What are the consequences of using the same BSAF and BAF across the study domain as compared with zone specific numbers?

Section 11.

15. Since field collected biota were used to determine the parameters (BAF and BSAF), it is a bit circular to compare the model results to the field data–not a cricitism, just a recognition of the limited biota data.

16. Pg 72 dioxin/furan in white perch and striped bass are overpredicted by 10 to 1000, even though the BSAF/BAF are unusually low. Does this suggest even a larger problem?

17. check van der Linde et al (2001) reference. CARP is using the upper end of estimated metabolism rate constants for dioxin/furans.

18. Reference for 6% of sediment organic carbon = black carbon?

19. Sensitivity analysis for dioxin/furan metabilism is not very convincing. Reduces overall modeled/measured concentration ration, but homolog differences remain.

Model Evaluation Group (MEG) Review of: "A Model for the Evaluation and Management of Contaminants of Concern in Water, Sediment, and Biota in the NY/NJ Harbor Estuary Sediment Transport/Organic Carbon Production Sub-Model" by HydroQual, Inc.

This is a very complex and ambitious modeling effort. For the most part the performance of the model is consistent with the inherent challenges of different modeling tasks, but there are also performance issues that reflect varying levels of experience of the Hydroqual team for different components. The organic carbon aspect of the model is very sophisticated, and it appears to be well calibrated based on comparisons with observations.

The performance of the sediment transport component is not as impressive, particularly insofar as it quantifies the trapping of sediment in different segments of the domain. In order to reduce the discrepancies between observed sediment trapping and the model results, various parameters were adjusted from one part of the domain to another. In some cases these adjustments were justified, based on consideration of physical processes not explicitly included in the model, such as wave resuspension. In other cases (such as adjustment of settling velocity and bottom stress), there was no physical justification other than to improve the model's agreement with data. Given the failure of the model to produce the observed sediment distributions with the a priori parameter specification, this fitting exercise was necessary and appropriate for the circumstances. However, it should be acknowledged in the report that the initial failure of the model represents a significant limitation to the predictive skill of the model, particularly with respect to the long-term distribution of sediment. This limitation may have a significant impact on the prediction of the long-term fate of contaminants.

The limitations of the model's ability to quantify sediment trapping is clearly illustrated in figure 4.7 (reproduced below). The red bars show how much sediment the model traps, whereas the blue bars indicate actual sediment trapping, based on dredging records. There are other sources of data that more-or-less support the dredging data. The distribution of sediment predicted by the model is uncorrelated with the observations. Notably, the model predicts about 5 times as much sediment trapping in the Hudson Highlands-to-Tappan Zee reach than in the Tappan Zee-to-Battery, whereas the observations indicate approximately the reverse ratio. This discrepancy occurs after the aforementioned adjustments of model parameters. Before these adjustments, the model trapped no sediment in the lower reach of the Hudson.

I am not sure what the model leaves out that would explain the discrepancy. There are a number of possibilities, including 1) inadequate lateral resolution to capture lateral trapping processes; 2) inadequate parameterization of temporal changes in sediment shear strength; 3) incomplete or inaccurate specification of spatial variability of erodibility; 4) inadequate resolution of frontal trapping processes.

The implications of this model limitation are as follows: 1) inaccurate quantification of the long-term distribution of contaminants; 2) underestimates of the exposure of contaminants to organisms in the lower Hudson; 3) overestimates of exposure estimates in the middle reaches of the river.

The model does appear to resolve certain aspects of the dynamics, such as the initial trapping of sediment during the freshet. Figure 4.5b shows the accumulation of sediment in the lower reach of the Hudson during the spring freshet, similar to observations by Woodruff et al., 2001. The problem is that during the subsequent period of low river flow, all of the sediment that was previously trapped is transported northward, out of the observed sediment trapping zone (fig. 4.5c).

The model does better in the prediction of water column suspended sediment than in trapping. The comparison of data from Poughkeepsie with model results indicates good agreement (fig. 4.2). In addition, the model reproduces the general trend of higher suspended sediment concentrations in the estuarine part of the river than upstream (e.g., figure from p 334 in appendix 4. These patterns vary a lot from season to season, sometimes showing a greatly attenuated turbidity pattern, inconsistent with observations (e.g. p 345 in appendix 4).

In spite of the limited ability of the model to reproduce observations, it does include the key processes responsible for the resuspension and transport of sediment. Thus I believe that it captures important aspects of the sediment-water exchange processes and their influence on the fate of contaminants.

ST-SWEM is an ambitious attempt to couple sediment transport calculations with a 3D hydrodynamic model of the NY/NJ Harbor estuary. This is a challenging problem – one that various groups, including the USGS, is investing considerable time and effort in. The best developed of these models at present may be the commercially available ones, like Delft 3D.

The fundamental quantities that must be characterized in a sediment transport model such as ST-SWEM are the resuspension (or erosion) rate, deposition rate and flux rate. Resuspension rate depends on sediment properties (grain size, density, cohesion, consolidation), which control the critical shear stress and erosion rate coefficient (a_0 in Eq. 8), and the flow-induced shear stress.

Deposition rate depends on sediment concentration near the bed and settling velocity, both of which can be affected by whether the sediment in suspension is aggregated (into flocs) or disaggregated. Sediment flux depends on the distribution of sediment in the water column and the velocity.

ST-SWEM relates deposition rate to sediment concentration C, layer depth h, a settling rate coefficient B and a correction for effects of salinity on aggregation rate. It isn't clear whether h is the flow depth or the depth of a thinner layer of the flow; presumably C is the average concentration over the depth h, but this also isn't stated. The settling rate coefficient B is related to the settling (fall) velocity of the sediment. A single value of B (Table 4-1) suggests that changes in deposition rate due to variations in grain size are unimportant in this system compared to changes associated with particle aggregation. The report doesn't discuss sediment size or its variability in the study area, so it is difficult to evaluate whether this assumption is reasonable. The salinity correction illustrated in Fig. 2-1, top panel, indicates a critical salinity of 20 ppt for aggregation. Most work I am aware of indicates a much lower value for the critical salinity – more like 5 ppt or less. There may be some other dynamics that become important at 20 ppt that this function captures. However, it might be worth investigating the effect of including another break at a salinity of 3-5 ppt.

ST-SWEM uses a standard erosion rate formulation to estimate resuspension rate, which depends on the bottom shear stress due to the flow, the critical shear stress, and a resuspension coefficient. In ST-SWEM, the resuspension coefficient is an adjustable parameter. This is reasonable, as it is difficult to predict in general and is often adjusted to provide the best fit between measured and calculated of suspended sediment concentration. The bottom shear stress is based on a near-bed log velocity profile, however the value of the von Karmen constant, κ, in the model is 25 times smaller than the canonical value of 0.4 (Table 4-1). No explanation is provided for setting κ to this value. A bottom roughness coefficient, z_0, of 0.001 (cm, I assume) is a reasonable for estimating skin friction stress, but neglects any spatial variability in sediment properties or bed morphology that might affect bed stresses. Are the results sensitive to this value? Critical shear stress is set to 0.5 dy/cm^2 in the upper 0-2 mm of the bed and 3.0 dy/cm^2 below that to reflect a fluff layer overlying a more consolidated bed. This suggests that net erosion on the time scale of the model calculations never exceeds 2 mm, while measurements by Geyer,Traykovski, and Sommerfield in the lower Hudson River show that a thick layer (many cm's) of the bed can be eroded and redeposited on tidal time scales during spring tides.

Nothing is said in the report about how suspended sediment is distributed vertically in the water column in ST-SWEM. The vertical distribution is important because sediment flux (advection) is the product of velocity and concentration. A balance between upward turbulent diffusion and downward particle settling produces a concentration profile with a maximum at the bottom and a rapid dropoff with increasing height above the bottom. Processes such as aggregation (increasing settling rates) and stratification (inhibiting turbulent mixing) tend to further increase the gradient in suspended sediment concentration. This means that concentration is high where velocity is low and vice versa. As a result, the total flux (integrated over depth), depends strongly on the vertical distribution of sediment in the water column.

I would have appreciated reading a little more about running the model. What time steps were used? Which of the 5 adjustable calibration coefficients in the sediment transport part of ST-SWEM (listed on p. 4-4) are the sediment transport calculations (TSS and deposition) most sensitive to? How much variation was there in model coefficients yielding best fits for the 4 segments of the system considered in the calibration. Bottom roughness was also treated as an adjustable parameter. Is this the only one that was allowed to vary in space? How much was it allowed to vary? No seasonal variation in calibration coefficients was included in the model. Is this reasonable? The calibrated model seems to do a reasonable job of capturing spatial and temporal variations in TSS – which is important for characterizing the organic carbon exchange processes – but appears to have more difficulty in reproducing the observed depositional pattern.

Overall, the results suggest that the model can capture some of the important variations in sediment loads, organic carbon, and patterns of sediment accumulation. However, in any model with multiple adjustable parameters, it is difficult to know whether the model is reproducing these variations for the right reasons. This is important if the model is to be used to forecast how the system might respond to changes in input or forcing parameters outside of the range of the calibration. In particular, the lack of any spatial variation in sediment properties in the model is not likely to be realistic. Identifying the parameters the model is most sensitive to could point future field efforts to mapping these parameters to improve their characterization in the model.

should be investigated, since the model's bottom stress calculation would magnify this discrepancy and thereby underestimate the potential for tidal resuspension in this region.

Salinity comparisons are useful indicators of the ability of a model to simulate constituent transport (a main issue here). Water temperatures calibrations are also useful in as much as temperature influences various modeled biogeochemical processes. Model-data comparisons for salinity and temperature are presented at two levels of temporal resolution: continuous and discrete. Comparisons to continuous observed data are provided for the Newark Bay/Kill complex using available CARP data. Comparisons to discrete observations (collected weekly, bi-weekly, monthly, seasonal, etc.) are provided at various stations outside the Newark Bay/kills complex using data provided by various sources. Note that the previous published study (Blumberg et al. 1999) provided comparisons to discrete salinity and temperature observations.

Overall, the model appears to track the observed sub-tidal salinity trends and seasonal stratification cycles in the Hudson Estuary and Upper/Lower Bay. The exception is the summer of 2002 -- when the model appears to underestimate observed salinities at most stations. Also, the model tends to smooth out some of the observed tidal oscillations (translations) in bottom salinity, especially at Newark Bay station 1 (upper Newark Bay). This result may be related to the coarse computational grid applied in this area (Figure 3-1) and uncertainty in model bathymetry. The Report does not provide an assessment of the model's ability to track tidal (semi-diurnal) salinity variations outside of the Newark Bay/Kills complex (e.g., at NOAA's Sandy Hook station).

At many stations (e.g., Hudson Estuary, Upper and Lower Bay), the model appears to simulate fortnightly variations in the vertical salinity structure, with neap-to-spring tide transitions from stratified to less-stratified conditions. This is an important feature that modulates the estuarine gravitational circulation. Unfortunately, the selected CARP data does not resolve this variability for model-data comparisons.

The Report's discussion of correlation analyses for salinity (figure 3-22a) and temperature lack some important details regarding method assumptions and limitations. Clearly, there is a mismatch in the time scales of observed (discrete) and simulated ("continuous" or hourly) salinity and temperature data outside the Newark Bay/Kills complex. Typically, this results in comparisons to only a limited number of points within each graph (e.g., Figs. 3-12 through 3-19). This reduces the significance of any statistical assessment of model error, although there are many such graphs. The fact that the discrete data are aliased with regard to tidal fluctuations further confounds the analysis. The Report does not state whether discrete observations are being correlated with instantaneous model data, low-pass-filtered model data, or both. If correlations are made only to filtered model data, then this should be stated (along with its inherent limitation of tracking only sub-tidal and seasonal trends and excluding short-term tidal variations). If this includes correlations with instantaneous model data, then small timing errors can result in relatively large model-data discrepancies (i.e., lower correlations) which would make the comparisons appear worse than they actually are. For these reasons, simple correlation analyses (alone) may not be the best technique to assess model skill. It is curious that other common metrics (e.g., RMS error, relative error, etc.) were not provided in the Report.

Also, the true regression lines are not shown (Figure 3-22a), only the zero-intercept, unit-slope line. The latter suggests that the model often underestimates the observed salinity data at several locations (Newark Bay/Kills, Raritan Bay, Inner Harbor). No correlation analysis is provided for Hudson River salinities (an obvious choice), though it probably would be favorable.

The model accurately tracks the observed subtidal and seasonal trends in water temperature at most stations. The model appears to follow the observed (continuous) trends in bottom temperature at the Newark Bay/Kills stations during 2001. Again, there is some evidence of tidal smoothing, but temperature trends are generally well reproduced in this region. The Report identifies some model discrepancies in the Hackensack Meadowlands region, but such inaccuracies are understandable given the complexities of this marsh system. Also, at estuarine monitoring stations outside the Newark Bay/Kills complex, the comparisons to discrete observations (collected weekly, bi-weekly, monthly, seasonal, etc.) demonstrate that the model reproduces low sub-tidal and seasonal trends.

Overall, the Report demonstrates that the CARP hydrodynamic model is a useful tool for simulating the hydrodynamics of the NY/NJ Harbor Estuary during the CARP years. The model is capable of passing accurate water surface elevations to other sub-models. Also, the model passes representative water temperatures for computation of temperature-dependent rate constants and kinetics. Bottom currents (and associated bottom shear) are generally well-reproduced in the Lower Newark Bay/Kills, where the model should be capable of simulating resuspension of particulate matter. However, the model tends to underestimated current speeds, and smooth tidal translations, in Upper Newark Bay – an issue that should be investigated further. Finally, the model appears to provide an adequate simulation of salt transport, except during the summer of 2002. Thus, the model should be capable of passing representative volume transports and diffusivities to other CARP sub-models.

In the future, the model's documentation may be enhanced by providing alternate quantitative measures of model skill (e.g., RMS error for all variables). Also, additional model-data comparisons may be performed at Hudson River and Sandy Hook monitoring sites where continuous temperature, salinity and/or velocity data are available. Such analyses would provide a more rigorous test of the model's fidelity, including the ability to simulate documented patterns of fortnightly variability in stratification/de-stratification (Warner et al. 2005).

Finally, the scientific merit of the Report would benefit from a discussion of the model results in light of the salient estuarine hydrodynamic features (e.g. tidal monthly variability, tidal straining, estuarine circulation patterns, meteorological forcing, etc.) and related transport processes. Unlike the previous paper (Blumberg et al. 1999), the present Report does not discuss model results in light of relevant physical transport processes. This omission limits the usefulness of the report in documenting the scientific tool developed in this study. However, this issue may be addressed in future publications.

References

Blumberg, A.F., L.A. Khan, and J.P. St. John. 1999. "Three dimensional hydrodynamic model of New York Harbor region." ASCE *Journal of Hydraulic Engineering*, 125: 799-816.

Blumberg, A.F., and D.W. Pritchard. 1997. "Estimates of the transport through the East River." *Journal of Geophysical Research*, 102: 5685-5703.

Blumberg, A.F., and G.L. Mellor. 1987. "A description of a three-dimensional coastal ocean circulation model." In: Coastal and estuarine sciences, Vol. 4, N. Heaps, ed. American Geophysical Union, Washington, D.C., 1-16.

Blumberg, A.F., and L.H. Kantha. 1985. "Open boundary condition for circulation models." ASCE *Journal of Hydraulic Engineering*, 111: 237-255.

Warner, J.C., W. R. Geyer, and J.A. Lerczak. 2005. "Numerical modeling of an estuary: a comprehensive skill assessment." *Journal of Geophysical Research*, 110, C05001.

F:\water\hudson\CARP2\review.doc