MODEL CALIBRATION:
MODELING STUDY OF PCB CONTAMINATION IN THE HOUSATONIC RIVER

PEER REVIEW

Model Calibration
Final Written Comments

Douglas Endicott
Great Lakes Environmental Center
June 3, 2005

RESPONSE TO CHARGE FOR THE HYDRODYNAMIC MODELING PEER REVIEW

I. General Overview of Response

As a reviewer, it is interesting to see how much this project has progressed in the 4 years since the review of the Framework document. The Housatonic River is a challenging system to understand and model. The PCB transport and fate processes in this river are unique in a number of ways, some of which I do not fully understand. As I mentioned in my preliminary comments, I am very impressed by the modeling work that is presented in the calibration report. I think that the project team has done a thorough job in assembling a suite of models that address the major processes affecting PCBs in the Housatonic River, and along the way have overcome a number of obstacles presented by site-specific data that challenge conventional wisdom. I remain optimistic that the modeling tools under development here will be valuable in terms of forecasting the outcome of remediation alternatives. I am also hopeful that concerns raised during the Calibration peer review will be considered carefully be EPA, and used to guide refinement of the modeling tools.

The modeling framework which we reviewed 4 years ago has changed considerably since then. The AQUATOX model biological/food chain component has been abandoned, and the QEA FOODCHAIN model has taken its place to predict PCB bioaccumulation. EFDC is now the sole framework for PCB transport and fate, although a number of potentially significant processes appear to be missing from this model. The role of HSPF in simulating the flow and solids boundary conditions to the PSA has been scaled back considerably. And perhaps most significantly, the modeling team has implemented EFDC using a 1-dimensional segmentation scheme in the water column for most of the PSA.

Calibration datasets have been generated for stage and flow, solids, organic carbon and PCBs in water, sediment, and biota compartments. An enormous number of sediment samples have been collected and analyzed to define initial conditions and characterize the variability of PCBs, organic carbon, grain size, porosity, etc. in the sediment bed. Unfortunately, much of the apparent variability in this data remains unexplained. Hints are provided in the RFI Report and Appendices that major components of this variability may be attributable to either measurement errors (including interlaboratory error) or the judgemental/focused bias applied in much of the sediment sampling. Variability arising from these factors was estimated, and could be used to better evaluate the sediment data. The observation of extreme spatial variability in sediment PCB concentrations, yet no apparent temporal trends is perplexing. Manipulations of the sediment data are in some cases (e.g., organic carbon content of sediment particle size fractions and TOC normalization of PCB concentrations) so complex and torturous that their descriptions are unintelligible. Although observations of erosion and deposition were made at a series of transects, this important information has not been directly utilized in the model calibration. Likewise, bank erosion was monitored but this data was not used in model calibration.

A less extensive sampling and measurement program was carried out for the water column, primarily based on monthly and event sampling at 5 fixed locations. The resulting data provide a good sense of the limited spatial trends in water column PCB concentrations through the PSA at low and moderate flow rates. These data were used primarily to calibrate a diffusive flux from sediment pore water. Because of the limited duration of monitoring in the water column, relatively few large flood events were sampled. In this and some other important ways (e.g., the boundary conditions and water column PCB partitioning) water column monitoring was limited to the point that the calibration suffered. In at least one case (the suspended solids composition at the upstream boundary), the modeling team creatively overcame this limitation.

To support calibration of the food chain model, biota were sampled and analyzed for congener-specific PCBs. Target fish species were sampled in 4 of 5 reaches in the PSA. Fish in each reach were treated as discrete populations, based on habitat and life history assessments. The data provide good measures of the variability in PCB concentrations between reaches and age classes of fish. The absence of data for the 5th reach (5D) is unfortunate, because the highest PCB concentrations were predicted by the bioaccumulation model for fish in this reach. The lower food chain (trophic levels 1 and 2) were underrepresented in the biota sampling effort.

Additional sampling and experimental activities were conducted to support various aspects of the models (SEDFLUME, bed load, pore and surface water partitioning). However, a number of other fairly standard water quality measurements were not conducted, including point and nonpoint source PCB monitoring. Supplemental studies to constrain several ambiguous parameters (e.g., pore water PCB diffusion flux and vertical extent of sediment mixing) appear to be necessary to support the calibration.

An important aspect of model calibration is the reduction of data to achieve consistency with the model inputs, parameters, state variables, etc. This involves spatial and temporal averaging, normalization, and sometimes more involved transformations. There was much data to reduce in this project, and for the most part the modeling team did a masterful job. In several instances, however, the data were not appropriately reduced. Two significant examples, which were problematic for the peer reviewers, were the PCB concentrations in the sediment (used to initialize the sediment bed) and in fish (used to confirm the bioaccumulation simulation). In both cases, the individual data were inappropriately compared to aggregated/averaged quantities in the models. Much confusion resulted regarding what the models were intended to predict and whether the residuals were indicative of model bias.

In general, calibration of the models was thorough, if limited by the relatively short time scale of the observations, the previously-mentioned lack of comparison to erosion and deposition measurements, and the omission of a number of processes (e.g., bank erosion) that appear to be significant to the mass balance. The calibrated suspended solids, water column total PCB, and fish PCB concentration appear to be reasonably accurate and unbiased. Many specific comments regarding the calibration are offered below. The major problem with the calibration as reported is that these are short time scale simulations, which are not sensitive to important features of the sediment and contaminant transport models. Overall, my sense is that the calibration report is an interim deliverable meant to satisfy the timetable of the Consent Decree. Many aspects of the calibration will necessarily be revisited, once a longer-term simulation is constructed and tested.

The other efforts documented in the Calibration report (sensitivity analysis and uncertainty analysis) seem redundant at this point, given that much of the model calibration is unfinished. The sensitivity analyses of the models are thorough and informative, however. Only the food chain model has been subjected to rigorous uncertainty analysis. I sense it may be the only one of the models for which this can be practically accomplished.

II. Response to Peer Review Questions

In considering the foregoing general issues and evaluating the EPA documents, the Peer Review Panel shall give specific consideration to the following questions. As modeling activities proceed, additional specific questions may be identified the panel to address.

A. ***Model Calibration***

1. Are the comparisons of the model predictions with empirical data sufficient to evaluate the capability of the model on the relevant spatial and temporal scales?

I can't answer this question with a "yes" or "no" for the calibration of the EFDC model, because I do not believe it is being applied on spatial and temporal scales that are relevant to the PCB contamination problem in the PSA and its remedy. I will address the issues of relevant spatial and temporal scales in the following paragraphs. In the case of the food chain model, it appears that comparisons made between predictions and observations are adequate, except in reach 5D where no fish were sampled.

I have a fundamental objection with the application of EFDC as a predominantly 1-dimensional model to simulate hydrodynamics, sediment transport and PCB transport in the PSA. Accurate simulations of velocities, shear stresses, erosion and deposition patterns, and streambank undercutting/erosion are only possible if significant lateral variations are resolved in the model. At least 3 lateral segments should be used in the main river channel, and should consider river features such as bathymetric profiles. This lateral segmentation should also be used in the sediment bed, with initial sediment conditions recalculated from data on the basis of this segmentation. Fortunately, EFDC is a 3-dimensional model so it should be able to accommodate this additional resolution.

The temporal scale of EFDC calibration is not particularly relevant to the PCB contamination problem in the PSA and its remedy. It should be noted that "calibration" in the context of this project has come to mean calibration of short-term (daily to seasonal) changes in model state variables. "Validation" now includes the calibration of long-term (annual to decadal) changes, which are the interesting changes in terms of managing toxic chemicals and making decisions about remedial alternatives. In other words, at this juncture we are unable to evaluate the model's capabilities in terms of its intended application. In order to fully evaluate the model's capabilities, it must be applied to a significantly longer simulation period (i.e., 10 years) and compared to data over this longer duration. This limitation also has an impact on how thoroughly we can address questions 4, 5 and 6 below.

Regarding the calibration of sediment transport in EFDC, there is too much emphasis on matching suspended solids concentrations, and not enough on scour and deposition in terms of changing sediment bed elevation. The calibration of suspended solids only fixes the magnitude of net settling/resuspension in the model, while the individual fluxes may or may not be correct. Confirming the change in sediment bed elevation can be much more revealing. In that regard, I have some trouble reconciling the 3 to 4 cm maximum change in sediment bed elevation predicted by the sediment transport model for the 14-month calibration period, with the several-foot change based on bathymetric transects and stratigraphic analysis. I am afraid the calibration results fail to demonstrate that the approach taken to calibrate settling and resuspension fluxes works. How do we know that the highly-nonlinear parameters describing things like cohesive sediment erosion, flocculation, and deposition can be determined by averaging data collected at a number of sites having different sediment properties? To a certain extent, the answer is obtained by running the model for a relatively long duration, and examining the results for anomalies in terms of the magnitude and pattern of sediment bed change.

In some other river systems, errors in deposition and erosion fluxes are revealed during calibration of water column PCB concentrations, because the concentration gradient between sediment and water amplifies the error. In the PSA this will not work so well, because there is only a small gradient between suspended and bedded particulate PCB concentrations. The lack of gradient makes it relatively more difficult to tell if the model is grossly in error.

The parameterized value of the mass transfer coefficient for pore water diffusion (K_f) is very high in comparison to most values I can find in the literature. K_f is being calibrated to reproduce the observed increase in water column PCB concentrations under low-flow conditions. In other words, all of the increase in water column PCB concentrations is being allocated to this mechanism. Whether the K_f calibration is correct depends upon this assumption. It would be most desirable to somehow independently confirm this value, either via measurement or by ruling out other potential PCB sources.

Rare flood events

The calibration report shows that EFDC is capable of predicting the extent of flooding. However, this is not an *impact* per se. The impact of concern is the remobilization of significant quantities of previously in-place pollutants. Sediment transport models like SEDZL and now EFDC are being used in a growing number of river systems to predict bed erosion under event conditions. However, the state of the art still requires extensive site-specific and model process-specific data. As far as I know, confirmation of model predictions under extreme events must still be demonstrated on a site-specific basis. Based on the calibration report, this confirmation is lacking in the Housatonic River. It is unclear whether the EFDC scour and deposition predictions are reasonable at very high flow rates, and the comparisons to data are problematic due to the spatial resolution of the model sediment bed. Although bathymetric data showing scour and deposition patterns were collected and reported at a number of transects in the PSA, this data has apparently not been used to confirm model predictions as it has in other river systems (e.g., Gailani et al., 1996.).

Discriminating between water-related and sediment-related sources of PCBs to fish and other biota

The transport/fate and food chain models address PCB bioaccumulation via both pelagic and benthic exposure routes. In principle, it would be simple to apply the models to discriminating between water-related and sediment-related sources of PCBs to fish and other biota. I have done this in other applications by running the model with two chemical state variables, one for the chemical initialized in the sediment bed and a second for the chemical originating from water column sources. Since the models are both linear with respect to chemical concentrations, the simulation can be decomposed in this manner to explicitly show the proportion of PCB body burden in different species and reaches contributed by PCB exposure originating in the sediment bed versus the water column. If this discrimination is important, the models should be rerun to make this diagnosis.

2. Is there evidence of bias on the model, as indicated by the distribution of residuals as a function of the independent variables?

The models appear to be fairly unbiased in terms of the principal state variables. The following exceptions were noted during review:

- The gradient in total PCB concentrations across Woods Pond indicates bias, due possibly to the magnitude of the diffusive flux from sediment. The net loss of PCBs from Woods Pond contradicts expectation and conceptual model;
- Dissolved PCB concentrations are consistently overpredicted in the storm event periods, suggesting either a weakness in the partitioning model or inadequacy in estimating low-end censored concentration data;
- There also appears to be an unexplained factor in the sediment pore water partitioning data, possibly some kind of solids effect. Cross-plots of particulate organic carbon versus apparent K_{oc} show that K_{oc} declines with increasing f_{oc} for both total PCB and congeners, regardless of whether data from some sediment cores are censored. Dissolved and particulate PCB predictions look OK in comparison to the partitioning study data, but this is a very limited number of measurements.
- Some mild bias is evident in predictions of PCB in fish. Predicted PCB concentrations in bullhead, sucker, sunfish and bass are generally lower than mean observations, while cyprinid PCB concentrations are overpredicted in all reaches except 5A. There is really too little data to check bias in lower food chain predictions.

I would not judge any of these biases to be so significant as to undermine the credibility of the models.

3. Does the model, as calibrated, based upon your technical judgment, adequately account for the relevant processes affecting PCB fate, transport, and bioaccumulation in the Housatonic River?

Much effort has been devoted to collecting a robust data set for model calibration. However, in any project of this complexity there are always opportunities for additional data collection to address important gaps in the dataset. Supplemental data can be critical in terms of strengthening the model calibration. The adequacy of the models could be strengthened in a number of ways, each involving the collection of additional data and other information:

Partitioning data:
Since the modeling team don't understand what is going on at 15% of the coring locations (many but not all are very-low organic carbon sediments), it would be appropriate to do some adsorption experiments using sediments that deviate from the equilibrium partitioning behavior.

Low-flow sediment-water flux:
Methods of measuring bioturbation activity and/or diffusive flux should be investigated and employed. Deploying benthic chambers is one option

Sediment mixing:
Experimental approaches that could be applied in the field should be investigated. At a minimum, the density of benthic invertebrates (including vertical distribution) could be measured in sediment core samples as a basis for evaluating mixing depth.

Potential PCB sources other than sediment bed and transport across confluence:
Neglecting the WWTP, tributaries, and groundwater as PCB sources in the PSA is not justified by the available data. The WWTP effluent has apparently not been monitored for PCBs, which I think may be an unfortunate mistake. Rationale for this offered by EPA (no gradient in water or sediment near outfall) is not compelling. We have observed PCB concentrations of 20-30 ng/L routinely in untreated sewage throughout New York and New Jersey. If, for example, the Pittsfield WWTP effluent were to contain PCBs at 20 ng/L, that would constitiute a source of 0.3 kg/yr to the river. That doesn't sound like a lot in comparison to what's flowing across the confluence currently, but what about after remediation upstream is completed?

Similar arguments apply for tributaries and groundwater. EPA defends no groundwater monitoring by stating that too many measurements would be required to meet modeling accuracy requirements. Is this a rationale for neglecting the process? I am not sure EPAs QAPP process recommends "do nothing" in this situation.

PCB volatilization:
Neglecting PCB volatilization in EFDC may not be justified. The accuracy of PCB flux estimates being used to justify the neglect of volatilization as a loss process in the PSA, depend upon good values of Henrys constant. The best experimental data I am aware of was published

by Holly Bamford (Bamford, Poster and Baker, J. Chem Eng. Data, 2000, 45, 1069-1074). She measured Hlc's over a range of temperatures for numerous congeners, and also generalized the results into predictions for all of the PCBs. She found that PCB Henrys constants depended more on the number of ortho-chlorines than on the homolog. If I substitute one of her Hlc values (a representative congener at 18 degrees C) for the value used in the RFI, and repeat the volatilization rate calculation, I get a rate and flux that is about twice as large. I suggest the modeling team evaluate Bamford's data and consider revising the PCB volatilization flux calculation accordingly.

Partitioning of PCBs to biotic organic carbon:
This was lost with the departure of AQUATOX. Instead, we get the assumption that suspended solids (with seasonally-invariant organic carbon content) has the same PCB sorbent capacity as phytoplankton, periphyton, macrophytes, etc. This is an assumption that should be stated in the PCB transport and fate model, and justified by data. I am not convinced that POC and phytoplankton/ periphyton carbon are interchangeable as sorbents for PCBs. I would like to see the evidence from the PSA that supports this assumption.

Monitoring of flow, solids and PCB concentrations at the boundary condition:
I am concerned that monitoring the flow, solids and PCB boundary condition above the confluence has not been emphasized enough. This results in unacceptable uncertainty in the upstream boundary condition. I think EPA and GE should consider adding more continuous instrumentation along with the pressure transducers, including ADCP and/or transmissometers. These would improve the flow measurements and allow continuous TSS monitoring, and could be used to make more robust estimates of the boundary conditions. Boundary conditions will become increasingly important as river remediation moves forward.

Streambank erosion:
It appears that a major component of the interaction between the floodplain and the river occurs via erosion of the streambank. Active undercutting of sediments deposited on the riverbank is evident in the upper half of the PSA. This mechanism should be included in the sediment and contaminant transport model during calibration, not after.

4. Based upon your technical judgment, have the adequate methodologies been employed to evaluate the sensitivity of the model to descriptions of the relevant processes, and to evaluate the uncertainties of model predictions?

As mentioned previously, I thought the sensitivity analyses were well done. A couple of important parameters were omitted from the sensitivity analysis of EFDC, the settling rate and composition of particle types at the upstream boundary. Significantly, the sensitivity of the EPDC model predictions to the grid resolution was not included in the report. The modeling team will need to revisit sensitivity at latter stages of the project, when longer-term simulations are run. At validation stage (for example), surficial sediment bed thickness should be explored in the sensitivity analysis.

Formal uncertainty analysis was only conducted on the food chain model. It may be the only one of the models for which uncertainty analysis can be practically accomplished. Any analysis of model uncertainty should address propagation of uncertainty between the models, and include uncertainty in statistical models (rating curves) predicting upstream boundary conditions, as well as uncertainty in flow measurements.

I do not favor "bounding" analysis (as suggested by GE) as a shortcut to uncertainty. It is too easily to subjectively manipulate such an approach to produce a desired outcome. Bounding analysis should be used only if absolutely necessary due to computational constraints.

5. Is the uncertainty indicated by model-data differences sufficiently inconsequential to permit use of the model to predict differences among remedial options?

This really is a judgment call that depends upon how ambitious the remedial options are. Uncertainty is "inconsequential" if it does not obscure the discrimination between outcomes (i.e., PCB body burdens in fish) of different scenarios. At this point, uncertainty of EFDC predictions have not been determined. The uncertainty of food chain model predictions of total PCB bioaccumulation appear to be acceptable to evaluate remedial options. For most fish, the ratio of predicted 90th-percentile to 10th-percentile PCB concentrations is a factor of 3 to 5. This matches my expectations from other systems, and indicates a high-quality model.

Of course, this is just uncertainty due to model parameters, which is only a part of real model uncertainty. I am told that the really fatal model uncertainties are the thinks you don't know about and cannot be anticipated. That is why (aside from curiosity) I would like to see EPA collect more data on this system, as identified above (response to question 3).

6. Are the processes in the model calibrated to the extent necessary for predicting future conditions including future concentrations of PCBs in the environment under natural processes and under potential remedial options for sediments and floodplains soils in the Housatonic River in the reach below the confluence? If not, what additional work needs to be done to calibrate the model?

As I mentioned above (see response to question 1), we can only anticipate the calibration of processes that are influential to long-term model predictions. In this context, I am primarily concerned with the calibration of resuspension fluxes, the sediment-water diffusion flux, and the surficial sediment residence time as defined by the mixed layer thickness. I also worry about unquantified PCB sources that will remain after remediation is completed.

There have been protracted arguments about whether the thickness of the surficial mixed sediment layer, or (alternatively) the bioavailable sediment layer, should be 6 inches as opposed to 3 or 4 inches. I think it is fair to say that each side prefers a number intended to produce a modeling outcome favorable to their own interest. EPA has not convincingly demonstrated that the current model parameterization of mixed layer depth is scientifically defensible. The truth is we have almost no system-specific data to guide the specification of this parameter, and there is little guidance currently available in the literature from other sites. All lines of evidence should be used to evaluate this parameter. That should include physical, chemical and biological data as well as what the model can tell us via calibration. The modeling team should consult with scientists familiar with the various biotic and abiotic benthic processes.

Finally, it should be recognized that simulations of long-term remediation alternatives will dramatically alter the sources and pathways followed by PCBs in the PSA. Instead of the dominant role played by advection of PCBs from the upstream boundary (the major flux pathway during the calibration period), PCB flux via diffusive and particulate fluxes from the sediment will eventually predominate. This will change the general sensitivity and uncertainty behavior of at least EFDC and may reveal errors that are not currently evident. I think some scenarios of this sort should be tested during model calibration, to ensure that the models behave in a way that is at least consistent with our overall understanding of the system. Models sometimes behave unusually when the major contaminant pathways are altered.

III. Specific Comments on the Model Calibration Report

I think a good technical writer could do a lot with this report in terms of improving readability and clarity.

PCB Fate and Transport Schematic:
I usually take the time to examine model schematics, and I find them particularly valuable in understanding how the conceptual model is applied. In this case (Figure B.4-1) there are a number of errors in the schematic which should be corrected. These errors should be obvious to the modeling team.

Bioaccumulation of coplanar PCBs:
As previously commented, I don't believe that the "correction factor" approach used in the food chain model to reduce the bioavailability of coplanar PCBs is scientifically correct. Based primarily on comparisons of the highest-quality measured PCB BSAFs for fish to model predictions assuming no metabolism (Burkhard et al., ES&T 2004), it appears that congener 77 and possibly congener 126 to a lesser extent are very slightly reduced through metabolism. Since the congener-specific metabolism would probably be aryl hydrocarbon receptor (AHR) mediated, metabolism in invertebrates is unlikely (i.e., it should only be taking place in the FISH). One cannot rule out a bioavailability effect associated with affinity of more planar PCBs with small amounts of black carbon in water and sediments. I think this should be addressed in greater depth in the Bioaccumulation section of the Calibration report.

PCB elimination by fish:
Biphasic (multicompartment or deep storage) elimination is sort of a hidden feature within the QEA foodchain model. By that, I mean the literature documenting the development and application of this model does not address this fairly important feature (as far as I am aware).

Again, this should be addressed in greater depth in the Bioaccumulation section of the Calibration report. In addition, computed elimination rates should be compared to rates measured in fish.