PEER REVIEW COMMENTS ON
THE MODEL VALIDATION REPORT

Introduction

I appreciate the opportunity to review the work of the modeling team for the Housatonic River. This has been an interesting learning experience for me. I would also like to compliment the modeling team for having conducted this project in a thoroughly professional manner. In addition, I appreciate their responsiveness in the short time frames allowed for the peer review process.

I do have several comments to make regarding the peer review process itself. In the *EPA Response to Questions from Model Validation Document Overview Meeting*, EPA correctly notes that models are uncertain and that a degree of uncertainty is inherent in “these types of studies”(question: What scientific study does not include a degree of uncertainty?). They go on to identify three principal sources of model uncertainty (models are simplifications/approximations of reality, parameters are uncertain, future conditions are unknowable). Although these can be significant sources of uncertainty, EPA chose not to highlight several others, which may be equally important:

- Inappropriate process representations/formulations,
- Errors resulting from discretization of the model or the scale of its application,
- Errors in programming, preprocessing, running, linking and postprocessing the models, and
- Untested model predictions.

These are “lurking” sources of model uncertainty, which are not typically revealed by uncertainty analysis. I think it is useful to acknowledge these pitfalls, and keep them in mind throughout the course of a project. I think much of the Peer Review deliberations, and more than a few of the comments, reflect concern about these factors.

It is unfortunate that the peer review process was conducted in more of a confrontational, as opposed to a collaborative manner. The scientific process is not well served by the former approach.
I think more would have been accomplished in the peer review sessions, had a freer dialog and more open exchange of information and ideas between the various parties been possible. Because of the constraints of the Consent Decree, this was not allowed. As a result, the relationship between the modeling team and the peer review panel was sometimes adversarial, and our recommendations rebutted inappropriately (from my view). Although this may be as the lawyers intended, it also resulted in some lingering issues that have not been appropriately resolved through the entire peer review process.

Questions for the Model Validation Report

1. Considering the changes implemented in the Phase 2 Calibration, does the model as calibrated and validated, based on your technical judgment, reasonably account for the relevant processes affecting PCB fate, transport, and bioaccumulation in the Housatonic River to a degree consistent with achieving the goal of the modeling study?

The modeling frameworks (HSPF, EFDC and FCN) applied in the Housatonic River are state-of-the-art. Within each model, all relevant processes are represented in a conventional manner. Therefore, I conclude that the models reasonably account for the relevant PCB transport, and bioaccumulation processes.

Of course, this question is somewhat ridiculous because a successful model must do more than just "account" for processes! Instead, it is necessary to reevaluate the modeling framework, its implementation, and its parameterization in light of our understanding of PCB fate, transport, and bioaccumulation processes as well as the behavior of the Housatonic River as an ecosystem. Two primary criticisms emerge from this reevaluation:

1. EFDC fails to address lateral variation in hydrodynamics and sediment transport processes, and
2. There appear to be parameterization errors in specific sediment transport and PCB fate and transport processes.

I should note here that, since the upstream and downstream models share the same frameworks, the same comments apply.

On the first point, EFDC appears to fail to address lateral variation in hydrodynamics and sediment transport processes, because the model grid in the river channel is one-dimensional over most of the Rest-of-River (ROR). Prior applications of SEDZL (the model upon which the EFDC sediment transport code was apparently based) to river systems have used a 2-dimensional, vertically-integrated grid, using 5 to 10 lateral grid cells to represent the channel cross section. The 2-dimensional sediment transport models have been demonstrated to predict scour and deposition in river channels with reasonable accuracy[1]. These applications have demonstrated that lateral variation in current velocities and bottom shear stresses result in considerable variation in sediment scour depending upon location in the channel. This behavior is often observed in natural channels as a pattern of erosion from mid-channel/deposition nearshore, or erosion on the outside of a bend/deposition on the inside. Because the relationships between current velocity, shear stress and sediment erosion are nonlinear, it appears likely that sediment scour

---

[1] See: Gailani, J.Z., W. Lick, K. Ziegler, and D. Endicott. 1996. Development and Calibration of a Fine-Grained Sediment Transport Model for the Buffalo River. *J. Great Lakes Res.,* 22(3):765-778.

predicted by the 1-dimensional model of sediment transport in the Housatonic River will be different than the scour predicted by a 2-dimensional model. How *much* different is unknown, because the EPA modeling team have not tested the sensitivity of the EFDC sediment transport model to the number of lateral grid cells in the river channel. This is a significant issue, because release of PCBs from eroding sediments is a major source of the contaminant in the ROR.

On the second point, there appear to be parameterization errors in specific sediment transport and PCB fate and transport processes. These include the parameterization of sediment scour and deposition, the thickness of vertical sediment segments and rate of bulk mixing between these segments, and the diffusive exchange between sediment pore water and the overlying water column. In addition, the appropriateness of the equilibrium partitioning assumption for PCBs on resuspended sediments has been repeatedly questioned. These issues are discussed below. Where there appears to be a tie-in between parameterization errors and lack-of-fit or bias in model predictions, this will be noted.

- The vertical representation of the sediment bed in the PCB fate and transport model: The thickness of the surficial sediment layers have been specified as either 4 cm (Reach 5A) or 7 cm (Reaches 5B thru 6). These layers are assumed to be well-mixed, via a combination of physical and biological processes. In addition, sediments from the surficial layers are mixed with deeper layers, down to a depth of 30 cm below the sediment-water interface. Analysis of benthic invertebrate data is offered to justify these assumptions. However, these data show that below Reach 5A the benthos are predominantly filter feeders or surficial sediment feeders/dwellers. Such organisms are unlikely to cause sediment mixing below the top couple of centimeters. 7 cm of surface sediment mixing is too deep, and mixing of sediment between the surface layer and deeper layers is inconsistent with the life history data for the resident benthos in these reaches. The consequence of too much sediment mixing (and too-thick surficial layers) is that the PCB fate and transport model will predict an unreasonably slow rate of decline in surficial sediment PCB concentrations, which is observed in both Phase 2 calibration and validation periods.
- Sediment transport parameterization: Cohesive sediment resuspension is parameterized based on the analysis of data from SEDFLUME experiments conducted on a number of sediment cores. Results of this analysis produced shear stress exponents of 1.59 for sediments in Reaches 5A and 5B and 0.95 in Reaches 5C and 6[2]. These values are considerably lower than shear stress exponents reported in the literature. For example, Lick, Ziegler and coworkers have reported that for cohesive sediments tested at a significant number of sites, the shear stress exponent is generally constrained within a fairly narrow range (n=2.6±0.3). Based upon this, it appears that parameterization of

---

[2] Model Calibration Report, Volume 3 - Appendix B Hydrodynamic and Sediment/PCB Fate and Transport Model Calibration, Attachment B.5 Analysis of Sediment Erosion Data

cohesive sediment resuspension in the model is erroneous, and should be corrected consistent with the guidance offered by Dr. Lick. As a consequence, it will also be necessary to recalibrate deposition rates.

- The diffusive flux (i.e. sediment-water exchange) of PCBs from sediment pore water to the overlying water column was parameterized as a constant mass transfer coefficient during Phase 1 calibration. Phase 2 calibration has altered the spatial profiles of PCB water column concentrations at low flow conditions, suggesting that sediment-water exchange should be recalibrated. In all likelihood, the mass transfer coefficient should be increased, and made a reach-specific parameter via a relationship between the mass transfer coefficient and the benthos density.
- EFDC uses the same equilibrium partitioning (EP) model as every other fate and transport model in use today. EP generally works quite well, and simplifies the model computation by requiring only a single state variable for each chemical. It has been suggested that desorption kinetics should be incorporated into the Housatonic River model, because PCBs desorbing from resuspended sediment might not reach equilibrium within the timeframe they remain in suspension. This disequilibrium may help explain the pattern of PCBs moving in "ribbons" of sediment downstream, and the high surficial PCB concentrations in Woods Pond sediment. I think this is mostly speculation and, although I am curious to see if it is true, it is well and beyond the objectives and goals of this modeling study.
- I have already commented on the misattribution of the assumption made regarding PCB partitioning to non-filterable organic carbon in the water column ($K_{doc} = K_{poc}/100$), and EPA has responded (less than satisfactorily) to this issue. Regarding the parameterization of the 3-phase equilibrium partitioning model for PCBs, a useful citation (i.e., the Burkhard references) is: USEPA, 2003. *Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health (2000), Technical Support Document Volume 2: Development of National Bioaccumulation Factors.* United States Environmental Protection Agency, Office of Water, Office of Science and Technology. EPA-822-R-03-03 (www.epa.gov/waterscience/criteria/humanhealth/method/tsdvol2.pdf)

This question also asks us to consider changes to the models resulting from additional (Phase 2) calibration. This specifically applies to EFDC calibration, since little or no changes were made to either the HSPF or FCN models. Previously, the Peer Review panel commented that the 14-month calibration was too short to reflect many significant long-term model processes, specifically the evolution of PCB concentrations in the sediment bed. So the EPA modeling team went back and extended the calibration to cover a much longer, 10-year duration. I expected to see some changes in the model parameters related to the residence time of PCBs in the surficial sediment layer. Instead, as a result of Phase 2 calibration we see some dramatic changes in the simulation of PCB transport and fate in the Housatonic River: There are now 3 times more PCBs entering the Primary Study Area (PSA) at the upstream boundary, although the difference (compared to the Phase 1 calibration & flux analysis) diminishes further downstream (2 times more PCBs at New Lenox Road, 1.5 times more at Woods Pond). Export of in-place PCBs from the sediments of Reach 5A is no longer the predominant contaminant source to the Rest-of-River. To me, this is a breathtaking change in the PCB fate and transport simulation, which rewrites the entire assessment going back all the way to the conceptual model presented in the MFD. I was disturbed to see such a major change taking place so late in the model development process, and was also somewhat surprised at the nonchalance with which this revision was presented and received.

Does the hydrograph or the longer duration of the Phase 2 calibration explain these differences? To some degree, maybe. In 10-year period, there are a number of high-flow years (e.g., ‘90, ’96,’98 and ‘00) as well as low-flow years (’91, ’92 and ‘99). The hydrograph impacts the balance of sediment erosion and deposition averaged over different durations. The MVR does not discuss this to any significant extent. However, the change in the PCB fluxes mostly reflects a new treatment of the boundary condition, apparently provoked by some new boundary condition data. Comments regarding the boundary condition used in the Phase 2 calibration are provided in my response to question 3. At this point, EPA should consider whether further refinements and improvements to the model will occur in the future, as more data are collected and the models are run. In other words, are the models only as good as the latest data point? How can the model calibration be considered “final” if, as the Phase 2 recalibration demonstrates, a few new data can so readily change the PCB flux analysis and possibly other model results?

Another change that has been undertaken by EPA is to extend the spatial domain of the models to cover the upstream reach of the East Branch through the area that has been remediated, as well as downstream from Woods Pond dam to Rising Pond dam. It is encouraging to see this development, because the modeling domain confined to the ROR was clearly too small to fully address the extent of PCB contamination in the Housatonic River. However, I am not convinced that either the upstream or downstream models are adequately calibrated and validated at this time. It is obvious that much less effort has gone into these models, and the presentations made by the EPA modeling team at the Document Review Meeting looked much more like progress reports than finished products.

2. Are the comparisons of the model predictions with data sufficient to evaluate the capability of the model on the spatial and temporal scales of the final calibration and validation?

Graphical and statistical comparisons between model predictions and data are the most objective measures of model performance. In general, sufficient comparisons have been presented in the MVR and supplemental material with which to evaluate the capabilities of each of the models. Specific comments are offered below, regarding the strengths and weaknesses of the various prediction-data comparisons, as well as identifying several comparisons that have not been made, but should be. In addition, errors in model predictions are revealed by some of the comparisons, and these are pointed out so that they can be corrected by the modeling team.

EFDC predictions of flow, TSS and water column total PCB concentrations were compared to data using both longitudinal profile and time series plots. Although the spatial profiles for TSS predictions appear reasonable for several selected days (Figures 4.2-34 thru 4.2-37), the time series plots indicate significant and persistent overprediction of low TSS measurements. This is particularly evident at Holmes Road (Figure 4.2-38), but also at New Lenox Road (4.2-39) and to a lesser degree at Woods Pond headwater (4.2-40). In the MVR (page 4-59), EPA offers that "Simulated TSS concentrations pass through approximately one-third of the data…", which is surprisingly poor model performance. Similar overprediction of low TSS measurements is apparent for the validation period as well. According to comments by both EPA and GE, the overprediction of low TSS measurements is due to significant bias in the specification of the East Branch boundary condition on days when TSS concentrations were not measured. Time series plots for specific flow events (Figures 4.2-42 thru 4.2-48) indicate a better fit of TSS predictions under high flow conditions.

Sedimentation rates predicted by EFDC for individual grid cells in the river channel over the 10-year Phase 2 calibration period are plotted in Figure 4.2-58. This figure shows a highly variable pattern of deposition and erosion, especially in Reach 5A, where deposition fluxes exceeding 5 cm/yr are predicted for cells in close proximity to others in which erosion fluxes of 1-2 cm/yr are predicted. Predicted sedimentation rates are not tested by any data at this scale, however. On a reach basis, relatively small deposition fluxes (0.2-0.5 cm/yr) are predicted throughout the ROR for both the calibration and validation periods. Figure 4.2-60 graphs the sedimentation rate in grid cells within Woods Pond, along with sedimentation rates measured in cores. There are many EFDC grid cells in Woods Pond, and a significant number of dated sediment cores. I specifically examined the 1995 sediment core data suggested to be most representative in terms of cesium-137 profiles, and found that both measured and predicted sedimentation rates are in the range of 1 to 10 mm/y. On a pond-wide average basis, sedimentation rates predicted for the calibration period agree fairly well with the sediment core data (Table 4.2-5). Over the 26-year validation period, the predicted sedimentation rates appear to be somewhat low in comparison to data (Table 6.2-7).

Predicted sedimentation rates were also compared to measurements of bathymetric change made between 2000 and 2006 at a series of transects in Reach 5A (Figure 6.2-33). The measurements of bed elevation change (positive and negative) were consistently larger and more variable than the corresponding model prediction although, since the predictions and data were both spatially averaged, I have trouble interpretting the significance of this comparison. As with other EFDC results for the sediment bed, grid cell predictions were aggregated to mile or sub-mile intervals. At finer spatial scales, there appear to be problems with the fidelity of the model. For example, the transect data from location XS153 show several feet of deposition while EFDC predicts significant erosion at this location, both during calibration (erosion=1") and validation (4") periods. The transect data themselves (presented at the Document Overview meeting: *09 - Erosion& Deposition - Garland.pdf*) show relative nonuniformity in sedimentation both longitudinally (between transects) and laterally (within transects). To me, the data show that lateral resolution is warranted for hydrodynamics (shear stresses) and sediment transport. Other subgrid features (lateral variations in bathymetry, bends, trees in river) are also significant physical features, which are not represented in the model.

Model predictions of total PCB concentrations in the Housatonic River water column were compared to data using time series plots and longitudinal profiles. The time series plots (Figures 4.2-65 thru 4.2-68 and 6.2-41 thru 6.2-44) indicate that, for PCB concentrations, there are significant gaps in data collection over the Phase 2 calibration and validation periods. As was the case for TSS, the model tends to overpredict low PCB concentrations, and the overprediction of these two state variables appear to coincide. This is most apparent at the Holmes Road station. Again, this is primarily due to significant bias in the specification of the East Branch boundary condition on days when PCB concentrations were not measured. Inspection of the timeseries of East River PCB boundary conditions (Figure 4.2-64) reveals that the only time East River PCB boundary concentrations are lower than 60 ug/L is when data have been measured below this value, yet probably more than half of the data fall below this concentration. There are sufficient PCB data to evaluate the storm event predictions on only one occasion (May 19-21, 1999 at New Lenox Road; Figure 4-2.76); the model prediction timeseries agrees well with this data. The spatial profiles for PCB water column predictions appear to be quite variable. The low-flow simulations have been criticized by GE as underpredicting the increase in PCB concentrations observed between the upstream boundary condition and New Lenox Road, as well as the decrease in PCB concentrations across Woods Pond. In addition, the profiles show a consistent "drop" in PCB concentrations immediately below the upper boundary. These features have been interpreted as suggesting that processes related to PCB fluxes between the water and sediment may be misrepresented or miscalibrated. In the *EPA Response to Questions from Model Validation Document Overview Meeting*, a statistical test (overlapping 95% confidence limits) is offered to show that the apparent decrease in PCB concentrations across Woods Pond may reflect sampling variability. It may be useful to use a statistical test to evaluate the gradient in PCB concentration data between two river locations; however, wouldn't it be better (more powerful) to test whether the *difference* in PCB concentrations between locations is significantly different than zero?

Timeseries plots comparing the predictions of PCB concentrations in the top 6" of sediment in mile reaches to data are shown in Figure 4.2-77 for the Phase 2 calibration period, and Figures 6.2-49 for the validation period. The predictions appear to exceed the majority of data at the end of both the calibration and validation periods, in some or most mile reaches. Although data prior to 1999 are not abundant, the model also appears to underpredict the rate of PCB sediment decline in many mile reaches, including Woods Pond (Figure 6.2-50) as suggested by the overprediction of the majority of 1999 data. The failure of the model to reproduce the rate of decline in sediment PCB concentrations is a major problem, since in the long-term, PCB concentrations in biota will respond to this rate of change. The MVR fails to address this discrepancy and it's possible explanations, even though this error is evident in both calibration and validation periods. It would also be useful to compare sediment PCB concentration predictions to data for deeper, sub-surficial layers. Slower changes in PCB concentrations would be expected in deeper intervals, reflecting slower transport in deeper sediments. Furthermore, I agree that the approach for developing initial conditions for sediment PCBs (circa 1990 for the Phase 2 calibration and 1979 for the validation) precludes a "robust test" of the long-term model predictions. Consequently, these initial conditions should become part of sensitivity and uncertainty analyses. Finally, the increase in surficial sediment PCB concentrations over time within the upper-most sediment reach (RM 135.13-134.89) seems problematic and should be corrected. Is this behavior related to sudden drop in water column PCB concentrations downstream of the boundary condition, for example in the way bed load is initiated? Without access to the model, it is impossible to sort this out.

Figures 4.2-79 and 6.2-51 are intriguing plots of the change in spatial profiles of the grid cell PCB sediment concentrations over the duration of the calibration and validation, respectively, superimposed with the "shotgun pattern" of data for individual sediment PCB samples. Although no interpretation of the model-to-data comparison is offered, it does suggest that the greater than 90% change in sediment PCB concentrations within many sediment grid cell is both plausible and may help explain the spatial variability observed in sediment PCB data throughout the river sediments. In other words, the model simulation suggests that sediment PCB variability arises from contaminated sediments remaining in locations where erosion is not significant, and (relatively) uncontaminated sediments replacing older contaminated sediments in locations where erosion is significant.

I remain concerned that we may be falsely confident in the erosion/deposition predictions made by this model, because the model is being applied and validated at relatively coarse spatial scales. Due to practical constraints, the modeling team did not simulate sediment transport at the resolution demonstrated in other successful sediment transport applications (Thompson Island Pool: HydroQual, 1995; Watts Bar Reservoir: Ziegler and Nisbet, 1995; Fox River: Lick et al., 1995; Saginaw River: Lick et al, 1995 and Cardenas and Lick, 1996; Buffalo River: Lick et al., 1995 and Gailani et al., 1996). The unstated assumption is that applying a similar model on a cruder resolution will correctly average the sub-grid scale phenomena, even though some of these phenomena are demonstrably nonlinear. This assumption is not supported by either data comparisons or numerical convergence testing using alternative grid resolutions. Predictions of net sediment accumulation can only be tested for Woods Pond, because that is the only portion of the ROR where a reach-wide average accumulation could be inferred from sediment data. Predictions of net sediment accumulation are untested in other reaches. This is of more than academic interest, because we are asked to believe in the model primarily on the basis of the water column predictions of TSS and PCB concentrations. Yet these only indirectly measure erosion and deposition. Reach 5A, for example, is described as a dynamic environment with rapid exchange of solids and associated PCBs. Within this and other reaches, there are locations where both deposition and erosion are taking place. If aggregated spatially (into either reaches or miles) much of this behavior is "averaged out".

Aren't there advantages to validating sediment transport predictions without this spatial averaging? In fact, that is the only way it can be done.

The food chain model (FCM) predictions should be compared to the mean or other central tendency of the fish species, reach, and age-specific data, because the model is designed to simulate "average" fish according to these categories. This is confirmed by analyses and graphics (e.g., Figure 4.3-8) displaying that the residuals tend to be smaller when the data are averages of 6 or more fish. Unfortunately, many calibration and validation graphics persist in using individual data. Nonetheless, there is overall good agreement between predictions and the averaged total PCB concentration data, with most residuals falling within a factor of 2 of the predicted values. EPA has revised Figure 4.3-7 to use the 95% confidence limits (or ± 2 standard errors) to better quantify the measurement precision, and the FCM predictions fall within these limits. FCM model performance is similar in calibration and validation periods.

Figure 3-9 (Comparison of mean measured biota tissue tPCB concentrations to FCM results [simulation using linked models]) provides a summary illustration of the FCM calibration. PCB concentrations predicted for invertebrates are in good agreement with the D-Net sample data, except for infauna in Reach 5A (which are substantially overpredicted). The model fits for bullhead and sunfish are very good; PCB concentration in these fish vary little between river reaches. Model-to-data comparisons are not as good for suckers, cyprinids and bass, although the PCB concentration data for these species is also considerably more variable. I assume that at least some of the variability of PCB concentrations in fish reflects the wide range of PCB exposure and/or failure of the food chain to spatially average these exposures on a reach basis. If I look for an overall pattern in the residuals on this figure, it would be a tendency for the predicted PCB concentrations to increase moving downstream, which is not reflected by the data. PCBs predicted in Reach 5D for cyprinids, sunfish and bass significantly exceed the measured (as well as predicted) concentrations for these species in other reaches. Since PCB concentrations were not measured in fish from this reach, these predictions are untested. Reach 5D is the only portion of the FCM where predictions based on PCB exposures derived from data meaningfully diverge from predictions based on EFDC exposures.

There are insufficient comparisons between data and predictions for both the upstream and downstream models. In the upstream model, there are too few TSS data to tell whether the model predictions are reasonable or not. For both TSS and PCBs, my impression is that the model predictions are much more variable than the few available data. PCB concentrations appear to be substantially overpredicted (e.g., a factor of 5 to 10 overprediction in PCB concentrations). In the downstream model, predicted TSS and water column PCB concentrations are in the same ranges as the data at the Rising Pond outlet, but there is not much fidelity beyond that. Initial downstream model PCB predictions included in the MVR clearly indicated a problem with the balance between erosion and deposition in high-gradient reaches. As presented at the Document Review meeting, this has been addressed by adding an additional noncohesive sediment class in the sediment transport simulation.

FCM predictions in the downstream model domain are compared to PCB concentrations measured in fish for Reach 7 and Rising Pond. These predictions look favorable overall, although there may be bias in the predictions for specific species. Results appear comparable in data-based and linked exposure predictions.

3. Is there evidence of bias in the models, as indicated by the distribution of residuals of model/data comparisons?

A number of significant model biases have been identified and discussed above (Question 2). The most significant are:

- The predicted changes in sediment PCB concentrations are slower than observed,
- Spatial gradients in water column PCB concentrations do not match data under low flow conditions, and
- The treatment of the upstream boundary condition results in serious overprediction of low TSS and PCB concentrations.

The first two issues appear related to parameterization errors, as noted previously, and should be satisfactorily addressed by the suggested corrections. The bias in TSS and PCB concentrations due to the boundary condition deserves further consideration.

At sampling locations in the upper reaches of the PSA, the model fails to match half to two-thirds of the TSS data. Predicted values at base flows are ~10 mg/L versus data in the 2-5 mg/L range. There is less discrepancy in the lower reaches. In general, the bias is most apparent at lower flow rates. Censoring TSS data <5 mg/L was offered by EPA as a way of correcting the model performance metrics. A similar problem with low-flow bias for PCB predictions is also apparent in the upper reaches of the PSA, and a similar censoring approach was suggested. However, I do not believe it is appropriate to ignore or discount the bias in model predictions under low-flow conditions. About 70% of an "average year" is low flow conditions, and significant PCB uptake by fish coincides with low flow periods. It is unclear whether the low-flow PCB bias has led to food chain miscalibration, as suggested by GE, because the food chain model was initially calibrated using data-based exposure concentrations. These would not exhibit the same low-flow bias as exposure concentrations predicted by the model. Nevertheless, I am concerned that bias due to the TSS and PCB boundary conditions may result in model errors that have not been addressed. We know very little about how the boundary condition will evolve over time following upstream remediation; hopefully, ongoing monitoring will be used to update and improve upon this situation.

The specification of the upstream PCB boundary condition is an important component of developing a reliable transport and fate model for the Housatonic River, for each of the time intervals over which the model is to be applied (calibration, validation and forecasting). EPA has developed different interpretations of the PCB boundary condition in the Phase 1 and Phase 2 Calibrations, and these differences produce dramatically different outcomes in terms of the PCB flux analysis generated by the model predictions. The PCB boundary condition developed for the Phase 2 calibration applies the partitioning model to estimate particulate and dissolved PCB concentrations from total PCB measurements, which are then independently regressed to flow rate. The PCB phase concentrations are "capped" to prevent the boundary conditions from grossly exceeding the data at high flows. This is a creative approach; I tried something similar in the Fox River. Unfortunately, it's not clear whether this approach does a reasonable job of describing the boundary condition data. Although this approach may improve the PCB boundary condition at high flows, in comparison to the Phase 1 calibration, at base flow rates the Phase 2 PCB boundary conditions are substantially worse (e.g., Phase 1: ~40 ng/L at base flow vs. Phase 2 : ~100 ng/L). I am not convinced EPA has made the best use of the boundary condition data, and offer some comments below regarding the approach they used to describe this data. Regardless of these issues, it appears that much of the variability seen in PCB concentrations at the upstream boundary cannot be explained, in the approaches taken in either the Phase 1 or Phase 2 calibration, or other alternatives that have probably already been explored. If the TSS and PCB boundary conditions cannot be improved, the best recommendation may be to add a random error component to the boundary conditions in the sensitivity and uncertainty analyses.

The boundary condition data eres provided by EPA in the spreadsheet "East_Branch_Boundary_Data.xls". Their interpretation of this data, and the descriptive approach used to model the boundary condition, is described in Section 4.1.5 of the Model Validation Report, with further discussion offered in supplemental responses. Figure 4.1-2 (reproduced below) shows the regressions of dissolved and particulate PCB concentrations versus flow that were used to specify the total PCB boundary condition for the EFDC transport and fate model (except for days when measurements were available). I have a few comments to make based on this figure and review of the boundary condition spreadsheet:

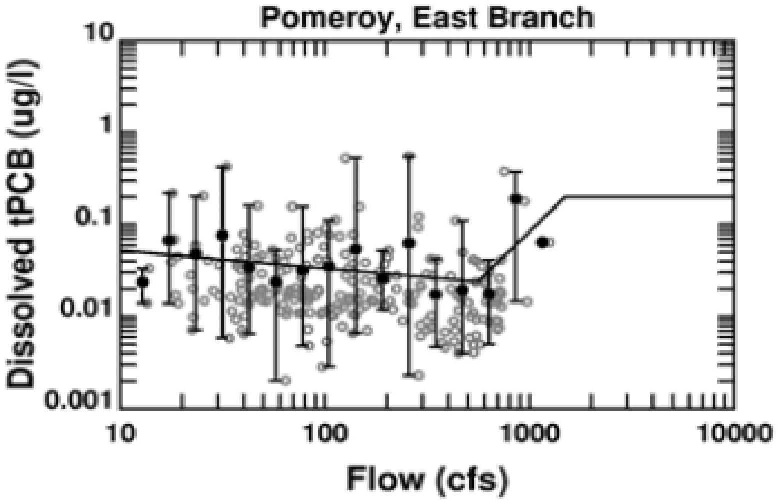
1. As a general observation, I would say that if one focuses on the actual data plotted in these graphs (the open circles), it appears that EPA's regressions describe relatively little of the variability observed in PCB concentrations at the boundary condition. The regressions are especially tenuous at low and high flows. In fact, the regressions exceed <u>all</u> of the data in the lowest flow "bin".
2. I cannot find the PCB concentration data collected at flow rates smaller than 24 cfs[3]. Therefore, I have no way to confirm the results shown in Figure 4.1-2 for the two lowest-flow "bins". At least for the lowest "bin", the regression line appears to be a poor fit. A noted above, it exceeds all of the measured data.
3. At high flows, the PCB concentrations are "capped" to prevent unrealistically high PCB concentrations during floods. The "cap" was specified as 25 mg/Kg for particulate PCBs, based

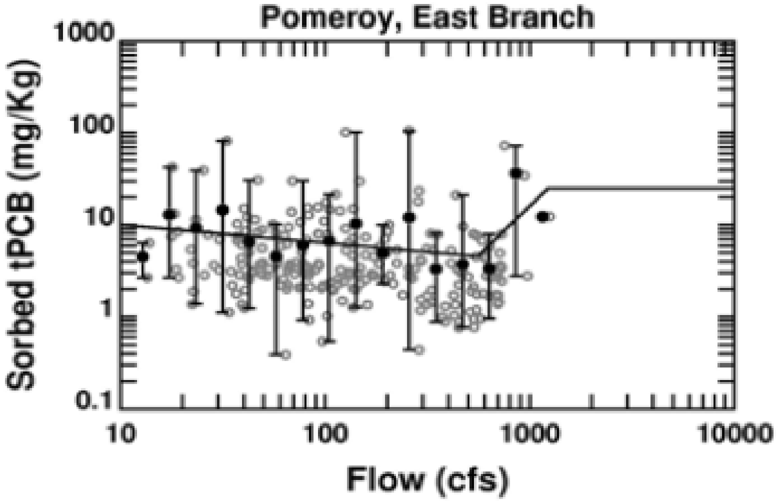
---

[3] I think this may be because East Branch daily flows were provided in the spreadsheet, while Coltsville (hourly?) gauge data were used in the regressions and possibly the graph. I couldn't find the source of the flow data in the MVR.

upon concentrations measured in sediments deposited on vegetation and banks. I think this value may be somewhat too high. My own calculation of the unbiased mean PCB concentration from the 13 sediment deposit samples is 20.5 mg/Kg, although I was measuring the data off of a graph. Please check this calculation. If 25 mg/Kg is indeed too high, it will bias the high-flow PCB boundary condition.

4. According to EPA, the detection limit was used as replacement values for censored PCB concentrations. While this has little impact on the distribution of total PCB concentrations, it does bias the particulate PCB concentrations that are calculated from total PCB measurements using the partitioning model. I didn't check, but I suspect that dissolved PCB concentrations may be biased as well. I think it is generally preferable to use one-half of the detection limit as replacement values for censored PCB concentrations.





5. Another thing I noticed when I reproduced EPA's partitioning calculation (which is used to estimate the dissolved and particulate PCB concentrations from the total PCB measurements), is that the dissolved PCB concentration estimates don't agree very well with the relatively few dissolved PCB concentrations measured at the boundary condition. This is shown below in Figure 2. If the partitioning model is being used to describe the boundary condition, shouldn't it fit the data?
6. Frankly, some of this data is mystifying. For example, 2 samples were collected on November 12, 1995. For both samples, the flow is reported to be over 1,100 cfs. In the first sample, [TSS] was 450 mg/L and total [PCB] was 0.2 ug/L. In the second sample, [TSS] was 17 mg/L and total [PCB] was 0.52 ug/L. Running these values through the partitioning model, one obtains particulate PCB concentrations of 0.44 and 23 mg/Kg. It is somewhat hard to believe that PCB concentrations on suspended solids could vary by a factor of 50 within one day.
7. It is also curious to me that there has been no shift or evolution in PCB boundary concentrations over the past 16 years, and that seasonality is not evident in the data. If PCB boundary condition data for all years are plotted together, we see the familiar shotgun target (Figure 3). There appears to be some underlying behavior, however, within individual years of data. Figure 4 shows total PCB data for years (especially 2002) in which a positive relationship with flow is evident. Figure 5 shows total PCB data for years (2003 in particular) in which the flow relationship is generally negative. The highest total PCB concentrations in Figure 5 come from the warm (and presumably higher organic carbon[4]) months of July-September, while the lowest PCB concentrations at very high flow were sampled in October and November. While I am confident EPA has looked at this data in great detail, I think the patterns in the data hinted at here suggest some underlying behavior which could be exploited in the regression models used to describe the boundary condition. I believe this might, in turn, provide better PCB boundary condition estimates than the current approach in which all measurements are lumped into one sample.
8. The MVR doesn't discuss what other alternatives may have been tested, but here are some suggestions:

---

[4] EPA is assuming a constant 7.3% organic carbon content on suspended solids, which cannot be confirmed from the data collected at the upstream boundary.

- AutoBeale (Beal's Stratified Ratio Estimator) is an alternative statistical approach that works well for boundary conditions when the data are stratified by season as well as flow;
- When regressing the PCB data, were autocorrelation and Lag-1 or 2 considered as factors?
- Check bias correction formula;
- Have untransformed PCB data been regressed? How do results compare to log-transformed regressions?

Figure 2

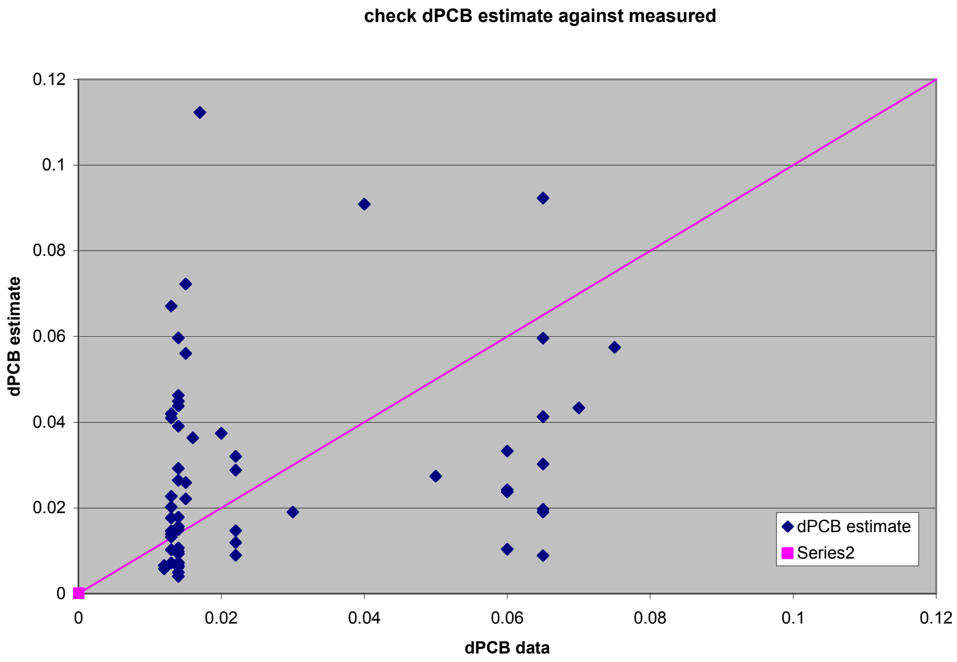


Figure 3

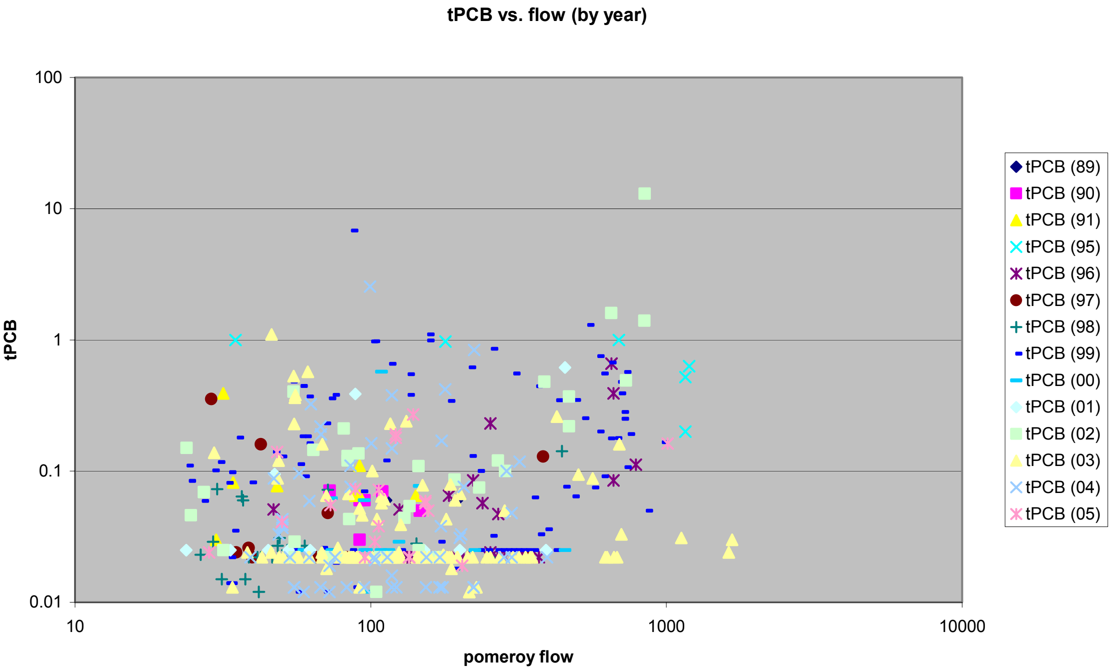


Figure 4

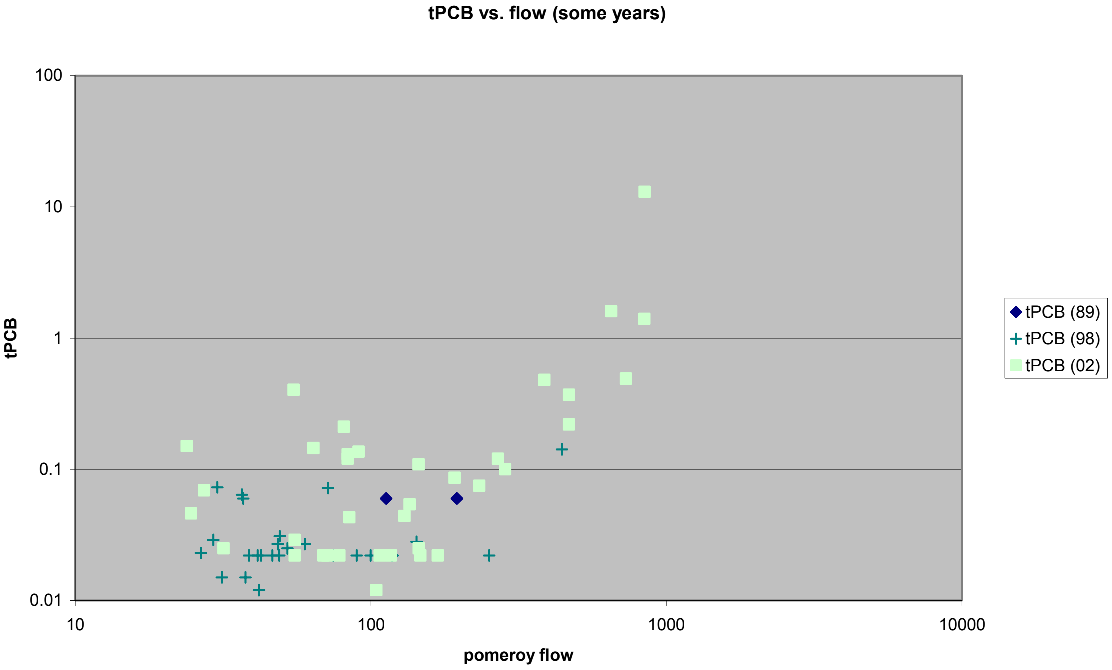
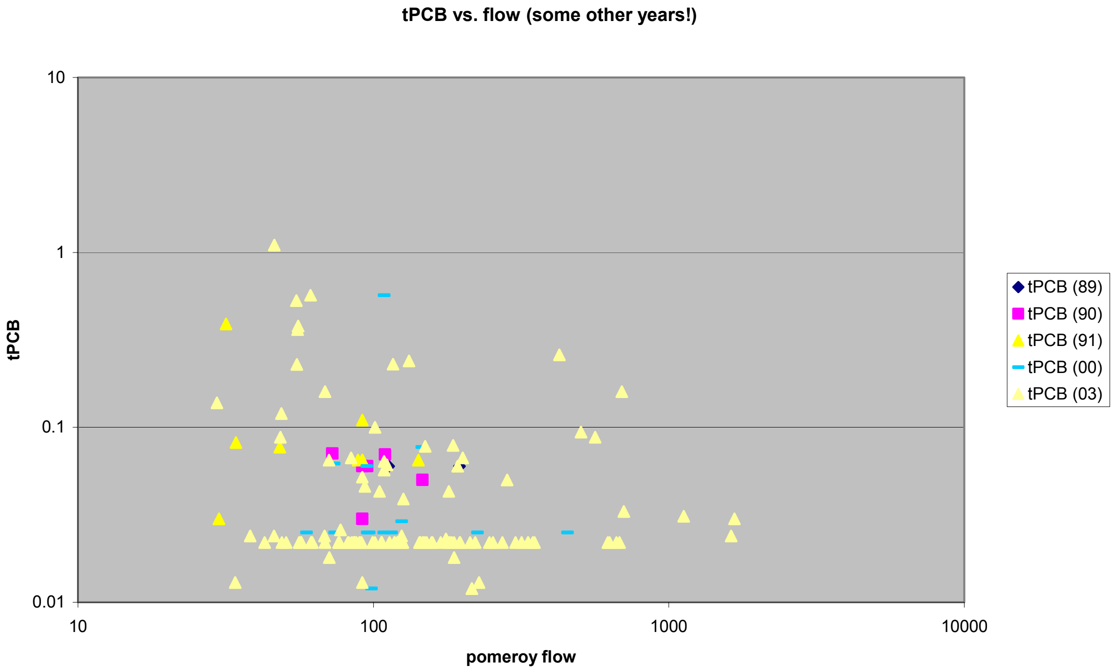
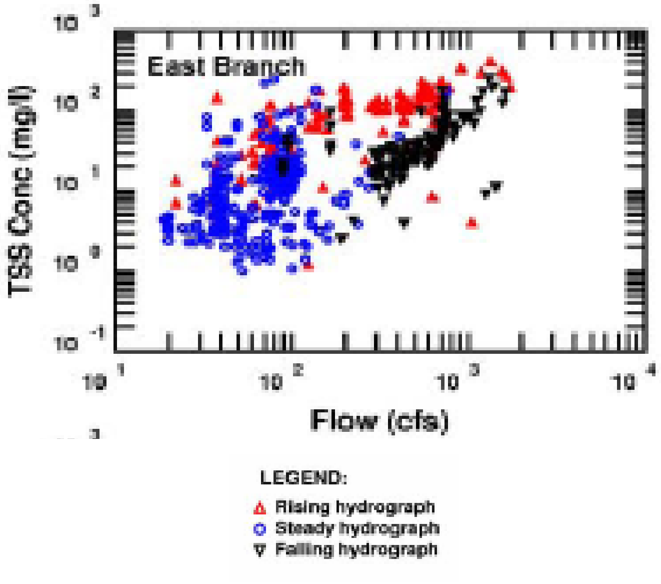


Figure 5



9. I also went back to look at the TSS boundary condition, which has not been changed from the Phase 1 calibration. Figure 6 is a reproduction of the plot displaying the East Branch TSS data as a function of flow (from: Volume 3—Appendix B Hydrodynamic and Sediment/PCB Fate and Transport Model Calibration, Attachment B.3 TSS and PCB Flux Analysis). The TSS data were partitioned according to 3 phases of the hydrograph – rising, steady and falling. TSS data are well-behaved during rising and falling hydrographs, but not at steady (and predominantly low) flow. This makes me wonder again about seasonal factors, intermittent upstream disturbances, or possibly sampling artifacts. It is hard to imagine TSS concentrations varying between 1 and 100 mg/L at base flow.

Figure 6



4. Have the sensitivities of the models to the parameterization of the significant state and process variables been adequately characterized?

The MVR does a good job of illustrating the sensitivity of the EFDC model predictions to input parameters. Several additional model inputs should be added to the sensitivity analysis, and these are identified below. As pointed out by GE, there are also some sensitivity analysis results that are counter-intuitive; these should be explored and explained.

Sediment transport sensitivity analysis is displayed in terms of peak and mean TSS fluxes, for the entire calibration period as well as 2 different flow events (Figures 5.1-3 thru 5.1-5). It would be useful to also see the sensitivity of the model displayed as changes in bed elevation.

Part of the sensitivity analysis for the sediment transport model should include the sensitivity to the grid resolution because, as discussed previously, the prediction of sediment bed scour may depend on the number of lateral cells in the channel cross-section. This is only needed for a “test reach” of the river channel, preferably the reach where bathymetry transects were measured. Note that Earl Hayter’s previous analysis was not informative on this issue.

PCB fate and transport sensitivity analysis is displayed in terms of peak and mean PCB fluxes, again for the entire calibration period as well as the two flow events (Figures 5.1-6 thru 5.1-11). The sensitivity of the Woods Pond PCB flux predictions at low flow (Figure 5.1-9) is puzzling, since it suggests practically no sensitivity to parameters associated with sediment-water exchange of PCBs. While there may be little exchange within Woods Pond itself, considerable PCB exchange in upstream reaches should be reflected in the sensitivities of these parameters in Woods Pond. Instead, Figure 5.1-9 show that low-flow PCB fluxes in Woods Pond are only mildly sensitive to the upstream boundary inflow, and really nothing else. It remains for EPA to explain how this result is evidence that "the model is properly representing the physics of the river as it relates to the movement of solids and contaminants".

As stated previously, the initial PCB concentrations specified for the sediment bed should be added as parameters in the sensitivity analysis, for all reaches where they are not independently determined from data. Likewise, the surficial sediment layer thickness should be included as a parameter in the sensitivity analysis

5. Are the uncertainties in model output(s) acknowledged and described?

Considerable effort went into the analysis of model uncertainty in this project, which is perhaps unprecedented for a suite of water quality models of this complexity. The MVR evaluates the uncertainty of the models using approaches based on Monte Carlo analysis. This approach basically assumes that the models are correct, and errors in model predictions arise from uncertainty or variability in parameter values. These are estimated based on subjective expert judgment using probability distributions, and parameters are generally assumed to be uncorrelated. While these assumptions are obviously suspect, they are nonetheless common practice in uncertainty analysis. Analyzing the uncertainty of the linked model predictions is a strength of this work.

The response surface modeling of EFDC uncertainty appeared to be very thorough and comprehensive. EPA has provided a descriptive field for the parameters listed in Table 5.2-8, which hopefully will be incorporated into the report. I would like to see EPA add initial and boundary conditions for PCBs to this analysis, since these are both uncertain factors. I disagree that the RSM approach is impractical or unworkable. The Responsiveness Document (p.2-110 and 111) discusses how uncertainty analysis can be used in context of evaluating differences between remediation scenario predictions. Hopefully, no more than one RSM will be necessary for evaluating uncertainty of these predictions.

The Monte Carlo analysis of the FCM is also quite thorough and informative. I agree with the suggestion to evaluate model uncertainty in terms of 10th and 90th percentiles, as being reasonably conservative, consistent with the distribution properties of the food chain model outputs. This is probably the simplest way to deal with inflation of predictive uncertainty. The critique of "unreasonable" parameter sets producing model predictions outside the bounds of the data, is really a criticism of the subjective estimation of parameter variability. This can be addressed using a Bayesian Monte Carlo approach, which builds posterior (informed) parameter distributions based on computing the likelihood function for each model realization[5]. The method is simple and works very well - although the computational burden of BMC

---

[5] Dilks, D.W., Canale, R.P. and P.G. Meier. 1992. Development of Bayesian Monte Carlo techniques for model quality model uncertainty. Ecological Modeling. 62:149-162.

(i.e., effort to achieve convergence in the likelihood function) can be heavy, unless there is some way to substitute the RSM for the numerical model. I don't know whether this is a practical option for computationally-intensive models such as EFDC.

There has also been much discussion about how residuals from comparisons between model predictions and data can be used to improve the uncertainty analysis. It has been suggested to use residuals to directly characterize uncertainty. This strikes me as a naïve approach, as I would normally consider residual error to be a "best case" estimate of uncertainty. Forecasts will seemingly always be more uncertain than calibration; the question is, how much more? It has also been suggested that reporting average bias (e.g., Tables 4.2-4, 4.2-10 and 4.3-2) masks the uncertainty due to variability in the PCB data itself. The remedy to this is straightforward: calculate and report the variance of model bias. Moreover, statistical model could be developed from the residuals to estimate the distribution of PCB concentrations from mean predictions. This would compliment, instead of replace, the Monte Carlo results. The difficulty I can foresee, is how do you determine the proper assumptions regarding homoscedacity (e.g., is the variance in a constant factor of the mean or simply a constant?).

No sensitivity or uncertainty analyses have been presented for the upstream and downstream models.

6. Upon review of the model projections of changes in PCB concentrations in environmental media in the example scenarios, are such projections reasonable, using your technical judgment, and are they plausible given the patterns observed in the data?

The MVR presented 25-year forecasts made with the linked models, in the Rest-of-River and Downstream domains, for 2 (or 3)[6] hypothetical remediation examples. These were intended to illustrate

---

[6] I am somewhat confused by what is referred to as a "base case" projection in a number of the time series figures (e.g., "Time series of water column concentrations for PCBs and TSS", "Time series of sediment PCB concentrations in each spatial bin" and "Time series of plots of FCM exposure concentrations" in *the EPA Response to Questions* document, and "Food chain model results" in the Document Overview presentation of Example Scenarios, which included a graph subtitled "Base case for the example runs - existing boundary loads"). The "base case" scenario is not mentioned or described in Section 7 of the MVR, or anywhere else I could find in the materials provided for our review. This makes it difficult to evaluate the example scenario projections, because I want to compare them to a "base case" (as is intended by the graphs mentioned above) but am not sure what this is. Initially I thought the "base case" was simply the validation prediction, since the hydrograph and boundary conditions are the same. Upon inspection, however, it appears that the initial conditions for PCBs are somewhat different, most noticeably in Woods Pond. A description of the "base case" scenario should be added to Section 7 of the MVR to resolve this ambiguity. Furthermore, EPA should consider whether a scenario should be added or modified which specifies the East Branch PCB boundary condition to be 20 ng/L, the assumed nominal concentration following cleanup.

the performance of the models in forecasting the outcome of remediation alternatives in terms of PCB concentrations in sediment, water and especially fish. These forecast predictions ("projections") are really the unique product of water quality models, and are the only defensible means of anticipating the long-term outcome of remedial action and the relative effectiveness of remedial alternatives. We are asked in this question to address the "reasonableness" and "plausibility" of the projections, which are both subjective criteria. Dom DiToro called this sort of evaluation the "laugh test", where we compare the numerical model's projections to those of the models we carry around in our heads. The modeling team is correct in pointing out that it is not possible to determine whether the example projections are "correct" in terms of accuracy (MVR, p 7-3), because there is really nothing to compare them to except one another.

Model projections for the example scenarios were presented a number of ways in the MVR and its supplements. These included the predicted time series of PCB concentrations in water, at the conventional monitoring locations (Holmes Road, New Lenox Road, …); PCB concentrations in surficial sediment in the mile reaches; and, PCB concentration in largemouth bass in the river reaches 5A thru 6. For EFDC, the projections were also presented as plan maps of the grid displaying PCB concentrations in surficial sediment and soil, and changes in concentration, at the end of the 25 year projections. For the top 6" of the sediment bed, the end-of-projection results are also shown as profiles. Each of these is discussed below.

The projected timeseries of water column TSS and PCB concentrations are plotted together in the figure "Time series of water column concentrations for PCBs and TSS" in *the EPA Response to Questions* document. The base case projection is qualitatively similar to the validation predictions, with TSS and PCB concentrations increasing to high levels during flow events, although over time the base case PCB concentration projections decline during low-flow periods. PCB concentrations projected for the example 1 scenario are lower than the base case, and projections for the example 2 scenario are lower than for example 1. In each scenario, there is a similar "event responsiveness" of PCB concentrations to fluctuations in the hydrograph. TSS concentration projections are comparable in each scenario, as would be expected since the TSS boundary condition and grain size distribution in the sediment bed are unchanged from the base case. I would consider the water column projections for each of these scenarios to be reasonable and plausible.

The projected timeseries of surficial (top 6") sediment PCB concentrations in river mile reaches are plotted together in the figure "Time series of sediment PCB concentrations in each spatial bin" in the *EPA Response to Questions* document. The base case projections for surficial sediment PCB concentrations are qualitatively similar to the validation predictions, and suffer similar errors. PCB concentrations are projected to increase in the first half-mile reach (an apparent artifact), and show almost no change after start-up in the other reaches. The rate of change in PCB concentrations in sediment is (again) unreasonably slow. Even 25 years is not long enough to see a discernable change in surficial sediment PCB concentrations, which I would certainly expect to see. Nothing shown in the MVR or at the peer review meetings convinces me that the model is correct in its predictions of the rate of change in sediment PCB concentrations. For example scenario 1, a decline in sediment PCB concentrations is projected in comparison to the base case, which is more encouraging. The difference is most dramatic in the first half-mile reach (RM 135.13-134.89), and is also apparent in reaches RM129.19-128.88, 128.88-128.69, and 126.99-126.47. In other reaches, especially in Woods Pond and its headwater, there is no difference in PCB concentrations between example 1 and the base case. For example scenario 2, negligible (1 mg/Kg or less; I can't tell the actual value due to the axis scaling) sediment PCBs are projected for reach 5A due to remediation. In the other reaches, sediment PCB concentrations are projected to be nearly the same or identical to example 1 projections. Although I don't believe that the model's projections of rates of change in sediment PCB concentrations are reasonable, at least the relative differences in projected sediment PCB concentrations between the scenarios are plausible.

The grid map diagrams (Figures 7-1 and 7-3) show the cumulative change of PCB concentrations in the top 6" of sediment and floodplain soil over the 25 year projection, for scenarios 1 and 2[7]. It appears that the sediments in reach 5A, which are cleaned up in example scenario 2, recontaminate to 1 or 2 mg/Kg over 25 years, due to continuing transport of PCBs from the West Branch boundary condition. It would be useful to see a similar grid map diagram in which the difference in end-of-projection PCB concentrations was plotted.

As pointed out by GE, predictions of PCB concentrations in flood plain soils are untested. There appear to be substantial increases in PCB concentrations in the top 6" of much of the floodplain soil in both example projections, which is somewhat surprising. However, since no data have been shown to establish rates of PCB concentration change in floodplain soils, the "reasonableness" of the flood plain projections cannot be determined. Really, this illustrates the absurdity of devoting so many EFDC model grid cells to the flood plain, which it seems dictated using a single footpath of tiles for the grid of the river.

The surficial sediment PCB concentration profiles offer a useful alternative view of the cumulative change of PCB concentrations over the 25 year projection. Projections for example 1 show the now-familiar redistribution of sediment PCBs in the high-energy reach 5A, but also significant changes (approaching ± 100% of the IC) in reaches 5B and the upper end of 5C. At the lower end of reach 5C and reach 6, sediment PCB concentrations decline by 20 to 80%. This seemingly contradicts what I interpreted from the "Time series of sediment PCB concentrations in each spatial bin" figures. Could it be that the combination of reach averaging, displayed on a log-axis plotting, makes it difficult or impossible to properly interpret the projection results?

At the Peer Review meeting last month, there were questions and discussion about whether it was reasonable that the EFDC model projected a significant reduction in PCB concentrations in the water column, but little or no change in sediment concentrations. In fact, such behavior is expected in water quality models incorporating both water column and sediment sedimentation. In such models, the concentrations of particle-associated constituents change much more rapidly in water than they do in sediment, whenever there is a change in the constituent loadings or boundary conditions. Since PCB

---

[7] I suppose the comparable graphic for the base case would be the validation figure (Figure 6.2-55) although I am forced to speculate here.

boundary conditions and loadings (actually fluxes from reach 5A sediments) are being manipulated in the example scenarios, the differences in the projections of PCB concentrations in the water column versus sediments are entirely expected and reasonable.

Example scenario projections made by the FCM were presented at the Document Overview meeting (13 - Example Scenarios.pdf) as timeseries of PCB concentrations in largemouth bass. In reach 5A, bass in the base scenario are forecast to have a body burden of 50 mg/kg at the end of the 25 year projection; for example 1, the comparable forecast is 30 mg/kg, and 0 concentration for example 2. In reach 5B, bass in the base scenario are forecast to have the highest PCB body burden at the end of the 25 year projection: 70 mg/kg; for example 1, the comparable forecast is 40 mg/kg, and less than 20 mg/kg for example 2. Further downstream in the Rest-of-River, the remediation examples are projected to considerably less effective in reducing PCB concentrations in comparison to the base case. In reach 5C, bass in the base scenario are forecast to have a body burden of 30 mg/kg after 25 years; for example 1, the comparable forecast is 20 mg/kg, while for example 2 the forecast is 10 mg/kg. At the end of the 25 year projections in reach 6 (Woods Pond), bass in the base scenario are forecast to have a body burden of 50 mg/kg; for example 1, the comparable forecast is 40 mg/kg, while for example 2 the forecast is 30 mg/kg. To me, this progression in the declining effectiveness of remediation as one moves downstream appears reasonable in the context of PCB water and sediment exposure projections for these example scenarios. In "no data" reach 5D, the least reduction in bass PCB concentrations due to remediation is projected. I suppose this may be because this reach is relatively isolated from the rest of the river.

No scenario projections were presented for either the upstream or downstream model.

7. Is the final model framework, as calibrated and validated, adequate to achieve the goal of the modeling study to simulate future conditions 1) in the absence of remediation and 2) for use in evaluating the effectiveness of remedial alternatives?

I interpret this to be more than one question, so I will give two answers. If the goal is to forecast PCB concentrations in the Housatonic River in the absence of further remediation, then the current models may be adequate (i.e., "good enough"). In their current form, the EPA models (and probably many much simpler models) can provide meaningful answers to many questions about the future of PCBs the Housatonic River. The model calibration and validation show the relative magnitude of PCB inputs to system, and the resulting predictions and diagnostics are instructive in terms of evaluating remediation targets. For example, eroding banks and sediment deposits in reach 5A are clearly targets for remediation. The 25-year validation and example scenario projections also address the need and possible effectiveness of remediation in the Housatonic River. Following the remediation of the East Branch reaches, upstream PCB concentrations of 20 ng/L and 0.4 mg/kg expected; these concentrations alone may support gamefish concentrations of 1 ppm or more in the PSA. In the absence of further remediation, PCB concentrations will decline very slowly from the current levels. Although water-column PCB concentrations will decline downstream of remediation conducted at the reach scale, surficial sediment PCB concentrations will be relatively unaffected in downstream reaches. The predicted response of PCB concentrations in fish to remediation is expected to be somewhat similar to the sediment. To substantially reduce (e.g., >50%) PCB concentrations in Reach 5D and 6 gamefish over 25 years appears to require remediation well beyond the spatial extent of hypothetical example 2. None of these conclusions require a highly accurate model.

However, if the goal is to forecast PCB concentrations to evaluate the effectiveness of further remediation, and to distinguish between the effectiveness of various remedial alternatives, then I believe that the current models are not adequate. Although they are good, they are not as good as they could be. I assume that GE will not willingly undertake additional remediation activity in the Housatonic River, unless EPA can convincingly demonstrate that such remediation is necessary and that the remedy will be effective. At their current stage of development, the EPA models are not ready to make such a demonstration. I am arguing that further model development and testing is warranted, and that EPA should be motivated to do so because the financial stakes are high. Otherwise, GE will likely exploit weaknesses in the models to argue against the need for further remediation.

Outlined below are the steps I believe are necessary to make the EPA models the best possible tools to accomplish the goals of the modeling study for the Housatonic River:

- Revise the MVR to incorporate supplemental material (*EPA Response to Questions* document , Document Overview presentations, etc.) and remove provisional results (mass balance diagrams, downstream model w/o boulders, …) that have been superceded since the MVR was released.

- Make a number of near-term corrections to EFDC, which can be addressed/resolved within a ~1 yr time frame)
  - Correct parameterization errors identified in EFDC model:
    - Reduce thickness of surficial sediment bed layers;
    - Parameterize vertical mixing rates as functions of benthos density & vertical position in sediment bed;
    - Increase diffusive PCB fluxes by calibration and parameterize spatially as functions of benthos density;
    - Reanalyze SEDFLUME data using resuspension parameters constrained by literature (e.g., shear stress exponent n=2.6±0.3) and recalibrate deposition rates. Lick, Ziegler and coworkers have published much guidance on the specifics of parameterizing sediment resuspension[8], some of which the modeling team has chosen to ignore;
  - Revise TSS and PCB boundary conditions to correct bias evident in low flow range.

---

[8] See, for example: A Quantitative Framework for Evaluating Contaminated Sediment Sites. SETAC 20th Annual Meeting, Philadelphia, PA, November 14-18, 1999.

  - Test sensitivity of EFDC hydrodynamics, sediment transport and PCB transport simulations to alternative grid resolutions in a "test reach" of the PSA. In all prior SEDZL applications I am aware of, at least a 2-dimensional model was used for hydrodynamics and sediment transport. In river systems, this was usually a vertically-integrated model. Furthermore, on rivers that bend as much as the Housatonic, a curvilinear grid has commonly been applied. Since the EPA modeling team has elected not to follow these conventions, they should at least demonstrate via numerical testing that their primarily 1-dimensional model of the Housatonic River channel produces comparable results for sediment and PCB transport.
  - Investigate methods to economize on hydrodynamic and sediment transport simulations (e.g., using steady state design storms and flow-duration statistics, such as the Saginaw River SEDZL application[9])
  - Consider using the calibrated and validated EFDC model to build a simpler box model of the river, if the two models can be shown to produce comparable results. The simpler model could be particularly useful for forecasting uncertainty. Don Mackay suggested this "second simplified model" approach for modeling hydrophobic contaminants in the Niagara River[10].

- Develop and implement a process of ongoing model validation, using data collected as remediation progresses. It goes with out saying that modelers want to continue modeling… However, the reality is that no model is truly ever "finished". So long as new data are collected, model refinement must be expected and accommodated by managers and decision makers.

I do not believe that either the upstream or downstream models are sufficiently developed to be adequate to address the goals of the modeling study. Their development appears to have been rushed, an impression reinforced by the fact that they apparently did not exist a year ago. Sensitivity and uncertainty of the upstream and downstream models have not been reported, and documentation of the development of these models has been inadequate. In its present status, the upstream model is incomplete and not fully validated. PCB fate and transport have not been simulated, and insufficient model-data comparisons have been made and/or reported. Incorporation of the upstream model in the MVR appears premature, given that substantial additional development efforts will be required before it can be used reliably. Likewise, the downstream model has not been sufficiently validated. Limited model-data comparisons suggest there may be significant bias in PCB concentration predictions. There also appears to be a lack of adequate and

---

[9] Cardenas, M. and W. Lick. 1996. Modeling the Transport of Sediments and Hydrophobic Contaminants in the Lower Saginaw River. *J. Great Lakes Res.,* 22(3):669-682**.**

[10] McLachlan, M. and Mackay, D., "A Model of Contaminant Fate in the Niagara River", report prepared for Environment Canada (1987).

appropriate data for this significant extension of the model domain. In their current states of development, neither the upstream nor downstream models are suitable for use in modeling future conditions in the Corrective Measures Study. I am hopeful that the inadequacies of the upstream and downstream models can be addressed through continuing monitoring and modeling activities.