Final Comments on Model Validation Report
E. Adams, July 2006

Question 1 Considering the changes implemented in the Phase 2 Calibration, does the model 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?

Sorption kinetics

The model assumes that three chemical phases of PCBs are in equilibrium. As EPA points out, this is a common modeling assumption, but I believe it is only reasonable within a stationary sediment bed, and not within the water column during sediment resuspension.

Following Wu and Gschwend (1986), and focusing only on sorption and desorption among two phases (sorbed PCB with concentration C_s and dissolved PCB with concentration C_d)

$$\begin{aligned} \frac{dC_s}{dt} &= k_1 K_p C_d - k_1 C_s \\ \frac{dC_d}{dt} &= k_2 C_s / K_p - k_2 C_d \end{aligned} \qquad (1)$$

where k_1 and k_2 are rate constants constrained by

$$k_1 = k_2 / \rho K_p \qquad (2)$$

where ρ is the solid-water phase ratio $\rho = (1-\phi)\rho_s / \phi$, ρ_s is the solid density, ϕ is porosity and K_p is the partition coefficient. For a stationary sediment bed, ρ is about 1, so for hydrophobic contaminants (large K_p), most of the contaminant is sorbed; hence the processes of sorption and desorption cause C_d to vary in the range $0 \le C_d \le C_s / K_p$, whereas C_s remains nearly constant. The characteristic time for sorption/desorption for highly particle-reactive species in a stationary sediment bed ($\rho K_p >> 1$) is thus k_2^{-1}.

Wu and Gschwend (1986) describe the sorption/desorption process as one of molecular diffusion through particles (or particle aggregates). For large K_p and an assumed intra-aggregate porosity, their effective (retarded) diffusion coefficient is given by

$$D_{eff} \cong \frac{0.2 D_m}{K_p \rho_s} \qquad (3)$$

where D_m is the molecular diffusivity of PCBs. Wu and Gschwend (1988) fit a first-order sorption kinetics model to the radial diffusion model over the (initial) time during

which half of the sorption/desorption takes place. Comparing the two models, and for $\rho K_p >> 1$

$$k_2^{-1} \sim \frac{(R/\rho K_p)^2}{D_{eff}} \tag{4}$$

where R is a characteristic aggregate radius. It can be seen that the time scale (k_2^{-1}) is proportional to K_p^{-1}. In other words, for highly particle reactive species, little of the initially sorbed mass is exchanged during desorption and that which is, is lost from the outside of the particle (over an effective length of order $R/\rho K_p$). Even for a large R (0.01 cm) and $K_p = 10^5$ cm^3/g, the time scale (k_2^{-1}) is less than one second, suggesting that desorption is very fast, and that the assumption of ***equilibrium partitioning is probably acceptable within the sediment bed***. See also following discussion under diffusive exchange.

During resuspension, where the sediment concentration in the water column (TSS) is small, equilibrium partitioning would require that most contamination become desorbed. Hence diffusion must take place through the entire particle, giving a time scale (k_1^{-1})

$$k_1^{-1} \sim \frac{R^2}{D_{eff}} \tag{5}$$

Even for a small R (0.001 cm), the time scale is over a day, which exceeds the duration during which most suspended particles remain in the water column. Hence the assumption of ***equilibrium partitioning is not really valid for resuspended particles***. Because much of the contamination on resuspended particles will not have time to de-sorb, assuming equilibrium partitioning overestimates the PCB flux from bottom sediments due to resuspension. Of course, during calibration, this could have been compensated for, in part, by assuming less sediment resuspension.

Biomixing and Bioavailable Depth

Originally, the model used a bioavailable depth of 15 cm over which the biomixing coefficient (bio-diffusivity) was $D_b=10^{-9}$ m^2/s, or approximately 1 cm^2/d. I commented that this seemed quite large, and EPA agreed. They have changed their formulation to utilize a subduction velocity, V_s, which they get from a literature review of the rates of sediment mass reworking per organism per time, and site-specific data on organism abundances. Their chosen velocities are approximately $V_s \sim 10^{-9}$ m/s over the top 4-7 cm, and $\sim 10^{-10}$ m/s over the next 6-8 cm. In the absence of direct measurements of mixing, this may be the best approach, but it is noted from figures presented at the May 10 Document Review Meeting (DRM) that there is tremendous variability in organism reworking rates, suggesting much uncertainty.

Assuming a vertical distance of 5 cm between the top two sediment layers, their equivalent new values of bio-diffusivity are $\sim(10^{-10}$ to 10^{-9}m/s)(0.05m) ~ 5×10^{-12} to 5×10^{-11} m^2/s or ~20 to 200 times smaller than previous. These are probably more reasonable,

but it is difficult to assess whether or not they are right, because there is not much vertical variation in the existing sediment PCB concentrations. However, following remediation, the vertical concentration gradients could increase substantially, so this is an important process.

This also raises the issue of vertical resolution. If, following remediation, clean sediment is overlain by a thin layer of contaminated sediment, the numerical model will immediately mix the contaminant over the top layer (say 5 cm), whereas, with a bio-diffusivity of 5×10^{-7} cm^2/s, mixing will take $(5cm)^2/5 \times 10^{-7}$ cm^2/s or about 1.5 years to achieve this mixing. Thus, from a numerical modeling standpoint, ***the vertical grid size is too large.***

The sediment-water interface can rise or fall due to deposition and erosion, and EPA argues that this should not affect the rates of mixing. This may not be true since you could expect different rates of organism mixing in relatively fine grained sediments that have been recently deposited, versus older, more consolidated sediments that have been eroding. Indeed, Figure 13 of the DRM handouts on sediment mixing shows a strong increase in V_s with percent fines. However, to honor their assumption, they make V_s dependent only on the depth below the (moving) interface. But because the individual layer thicknesses are changing, the amount of vertical mixing will change. (A constant value of V_s will result in more mixing between thick layers having the same concentrations as thin layers.) Using a bio-diffusivity (dimensions of L^2/T) would take this effect into consideration since it is effectively V_s times the mixing length.

Diffusive exchange

The calibrated mass exchange coefficient is $K_f = 1.5$ cm/d, which actually seems to be on the small side, since it incorporates a number of processes in addition to strictly pore-water diffusion. The coefficient for pore-water diffusion by itself should reflect the rate at which bioturbation brings PCB-sorbed sediment to the interface, the rate at which PCBs are desorbed to the porewaters, and the rate at which diffusion transports the dissolved PCBs into the overlying water column. These processes work like resistors in series. Chen (1993) showed that the sediment-water flux can be expressed as

$$J = \frac{C_{sL} / K_p}{\frac{\delta}{D_m} + \frac{2.2R}{(1-\phi)[(D_b + D')D_m \rho_s K_p]^{1/2}} + \frac{L}{(1-\phi)\rho_s K_p D_b}} \qquad (6)$$

where C_{sL} is the sorbed phase PCB concentration at the depth of the mixed layer L, δ is the water side boundary layer thickness, and D' is the effective diffusivity of PCBs within the sediment (molecular diffusion as affected by tortuosity).

Jorgensen and des Marais (1990) suggest $\delta = 0.02$ to 0.1 cm, and Shaw and Hanratty (1977) suggest

$$\delta = \frac{12 D_m^{0.3} \nu^{0.7}}{u^*} \quad (7)$$

where ν is the kinematic viscosity of water (~10^{-2} cm^2/s) and u* is the friction velocity. Using u* = 0.25 cm/s gives $\delta \sim 0.05$ cm. Assuming L = 5 cm, $K_p = 10^5$ cm^3/g, D_m = 0.5×10^{-5} cm^2/s, D' = 0.3×10^{-5} cm^2/s, $D_b = 5 \times 10^{-7}$ cm^2/s (EPA's surface value), ρ_s = 2.5 g/cm^3, ϕ = 0.6, and R = 0.01 cm, gives values for the three terms in the denominator of Eqn 6 of roughly 10^4, 3×10^1 and 10^2 s/cm respectively. This suggests that the flux is indeed water-side controlled (i.e, biomixing supplies contaminant to the interface sufficiently fast, and the contaminant desorbs sufficiently fast, that diffusion on the water side limits the transport) and that the last two terms can be ignored. Even if D_b were smaller by an order of magnitude ($D_b = 5 \times 10^{-8}$ cm^2/s), the three terms in the denominator would be 10^4, 3×10^1 and 10^3 s/cm, leading to similar conclusions, albeit by a smaller margin. The second term represents the "resistance" due to desorption, and the fact that it is small suggests that equilibrium partitioning can indeed be assumed in computing the flux. The reciprocal of δ/D_m is K_f which, with the above numbers, is ~ 10^{-4} cm/s (8.6 cm/d). This is in the range of the values computed directly from EPA's flux analysis (e.g., Figure B.4-30 of the MCR, which includes values between 0.8 and 250 cm/d, with the majority between 3 and 10 cm/d). However, it is significantly above the value of 1.5 cm/d identified in the Phase 1 model calibration. One possible reason for the calibrated value being significantly lower is that the PCB concentrations used for the upstream model boundary conditions are generally higher than the data, at low flow, which may cause the calibration to underestimate the sediment-water exchange flux downstream.

Although the flux analysis indicated significant temporal variability in K_f that seems like it is correlated with stream flow rate, EPA's values of K_f based on complete model calibration appeared to be independent of flow rate. There seemingly should be some be some dependence, since increased flow would increase stream turbulence, decreasing δ and increasing K_f. As EPA acknowledges, as flow rate increases, it dilutes the water column concentration of PCBs, making it difficult to test for flow dependent fluxes. But observations do show that the model over-predicts water column PCB concentrations during low flow (when diffusive fluxes would dominate) which could reflect, at least in part, the lack of flow-dependence.

The time scale for natural recovery is the mass inventory per unit area, $C_s(1-\phi)\rho_s L$, divided by the flux, $J = C_{sL}K_f/K_p$, or $\tau = (1-\phi)\rho_s L K_p/K_f$. For $K_f = 10^{-4}$ cm/s (my value), and using other parameter values from above, the time scale is 160 years. Thus diffusive exchange may not be very important for contaminated sediments that are buried at or below the assumed level of bio-mixing (i.e., most existing sediments). However, for recently transported sediments with smaller L (e.g., following remediation), the time scale could be much smaller, especially if K_f were even larger. For example, if L were only 1 and K_f were simply twice the above value (2×10^{-4} cm/s), the time scale would be reduced to 16 years. However, the model would not be able to resolve the resulting sharp gradients with the current, relatively coarse, vertical grid scheme.

Erosion
The erosion formulation and parameters come from analysis of SedFlume data. W. Lick argues that the exponent n (denoting dependency of erosion on shear stress) should be significantly greater than the chosen value, which would produce proportionally more erosion during high flow conditions. I agree with him, but am also leery of the fact that shear stresses in the model are computed based on a grid width that is essentially equal to the river width. Thus the computed hydrodynamics will yield cross-sectional average velocities, which ignores regions within a cross-section with relatively high velocity and erosion (refer to later discussion under Question 7). Unlike conditions in SedFlow, the real river is non-uniform. Thus the formulation of erosion can not be divorced from the question of grid resolution. Indeed, it would seem incorrect to simply import an erosion calculation from SedFlume.

Spatial Variability
Much has been was said about the tremendous spatial variability in sediment PCB concentrations over space scales of order one meter and the fact that the model can not reproduce this variability. Whether or not this is a model failure, *per se*, or simply unresolved variability in model input and output (sediment bed concentrations), depends on what has caused the variability. I believe the variability was caused mainly by the stochastic method in which the PCBs were introduced in the first place. In such case, we cannot expect the model to predict this variability and the fact that the model averages concentration over relatively large grid cells is not a problem (with the mean) unless sediment-water exchange of PCBs varies non-linearly with concentration (and some non-linearity is inevitable, given the averaging associated with the coarse grid). Of course, we can not expect the model to tell us anything about the future variance of sediment bed concentrations and to the extent this is important, we should rely on the observed variability. The PCBs have been in the sediments for several decades, and to a first approximation the variability expected in the next decade or two (presumably our focus) will not be very much different from the variability observed historically (at least for natural attenuation).

On the other hand, if the variability is due to active sediment transport processes that are sorting the sediments and their contaminants, then the failure to pick this up could be a significant model deficiency. For example, natural attenuation could conceivably increase local PCB concentrations. The available time series data of PCB concentrations within surficial sediments suggest a decrease in concentration (indeed more so than is being modeled), so I don't believe this is a significant process.

Boundary Loads
The trend of the PCB concentration versus flow data for the E. Branch, shown in Figure 4.1-2 seems strange (why the sudden change at 550 cfs?). I guess this is simply what the data suggest. Also the model fit exceeds the data at low flow (10 cfs) by a factor of 2-3 for both dissolved and particulate PCBs. This excess upstream load may have led to an underestimation of calibrated fluxes from other sources downstream. Given the prevalence of low flow periods, this is important.

Initial conditions for sediment PCB concentrations
I agree with GE that hindcasting predicted trends is not a robust way to establish initial conditions. By playing the tape backwards, then forwards, they will get the same conditions they started with. I am not sure what the alternative is, but the lack of independence should be acknowledged.

Summary for Question 1
I believe that several processes could be better represented. But I recognize that several of these processes have been calibrated, so that ***if one parameter were changed, others would have to be changed*** as well. If there is time, I would like to see the model recalibrated using more appropriate values, within physical/chemical/biological constraints. (I do not think an independent validation is necessary.) In the absence of re-calibration, the model users must exercise considerable judgment in the interpretation of model output.

Question 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.

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

These questions are clearly related, so their answers are combined. The "upstream" and "downstream" models are discussed at the end.

Flows and Velocities
Flows were simulated for the 10.5 year Phase 2 calibration period (Figures 4.2-4 through 4.2-14). Unfortunately there is not much data to compare with. The most data are for early 1999 when the model consistently under-predicts flow. (Yet it seems to do well in the later validation periods, based on the pressure transducer data.) Similar agreement with stage suggests that average velocities should be pretty good. (Figure 4.2-26 shows reasonable agreement between measured and predicted velocities, but the predictions are generally too low; EPA argues that discrepancies are due in part to coarse grid resolution.) During storms there was reasonable agreement, with moderate overprediction of flow at low flow. Agreement during storms shows times of over- and under-prediction, but they seem OK on average. The model seems to do reasonably well simulating the extent of overbank flow during August 1990 (Figure 4.2-25). Model statistics seem generally within ranges specified by the QAPP and there are few indications of bias.

In the validation period, flows and velocities look similar to or better than the calibration period. There are some errors in the timing of flows, but these are not important. No major bias is seen.

TSS

The model uses flux analysis for upstream TSS when there is no data and often this analysis overestimates TSS by a factor of 2-4 (Figures 4.2-31 and 6.2-12). There is some underestimate of TSS at high flow.

The model tends to overpredict TSS during the calibration period (Figure 4.2-39), especially at lower TSS concentrations. This is possibly because of the overprediction of upstream loading. The model seems to do better with the storm events, which EPA claims are more important. While bias in TSS at low flow may be insignificant in terms of the sediment budget, it does affect the water column PCB concentrations, and hence the relative uptake of PCBs from the water column and the sediment in the Food Chain Model. In general it would have been nice to have had more data on TSS.

The statistical summary (equivalence plot, Figure 4.2-61) indicates generally good agreement, but the model shows less variability than the data; i.e., the model overpredicts low concentrations and underpredicts high concentrations. There appears to be no significant bias with flow rate. Sedimentation rates agree reasonably well with measurements using Cs-137.

Bedload concentration is computed as bed load mass rate divided by river flow (pg. 4-56). Given that the bedload is traveling slower than the average stream velocity, this would seemingly give too low concentrations, but I suspect velocities are more important than concentrations.

PCBs

The spatial plots at low Q (Figure 4.2-69) and moderate Q (Figure 4.2-70) show that upstream concentrations of both TSS and PCB are too low (a consequence of the BC) and that there is more local variability (indicated by shading) than variability between reaches. There is no way to verify the local variability. It is hard to say whether the overall spatial trend is correct or not, because of variability in the data. The temporal trends look OK.

The temporal changes in sediment PCB (top 6 inches) show that the model does not reproduce the significant decline in Woods Pond over time (Figure 6.2-50). This is one of GE's main points and suggests that the model might also underpredict the response to remedial actions, including natural attenuation. ***I see this as a major concern***.

The model misses the decline in PCBs across Woods Pond that is both observed and intuitively suspected. This would mean the model overpredicts loading downstream (downstream model).

Like TSS, the equivalence plots for PCBs (Figure 4.2-85) shows less variability than the data. However, there is little indication of (overall) bias.

And also like TSS, it would also be nice to have had more data on PCBs. Again, there is better model-data agreement at relatively high (presumably more biologically relevant)

concentrations. As with other variables, the model shows less variability than the data (i.e., it underpredicts the highs and overpredicts the lows).

Except for the transducers which record flow (through a stage discharge relationship), better data seem to be available for the calibration (especially the original calibration) period, than for the validation period. However, I am less concerned with traditional model validation (the model's ability to predict absolute concentrations under different hydrologic conditions) than I am with the model's ability to distinguish among different remedial alternatives that would presumably be evaluated under the same hydrologic conditions. See comments under Question 6.

Upstream Model

The model was extended two miles upstream to include the 0.5 and 1.5 mile remediation areas. Upstream Q and TSS were obtained from rating curves, while upstream tPCB = 0.019 microgram/L from boundary measurement. Comparisons were attempted against TSS and tPCB measured at Lyman St. in 2001-2004 (some of this is reported in DRM handouts and not in the MVR).

Predicted TSS varied from ~$3x10^{-4}$ to >10^{2} mg/L while data were in the range of 3 to 30 mg/L (mostly 3-10). [I am assuming these are the correct units for TSS; the vertical axis on the handout distributed at the DRM says ng/L.] Can one really predict TSS as low as $3x10^{-4}$ mg/L? Some data are in the range of the model, but model predictions show much more variability? Where does this come from?

Predicted PCBs were $3x10^{-5}$ to 10^{-3} mg/L while data were in the range of ND up to $3x10^{-5}$. [Again, I am assuming these are the correct units; they are consistent with the 30 to 1000 ng/L indicated on the handout GE provided, though EPA's handout during the document review meeting indicates $3x10^{-5}$ to 10^{-3} ng/L.] In any case, the agreement is not very convincing, and ***it is hard to say that the upstream model is validated***.

The downstream concentrations of TSS and PCB from the upstream model are much lower than the corresponding upstream boundary values used for the main model. Given that output from the main model depends so heavily on its upstream boundary condition, more work should be done on the upstream model, or at least more interpretation of the results. There is scant discussion in the MVR.

Downstream Model

The model was extended 19 miles downstream to Rising Pond Dam (not described in MVR; only in DRM handouts). Upstream boundary conditions were taken from output of the main model at Woods Pond. Model-data comparisons for TSS and tPCB were made between 1990 and 2004 ~ 1mile downstream from Rising Pond Dam.

Predicted TSS varies widely as do the data, with the peaks corresponding with peaks in flow. They are in the same ballpark, but the variability is so great that you need a statistical comparison to determine how well they match on average.

Predicted PCB concentrations also vary substantially (~$3x10^{-3}$ to ~3 microgram/L), with predicted peak concentrations corresponding with peak TSS and peak flow, and non-peak concentrations hovering generally around 0.03 micrograms/L. (But why are some up to 10 times lower?) Measurements show much less variability with an average of around 0.1 microgram/L, or about 3 times the predicted non-peak values. The data are too sparse to see if they respond to peaks in flow. Again a statistical comparison would be helpful to quantify model-data agreement. Both measurements and predictions seem to be 2-3 times lower than corresponding values at Woods Pond Outlet, which makes sense given the 20 mile separation.

GE's plot of simulated versus measured PCBs at Rising Pond Dam shows very little correlation between simulation and measurement ($r^2 = 0.01$). ***At this point it is hard to say the downstream model is validated***.

Question 4 Have the sensitivities in the models to the parameterizations of the significant state and process variables been adequately characterized?

Most sensitivities seem reasonable. As expected, flows show strong sensitivity to upstream boundary flows, TSS is sensitive to upstream boundary flow and TSS, and PCBs are sensitive to upstream boundary flow, TSS and PCB concentration.

At low flow PCB there is strong sensitivity to K_f and partitioning (at least at New Lenox Rd), which makes sense, since "diffusion" is the only important process, but this effect is largely diluted out in the 10.5 year simulation. At high flow, there is strong sensitivity to parameters dealing with particulate phase transport (settling, erosion).

While the model shows substantial sensitivity to diffusion parameters during low flow at New Lenox Rd, GE points out that virtually no sensitivity is indicated at Woods Pond Footbridge. Since the flow must pass from NLR to WPF, this is illogical. Unless there is a mistake in plotting (and EPA suggested there was not), this could indicate a potential problem in model formulation.

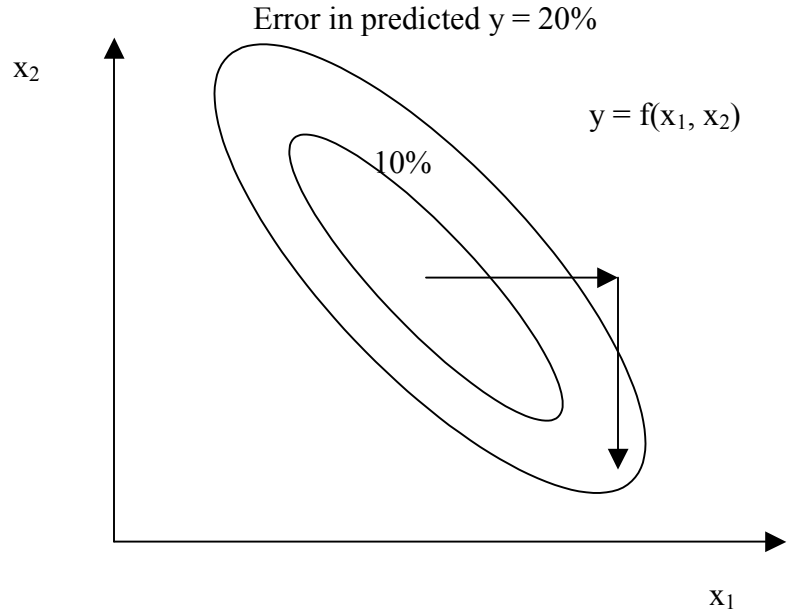
Question 5 Are the uncertainties in the model output(s) acknowledged and described?

In general, model uncertainty comes from two sources: imperfection in the basic model(s) (including unknown unknowns) and imperfection in the selection of the (known) model parameters. The former has been discussed to a certain degree under Question 1, which relates to processes, so the following relates mainly to the latter issue.

Although they use some sophisticated approaches, the modelers performed a rather traditional uncertainty analysis, varying known model parameters largely ***independently*** in a simulation of ***existing conditions***. If the model and data were both perfect, and the only uncertainty involved the values of model parameters, one could expect most of the parameters to be independent. (Some parameters that have a physical relationship, such as erosion parameters would still be correlated.) But the model has been calibrated to

observed data, and the calibrated model parameters reflect both imperfection in the model (e.g., a parameter being used to cover more than one real process) and imperfection in the data (e.g., due to sampling or analytical error, insufficient sample resolution, etc). Thus, during uncertainty analysis, when one parameter is changed, other parameters should also change (i.e., they are not independent), as is illustrated in the following sketch. Here the dependent variable y is a function of two independent variables x_1 and x_2 and the contours are plots of some measure of error between measured and predicted values of y. If, as part of an uncertainty analysis, variable x_1 were changed (horizontal arrow), one would expect a correlated change in x_2 (vertical arrow) in order to minimize the error between measurements and predictions. Failure to take this correlation into account tends to overestimate the uncertainties in the model prediction. And, as suggested in the following paragraph, I suspect this issue becomes more extreme when the future conditions being predicted are substantially different from those under which the model was calibrated.

Apart from the question of parameter independence, there is the question of whether there are certain variables to which the model is more or less sensitive as a function of the remediation scenario. For example, following remediation there may be thin patches of clean/dirty sediment overlaying thicker regions of dirty/clean sediment, making calculations more sensitive to diffusion and partitioning than they were in the base case scenario. If this is the case, we may get a distorted impression of the effectiveness of remediation.



Based on sensitivity analysis, EPA tested the uncertainty of model outputs (e.g., PCB concentrations in water, bed and fish) to uncertainties in model input parameters. As GE points out, the model uncertainty is substantially greater than the uncertainty in the model

(e.g., the s.e.m.) If the goal were simply to predict the uncertainty in present conditions, the uncertainty in the individual parameters could be scaled back until the uncertainties in the model and data matched. But it is also important to see the ***sensitivity of predicted model changes due to possible remedial alternatives*** to these same model input parameters. This is not difficult and should be performed either before or after the model is transferred to GE.

Finally, one use of the model is simply to compute relative fluxes, as EPA as demonstrated recently. Remedial decisions could be made on the basis of how much less PCB would be transported, rather than on (or in addition to) the basis of quantitative predictions of concentrations. Thus it would also be nice to see the ***sensitivity of predicted fluxes to the same model parameters***.

In summary, because of the lack of parameter correlation in the sensitivity study, and the fact that predicted uncertainty exceeds the uncertainty in the existing data, I believe the predicted uncertainty, when applied to future remedial conditions, will be too large.

Question 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?

I appreciate the inclusion of the two example simulations that shed some light on how the model might perform when simulating remediation options. This is clearly a start. There would seem to be two issues to consider. The first is how well the remediation measures themselves will work. For instance, will a cap erode? The second is what will happen downstream (in space and time) after remediation? For example, will dirty sediments cover up capped or dredged sediments? The example simulations don't consider the first question (though it is very important), but partially address the second. However, several considerations come to mind. First, the example simulations were begun with "cold start" initial sediment conditions, which seems to explain, at least in part, why sediment concentrations increased in some areas downstream from simulated remediation (i.e., places where there was no PCB source). Better that the model be spun up to initial sediment conditions that reflect an equilibrium with the modeled hydrology and bathymetry. Also, the model seems to ignore any particulate PCBs that might have wanted to deposit in the channel of Reach 5A. (What happens to these in the model?) Ultimately, the model could be generalized to look at "different colored" PCBs emanating from different source locations and to see where they end up, with implications as to which would be the best sediments to remediate. This might require subscripting certain variables in the code so the calculations could be run in parallel. EPA has started to do some of this, but more would be instructive.

As I expressed at the DRM, I am concerned with the absence of validation data appropriate to remedial measures. Dredging would result in horizontal gradients in sediment concentration over 10s to 100s of meters, while capping would result in vertical gradients in sediment concentrations over a few cm. The current sediment concentration

distributions do not show variability over these scales, so it is difficult to tell whether or not the model will be able to successfully predict the effects of such remedial measures.

Figure 4.2-53 shows areas of net deposition indicating ~ 42% of net deposition is to floodplain, and ~ 10% is to Woods Pond. The remaining ~48% is to the sediment bed upstream of Woods Pond. It would be nice to keep track of the deposited sediments. My guess is that those deposited to the flood plain and to Woods Pond stay put, or are slowly released, while those in the sediment bed move on. Can this be quantified?

Question 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 measures?

Comments focus on the related questions of grid resolution and computational feasibility. Recommendations follow at the end.

Like most other review panelists, I continue to be ***concerned about the lack of lateral resolution in the computational grid***. Having only about one grid cell over the river width means that important processes must be parameterized. Because the existence and magnitude of erosion and deposition depend non-linearly on stream velocity, a trapezoidal section with depth varying linearly between 1 and 3 meters at the two banks will look to the model like a section with uniform depth of 2 m, yet in practice the former may have regions of strong erosion and deposition (as in re-surveyed cross-section XS061 shown in the DRM handouts), whereas the latter might be marginal one way or the other. Even if the model gets the net erosion/deposition correct (doubtful due to the non-linearity), the failure to capture gross erosion/deposition is problematic, both for evaluating remedial measures (such as capping) and natural attenuation. Assume an entire channel cross-section is capped. If half of the channel is erosional and the other half is depositional, the first half will require added protection to keep the cap intact, while the second half will get covered with contaminated sediments from upstream. Neither process would be forecast with a model that predicts marginal net erosion/deposition. As for natural attenuation, contaminated sediments from an upstream area with gross erosion can contaminate (or re-contaminate) downstream areas, but the model would not predict this.

The issue of model resolution conflicts directly with the issue of computational time. Clearly, with the current coupled in-channel/overbank modeling system, one cannot afford the extra computational cost of reducing Δy, and still be able to afford multiple simulations of multiple scenarios each over multiple decades. I share GE's concern that the existing model is already too cumbersome to be used effectively to study corrective measures. Indeed, this is probably the biggest model issue: as Yogi Berra might say, "If you can't use the model, you can't use the model." The following are some thoughts on the related issues of grid resolution and computational cost.

Grid resolution
Too begin with, it must be recognized that the lateral grid size is a model parameter as well as a numerical parameter. Ideally, when a set of equations are solved using a discretized numerical model, one likes to reduce the grid size until results converge. ***But, we are no where near that here.*** Hence if the model grid size were to be reduced, this would affect other model processes (notably erosion, which depends heavily on velocity, which would vary significantly across the river width). Thus ***if the model grid were to be reduced, the entire calibration would need to be redone***.

It is not clear how much resolution would be needed to properly compute transport. Because of this uncertainty, and the time and effort involved, it may be too late in the game to make major changes to the model grid. However, at a minimum, some sensitivity tests could be conducted using a finer grid over a small portion of the (in-channel) domain and run over a short duration of the 26 year simulation period. The output from the more highly resolved model could be compared with that from the coarser model and then parameterized. For example, it might be determined that the net erosional flux with a resolved grid is X times that computed with the coarse grid, in which case predicted erosional fluxes with the coarse grid would be multiplied by X.

Computational costs
A number of options have been mentioned by the Peer Review Panel and by GE to improve model efficiency. These included splitting the calculation of hydrodynamics, sediment transport and PCB fate; running synthetic hydrological sequences; and employing alternate grid schemes. EPA did implement some efficiency measures (dynamic and split time steps, by-passing sections of the model that change slowly and running simulations of selected portions of the calibration period), but seemed to have dismissed the bigger ticket items. I think some of these have to be revisited.

There is no reason, in principle, why the model calculation of hydrodynamics, sediment transport and PCB transport/fate could not be de-coupled. In particular, the hydrodynamics could be run first. These results could be stored and used to transport sediment and PCBs. Since the hydrodynamics would only need to be computed once (hydrology does not depend significantly on scenario), considerable saving could be obtained. The devil is in the details of course, and this would probably take a few months of time. But since EPA is a model developer, and such a decoupled model would have future applications, EPA may want to invest in such a project, possibly calling in the model developer (J. Hamrick).

Further savings could be obtained ***using synthetic hydrologic flows***. Based on the relatively large data base, a suitable discrete distribution could be generated that includes representative flows of different magnitudes and recurrence intervals. The model (with or without the hydrodynamics disaggregated from transport) could then be run for a synthetic year (or a short period of years large enough to include the largest/least frequent flow of interest), and the changes per year documented. Perhaps, without rerunning the model, ***the long term effects could be computed by extrapolation of the one year results***.

Also, it still seems to me that in channel and over bank calculations could be de-coupled, saving additional computational time, and/or allowing more detail during the vast majority of time when the flow is within the banks and essentially one-dimensional. EPA claims there are coupling issues such as loss of momentum conservation, which may be true. But exact coupling is not critical. We are only expecting the model to predict gross trends in PCB concentrations (over time and the longitudinal direction), so a temporary mis-match might be acceptable. In general, I would support more a somewhat more approximate model (e.g., with parameterizations) that could be more efficiently used.

In this regard, I like M. Garcia's idea of a flood plain number to quantify deposition of PCBs within portions of the floodplain. PCBs appear to be on a conveyor belt whereby they are resuspended from within the channel at high flow, partly deposited on the banks, then gradually eroded from the bank back into the river, until some of them are deposited in Woods Pond. These processes are all in the model, but it might be nice to document the life history of a "numerically marked" cohort of sediment, and to test the sensitivity of the ultimate fate of this cohort to processes such as the bank erosion rate, frequency of over bank flows, etc. By decoupling the sediment transport and fate, this could lead to a model simplification, whereby the over bank processes (deposition and bank erosion) are parameterized as simple first order sinks and sources whose rates are determined by calibration to detailed simulation over the calibration period. These rates could vary with flow conditions, but would be independent of PCB concentration, allowing simple calculations during the Corrective Measures Study.

In summary, it is a shame that the longitudinal and transverse grid sizes (Δx and Δy) must be similar since the former need not be nearly so small (fish average their exposure over large distances) and the latter should be smaller. In my previous comments I suggested eliminating Δx as a dependent variable (substituting, instead, grain size by reach), but I recognize this was a radical idea that might take too much time to implement.

Finally, while the above discussion refers to lateral resolution (and its relationship to computational cost), I am also concerned that the vertical resolution within the sediment bed that may be insufficient to resolve near bed gradients that might result from remedial efforts (see Question 1). Clearly, calculations in the sediment bed can use a different (much longer) time step than those in the water column. It is not clear how costly additional vertical resolution in the sediment bed would be, but I think it should be considered.

Recommendations

Considering the difficulty of the modeling task, the current state-of-the-art, the amount of effort that has already gone into the modeling, and the difficulty in arriving at a consensus (when our review panel is not allowed to), I would say that ***the current work is "acceptable"***. However, in view of the many reservations that many of us have, the model users (GE) will have to exercise considerable professional judgment in their use of the model, and EPA should grant them this judgment. And depending on how much time is available, ***I would recommend some modest additional effort***, which I organize into three categories: easy, intermediate and difficult. I believe that GE should be involved

with this as much as possible (rather than simply be given a black box), because it is their model judgment that will be relied upon, and this judgment can only come with experience using the model.

Easy

These are things that should be easy to accomplish within a framework of a few weeks to a couple of months. They are probably also the most important.

1. Develop a strategy to improve model efficiency so that remediation scenarios can be evaluated efficiently. I suspect that the best approach here is to make use of synthetic hydrological sequences that are repeated.
2. Perform additional uncertainty analysis that looks at the uncertainty of future remediation scenarios (as opposed to simply past natural attenuation), both in terms of absolute concentrations and in terms of PCB fluxes from different sources.
3. Improve the "upstream" and "downstream" models through further calibration (GE indicates that additional data are available for this purpose).

Intermediate

The time frame here might be several months to half-a-year.

4. Develop a more highly resolved grid within at least a portion of the channel and use the grid in a short term simulation to understand the sensitivity of model predictions (again, actual concentrations as well as fluxes) to grid resolution. Information parameterized from this sensitivity test could be used to adjust the coarse grid output, as suggested above.
5. Decouple the model of the channel and the floodplain (to improve model efficiency).
6. Decouple the hydrodynamics from the sediment transport and PCB fate within EFDC (to improve model efficiency).
7. Re-calibrate the model based on suggestions made by reviewers under Question 1, and in consideration of the several instances of bias noted under Question 3. (I don't believe a separate validation is required.)

Difficult

This step would lead to the greatest model accuracy and robustness, but might take a year or so to accomplish.

8. Restructure the grid to allow more lateral resolution. This would have to include computational efficiency measures so that extended model runs are feasible, as well as a complete model re-calibration.

References

Chen, H.-W. (1993). Fluxes of organic pollutants from the sediments of Boston Harbor. S.M. Thesis, Dept of Civil and Environmental Engineering, MIT, Cambridge, MA.

Jorgenson, B.B. and D.J. des Marais (1990). The diffusive boundary layer of sediments" oxygen microgradients over a microbial mat. *Limnol. Oceanogr*. 35(6): 1343-1355.

Shaw, D.A. and T.J. Hanratty. (1977). Influence of Schmidt number on the fluctuation of turbulent mass transfer to a wall. *A.I.Ch.E.J*. 23, 28 pp.

Wu, S.C. and P.M. Gschwend. (1986). Sorption kinetics of hydrophobic organic compounds to natural sediments and soils. *Environ. Sci. Technol*. 20: 717-725.

Wu, S.C. and P.M. Gschwend (1988). Numerical modeling of sorption kinetics of organic compounds to soil and sediment particles. *Water Resources Research* 24(8): 1373-1383.