Response to Charge Questions

Housatonic River

Frank Gobas, Ph.D., Professor

School of Resource & Environmental Management

Simon Fraser University

June 3, 2005

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

Tables 2.5 and 2.6 demonstrate that *mean* water flow rates are very well predicted over time and space by the **HSPF** model. Despite the good agreement of model predicted and observed *mean* flow rates, daily and monthly scatter plots at Coltsville illustrate that there are considerable variations around the *mean* flow rates. The variability around the mean is almost one order of magnitude. This variability around the mean value is not represented in the measures used to characterize the quality of model calibration. I recommend that they are added. They it should be considered when applying the model under scenarios where temporal variations as well as maximum and minimum flow rates are important. The *mean* flow estimates of the HSPF model are likely sufficient for addressing the most important management questions such as the response time of contaminant concentrations following remediation options.

Table 2-12 and 2-13 illustrate that predicted and observed TSS loading rates are also in good agreement both on a spatial and temporal scale. In comparison to the water flow rates, the TSS model predictions show larger discrepancies between observed and predicted values. Differences of up to 139% are reported. However, the average difference is approximately 10%. The comparison of model predictions and empirical data appears to be sufficient to make estimates of *mean* TSS loads under normal conditions. But again, I recommend that additional detail is provided in the report to better represent the capability of the model to make predictions on spatial and temporal scales.

Tables 2-15 and 2-16 show that differences between observed and predicted water temperatures are very small. These differences are essentially insignificant and the model's capability to predict temperature is very good.

With regards to the **hydrodynamic** model, empirical observations are to a large degree internalized in the model. The comparison of observed and predicted is therefore not an independent test of the capability of the model. The model was tested for two extreme events and showed good results. This is promising, but it is not sufficient to conclude that the hydrodynamic model has the capability to predict the hydrodynamics at the relevant spatial and temporal scales. In my view, it is premature to comment on the capability of the model at this point. The real capability of the model will be revealed in the model validation phase, which will provide a relatively independent test of the capability of the model.

There is a reasonable data base available to test the **sediment transport model** at two locations (i.e. New Lennox Road and Woods Pond Outlet). Data from other locations (i.e. Holmes Road and Woods Pond Headwater) exist but the sample size is not large. Figures 4-34 and 4.35 illustrate that the sediment transport component of the EFDC model has reasonable central tendency characteristics. However, there are also significant discrepancies between observed and predicted data. Differences between measured and simulated TSS data show that predicted TSS produce a narrower range of concentrations of TSS concentrations than observed. Also, there appears to be a considerable variability in the measured TSS data at New Lennox Road and Woods Pond Outlet that is not explained by the model. In terms of assessing the spatial capabilities of the model, it would be beneficial to have access to more data for model-data comparison but the currently available data sets can be considered adequate as long as the magnitude of the uncertainties are recognized by the model and considered when remedial options. I think that the reported analyses can be improved upon by explicitly recognizing the variability among the individual data/prediction comparisons.

The capabilities of the sediment transport model on a temporal scale are tested over a 14 month period starting in May 1999. A reasonable number of data is available for model calibration. However, Figure 4-25 shows that significant discrepancies exist between observed and predicted TSS concentrations. Therefore, some doubts remain with regards to the temporal capability of the model. At this point it is unclear whether this quality of agreement between observations and predictions is due to systematic errors in the modeling approach or reflects statistical variability or uncertainty in TSS concentrations. In my view, much value will be added to the modeling effort if in addition to the central tendencies of the model, variability and uncertainty are recognized and explicitly stated.

Comparisons of predictions of the sediment transport model and empirical data for several storm events are also presented. The agreement between measured and simulated TSS and flow data are with some exceptions are quite reasonable.

Regarding **EFDC**, Figure 5.11 shows the comparison between predicted and observed concentrations of PCBs in pore water. The comparison is quite good, suggesting that the assumption of equilibrium between PCB concentrations in sediments and pore water is justified. Figures 5.17 to 5.19 show that the comparison between predicted and observed PCB concentrations in the water column over 14 months at 3 locations. Figures 5-20, and 5.21 illustrate the comparison of observed and predicted PCB concentrations in the water column at various locations in the River. Figure 5.23 illustrates the comparison of PCB concentrations in water column after a storm event. The agreement of the model with the data appears quite reasonable. This is in some contrast to the results depicted in Figure 5-30, which illustrates a reasonable central tendency of the model in predicting PCB concentrations in the water column, but also considerable discrepancies between model predictions and observations. Additional attention could be devoted to the comparison of observed and predicted concentrations of PCBs in bottom sediments of the river. These data are likely to be very useful in assessing the fate of PCBs in the River.

In terms of the adequacy of the model-data comparisons to evaluate the capability of the EFDC model on the relevant spatial and temporal scales, there appears to be a reasonable amount of data available to evaluate the capability of the model to assess water column transport. The capability of the model to assess some other key aspects of the fate of

PCBs, such as long term response times of the PCB concentrations in the River, is not convincingly demonstrated in my view.

As for the **bioaccumulation** model, comparisons of model predictions and empirical data on a spatial scale are shown in Figures 6.5 to 6.16. There are additional comparisons presented in Figures 2-34 of attachment C15. The concentrations of tPCBs in sediments and suspended solids show small differences among the reaches 5A to 6. Hence, the model calculations of the tPCBs in biota do not show a strong spatial dependence. As a result, the capability of the model to make spatially explicit estimates could not be fully explored in this study. However, this is not of great importance for the development of the bioaccumulation model as the spatial (and also temporal) differences in concentrations are predominantly determined by other components of the model.

The temporal capability of the bioaccumulation model is tested in terms of the relationship of the PCB concentration in fish species with age. Other temporal effects (e.g. summer vs. winter) are not explored.

Overall, though, there appears to be a good PCB concentration data set available to assess the capability of the model. There is a lack of fish tissue concentration data for Reach 5D. However, I do not think that this should preclude the calculation of fish concentration data for fish in this reach. One of the goals of the model is to make estimates where concentration data are not available.

2. **Is there evidence of bias in the model as indicated by the distribution of residuals as a function of the independent variables.**

Figure 3-18 shows no bias in the distribution of residual flows as a function of the measured flow rates. Figure 4.36 also shows no significant systematic bias for the calculations of TSS by the HSPF model although some considerable variation between observed and predicted TSS values was found in some cases.

Estimates in the tPCB water column concentrations show no significant systematic bias at Holmes Road and Woods Pond Headwaters but some bias is apparent from the distribution of the residuals for data collected at New Lennox Road and Woods Pond Outlet (Figure 5-31).

Figures C3-26, C3-27 and C3-29 9and C3-49, C3-50 and C3-52 (for the linked model) plot the residuals against the measured PCB concentrations. The plots do not lend themselves to explore issues of bias. A statistical treatment of the data would be more useful. Hence, it is difficult to confirm the statement on p. C.3-31 that there is no model bias across the range of PCB concentrations evaluated. Just looking at Fig C3-26, it looks as if there are more data points below the zero line than above it. Figure C.3-27 appears to confirm this for coplanar PCBs. But again, this may not be so.

There are various ways to explore the issue of bias for PCBs on a congener or total-PCB basis. We have used the model bias MB, which is derived on a species-specific basis as:

$$MB_j = 10^{\left(\sum_{i=1}^{n} \frac{\left[\log\left(C_{P,i} / C_{O,i} \right) \right]}{n} \right)}$$

In essence, MB_j is the geometric mean (assuming a log-normal distribution of the ratio $C_{P,i} / C_{O,i}$) of the ratio of predicted ($C_{P,i}$) and observed ($C_{O,i}$) for all PCB congeners i in a particular species j included in the analysis. MB is a measure of the systematic over- (MB>1) or under-prediction (MB<1) of the model. It should be stressed that in the calculation of MB, over- and under-estimations of the observed concentration values for individual PCB congeners have a tendency to cancel out. Hence, MB tracks the central tendency of the ability of the model to predict PCB congener concentrations. It is a useful measure of model performance if total PCBs (SPCB) are of primary interest. The variability of over- and under-estimation of measured values can be represented by the 95% confidence interval of MB (i.e. 95% CI = antilog(geometric mean ± ($t_{\nu,\,0.05} \times$ standard deviation)). The 95% confidence interval represents the range of concentrations that includes 95% of the observed concentrations. It can be viewed as a measure of the uncertainty of the model predictions. The same approach can also be applied to total PCB as well. In that case, model bias MB* is:

$$MB_j^* = 10^{\left(\sum_{i=1}^{m} \frac{\left[\log\left(C_{P,\,\Sigma PCB} / C_{O,\,\Sigma PCB} \right) \right]}{m} \right)}$$

MB_j^* is the geometric mean (assuming a log-normal distribution of the ratio $C_{P,\,\Sigma PCB} / C_{O,\,\Sigma PCB}$) of the ratio of predicted and observed concentrations for ΣPCB in species j (Arnot and Gobas, Environ. Toxicol. Chem. 23, 2343-2355 (2004)).

I do agree with the authors that the vast majority of the observed concentrations are within an order of magnitude of the mean simulated by the FCM. However, it is not clear why Figure C.3-29 only suggests a range of an order of 2 rather than 10 (in Fig C.3-26). The latter may be due to the fact that Fig C.3-26 refers to the combined data set while Fig C.3-29 refers to means of subsets of samples. (This needs to be clarified in the Figure legend).

While the residuals may not show a bias, Fig C.3-27 illustrates that the uncertainty in the model predictions can be large. It is possible that this uncertainty is to some degree caused by difficulties in modeling the bioaccumulation behavior of certain congeners. Metabolism can be a significant process for some congeners and not for others. It is therefore instructive to explore the issue of bias on a congener specific basis rather than combining many congeners in one analysis. I recommend this is done. Also, I recommend that, in addition to the mean model bias, the uncertainty around the mean model bias is explicitly stated. A high degree of uncertainty of the model calculations should not be viewed as criticism of the model but a reflection of the actual state of the modeling capability.

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

With regards to the EFDC model, the model contains several processes controlling the fate of PCBs. The key processes that are included in the model are sediment-water diffusive exchange, solids settling & resuspension and flow. There are some key processes that are acknowledged in the model architecture (Figure 5-1) but which are not fully considered in the application of the model.

For example, degradation in the sediment is not considered in the model. The authors state that the rate of dechlorination is too small to be significant. They base this conclusion on the lack of a change in CL:BP ratios between originally discharged and current Aroclor 1260 in Reach 5A. However, this ratio does decrease in the lower portion of Reach 5A (Figure 5.15), hence suggesting dechlorination. To shed more light on this issue it is beneficial to explore changes in PCB composition over time on a congener specific basis. Although dechlorination may be a slow process, it can have a significant effect on the overall fate of PCBs in the Housatonic River if other loss rates of PCBs from the River are also slow. The latter appears to be case since net loss of PCBs from the 5A to 6 reaches is very small when expressed as a fraction of the mass of PCBs present in the River. This means that the river's response time to changes in PCB loadings is very long, i.e. it takes a long time for sediment concentrations to respond to a new loading regime. In slowly responding systems, slow processes can have an impact on the overall response time of the PCB concentration in the system and even be rate limiting. In that light I recommend that the authors include the degradation rates of PCBs in the model, preferably on a congener specific basis.

I also recommend that the volatilization of PCBs to the atmosphere is considered in more detail as it may be a significant loss rate for PCBs in the River. While volatilization from the river may be small due to small surface area, this surface area is significantly increased during flooding events when particulate material and water are distributed over large areas of floodplain. When flooding subsides, these particulate materials will be in contact with air for considerable times and PCBs may volatilize.

The decay of organic matter in suspended sediments is recognized in Figure 5, but it is not clear whether it is actually considered in the application of the model. In the Housatonic River, where PCBs have been associated in sediments for a long time, it is possible that as a result of the relatively rapid decay of organic matter compared to a slow desorption rate of PCBs to the water, organic carbon normalized PCB concentrations increase relative to the water concentration, causing a suspended sediment-water disequilibrium that affects the PCB concentration in the water available for respiratory uptake in biota. Evidence of this process has been observed in suspended and bottom sediments in some other system (Environ. Sci. Technol. 37(4): 735-741).

In summary, the model is focused on the description of sediment dynamics and sediment-water partitioning of PCBs but does not fully explore several other fate controlling processes that, considering the slow temporal response of the system, may have a significant effect on the outcome of the model.

Another significant limitation of the EFDC model is its inability to model PCB congeners. The representation of PCBs with average properties (e.g. Kow) can produce a significant error in the calculations of PCB concentrations. The modeling of PCBs in terms of total PCBs has merits but it is not a state of the art modeling methodology. The tPCB modeling becomes a limitation when the model results of the EFDC model are transferred to the FCM model and used to assess ecotoxicological effects. With regards to assessing the ecotoxicological effects of PCBs, the current practice relies on assessing risks of effects based on congener concentrations. I strongly recommend that the EFDC model conducts congener specific calculations that can take advantage of available congener specific physical-chemical and biological data.

The bioaccumulation model contains the key processes controlling the uptake and elimination of PCBs in fish and invertebrates. Uptake from water and diet are included along with elimination to water and other excretion processes and growth dilution. The model also includes a de facto mechanism for biomagnification, apparently through a resistance factor C_R that applies to the gill elimination rate but not to the gill uptake rate. The fact that the resistance does not apply to both uptake and elimination for this reversible process is not correct in my view. However, this practice produces a biomagnification effect that is not explicitly included in the model. The resulting model with the resistance factor can be expected to work well as it appears to do. There are some processes that could be included such as egg and sperm deposition for spawning fish. However, I do not recommend this. The model is calibrated to quite a significant extent and adding further parameters that are included in the calibration recipe makes the model less transparent while any improvements in predictability are unlikely to be significant.

A gap, perhaps in the reporting only, concerns the model for accumulation in aquatic macrophytes and algae. A simple lipid-water type partitioning model is unlikely to be successful in describing the bioaccumulation of PCBs in algae & macrophytes. Adding this component to the model may not have a big effect on the model outcome given the apparently strong linkage of the food-web to the sediment. However, it is important to ensure that the reporting of the modeling approach is complete.

While not a process, I question the wisdom of not including some other target species in the model such as muskrat, waterfowl and raptors. These organisms are susceptible to high concentrations of PCBs due to bioaccumulation and "dose-response" relationships exist for risk analysis purposes. This may have been addressed in an earlier bounding exercise.

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

The methodologies used for the sensitivity analysis of the HSFP, EFDC and the Bioaccumulation models were carried out in an appropriate fashion. However, the sensitivity analyses would be more insightful if it they would focus on the key issues that the model needs to address. For example, the temporal response of PCB concentrations in water, sediment and biota to remediation efforts (and associated PCB loading reductions) is a key objective of the model. It would be helpful if the sensitivity analysis could report on the effect of various model parameters on the temporal response of the PCB concentrations in the area of concern. This is particularly important for the model parameters dealing with the sediment mixing (i.e. depth of active sediment layer, bed sediment mixing, resuspension and sedimentation) which have the largest effect on the time response of the PCB concentrations. However, it also important for other model parameters such as flow rates and lipid contents of fish. I therefore recommend that the sensitivity analysis is further developed. The current analysis provides useful information about which parameters are the most sensitive. The second phase of the sensitivity analysis can focus on these parameters and address how they affect key characteristics of the model, such as the temporal response of PCB concentrations in water, sediment and biota as a result of remediation options.

The methodology used to conduct the uncertainty analysis contains some significant limitations throughout the entire modeling effort. There are several issues. First, the basic modeling strategy relies heavily on model calibration. In model calibration, the observed data are used to parameterize the model. This produces a model where observed data cannot be used as an independent data set to test the performance of the model. Comparing model predictions against independent data is probably one of the best and simplest ways to assess the uncertainty of the model. However, this method cannot be used to its fullest advantage in the model due to the reliance of calibration to make the models work. Even without access to independent data (e.g. PCB concentration data in sediments and biota), there is still considerable merit to using differences between observed and simulated data as a measure of model uncertainty. I recommend that this is added to the modeling strategy given the limitations of the Monte Carlo Simulation technique that is the main method used to assess model uncertainty in this study. There are various statistical methods to do this such as mean squared error or calculating confidence limits of the model bias discussed above. The resulting uncertainty calculated should be treated with some caution as the uncertainty has a tendency to underestimate the actual uncertainty.

The second issue relates to the application of Monte Carlo simulations, which was conducted in the bioaccumulation model. The application of Monte Carlo simulations to complex models like EFDC and the bioaccumulation model is difficult. There are two conditions that need to be met for the Monte Carlo simulations to be informative. One is that the model variables included in the Monte Carlo simulations are independent and not correlated. This was done for the bioaccumulation model but the report does not provide details on how this was done. This issue could therefore be expanded and perhaps improved upon in future work. A second condition for an informative Monte Carlo simulation analysis is that the variability and error in the model variables can be determined or are known. For some model variables this can be done relatively easily, while for others (e.g. feeding preferences, growth rates) this is very difficult. The report does provide information on this issue and the authors are doing a good job to deal with this difficult issue

An issue that requires further investigation, in my view, is why the MCS calculated uncertainty in the concentrations of PCBs in biota is considerably less than the observed uncertainty. The latter is indeed not impossible as the MCS method does not capture all sources of uncertainty. However, it raises issues about the value of the uncertainty analysis and how the results of the MC simulations should be interpreted when applying the model.

In terms of uncertainty analysis for the EFDC and FCM models, there is considerable room for improvement and additional work. For example, there is no uncertainty analysis for the EFDC model at this point. As for the FCM model, the MC simulations provide useful information but the report makes qualitative statements (i.e. “two-fold” in an number of places e.g. p.C.4-24), which do not appear to be representative of the real model uncertainty as demonstrated by differences between model predicted and empirical PCB concentrations in biota. My recommendation is:

1. To include an assessment of model uncertainty based on a comparison of observed and predicted concentrations (The 95% confidence intervals of the model bias, discussed earlier, can be a useful tool to do this).

2. To conduct the planned MC analyses considering the importance of conducting the analyses with non-correlated state variables and supporting the distributions of state variables with scientific data or appropriate and documented judgment.

Finally, it is important to stress that both approaches have their pros and cons and that they only arrive at *estimates* of model uncertainty. In the application of the model this should be recognized.

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

As for the HSPF model, the model-data differences are sufficiently small to use the model to predict mean flow, TSS and temperature in the River following remedial actions.

On balance, the hydrodynamic model and sediment transport model appear to produce relatively small differences between model predictions and observations. The small differences between observations and predictions are partly caused by the calibration methodology which uses the observed data to make the model predictions. Hence, a good agreement between observations and predictions should be expected. It is unclear from the study sofar how predictive the model really is and hence what the model's uncertainty is. This can de determined by conducting a model validation (better is model performance evaluation), where the observed data are not used to make the model predictions. However, despite some limitations in the approach sofar, the model is a reasonable tool to start making certain predictions among remedial options.

The performance of the PCB fate model is only tested in its ability to estimate PCB mass in the water column. While the performance of the model as characterized by differences between observed and predicted concentrations are reasonable they do not shed much light on the ability of the model to estimate spatial differences in PCB concentrations (as concentrations of PCBs do not appear to show statistical differences among the stretches of the River of concern), or the model's ability to estimate the temporal response of the PCB concentrations in water and sediments in the River. If it is further considered that the model may not have fully represented some key fate processes, I recommend considerable caution in the application of the model in a predictive sense, in particular if the long term temporal response of the model is important. I think uncertainty analyses need to be added.

Model – data differences in PCB concentrations in the bioaccumulation model are considerable despite significant calibration efforts. Biological data often exhibit a large degree of variability. Hence, it is not uncommon in bioaccumulation modeling efforts that there are significant discrepancies between predicted and observed concentrations. The latter should not be viewed as criticism of the model or an impediment in the application of the model. As long as the uncertainty in the model calculations is appropriately recognized, the results of the model can be interpreted accordingly and the model can be used productively to assess the impact of remedial actions. The issue of uncertainty requires further attention in the development of the bioaccumulation model. Currently, the report contains statements about "*the majority of PCB tissue concentrations being within a factor of 2 of the deterministic values*". These statements does not appear to be representative of the real model-data differences shown in Figure C3-27 or even Figure C3-28 (which I presume are mean concentration values). Only, when uncertainty is appropriately recognized, application of the model should be considered.

Although I do not think that the uncertainty of the FCM model is correctly represented in the report, I do think that when this is done, the model can be used to predict differences in PCB concentrations in biota resulting from remedial options despite the fact that differences in observed and predicted concentrations are considerable. It is possible that, in some cases, there may not be statistically significant differences in PCB concentrations in fish resulting from different remediation scenarios but, if so, this is important information to know.

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

The report documents that model calibration has been carried out to a significant degree in the hydrodynamic, sediment, chemical fate and bioaccumulation models. In my personal view, the model development has embraced calibration a little too strongly at the expense of evaluating model performance and model uncertainty. The calibration of the model has produced a model that has reasonable central tendencies and produce reasonable values for mean conditions such as flow rates, TSS and PCB concentration on TSS and in biota. However, the uncertainties in the model predictions require further attention before the model can be used productively to explore remedial options. In terms of additional work, it is possible to collect new PCB concentration data sets to carry out a model performance analysis that is not dependent on the collected data. Alternatively, it may be possible to revisit existing data sets and calibrate the model to certain data while using other available data for model performance analysis and uncertainty analysis. A more daring approach is not to use PCB concentration data at all in the model calibration phases. This should be possible for the PCB fate and bioaccumulation model.

One area where the model calibration is lacking is in the temporal behavior of the PCB concentrations in sediment and biota in the River. This characteristic could not be calibrated very well because PCB concentrations did not show significant changes over time during the study period. As a result, there is little information on the performance of the model in terms of predicting future PCB concentrations in response to remedial options. There is not a simple solution to this problem. One approach that could be pursued is to better characterize some key loss processes of PCBs in the River. This would involve characterizing PCB degradation rates and volatilization rates rates. These rates may have a significant effect on the temporal response of PCB concentrations in the

River. Although this work would not actually test the temporal response of the model, the credibility of the model would be improved by a better presentation of mechanisms of chemical loss.

Despite the large amount of effort that has been devoted to modeling and data collection, I am not convinced that, at this point, a holistic understanding of the fate of PCBs in the River has emerged. The report is unclear about what the key processes are controlling the fate of PCBs in the River. One, in my view, useful approach is to add PCB flux diagrams to the report. Flux diagram are a useful and simple tool to integrate a lot of information with the goal to determine the controlling processes in the River.