Response to Charge Questions

Housatonic River

Frank Gobas, Ph.D., Professor

School of Resource & Environmental Management

Simon Fraser University

April 22, 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.

As for the HSPF model, Tables 2.5 and 2.6 demonstrate that mean water flow rates are very well predicted over time and space and that as a result measured and predicted mean flow rates are in very good agreement. Despite the good agreement of model predicted and observed mean flow rates, daily and monthly scatter plots at Coltsville (p.17, handout) illustrate that there are considerable variations around the mean flow data. 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. However, 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 modeling the impact of remediation options. The comparison of model predictions and empirical data appears to be sufficient to make estimates of mean flows under normal (not unusual) conditions.

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. 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.

Tables 2-15 and 2-16 show that differences between observed and predicted water temperatures are very small. These differences are insignificant and temperature is not a parameter of concern in the modeling.

With regards to the HSPF model, the model predictions with empirical data are sufficient to evaluate the capability of the model on the relevant spatial and temporal scales.

With regards to the hydrodynamic model, empirical concentrations are to a large degree internalized in the model and a comparison of model predicted and observed flow rates does not appear to be an adequate test of the spatial and temporal performance of the model. The model was tested for two extreme events and showed promising results, but this test cannot be considered sufficient to evaluate the capability of the model at this point. The latter is likely to be accomplished in the model validation phase, which has not been completed yet.

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 and considered acceptable for modeling remedial options.

The capabilities of the sediment transport model on a temporal scale are tested over a 2 year period starting in May 1999. A reasonable number of data is available for model calibration. However, the agreement between simulated TSS and observed values is not strong. In my view, the availability of data and the resulting comparisons of the model predictions with empirical data may therefore not be sufficient to demonstrate the temporal capability of the model. Perhaps, better insights into the temporal capability of the model are obtained when additional data are collected in a model validation phase of the study.

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. The results indicate that the available data and models comparisons are sufficient to evaluate the capability of the model for certain storm events.

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. Further investigation does not appear necessary. Figures 5.17 to 5.19 show that the comparison between predicted and observed PCB concentrations in the water column over 2 years 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 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. An apparent gap in the presentation of the data is the comparison of observed and predicted concentrations in bottom sediments of the river. These data are likely to be useful in assessing the fate of PCBs in the River. Perhaps the new data that were mentioned in the presentation by Dick McGrath will be used for this purpose. 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, it can be concluded that there is 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 other aspects of the fate of PCBs such as bed load transport, degradation, volatilization, losses through "burial" cannot be assessed based on the available data.

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 and there are additional comparison 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 explored in this study. The temporal capability of the 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.

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 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 5-60).

Figures C3-26, C3-27 and C3-29 9and C3-49, C3-50 and C3-52 (for the linked model) show that there is a considerable bias (the residuals range over approximately two orders of magnitude) in the residuals of observed and predicted biota concentrations with measured concentration.

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. I think this needs to be revisited. In response to my question, Dr. Ed Garland reported that the 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. 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 in flooding events when particulate material and water are distributed over large areas of floodplain. When flooding subsides, these particulate materials may be in contact with air for considerable times and PCBs may volatilize. PCB losses to deep bed sediments are recognized in the Figure 5.1 but this process is inadequately treated in the model. It is possible that this process is responsible for a considerable loss of PCBs from the River. Perhaps of less relevance is the lack of characterization of decay of organic matter which again is recognized in Figure 5, but its influence is not considered in the application of the model. 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 EDFC 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 may become a limitation when the model results are to be related to ecotoxicological effects, which are related to concentrations of specific PCB congeners.

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. 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. Equilibrium partitioning between POM and water may overestimate the water concentrations somewhat as any disequilibrium between POM and water concentrations is not included. Hence, disequilibrium could be added although it may not to have a significant effect on the outcome for tPCB of the model as uptake of PCB in biota is predominantly via the diet. I do not recommend adding an equilibrium factor as its effect is likely unnoticeable given the calibration methodology and estimates of the disequilibrium factor are likely not available.

A gap, perhaps in the reporting only, concerns the model for accumulation in aquatic macrophytes and algae. A simple 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 apparent 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 key 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 a bounding exercise.

On balance, the reliance on calibration of the bioaccumulation model makes further refinement of the model somewhat redundant. The only argument against this is the behavior of the residuals of observed and simulated concentrations which show a strong a relationship with concentration (figure C.3-26). This behavior of the model contributes uncertainty in the model’s capability to estimate concentrations correctly under scenarios where concentrations are going up or down. The latter is likely an important scenario.

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 were adequate for the HSFP, EFDC and the Bioaccumulation model.

The results of the sensitivity analysis of the bioaccumulation raise some questions about the proper functioning of the bioaccumulation model. For example, Table C-4-1 shows a less than proportional relationship of the infauna PCB concentrations with the sediment concentrations. Given the functional relationships for the infauna, I would not expect this as pore water and diet concentrations are the inputs for the calculation and they are linear with the concentration in the sediments. Also, the sensitivity analysis shows a very low sensitivity of the model outcome to the lipid content of the organism (2.2, 3.5 and 6.7% in Table C4.1). On the other hand, Table C-4-4 shows that the sensitivity of the chemical assimilation and food assimilation efficiencies are greater than 100%. This implies a non-linear relationship between the PCB concentration and the assimilation efficiencies, which is not consistent with equation 9 in attachment C1.

The methodology used to conduct the uncertainty analysis contains some significant limitations throughout the entire modeling effort. There are several issues. First, is the basic modeling strategy which 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. The test of the model against an independent data set 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. Although, none of studies applies an independent concentration data set for model performance analysis, there is still considerable merit to using differences between observed and simulated data as a measure of uncertainty. I recommend that this is added to the modeling strategy. There are various statistical methods to do this such as mean squared error or calculating confidence limits of the model bias. The resulting uncertainty calculated should be treated with some caution as the uncertainty has a tendency to underestimate the actual uncertainty. Much of this work has already been done in the model calibration phase and I think it will be considerable use in assessing model uncertainty given the limitations of the Monte Carlo Simulation technique that is the main method used to assess model uncertainty in this study.

The second issue relates to the application of Monte Carlo simulations, which was attempted in the bioaccumulation model and planned for the EFDC model (B.5-4). The application of Monte Carlo simulations to complex models like EDFC 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. The latter is not easy to confirm. It may need considerable experience with the inner-working of the model as well as numerical techniques to identify occurrences of correlated parameters. From the report and the response to my questions on this issue, I have not been convinced that this was done adequately. A second condition for an informative Monte Carlo simulation analysis is that the variability and error in the model variables can be determined or is 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 not provide adequate information on this issue and it is not possible to confirm at this time whether the variability and error in the many model variables was appropriately accounted for.

Tables C4-7 to C4-9 show the results of the uncertainty analysis. It is not clear to me what the 95% confidence limits in the Tables represent, i.e. whether the values presented are the actual limits or whether the values need to be applied to the mean to get the confidence limits. However, either way I interpret the results, I have some concerns. Following a question on this issue during the Panel meeting, I did not receive an answer as of yet. Hence at this time, I cannot confirm that the uncertainty analysis was executed correctly and that the results are meaningful.

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

This is a good question, but hard to address, as it depends on the magnitude of the differences among remedial options, which are not known at this point.

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 certain 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 to refer to this as a 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 temporal response of the model is important.

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 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. An issue that requires some further attention in the development of the bioaccumulation model is the characterization of the uncertainty of the model. Currently, the report contains some rather qualitative statements (i.e. "two-fold" in an number of places), which do not appear to be representative of the real model-data differences). Only, when uncertainty is appropriately recognized, application of the model can be considered. Another issue of concern is the behavior of the model. As discussed earlier, the apparent concentration dependence of the model-data residuals is cause of concern with regards to the model's proper functioning and its ability to make accurate predictions under new scenarios. Also, the results of the sensitivity analysis indicate that the internal mechanics of the model may require some further investigation before the model is ready to be applied to explore remedial options.

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. Sediment core data could be useful, but given the high degree of mixing that appears to occur, the sediment core data are likely to be uninformative on this issue as well. 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, volatilization rates and burial 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.