Comments on Model Calibration

E. John List, Ph.D., P.E.
Principal Consultant
Flow Science Incorporated
Professor *Emeritus*
California Institute of Technology

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?

My opinion is that the watershed modeling program HSPF is adequately calibrated for use in a statistical sense, although there appears to be insufficient rain gauge information available to apply the model in a storm-by-storm prescriptive way. In other words, the model cannot be used to generate stream flows from a specific storm, but it will be satisfactory when applied to generate synthetic stream flow sequences appropriate for use in the comparison of remediation options. The reason for this is that remediation options will be compared based on a series of stream flow sequences and it is not relevant that these stream flow sequences will ever actually occur, it is enough that they properly represent the statistical variation for the watershed, which it appears they do. The overall strategy of producing a synthetic stream flow record from a rainfall record is a well-founded technique in hydrology and this work appears to have been done well. However, it is of concern that the HPSF model, apparently because of a lack of adequate rainfall data, was unable to properly represent the two large flow events of May and September in 1999.

On the other hand, the PCB fate and transport model EFDC application appears to have limitations in that the variance of the model output for water column tPCB does not reflect the variance in the measured data. This failing of the model is encapsulated in the comparisons presented in Figure 5-30, where it can be seen that the variance in the measured data is about two orders of magnitude larger than the predicted variance. Part of this problem is most likely associated with the high degree of variance that exists in the sediment PCB concentrations throughout the study area, as illustrated in Figure 5-26. Another issue with the EFDC modeling calibration is the relative lack of data that appear to exist in the last nine months of the calibration period. In reality, it is a stretch to claim that the model has been calibrated over a 14 month period when so few data exist for the last nine months, (see Figures 5-17, 5-18, and 5-19), especially since it appears that the intention is apply the model to time scales at least 20-30 times longer than the actual nine month calibration period. A longer calibration period is clearly appropriate.

The comparison of average PCB mass flux with predicted mass flux, over the 14-month calibration period that is shown in Figure B.4-46, is essentially meaningless because there is no indication of the confidence limits that apply to either the field estimates of the flux or the predicted flux. On the face of the data presented it appears that the prediction is valid but in the absence of any error margins for both the field and modeling estimates the results in Figure B.4-46 cannot be used legitimately as a calibration basis for the EFDC model.

A similar disparity in the variance of the predicted and measured PCB tissue concentrations is an outcome of the FCM modeling, as illustrated by Figures C.3-50 and C.3-51.

However, the overall performance of the FMC model, as depicted in Figure C.3-51, is certainly encouraging, given the difficulty associated with such food chain modeling. Nevertheless, it would be very nice to have this disparity resolved. The basic argument, if I understand it correctly, is that the model is only modeling average fish and not individual fish so the field variance in individual fish cannot be properly represented. I see this as a good reason to adopt a Monte Carlo approach that samples from populations of fish and the creatures that fish feed on. This Monte Carlo method has now become the norm in human health risk analysis and could easily be applied in this context.

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

There certainly appears to be bias in the FCM as indicated by the lipid-normalized PCB concentrations predicted by the FCM in Reach 5D of the PSA, as indicated by Figure 19 of Attachment C.16, and other portrayals of the FCM predictions for Reach 5D that are included in Attachment C.16. It is unfortunate that there appears to be no field data to substantiate the predictions of the FCM in this reach of the river, especially since the predictions appear to be incongruent with those for the rest of the PSA. In the absence of any explanation and/or correction for this phenomenon it is not clear that it is real, in which case it makes the application of the model problematic, at least until the reason for this indicated bias is understood.

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

In short, the answer is no. The basic problem is with the specific application of EFDC model to the sediments in the Housatonic Valley. The difficulty arises in part because the PCB concentrations within the sediments show an extreme spatial variability. This variability exists on a very small length scale (see slides Number 3 and 4 of the presentation by Dick McGrath to the Peer Review Panel on April 13) and, what is even more important, it seems that this small-scale spatial variability is carried over essentially uniformly to the large scale, as is evident on Figure 5-26 of the calibration report. These figures show that PCB concentrations in the top six inches of sediment range almost uniformly from 0.5 mg/kg to 200 mg/kg (approximately three orders of magnitude) over 11 miles of the river valley. There is no explanation for this essentially uniform distribution of extreme variability in the calibration report. When Ed Garland (presumptive leader of the modeling team) was directly asked for an explanation for this variability at the Peer Review Panel Meeting he responded that he did not know, and nobody from the EPA consulting team volunteered an explanation, other than to state it was a commonly observed phenomenon with PCB contaminated sediments. In the absence of any explanation for why this variability is present it is difficult to believe that the modeling exercise, which deals only in spatial averages across the entire width of the basic river channel, can properly represent the fate and transport of the PCB.

The key point here is that there is no reason to believe that this extreme variability in the sediment concentration of PCB has not existed for many years; for it seems unreasonable to believe that it is a recent phenomenon. It is also unlikely that it is a mere sampling artifact. In other words, all of the sediment erosion and deposition since major PCB releases occurred has not caused the concentration of PCB's to average out, even over relatively short horizontal scales of a few meters. The importance of this observation is that the EFDC fate and transport model uses *spatially-averaged* PCB concentration data and therefore cannot possibly hope to reproduce either the observed current spatial variability or predict any future variability. A long term application of the model is therefore simply going to smooth out the variance in concentration to produce a uniform concentration distribution; something that has not occurred naturally, at least so far. In some respects, it is somewhat analogous to trying to predict the maximum force on a structure in a water body using a hydrodynamic model that includes tides but no waves. In this case a model is being used to project changes in average concentration over a relatively large spatial element when this large element includes sub-elements that have almost three orders of magnitude variability in concentration that are "washed out" in the averaging process. Since the output of the model provides concentration averages there can be no hope of it reproducing the observed spatial variation. In fact, since it is not known exactly what process sustains the spatial variability it is seems entirely possible that the EFDC model does not even have that transport process properly represented. It may well be that the spatial variation is a legacy of the manner in which PCB releases occurred in the past and will not occur in the future; in the absence of any explanation we simply do not know.

It is recognized that if, in the model, "average" particles are entrained by the stream, transported and dropped at some new location, then it should lead in time to the generation of a sediment and water column with average concentrations of PCB, which are then input to the FCM model. So, why not simply use the average concentrations and forget about the variability? The point is the use of spatially-averaged input data leads to a reduction in the variance of PCB concentrations that is solely an artifact of the modeling process, something that we have already seen to occur. On the other hand, the modeling process goes to great lengths to partition the PCB by sediment size while at the same time ignoring the spatial variability of the erosion/deposition processes across the river, where in fact there is significant particle size variability. Given the inherent non-linearity of the sediment transport process, it is not at all clear that the approach of taking a single modeling element across the river is going to lead to the correct results.

Regardless of the reason for the PCB spatial concentration variations, this failure of the model to directly address the spatial variability in PCB concentration, which appears to be uniformly distributed along the river (see Figure 5-26), is a serious problem. It is highlighted by the fact that most of the calibration exercises performed for the EFDC model to date simply address the prediction of *averages* and the comparison of *averages* of field data and *averages* of model output (e.g., Figure B.4-46). The ability of a model to predict the *variance* in a distribution can be just as important as predicting the mean, especially where confidence limits on the predicted result, or the resolution of comparative remedial hypotheses, are important. There is nowhere (at least nowhere that I could find) in the PCB fate and transport calibration presentation that addresses the predicted *variance* in PCB concentration or flux (e.g., Figure 5-24) and calibrates this prediction with field data. It is implicitly presented in Figure 5-30 for tPCB water concentrations, but there is no formal comparison that I could find for the PCB fluxes, which are surely a very important part of the remediation modeling.

One PCB transport mechanism that appears to have been neglected is the contribution from groundwater inflow into the stream. I did not see where the mass flux of PCB to the stream by this mechanism had been properly quantified. My own experience in measuring groundwater inflow to streams indicates that this contribution can be quite substantial, and at times can far surpass any diffusive flux out of the river bed sediments.

Another mechanism for PCB loss that should be included in any analysis of the long term remediation process is the volatization of PCB from the river flood plain sediments. Experience with other chlorinated hydrocarbons (DDT) has shown that the half-life for such compounds in soils exposed to solar heating can be in the range of 10-12 years.

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

The sensitivities of model predictions to the relevant processes that have been included within the models seem to have been exhaustively evaluated. The issue is the uncertainties in the model predictions, which are of two types. One type is the inherent uncertainty associated with the uncertainty in the input data, which of course arises from the precision and accuracy of sample laboratory work and the inability of a limited number of samples to exactly define the statistics of a population. All of these uncertainties are quantifiable, although this has not been done to the level one would have liked to have seen. For example, the uncertainty in the predictions of PCB and sediment stream fluxes appears to have been omitted (e.g., in Figure 5-24). Another omission that could have easily been included is whether the mean PCB concentrations in the upper 6 inches of sediment show any statistically significant difference along the 11 mile river reach. This is exemplified by the model input data in Figure 5-26. Because there are no confidence limits on the estimates of the mean concentration of tPCB along the river reach it is not clear that the variations shown in the tPCB profile along the river have any statistical significance.

The second type of uncertainty is created by the inability of a model to exactly replicate the processes active in the field and the very distinct possibility that a model may have inherent, and unrecognized, flaws. This type of uncertainty is very difficult to quantify and can only be properly defined by a careful calibration and validation procedure. If the calibration and validation procedures indicate, over a large number of evaluations, that the model is reproducing field results, then some degree of confidence can be ascribed to the output of the model. Of particular concern here is that despite reproducing the mean of several measured parameters adequately, the model is not indicating the proper degree of variance in the output, as was discussed above in the response to Charge Question No. 1. On the other hand, as pointed out by another member of the Peer Review Committee (Dr. Lick), the model may be getting the "right" results despite incorrect science, simply because there are two compensating basic errors.

Another issue in this regard was also commented upon above. The fact that the sediments show a very wide variability in concentration of PCB, three orders of magnitude on horizontal spatial scales of 2 meters or so and uniformly over the 11-mile river reach, is disturbing for two reasons. One is the fact that there is no known explanation for this and therefore the appropriate fate and transport mechanism cannot be included in the model. And the second is that despite the fact that the mechanism is not included in the model (because it is unknown), there is no proof that its omission is of no consequence to the outcome of the modeling exercise.

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

In order to resolve a formulated hypothesis that one remedial option has a better "performance" than another, (however that performance may be quantified), there has to be a comparative statistical test developed. The test must be capable of resolving the probability of both false positives and false negatives and knowledge of sample means alone is insufficient information to resolve any formulated hypothesis. There are certainly non-parametric tests that can be used, but these require some knowledge as to the nature of the distributions and the ability to resolve false negatives may be impaired. As previously discussed, there is definite uncertainty in the model predictions and some of this has not been quantified, which will definitely impair the ability to resolve statistical hypotheses regarding the relative performance of remedial options. More attention should be directed at developing the data required for hypothesis testing

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?

Overall I am impressed at the depth and breadth of the work that has been completed and feel much more positive about the outcome than I did four years ago. The shortcomings in the modeling that have been identified here, and by others on the Peer Review Committee, can most probably be rather easily overcome. There is some additional calibration work that is necessary before the model can be used to predict future conditions with any degree of confidence. The issues that I see that need resolving are:

1. Attempt to resolve why PCB concentrations within the river reach have such extreme spatial variability on a horizontal scale. Is it because the PCB was initially released as free product and formed droplets that were carried by the stream and deposited out of the water column, or is it some other reason? Does the spatial inhomogeneity of PCB in the sediments reflect spatial inhomogeneity of organic carbon, and if so, why is the organic carbon distribution spatially inhomogeneous? Now that no further new releases of PCB are occurring will the erosion-deposition processes in the river lead to homogenizing the PCB concentrations in the sediment? If not, what is the mechanism by which the inhomogeneity is being maintained? How can this mechanism be included in the fate and transport model? If it cannot be readily incorporated is it still possible to use such a model to describe the fate and transport of the PCB? It is not a satisfactory response to simply ignore the issue. Given that the modeling exercise is chartered with predicting the outcome of remediation strategies it is my opinion that the fate and

transport modeling process applied so far, which focuses only on spatially-averaged concentrations in a situation where there is extreme spatial variability, is inappropriate. The failure of the study team to investigate and understand the basis for this extreme, and apparently temporally sustained, spatial variability in concentration, is somewhat disturbing, especially when it appears to be of such importance to understanding the fate and transport of the PCB.

It is not sufficient to simply acknowledge this deficiency with the statement (pg. 5-58): "This variability, which is a combination of stochasticity and analytical variability, is not represented in the model inputs, and therefore it is not represented in the model output."

2. Representation of a section of river by a single element that averages over the lateral extent of the basic channel is unlikely to provide the resolution of erosion and deposition that is required. In my opinion at least three basic elements are needed to represent the potential for erosion and deposition to occur at the same river mile, which is what really happens in most streams with which I am familiar.

3. The issue of why the modeling does not adequately reproduce the variance of tPCB concentration in the water column and in the biota needs resolving. If the model cannot resolve the variance of the distributions of PCB it is difficult to see how any statistical hypothesis regarding the relative efficacy of different remediation strategies can be properly resolved by the output of the modeling.

4. Some basic statistical analysis on both the data input to the EFDC model and its output appears not to have been completed. In the absence of this analysis it is not clear if variations in the input data (e.g., mean PCB concentrations in the sediment elements along the river) have any statistically significant variation from one end of the 11-mile river reach to the other. The predicted fluxes of PCB and sediment similarly appear to have no uncertainty analysis. These are serious omissions that can easily be corrected.

5. The apparent bias in the food chain model in Reach 5D needs resolution. In the absence of any field data in this reach it is difficult to make an unequivocal judgment that it is actually model bias, but the results are so far out of congruence with the other reaches that it is very suggestive of something amiss in the modeling.