Model Calibration:
Modeling Study of PCB Contamination in the Housatonic River

A Peer Review

By

W.Frank Bohlen
University of Connecticut
Department of Marine Sciences
Groton, Connecticut 06340

Introduction

Agreements developed between the General Electric Company and the U.S Environmental Protection Agency regarding means to optimize remedial activities intended to reduce or eliminate environmental and/or public health effects thought to be associated with exposure to PCB contaminated sediments include the development of a predictive numerical fate and transport model for application in the Housatonic River system. Given the complex of processes affecting this transport, this represents an ambitious project. The proposed model consists of three primary components, a watershed model (HSPF), a hydrodynamic/sediment-contaminant transport model (EFDC) and a bioaccumulation model (FCM). The models are linked but non-interactive. Details of each of these models and the results of the calibration phase are presented in a lengthy master volume and three detailed Appendices. A Peer Review Panel was asked to review this work and to answer a series of specific questions. Before I get to these questions I'll begin with a number of more general observations and recommendations.

General Comments

As discussed by several reviewers, assessments of model adequacy require a clear understanding of the intended application. Beyond the fact that the developing framework is to be used in the assessment of remedial alternatives, little detailed information is provided. It is not clear whether the model is to be used to assess benefits of one scheme relative to another or to provide an absolute assessment of selected schemes. One might argue that such specification is premature and in fact will be based on model results detailing the relative importance of particular source areas or processes. While this may ultimately prove to be the case, it still would be well to begin with a defined set of possibilities. In the case of the Housatonic River Study Area (PSA) this seems entirely possible due to the depth and breadth of data available detailing all primary site characteristics and the associated contaminant distributions. We know, for example, that a significant fraction of the total PCB mass in the river downstream of the confluence to Woods Pond is located in the floodplain. Given our less than perfect understanding of the range of processes affecting contaminant transport to and from the floodplain this fact might suggest that models would be best used to assess relative benefits so as to favor cancellation of errors. Alternatively, one might argue that the floodplain is primarily a trap and it is the more mobile components of the system such as the stream bed sediment column that is responsible for continuing exposure and the downstream flux of PCBs. This system, while still complicated may be more amenable to absolute assessment. Future discussions of model adequacy would benefit from a clear concise definition of the most probable primary use of the suite of models.

Beyond care in the definition of model application, this entire exercise would benefit from a careful (i.e. brutal) editing of the reports in the interest of clarity, understanding and the retention of the reader's interest. This is a complicated multi-faceted effort with a large number of investigators spread across the country (at least). I recognize that the coordination of the resulting writing effort is itself something of a herculean task. This must be faced however, if the goal is to produce a product that is at once comprehensive and amenable to detailed review and evaluation. The Executive Summary provides the framework to build on. It would however benefit from additional detail regarding boundary conditions applied for each model. These were spread through each of the individual model presentations and as such, easy to loose track of. A summary statement up-front would help.

The individual sections in Volume 1 dealing with each model would benefit from editing by a single hand to provide a consistent "story line" throughout. In general I found that the Appendices added little to the discussion. Much of the material presented was already available in Volume 1. Future editions might use the Appendices as the site for the majority of the data plots and in some few cases for detailed elaboration of some particular aspect of model formulation. This combination of efforts would reduce the overall size of the report and contribute to the acceptance of the model predictions by the broadest possible user group.

Moving now to the specific charge questions.

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 ?

The calibration period extends from May 1,1999 to June 30, 2000. Although this was a period that allowed for the sampling of a range of average ambient and aperiodic storm conditions it is, from the standpoint of sediment/contaminant transport in the Housatonic River study area, a very short period of time. Erosion processes affecting the side banks and the associated channel migration, sediment deposition in the backwaters and Woods Pond and many of the transport processes affecting the floodplains operate on time scales long compared to the calibration period. As a result the comparisons conducted over the calibration period provide only limited indication of the model's ability to accurately predict longterm change. There is some indication that this is recognized by the model developers and will be addressed during the verification phase. Such use of the verification phase for calibration purposes is not recommended.

If 14 months is too short what might be an adequate calibration period ? This is a question that would benefit from some amount of discussion by those most familiar with the study area. At present the reports provide relatively little discussion of the reasoning that lead to the selection of the 14 month period. Although I've been trying to encourage a shorter report this is a subject that would benefit from additional discussion.

My brief review of available data detailing sedimentation in the study area as well as contaminant concentrations suggests that a five year period of calibration would result in a more robust test of model capability and complement the longer term validation runs.

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

function of the independent variables ?

The distribution of residuals provides no indication of model bias.

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

As noted , **The Model** consists of three primary components, a watershed model (HSPF), a hydrodynamic-sediment/contaminant transport model (EFDC), and a bioaccumulation model (FCM). This combination of models accounts for all relevant processes affecting PCB transport and fate in the Houstonic River study area. That is, all relevant processes can be accommodated in the models and have received some consideration in model development. Unfortunately this does not mean that this combination of models, as presently structured, will provide the accurate simulations of PCB transport and fate needed to facilitate remedial designs.

The watershed model (HSPF) has the potential to provide both estimation of surface water volume inflows as well as water temperature and the associated sediment loads entering the study area. The development of the model and subsequent comparisons of model outputs vs. measured flows indicated close agreement, well within the QAPP specified “very good” category of +/-10%. The agreement for the case of the suspended solids load was poorer which is not entirely unexpected. Unfortunately the discussion of the reasons for these latter differences was weak. Moreover, the comparisons relied on relatively long term averages (mean annual load). Examination of higher frequency TSS data/model comparisons for several storm events (May and September, 1999 e.g.) show differences exceeding 100% in concentrations as well as significant differences in timing. these differences make one wonder how it is that the annual average data is so well simulated. This was not subject to sufficient discussion in either Volume 1 or Appendix A. The significance of these differences for purposes of overall model calibration are impossible to assess since the upstream inputs of sediment used during the calibration runs relied on empirical data rather than the HSPF output. Since more than 75% of the stream associated input to the study area crosses this boundary this leaves HSPF inaccuracies affecting only 25% of the inputs an effective further diminished by the fact that these tend to be distributed along the length of the study area. Some consideration and subsequent discussion of this issue is recommended.

In addition to consideration of high frequency TSS characteristics the presentation of HSPF would benefit from a more detailed discussion of the methodology leading to the specification of water temperature, a parameter passed to the bioaccumulation model. It appears that a moderately complex heat calculation is performed but details are not included in the calibration reports. A quick search indicates that they are also limited in the MFD. For the most part I'd believe that water and air temperatures are equal through much of the study area. Where deviations might occur however, is in the deeper, lower energy backwater and/or pond areas. Are these areas stratified ? How is that to be handled. The discussion of water temperature suggests that the water column throughout the study area is always well mixed. If this is the case this should be stated. If not the case, how will the water temperature near-bottom in Woods Pond be specified for bioaccumulation purposes ?

The discussion of HSPF also provides no indication of how the model calculates in-stream sediment transport. The majority of the streams entering the study area appear to be treated as simply conduits for the transport of sediment supplied by surface water drainage across sections of the watershed. Is this so ? This has implications with regard to the amount of sediment delivered, its quality and the timing of delivery.

Moving next to EFDC beginning with the hydrodynamic model. Here I'm completely at a loss to explain why so much emphasis is placed on stage/discharge relationship and so little on the associated velocity field. The few direct current data presented in the report indicate that the model does a relatively poor job of specifying velocity despite its ability to define water level elevations. Review of the data presented indicated that this is most likely the result of two factors; the timing or phase of the stage along the study reach and/or the spatial segmentation used in the model. In discussing the model results no mention is made of the fact that while the model does provide a reasonable simulation of water level magnitudes, the timing of these elevations often differs substantially from that observed. This produces substantial differences in free-surface slope affecting both absolute velocity as well as the structure of the boundary layer. Since it is these characteristics that ultimately go into the calculation of boundary shear stress any error at this point may have profound effect on subsequent sediment transport calculations. Without care in the discussion of these factors this reviewer finds it hard to believe that this model can accurately simulate sediment/contaminant transport in the study area. This situation is best corrected by a careful analysis of the current meter data and comparisons with modeled velocity and a reasoned presentation of the results. If the existing data are not sufficient for this task an additional field effort should be initiated.

In addition to questions regarding the temporal characteristics of the stage along the study area, the differences between modeled and observed velocities could be simply the result of the relatively coarse spatial segmentation selected for use in the model. Use of 20m square grids within the main stem channel results in nearly the entire channel width being covered by a single cell. this coarse segmentation will not accommodate the lateral variations in flow known to exist in a meandering channel such as that found in many portions of the study area. As a result it's not surprising that the cell average velocity can differ substantially from values observed at discrete points across the channel. Examination of channel bathymetry, planview contours and the associated sediment distributions indicates that a minimum of three grid cells should be used across the main stem channel in order to adequately simulate hydrodynamics for use within subsequent transport estimates. If this results in an unacceptable increase in computation times consideration should be given to alternatives to batch runs including separating the individual models and developing discrete runs based on a well defined series of scenarios.

The combination of coarse segmentation and questionable hydrodynamics makes it hard to believe that the model being tested is able to accurately simulate sediment/contaminant transport in the study area. Adding in the questions raised by Dr. Lick regarding the transport formulations being used, the interpretation of the empirical data and the vertical segmentation of the sediment column only adds to this concern. How is it then that the model outputs dealing with TSS are in reasonable agreement with observations ? This brings us back to the matter of the formulation of model boundary conditions and in particular how the modelers specify the boundary conditions at the confluence. As discussed above, these are based primarily on empirical data and can be assumed to be reasonably accurate. Given the short calibration period of 14 months it's likely that this input represents the majority of the sediment moving through the area. The bed representing a small and generally negligible source. With this possibility, inaccurate specification of shear stress and the subsequent sediment transport would have little effect on model results be they TSS or estimates of deposition in low energy areas. This lead to the uncomfortable postulate that what had been produced by all this work was in fact little more than a very complicated advection-diffusion model dominated by average cross-sectional flows and particle settling velocities.

To correct this situation several steps are required. First, the postulate might be tested by a test of the sensitivity of the model to upstream boundary conditions. This was not included in the present series of sensitivity tests. Next, whatever the sensitivity run suggests, the structure of the velocity field must be better accommodated in EFDC. This will begin with a decrease in grid size followed by adjustments in channel friction factors to reduce the time differences between observed and modeled stage. With the hydrodynamics under control attention must turn to the sediment transport aspects of the model. I second all of Dr. Lick's comments with the additional proviso that even the sedflume results must be applied with care due to the differences in spatial scales affecting the flow regime in the river versus the flume. This is where some additional field data such as bathymetric change and/or radionuclide based estimates of sedimention rates come in providing a "weight of evidence" corroboration of the empirical measurements. This is also a subject that would benefit from an increase in calibration period.

The ultimate accuracy of the sediment transport formulations would also benefit from inclusion of side bank erosion in the model. It appears that it was left from the initial calibration runs because of time considerations. If so this is another reason for an increase in the calibration period. I do not advocate using portions of the verification runs for calibration purposes. Too often, this is self defeating.

In addition to the inclusion of the sidebanks I'd recommend additional discussion of the dynamics to be applied to the floodplains. At present, they appear to be being treated simply as sinks. I believe that they also have the potential to serve as sources particularly during immediate post-storm periods when rainfall-runoff may displace sediments freshly placed on vegetation and/or the adjoining soil surfaces. Such processes may become increasingly important as alternative source areas are eliminated.

With the model including the proper range of dynamics and the range of source sink areas a mass balance calculation must be conducted to insure that numerical artifacts neither produce nor consume mass. There was no demonstration in the present reports that the models were tested for mass continuity. This should be considered for both sediment and PCBs.

An additional factor missing from the sediment/contaminant model is the matter of PCB volatility. A variety of studies have shown that this factor can result in significant PCB flux (see Thibodeaux et.al, 2002 - ACS Symposium Series 806:130-149). It is not clear from the discussion provided why it was neglected in the present models.

4. Based upon your technical judgement, 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 methodologies used to evaluate sensitivity are generally accepted and adequate. Any deficiencies in the sensitivity analysis have more to do with omission. e.g. a test of the model sensitivity to upstream boundary conditions, noted above.

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

For the reasons discussed in Question 3 above answering this question seems premature.

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 ?

No. The fundamental problem (beyond the issues discussed above) with the current model calibration is its limited duration. Since data seem to be available for the extension of this period it is my recommendation that the period be extended in year steps out to five years. As suggested above, the five year period appears to be sufficient to allow measureable change in sedimentation and the associated sediment/contaminant concentrations to occur.