ATTACHMENT H

MONTE CARLO METHODOLOGY FOR UNCERTAINTY ANALYSIS ON THE EMBM AND FORECAST TRAJECTORY

Attachment H: Monte Carlo Methodology for Uncertainty Analysis on EMB and Trajectory Forecasts.

1.0 Introduction

Environmental systems generally have several sources of uncertainties, and these uncertainties are not merely due to a lack of proper measurements, but primarily due to the randomness inherent in real ecosystems. The implications of these uncertainties are particularly important in the assessment of several potential regulatory options, for example, with respect to the selection of a strategy for the control of pollutant levels. Incorporating these uncertainties into the modeling process, could potentially result in providing useful information that can aid in decision making.

The EMBM best estimate scenario assumed average values for all model inputs in determining the solids contribution, fate and transport of chemicals, as well as the forecast of future surface sediment concentrations under various remedial scenarios. To incorporate uncertainties in model parameters, a Monte Carlo[1] sampling approach was used to develop 10,000 iterations of each input. These 10,000 inputs were optimized in the EMBM and the optimized results were carried through the trajectory forecast calculations. A combination Microsoft Excel® Solver and the Crystal Ball® 7 (Decisioneering, Denver, CO, USA) add-on for Microsoft Excel® (a tool typically used for solving optimization problems), was used to perform this analysis. The objective of the uncertainty analysis was to provide an insight into the level of confidence in the model estimates for the best estimate scenario. This attachment presents the detailed methodology for the Monte Carlo analysis for the EMBM and Trajectory forecasts.

2.0 Methodology

The following stages were involved in the uncertainty analysis of the solids and contaminant mass balances, and contaminant forecasts presented in the CSM: (a) characterization of uncertainties in EMBM input chemical profiles, (b) estimation of the uncertainty in EMBM optimized outputs resulting from the uncertainty in chemical profiles, and (c) characterization of the uncertainties in model forecast resulting from uncertainties in the input profiles, EMBM outputs of solids contribution, decay of excess contaminant concentrations (lambda), and depth of resuspension reservoir/mixed layer (uncertainty propagation). A schematic diagram illustrating the Monte Carlo methodology is given in Figure H-1 and detailed description is presented below.

[1] Monte Carlo simulation is categorized as a sampling method in which the trails or realizations are randomly generated from probability distributions to simulate the process of sampling from an actual population.

2.1 Uncertainties in EMB chemical input profiles

Thirteen chemicals (copper, chromium, mercury, lead, trans-chlordane, 4,4'-DDE, 2,3,7,8 TCDD, total TCDD, total PCB, benzo(a)pyrene, fluoranthene, iron and TOC) were optimized in the EMBM to determine the solids balance. The uncertainties in the concentrations of these 13 chemicals for the external sources, and the resuspension source were defined by parametric and non-parametric statistics, respectively. These are described below.

2.1.1 External Sources and Receptor Profiles

The observed concentrations for the 13 chemicals for the external sources (Upper Passaic River, Newark Bay, Saddle River, Second River/SWO, Third River and CSOs) were generally normally distributed. For each external source and the receptor, a bounded normal distribution defined by the mean, standard deviation, minimum and maximum of each chemical was used to perform Monte Carlo simulation in Crystal Ball® 7. In performing these Monte Carlo simulations, it was important to maintain the relationship amongst the variables. Therefore, for each source, the correlation matrix was also specified in Crystal Ball® 7 to ensure that the 10,000 iterations of chemical profile represented the variability, inter-dependencies, and uncertainty for each external source and the receptor. Figure H-2a though H-2g presents the statistical distributions of chemical concentrations for the 13 chemicals optimized in the EMBM, for the external sources and the receptor.

2.1.2 Resuspension Source Profiles

The chemical profiles for the resuspension source were generated based on the TSI 1995 observations. The concentrations of each chemical in this data were neither normal nor log-normal distributed. None of the complex parametric distributions in Crystal Ball® 7 could adequately fit the data set. Therefore, to create the 10,000 iterations of concentrations for the resuspension source profile, a non-parametric simulation method called a bootstrap[2] was used.

The basic bootstrap approach uses Monte Carlo sampling to generate an empirical estimate of the sampling distribution of interest. In the bootstrap method, the 1995 data set was treated as the population and a Monte Carlo-style procedure was conducted on it to 10,000 iterations of the mean of the 13 chemicals optimized. This was done as follows:

1. The original sample locations of size 92 from the 1995 TSI data set were assumed to define the population of data set in surface sediments for resuspension. Note that in performing this analysis, TSI Location 246 was removed from the data set, because the PAH concentrations at this location were not representative of PAH values generally reported in the 1995 TSI data set.

[2] Bootstrap is a powerful Monte Carlo method that re-samples the original sample set with replacement to generate a distribution of sample's statistics. It is a non-parametric method.

2. The original locations were resampled with replacement to generate a bootstrap sample of size 91. This creates a bootstrap data set of the same size as the original, excluding Location 246. By resampling the locations rather than each chemical independently, the correlations amongst the chemicals were maintained. Note that this bootstrap sample set may include some sample numbers in the original sample several times, and at the same time other sample numbers may be excluded.

3. Using the chemical concentrations for the locations selected in the 91 bootstrap samples, the average concentration for each chemical was calculated.

4. Steps 2 and 3 were repeated 10,000 times to generate the empirical distribution of the resuspension source profile (Figure H-2h).

The 10,000 average concentrations generated for each chemical via bootstrap for resuspension were used along with the 10,000 iterations for the external sources and receptor to represent the uncertainty in the inputs for EMBM optimization.

2.2 Estimation of uncertainty in EMB Output

A Microsoft Excel® macro[3], which calls the SOLVER routine, was developed to perform the EMBM optimization with the aim of determining the relative solids contributions from the various sources and the mass balance for the chemicals optimized. The macro was used to solve the 10,000 optimizations using the 10,000 iterations of the sources and receptor generated by the Monte Carlo simulation. The results of the optimization run were used to understand the uncertainty in the relative source contributions and chemical mass balance for the Lower Passaic River. The 10,000 EMBM optimized results were also used as input to the trajectory forecast calculations.

2.3 Uncertainties in Trajectory Forecast

Uncertainties in forecasted chemical concentrations were defined by the results of 10,000 iterations of forecasted values. The chemicals forecasted included: 2,3,7,8-TCDD, mercury, copper, lead, 4,4'-DDE, Total PCB and gamma chlordane. Three remedial scenarios were considered including: natural recovery, remediation of Primary Erosion and Inventory zones, and remediation of sediments from RM0 to 8 (see Chapter 20 for complete description of these scenarios). Forecasting the future concentrations of chemicals under the various remedial scenarios required inputs of (i) chemical concentrations, (ii) solids contributions for the various sources determined by the EMB optimization, (iii) decay of excess contaminant concentrations

[3] A Microsoft Excel® macro is a set of instructions written in Visual Basic programming language for Application that can be triggered by a keyboard shortcut, toolbar button or an icon in a spreadsheet. Macros are used to eliminate the need to repeat the steps of common tasks over and over.

(lambda; λ), net sedimentation rate, and (iv) the depth of the sediment mixed layer. Uncertainties in these inputs were defined as follows:

1. Uncertainties in the chemical concentrations were defined by the 10,000 iterations used as inputs to the EMB optimizations (Figure H-2). For each forecast calculation, the source and receptor profiles were represented by the Monte Carlo generated values as described in Section 2.1 above.

2. Uncertainties in solids contribution from the various sources were obtained from the uncertainty in the solids contributions determined by the EMBM optimization results. This was implemented by using the 10,000 solids contribution results from the EMBM.

3. Uncertainties in decay of excess sediment contamination were defined by the regression between the natural logarithm of the excess concentrations versus time (see Figure 20-1 to 20-9 in Chapter 20). Using the slope (λ), standard error and confidence bounds from the regressions, 10,000 iterations of λ were simulated using Monte Carlo sampling from bounded normal distributions (Figure H-3). Note: $Half\ life = \ln(0.5)/\lambda$

4. Uncertainties in the sedimentation rates were generated by bootstrap analysis of the differences between the 1989 and 2007 bathymetric surfaces (Figure H-4).

5. The uncertainties in depth of the sediment mixed layer were generated by 10,000 random numbers between 10 cm to 20 cm in Microsoft Excel®. Note that Microsoft Excel®'s random number generates uniform distributions of the parameter of interest (Figure H-5).

The Microsoft Excel® spreadsheets designed to perform forecast calculations using the best estimate for all inputs were modified to perform the calculations for the 10,000 iterations through a macro. For each iteration, the macro reads the input values of chemical concentrations, λ, sediment deposition rate and mixed layer depth, updates the forecast spreadsheet with these values, and then saves the results of the forecast calculation for all the remedial scenarios.

3.0 Results

Uncertainties in the EMBM solution and trajectory forecasts were defined by the confidence interval (5th and 95th percentiles) of the 10,000 optimized solutions. All the results are presented and discussed in Chapters 19 and 20.