NISS

A Bayesian Model for the Fate and Transport of Polychlorinated Biphenyl in the Upper Hudson River

Laura J. Steinberg, Kenneth H. Reckhow, and Robert L. Wolpert

Technical Report Number 37
August, 1995

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

A Bayesian Model for the Fate and Transport of Polychlorinated Biphenyl in the Upper Hudson River

Laura J. Steinberg[1], Kenneth H. Reckhow[2], and Robert L. Wolpert[3]

(Submitted to the *Journal of Environmental Engineering*)

Revised 8/25/1995

Abstract

Modelers of contaminant fate and transport in surface waters typically rely on literature values when selecting parameter values for mechanistic models. While the expert judgement with which these selections are made is valuable, the information contained in contaminant concentration measurements should not be ignored. In this full-scale Bayesian analysis of PCB contamination in the upper Hudson River, these two sources of information are combined using Bayes Theorem. A simulation model for the fate and transport of the PCBs in the upper Hudson River forms the basis of

[1]Research Fellow, National Institute of Statistical Sciences, 200 Park Office Drive, Research Triangle Park, NC 27709-4162

[2]Associate Professor, School of the Environment, Duke University, Durham, NC 27708

[3]Associate Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708

the likelihood function while the prior density is developed from literature values.

The method provides estimates for the anaerobic biodegradation half-life, aerobic biodegradation plus volatilization half-life, contaminated sediment depth, and resuspension velocity of 4400 days, 3.2 days, 0.32 meters, and 0.02 m/year, respectively. These are significantly different than values obtained with more traditional methods, and are shown to produce better predictions than those methods when used in a cross-validation study.

1 Introduction

Predictions from water quality models are largely dependent on the values of the parameters in the models. In mechanistic models, examples of model parameters include chemical and physical reaction rates, sedimentation and resuspension velocities, and environmental variables such as water clarity. In empirical models, parameters are coefficients or exponents in the model, and do not necessarily have particular physical meanings.

Within the context of a particular model specification, a modeler may be uncertain about which value to choose for the parameter. For example, it frequently happens that while there is some information which indicates the expected behavior of the chemical in the environment:

- there are no site-specific data,
- there are no chemical-specific data,
- the available data conflict with each other, or
- the parameter is an aggregated one representing conditions during many time periods or at many sites.

In these cases, it makes sense to adopt a probability distribution in which both the uncertainty the modeler has about the parameter value and its inherent variability can be quantified. The data upon which the investigator bases the probability distribution as well as the judgement exercised in combining these data are referred to as "prior knowledge". The term refers to the information which the investigator has prior to reviewing and analyzing the data set for the particular study of interest.

Sometimes modelers estimate parameter values by fitting the model parameters to the data. In this case, the resulting parameter values are uncertain because the modeler must use some (often arbitrary) criterion to determine the best fit. Even parameter values obtained from a nearly perfect fit with the data are uncertain when the calibrated model is applied to a new input data set. There are many possible reasons for this behavior; Beck (1987) presents a particularly lucid discussion of this and other issues related to uncertainty.

When model fitting is done in the context of a statistical model for the mea-

sured concentrations, and the fitting criterion is to maximize the likelihood of the observed measurements, the method is known as maximum likelihood estimation. For example, the measured concentration may be modeled as a normally distributed random variable with mean, μ, equal to the prediction from a simulation model, and unknown variance σ^2. Let $\boldsymbol{\theta_p}$ represent the vector of unknown parameters in the simulation model, and $\mu(\boldsymbol{\theta_p})$ the model prediction. The probability density function for observing a particular concentration, y, given $\boldsymbol{\theta_p}$ and σ, is the likelihood function and is given in Equation 1:

$$f(y \mid \boldsymbol{\theta_p}, \sigma) = \frac{1}{(2\pi\sigma^2)^{0.5}} \exp\left[-\frac{1}{2}\left(\frac{y - \mu(\boldsymbol{\theta_p})}{\sigma}\right)^2\right] \tag{1}$$

Together, the parameters $\boldsymbol{\theta_p}$ and σ comprise $\boldsymbol{\theta}$, the vector of unknown parameters to be estimated. If a set of independent observations, $\boldsymbol{y}$, is taken, the likelihood function for the set is the product:

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}) \tag{2}$$

of the normal density functions from Equation 1. Usually, the logarithmic forms of Equations 1 and 2 are used. In that case, Equation 2 becomes:

$$\ln f(\boldsymbol{y} \mid \boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i(\boldsymbol{\theta_p}))^2 \tag{3}$$

Maximum likelihood estimation would maximize $\ln f(\boldsymbol{y} \mid \boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

One of the problems with maximum likelihood estimation and many other methods which estimate parameters from the data is that a global optimum may be difficult to identify when many local optima exist. A second deficiency of the optimization approach is that it offers no formal route for prior knowledge to enter the analysis — as a result, the optimized values may be best for the calibration data set, but may not simulate the process well in other applications of the parameterized model. One way to incorporate *both* prior information (in the form of probability density functions) and empirical evidence (in the form of the likelihood function) is to use Bayesian parameter estimation. An added benefit is that the resulting point estimate of the parameter is unambiguous since the resulting function typically has only a single optimum.

In this paper, we describe how Bayesian parameter estimation can be used to estimate parameter values for a contaminant fate and transport model. This is undertaken for a mechanistic model of polychlorinated biphenyl (PCB) in the upper Hudson River. Following a discussion of Bayes Theorem, we describe the simulation model used and the model inputs. Results of the Bayesian analysis are presented, and compared with those resulting from the use of prior knowledge solely, and from maximum likelihood estimation. Finally, we conclude with some observations about the advantages of Bayesian parameter estimation in water quality modeling (see Berger, 1985 for a thorough discussion of Bayesian

techniques for parameter estimation).

2 Bayesian Parameter Estimation for Water Quality Models

2.1 Bayes Theorem

Bayesian parameter estimation is based on the use of Bayes theorem, published by Thomas Bayes in 1763. This theorem is used to infer probabilities of events or distributions of parameters, given data which are indicative of them. As Berger (1985) writes:

> The typical phrasing of Bayes's theorem is in terms of disjoint events $A_1, A_2, \ldots, A_n$, whose union has probability one (i.e., one of the A_i is certain to occur). Prior probabilities $P(A_i)$, for the events, are assumed known. An event B occurs, for which $P(B \mid A_i)$ (the conditional probability of B given A_i) is known for each A_i. Bayes's theorem then states that:
>
> $$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)}$$
>
> These probabilities reflect our revised opinions about the A_i, in light of the knowledge that B has occurred.

When Bayes theorem is extended to estimation of parameters, rather than the estimation of probability of events, it takes the form:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})\, d\boldsymbol{\theta}} \tag{4}$$

In this equation, $\pi(\boldsymbol{\theta})$ represents the probability density function for $\boldsymbol{\theta}$, the vector of unknown parameters, before any data are collected. This is called the prior distribution for the parameters. $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ is the revised probability density function after data, $\boldsymbol{y}$, have been collected and is called the posterior density function. $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})$ is the likelihood function for y given $\boldsymbol{\theta}$. It is the probability density of $\boldsymbol{y}$ given $\boldsymbol{\theta}$, considered as a function of $\boldsymbol{\theta}$.

The numerator of Equation 4 is regarded as the joint density of y and $\boldsymbol{\theta}$, and the denominator is the marginal density of $\boldsymbol{y}$; the division of the former by the latter creates the revised conditional density of $\boldsymbol{\theta}$ given $\boldsymbol{y}$. Since the denominator is always a constant (i.e., not a function of $\boldsymbol{\theta}$), the posterior distribution is known up to a normalizing constant when only the numerator is evaluated. Thus, the shape of the posterior density function may be found from evaluating the numerator only, while the scale of the distribution requires calculation of the denominator as well.

In this paper, the mode of the posterior density function is found and it is used as a point estimate for the parameter vector (rather than the more common maximum likelihood estimates). Ultimately, the full value of a Bayesian analysis

should be realized with the use of the posterior distribution to yield probabilistic statements about the parameters and predictions; the present analysis gives only an approximation of this goal.

2.2 Fate and Transport Modeling

Bayes Theorem can be applied to fate and transport models in surface water. These models typically take their form from the advection-diffusion-reaction equation (Thomann and Mueller, 1987). Consider a one-dimensional form of this equation for the fate and transport of a contaminant in a river with known loading rate W:

$$\frac{\partial C(x,t)}{\partial t} = -u\frac{\partial C(x,t)}{\partial x} + \frac{\partial}{\partial x}\left(D\frac{\partial C(x,t)}{\partial x}\right) - rC(x,t) + W \qquad (5)$$

where u = water velocity, D = the longitudinal dispersion coefficient, r = a first order rate constant for loss of the compound, and $C(x,t)$ = concentration at location x at time t. Assume that the velocity and dispersion are known (or can be fairly well approximated from measurements and empirical formulas) and that interest centers around estimating the first order rate constant which governs the loss of the contaminant. Then Equation 4 can be solved for $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ with σ and r comprising $\boldsymbol{\theta}$, and $\boldsymbol{y}$ being the vector of known concentrations, $C(x,t)$. $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})$ is the likelihood of observing $\boldsymbol{y}$ given the unknown parameters; its logarithm may computed from Equation 3 with $\boldsymbol{\mu}$ equal to the predictions from Equation 5. The

joint prior density function for r and σ can be computed from the multiplication of individual prior densities for r and σ assuming that the investigator's uncertainties about each are independent.

In assessing the prior density function for r, the investigator uses all the information available in the literature which can provide guidance on expected loss rates. This could include the results of laboratory experiments, field tests on the investigated water body or elsewhere, empirical equations whose structure and parameters are fit from other data sets, and the investigator's own intuition based on previous experience.

The error standard deviation, σ, originates from the concept that any one of a range of PCB concentrations might be measured at a given site in the river at a particular moment. Choosing a prior distribution for σ requires the investigator to assess the lack of fit in the model and the two main sources of randomness in the contaminant data: the degree of error in the measurement technique and the inherent variability which characterizes the water body. The assessment may consider model specification error, the reliability of the laboratory analytical techniques employed, the spatial heterogeneity of the data and the coarseness of the grid of measurements, and the temporal fluctuations in the concentrations.

3 The Hudson River Case Study

3.1 Fate of PCBs in the Hudson River

Polychlorinated biphenyl (PCB) contamination of the Hudson River dates from approximately 1950 when General Electric began discharging PCBs from its electrical equipment manufacturing plants in Fort Edward and Hudson Falls, NY. (Schroeder and Barnes 1983). Over 230,000 kilograms of PCBs (mostly in the commercially available forms of Aroclor 1016 and Aroclor 1242 (Bopp, 1979)) were discharged into the Hudson River at these sites between 1950 and 1977 (Tofflemire et al. 1979). Due to the hydrophobicity of PCBs[4], most of the chemical sorbed to river sediments. This material has migrated downstream, contaminating the Hudson River sediments with PCB from Fort Edward to New York Harbor (Figure 1) and serving as a source of PCB loading to the water column. Figure 2 shows measurements of PCB concentrations (ppm) from 243 sediment cores taken in the upper Hudson River[5] from 1976–1978.

Besides downstream transport into the lower Hudson River and into the New

[4]There are 209 different forms (congeners) of PCBs; Hawker and Connell (1988) estimate that the octanol-water partition coefficients for PCB congeners vary from 2.9×10^4 to 1.5×10^8, with hydrophobicity generally increasing as the number of chlorine atoms in the molecule increases.

[5]The upper Hudson River is defined as the Hudson River from its origin in the Adirondack Mountains to Troy N.Y. (just south of Waterford). Between Troy and New York Harbor, the river is tidally influenced.

York Harbor, loss of PCB mass in the upper Hudson River is due primarily to three mechanisms: volatilization, aerobic biodegradation, and anaerobic dechlorination. A brief discussion of the importance of these mechanisms follows. A more complete discussion can be found in Steinberg (1994) and in the referenced publications.

Volatilization of PCBs in the laboratory has been demonstrated by Doskey and Andren (1981), Dunnivant (1988), and Warren et al. (1987). Investigations of the fate and transport of PCBs in the natural environment typically include a determination of the volatilization flux; Baker and Eisenreich (1990) calculated volatilization rates of PCBs in Lake Superior and Thomann et al. (1989) estimated the effect of volatilization on PCB mass in the lower Hudson River using a simulation model. Larsson et al. (1990) showed that volatilization of PCBs occurred in the River Em in Sweden using a floating sampler. Calculations using Whitman's two-film model indicate that half-lives on the order of 3.5 days are likely.

Aerobic biodegradation is a biological removal process in which PCB molecules are converted to intermediate products such as benzoic acids. These products may later be completely mineralized by other bacteria (Furukawa et al. 1979). Aerobic biodegradation of PCBs has been observed in the laboratory (Bedard et al. 1986; Bedard et al. 1987; Furukawa et al. 1978); and *in situ* under artificial stimulation in the upper Hudson River (Harkness et al. 1993). Half-lives computed from these

experiments are on the order of 0.08 days to 8 days.

Anaerobic dechlorination is a biological *transformation* process in which chlorine atoms in highly chlorinated PCB molecules are removed and incorporated into metabolic products or by-products. This results in an increase in the number of less-chlorinated PCB molecules and a decrease in the number of highly-chlorinated PCB molecules. Anaerobic dechlorination has been observed in laboratory experiments (Quensen et al. 1990; Rhee et al. 1989), and existence in the natural environment has been deduced from field data in the Hudson River (J. Brown et al. 1987a,b). Half-lives for the loss of PCB mass (due to the loss of chlorine atoms attached to PCB molecules) of 200 to 2000 days have been calculated from the published data (Steinberg et al. 1994).

3.2 The simulation model

The simulation model for PCB fate and transport in the upper Hudson River is derived from the advection-diffusion-reaction equation . A number of simplifying assumptions were made in deriving the form of the simulation model. These assumptions were driven by the lack of observational data to support more complex model formulation. They are detailed below.

3.2.1 Simplifying Assumptions:

Overall PCB physical/chemical behavior PCB fate and transport models are always complicated by congener and homologue specific rates of sorption, degradation, and volatilization. While the original mixture of PCB congeners discharged into the river may be crudely estimated from General Electric records of Aroclors bought (Bopp, 1979), there are no observational data until 1990 with which to calibrate congener or homologue specific concentration predictions. Furthermore, reaction rates and removal rates for individual congeners and homologues are largely unknown and would have to be estimated from the data; the Hudson River data set is not sufficiently large to identify so many additional parameters.

In this simulation model, the model parameters reflect the overall behavior of PCBs in the upper Hudson River. Individual congener and homologue behavior is accounted for in the creation of the prior density functions for degradation and volatilization (see Steinberg et al. 1995). Experimental results and empirical predictions used to derive the prior density functions were weighted by the mix of Aroclors believed to have originally been discharged into the Hudson River as documented in Bopp, 1979. The partition coefficient for PCB is calculated using the same weighting (see Section 3.3).

Sedimentation and Resuspension Few measurements of resuspension and sedimentation in the upper Hudson River have been reported. Data to parameterize a

model with segment-specific sedimentation and resuspension rates have not been collected in the upper Hudson River. In light of this difficulty, a single overall resuspension rate was chosen to represent the behavior of sediment in the upper Hudson River. The rate of sedimentation was calculated based upon the assumption of no net build-up or removal of sediment. Thus, the flux of sediment onto the river bottom was set equal to the flux of sediment from the river bottom into the water column. Although there may be net deposition in some regions and net scouring in others, the delineation of these areas is not known, the amount of yearly change is not known, and the movement does not lend itself to one-dimensional modelling (in which we average over cross-sections) because it tends to be a localized phenomenon.

Interstitial Diffusion The model is also simplified in that it does not include a term for interstitial diffusion of dissolved PCB between the sediment and the water. This process may be of particular importance in returning PCB to the water column during low-flow periods. In trying to reduce the number of parameters in the model, we did not include this term specifically because we believe that the effects of interstitial diffusion are reflected in the resuspension velocity parameter estimate. The resuspension velocity is responsible for moving PCB from the sediment to the water column and therefore its posterior value reflects all processes which do this, including the interstitial diffusion.

A Single Sediment Layer The sediment was modelled as a single, mixed layer of PCB-laden sediment. Core data were averaged vertically over the depth of the PCB-contaminated layer to set initial PCB concentrations in the sediment.

We had expected core samples to reveal a consistent profile of concentration variation with depth, reflecting the time pattern of deposition and degradation. The four core sample data sets used in this study showed no such profile, even for multiple cores taken in a single segment of the upper Hudson River. This absence of a characteristic concentration profile suggests that such patterns either are widely variable, or else do not persist over time. This leads us to suspect that mixing arising from storm events, other natural phenomena, or perhaps human intervention has destroyed any continuous record of PCB accumulation at the core locations. For this reason, we did not choose to model variations in the PCB concentration with depth in the benthic sediment; rather, we elected to model the benthic sediment in aggregate as (locally) well-mixed.

3.2.2 Model Specification

The simulation model for PCB fate and transport in the upper Hudson River is derived from the advection-diffusion-reaction equation. It consists of two coupled differential equations, Equation 6 and Equation 7. Equation 6 simulates c_a, the total PCB concentration in the water column, and Equation 7 models c_b, the PCB concentration in the sediments. In addition to transport terms (1), Equation 6

includes a first order reaction term for volatilization and aerobic biodegradation in the water column (2), a term for increases in c_a due to resuspension of bottom sediments (3), and a term for loss of PCB mass due to sedimentation from the water column (4). Symbols in Equation 6 include: u, the water velocity; D, the longitudinal dispersion coefficient; k_{vol}, the volatilization rate constant; k_o, the aerobic biodegradation rate constant; F_w the fraction of PCB which is dissolved; v_r the sediment resuspension velocity; z, the water column depth; and v_s the sediment sedimentation velocity.

Changes in the concentration of PCB in the bottom sediment, c_b, are modeled with a term for concentration increase due to sedimentation of PCB-contaminated particles from the water column (5), and a term for concentration decrease due to anaerobic dechlorination and resuspension of PCB-contaminated sediment (6). The sedimentation velocity is modeled as the complement to the resuspension velocity. Its value is set as required to ensure that the flux of material resuspended is compensated by an equal flux onto the bed. In this way, no net scour or fill at any cross-section is simulated[6]. Symbols in Equation 7 include: ζ, the thickness of the contaminated sediment layer and k_a, the anaerobic biodegradation rate constant.

[6]This approach is suggested by the results of Lane and Borland (1954), Culbertson and Dawdy (1964), Colby (1964), and Andrews (1979), each of whom reported cycles of fill and scour and little net deposition or fill over the time period studied.

$$\frac{\partial c_a}{\partial t} = \overbrace{-u\frac{\partial c_a}{\partial x} + \frac{\partial}{\partial x}\left(D\frac{\partial c_a}{\partial x}\right)}^{1} - \overbrace{(k_{vol} + k_o)F_w c_a}^{2} + \overbrace{\frac{v_r}{z}c_b}^{3} - \overbrace{\frac{v_s}{z}(1 - F_w)c_a}^{4} \quad (6)$$

$$\frac{\partial c_b}{\partial t} = \underbrace{\frac{v_s}{\zeta}(1 - F_w)c_a}_{5} - \underbrace{(\frac{v_r}{\zeta} + k_a)c_b}_{6} \quad (7)$$

The simulation model was solved numerically using implicit finite differences with timesteps of approximately one day and spatial gridding of 1000 m (see Smith (1978) and Richtmyer and Morton (1967) for a discussion of finite difference methods for solving partial differential equations).

In the next section, we present the basis for calculation of input values for the solution of Equations 6 and 7, including the river velocity, the longitudinal dispersion coefficient, the fraction of dissolved PCB, suspended sediment concentrations, and boundary conditions.

3.3 Simulation Model Inputs

River velocity was calculated from the equation of continuity using flow rates reported by the United States Geological Survey (USGS) and assuming rectangular channels with width measured from USGS topographic maps.

River depth was calculated from the empirical weir equation:

$$Q = KLH^{1.5}$$

with K equal to 1.77(SI units), Q equal to the flow rate, L equal to the weir length, and H equal to the height of the water over the weir. Weir lengths and heights were obtained from the New York State Department of Environmental Conservation.

The longitudinal dispersion coefficient was calculated using Fischer's equation (Fischer et al. 1979).

$$D = 0.011 \frac{u^2 w^2}{z\sqrt{gzS}} \tag{8}$$

where w is river width (estimated from USGS topographic maps), g is gravitation acceleration, and S is the slope of the water surface (estimated from 10-year flood elevations from Federal Emergency Management Agency floodway studies).

F_W is the fraction of PCB dissolved in the water column, and is calculated from:

$$F_W = \frac{1}{1 + sK_p}$$

where K_p is the partition coefficient and s is the suspended solids concentration. The partition coefficient, K_p, is calculated using the K_{ow} estimates of Hawker and Connell (1988) for PCB congeners, and averaging the estimates to reflect the composition of the Aroclor 1016/1242 mixture originally discharged into

the Hudson River. K_{oc} was estimated from the Karickhoff (1979) relationship: $\log_{10} K_{oc} = \log_{10} K_{ow} - 0.21$. This yielded a log K_{oc} of 5.36. The fraction of organic carbon, f_{oc}, was taken as the median f_{oc} measurement of the upper Hudson River core samples. This resulted in an estimate of f_{oc} of 6.3%. Multiplying $10^{5.36}$ by 0.063 yields a K_P value of 14400 and this was the value used in the simulation model.

Suspended sediment concentrations are sampled daily at Fort Edward, Stillwater, and Waterford. Missing samples were imputed using regressions of the logs of suspended sediment concentration on flow or using regressions of suspended solids concentration on suspended sediment concentrations at adjacent stations.

The simulation model also requires boundary PCB concentrations in the water column. In order to predict missing days, the log of PCB water column measurements taken at Fort Edward and at Waterford were regressed against flow and year.

Initial PCB concentrations for use in the simulation model were calculated from 243 core samples taken from the upper Hudson River in 1976, 1977, and 1978. Each core had been sectioned; the PCB concentrations of the sections were averaged after removing all bottom sections of 0 ppm concentration. The average concentrations were assigned to 50-meter reaches of the river based upon where the sample had been taken. All cores belonging to the same 50-meter reach were

averaged, and this value was assigned as the reach PCB concentration. Reaches where no PCB measurements were taken were assigned 0 ppm concentrations, on the understanding that cores had been taken at locations where it was thought that PCB would most likely be found. Initial concentrations at each node of the model were estimated as the average of the nearest 10 reaches on either side of the node.

4 The Bayesian Model

It is assumed that several of the parameters in the model are uncertain and should be characterized by prior probability density functions. These include the anaerobic dechlorination rate constant (k_a), aerobic biodegradation rate constant (k_o), volatilization rate constant(k_{vol}), resuspension velocity (v_r), and the thickness of the contaminated sediment layer (ζ). The estimation of the prior probability density functions for each of these parameters is discussed in Steinberg et. al 1994. In summary, the distributions were estimated as follows:

- k_a and k_o: estimated from the results of laboratory studies on dechlorination and biodegradation,

- k_{vol}: estimated from empirical equations relating chemical properties and environmental conditions to the volatilization rate,

- v_r: estimated from resuspension data collected in other rivers (no data available for the upper Hudson River), and
- ζ: estimated from sediment cores taken from the upper Hudson River.

Also in Steinberg et. al (1994) is a discussion of how the prior distributions for k_{vol} and k_o were combined into a single distribution for k_d, a new parameter representing the combined effects of k_{vol} and k_o. This reparameterization was required because neither parameter is individually identifiable from Equations 6 and 7. Then, for greater interpretability, the density functions for k_a and k_o were transformed into distributions for half-lives due to anaerobic dechlorination (t_a) and aerobic biodegradation plus volatilization (t_d).

A fifth unknown parameter, σ, appears in the likelihood function (Equation 3). A discussion of the derivation of the probability density function for σ, the unknown variance of the PCB concentration, may be found in Steinberg et. al (1994) and Steinberg (1993).

Under the assumption of independence of these distributions, i.e. that knowledge of one parameter is not informative about the distribution of another, they may be multiplied together to form the joint probability density function, $\pi(\boldsymbol{\theta})$, for the suite of parameters:

$$\pi(\boldsymbol{\theta}) = \pi(t_a)\pi(t_d)\pi(v_r)\pi(\zeta)\pi(\sigma) \tag{9}$$

The result is a multivariate lognormal density function with mode located at the parameter values shown in Table 1. These are the same parameter values found at the mode of the individual prior density functions.

The likelihood function is constructed from Equation 3. We used measured PCB concentrations in the sediment taken in 1984, 1985, and 1990 as $\boldsymbol{y}$ and predictions of sediment PCB concentrations from the simulation model as $\boldsymbol{\mu}$. The suite of five unknown parameters (t_a,t_d,v_r,ζ, and σ) comprises $\boldsymbol{\theta}$.

5 Results

5.1 The Posterior Distribution

By Bayes' theorem the posterior density function is proportional to the product of the prior density function and the likelihood function. Even without knowing the proportionality constant (which would require a high-dimensional integration) the shape of the posterior distribution can be explored, particularly in a neighborhood of the *posterior mode* where it attains its maximum. The mode itself can be used as a parameter point estimate, replacing the point estimates commonly used by water quality modelers that are based solely on expert opinion (i.e., prior information) or on *ad hoc* model calibration exercises; see Schnoor et al. (1987). As computing hardware and integration algorithms improve it will soon become practical to base

parameter estimates and model predictions on the full posterior distribution, better reflecting all sources of variability and uncertainty.

Even locating the posterior mode is difficult in a statistical model this complex. Evaluating the posterior density function at each single point θ requires approximating the numerical solution to coupled partial differential equations, taking about half a minute on a 20 MIPS computer; both inherent variability and numerical approximations lead to non-smooth posterior density functions, making efficient derivative-based optimization routines ineffective. We used a polytope direct-search method (DUMPOL; see IMSL, 1985).

Achieving convergence with DUMPOL when searching for a maximum in the five-dimensional parameter space required good starting values. These were obtained by methodically varying 2 parameters at a time (a total of 40 combinations for each pair of parameters was used), and using DUMPOL to search the remaining three-parameter space for the maximum value of the posterior density function.[7] For each pairing of parameters, a contour plot of the maximized posterior density function was constructed by interpolating the 40 calculated posterior density function values. The contour plots indicated the regions in two-dimensional parameter space where the highest values of the posterior density function were found. These regions were used to establish several good starting points for the

[7]The optimization is actually performed by minimizing the negative log of the posterior density. This is equivalent to maximizing the posterior density itself.

complete five-dimensional optimization of the posterior density function. When the optimization was performed with these starting points, the optimal parameter values shown in Table 2 were found. Since these are the parameter values at which the posterior density function is maximized, these are the values of the parameters which determine the mode of the posterior density function.

It is interesting to explore the shape of the posterior distribution in the vicinity of the mode to determine if there is a sharp drop-off in the posterior density function around the maximum, or whether the function tends to be very flat in this area. A flat posterior would indicate that many combinations of the parameters will yield nearly the same posterior density, and hence all would be equally valid to use as a point estimate. This analysis also indicates which parameters are especially influential in determining the posterior density function in the area of the maximum. The results of the analysis are shown in Table 3. The table shows the change in the value of the negative log of the posterior density function. The changes are generated by holding four of the five parameters constant at the values shown in Table 2 while varying the value of the fifth parameter by $\pm 5\%$, $\pm 10\%$, and $\pm 50\%$.

It can be seen from Table 3 that the error standard deviation has the greatest impact on the posterior density in the neighborhood of the posterior mode. The combined volatilization and aerobic biodegradation rate constant shows little im-

pact on the posterior until the change becomes greater than $\pm 10\%$. The three remaining parameters appear to be approximately equally influential in the neighborhood of the posterior mode. Thus, the table indicates that in the neighborhood of the posterior mode, the posterior density function is relatively flat in the dimension representing volatilization and aerobic biodegradation, very steep in the dimension representing the error standard deviation, and mildly steep in the dimensions representing the anaerobic biodegradation rate constant, the contaminated sediment depth, and the resuspension velocity.

5.2 The Likelihood Function

Maximization of the likelihood function is another method for selecting parameter values. This method considers only the information contained in the data, and not the knowledge that an expert or well-informed investigator might have about likely values for the parameters. Maximum likelihood estimation is concerned with finding the values of the parameters which maximize the probability of y given θ. Thus, it is expected that the maximum likelihood estimator for θ will provide better fits to this particular data set than either the posterior or prior mode, but that the posterior mode would provide better fits to future data sets collected from the Hudson River because of the additional information supplied by the prior distribution.

Finding the maximum likelihood estimate of the parameter involves a series of optimizations similar to that undertaken to find the posterior mode. Two alternative explorations of the likelihood were used. First, the shape of the likelihood function in the neighborhood of the posterior mode was investigated by changing one parameter value at a time. The results of this analysis are shown in Table 4. They indicate that changes in any one of the parameters except the combined volatilization and aerobic biodegradation rate constant will improve the likelihood function. This emphasizes the point that the posterior mode is *not* a maximizing point for the likelihood function.

For the second exploration of the likelihood function, the locations of local maxima are found by using a variety of starting points for the optimization. It is not expected that this likelihood function will have a single optimum for two reasons. First, the effects of one parameter may be compensated for by the effects of another parameter. For example, a large value for the contaminated sediment depth will minimize the effects of resuspension on the benthic sediment PCB concentration, while a fast anaerobic biodegradation rate will decrease the benthic sediment PCB concentration quickly. This was not a problem with the posterior distribution, since unlikely values of the parameters yielded small values of the prior density, and hence two extreme parameter values could not cancel each other out. Second, since the likelihood function is determined by σ and the difference

between $\boldsymbol{y}$ and the predictions, when two parameter vectors have nearly the same σ and one produces concentrations x ppm greater than the measurements while the other produces concentrations x ppm less than the measurements, both will yield the same likelihood function value.

When a starting point close to the posterior mode was used, a local optimum was identified nearby at:

$\sigma = 39$ ppm

$t_d = 0.35$ days

$t_a = 5200$ days

$\zeta = 0.61$ m

$v_r = 0.001$ m/yr

When a different starting point was used, the DUMPOL algorithm failed to identify a minimum after 200 iterations, but did locate several points which appear upon investigation of nearby points to be local optima and which yield the same likelihood function value. Two such points are shown in Table 5.

5.3 Fitted Values and Predictions

The fitted PCB concentration values obtained from the modes of the prior and posterior density functions, and from a local maximum of the likelihood function

were compared with measured PCB concentration values. In general, the fitted values from the prior density function seriously underestimated the measured values while the parameter values from the posterior density function provided much closer fits. Fitted values from the likelihood function were closest (in the root mean square error sense) to the measured concentrations, since minimizing the squared error term is the sole criterion for fitting the likelihood function.

These results are summarized in Figure 3 which presents boxplots of the absolute value of the residuals obtained from each of the three parameter estimation techniques. The shaded boxes indicate the interquartile range. The upper whiskers are 1.5 $\times$ the interquartile range and the lower whiskers are located at the smallest residual value. The horizontal lines represent outliers.

The median values of the residuals in the boxplots do not differ much (10.5 ppm, 8 ppm, and 12 ppm for the prior density, posterior density, and likelihood function, respectively). However, the inter-quartile range is considerably larger for the prior density function than for the other two methods. This effect is also apparent in the values of the root mean square errors for the fitted values which are 52 ppm, 45 ppm, and 43 ppm for the prior, posterior, and likelihood, respectively.

Although the likelihood function produces fitted values closest to the actual measured values, it is expected that parameter estimates obtained from the likelihood function will not predict concentrations at *new* points in time or space as

well as those obtained from of the posterior density function. In other words, it is believed that existing scientific knowledge concerning processes and reaction rates, as expressed in the simulation model and the prior distribution, is of value for prediction. When this prior information is combined with that from the observational data set, the resultant posterior distribution is expected to provide the best fits to other (including future) data sets collected from the upper Hudson River. Ideally, one would like to have a validation data set on which to test this assertion. In the absence of a validation data set, the existing data set was divided into two sets. One set consisted of 1984 data and a second set of 1985 and 1990 data. The first set was used to find a local maximum of the likelihood function and the posterior mode. Then, the resulting parameter values were used to predict concentrations in 1985 and 1990.

When the predicted values are subtracted from the measurements and squared, the squared error obtained with predictions from a local maximum of the likelihood function is almost double that obtained with predictions from the posterior mode (3899 ppm^2 *versus* 7256 ppm^2). This indicates that under the squared loss criterion, the posterior estimator gave better matches to the 1985 and 1990 data set than did the likelihood estimator.

6 Conclusions

Reliance on complex mechanistic water quality models has tended to divert attention from underlying uncertainties about the aquatic systems under investigation. While models grow more complex, scientists' ability to provide rate constants and other process values which are appropriate for the particular chemical and site under study may not keep pace. Bayesian statistics provide a method for selecting parameter values for water quality models which updates limited information from experimentation with chemical-specific and site-specific data. The method also explicitly incorporates the spatial and temporal variability in "calibration" data (observations of concentration) which results from the heterogeneity and non-stationarity of environmental processes.

In the Hudson River PCB case study, it was shown that information on the behavior of PCBs in the natural environment is often contradictory (for example, PCB half-lives of 200 to 2000 days) or is dependent on site-specific processes for which data are not available (for example sediment resuspension rates). The Bayesian paradigm of beginning with prior information, updating with observations, and recalculating parameter values offered a way to combine all existing information, producing more chemical-specific and site-specific parameter values. The updating was reflected in the difference between the modes of the posterior density function and the prior density function. The posterior density was simple

to calculate, requiring just the multiplication of the likelihood function by the prior density function. The posterior density was found, as expected, to have a global maximum, and the location of this maximum was found with little difficulty. The likelihood function, on the other hand, had multiple optima. Furthermore, cross-validation analysis showed that predictions made with the posterior estimator were better (in the squared error sense) than those obtained with the maximum likelihood estimator.

The updated parameter estimates (located at the mode of the posterior density function) were: 37 ppm for the measurement standard error, 3.2 days for the combined volatilization and aerobic biodegradation rate constant, 4400 days for the anaerobic biodegradation half-life, 0.32 meters for the contaminated sediment depth, and 0.02 m/year for the resuspension velocity. In the neighborhood of the mode, the posterior density function was steepest in the directions of the error standard deviation and the anaerobic biodegradation half-life, indicating that these two parameters were the most influential in determining the posterior density function value at the mode.

Modelers should be aware that data on which to base parameter estimates may be sparse, contradictory, or irrelevant upon closer examination of the modeling problem at hand. Use of the Bayesian paradigm requires the modeler to sort out these discrepancies and inconsistences, and to account for them in a prob-

ability density function. By bringing in calibration data in a rigorous way, the Bayesian analysis produces parameter estimates which incorporate this parameter uncertainty as well as spatial and temporal variability in the measurements. As in this study, the resulting parameter estimates may be significantly different than estimates obtained through more traditional methods.

One result of the Bayesian analysis may be the realization that parameter uncertainty is quite high, and that the resulting predictions will be highly uncertain, leading to management decisions with little scientific support. In this case, it may be more useful to pursue a simpler model, or an empirically-derived model, from which parameters and predictions can be calculated with greater certainty. Ultimately, the most useful models are likely to result from a compromise between scientific feasibility, as characterized by uncertainty, and management desirability, as characterized by decision maker information needs.

7 Acknowledgments

The authors would like to thank the National Science Foundation (SES-8921227), the United States Environmental Protection Agency (cooperative agreement CR821439-01-0), the Hudson River Foundation, and the National Institute of Statistical Sciences through NSF grant DMS-9208758 for their support of this work. We would also like to acknowledge the helpful advice of the reviewers.

References

[1] E. D. Andrews (1979). *Scour and fill in a stream channel, East Fork River, western Wyoming*. Geological Survey Professional Paper 1117.

[2] J. E. Baker and S. J. Eisenreich (1990). "Concentrations and fluxes of polycyclic aromatic hydrocarbons and polychlorinated biphenyls across the air-water interface of Lake Superior," *Environmental Science and Technology* vol. 24, no. 3, pp. 342–352.

[3] M. B. Beck (1987). "Water Quality Modeling: A Review of the Analysis of Uncertainty," *Water Resources Research*, vol. 23, no. 8, pp. 1393–1442.

[4] D. Bedard, R. Unterman, L. Bopp, M. Brennan, M. Haberl, and C. Johnson (1986). "Rapid assay for screening and characterizing microorganisms for the ability to degrade polychlorinated biphenyls," *Applied and Environmental Microbiology*, vol. 51, no. 4, pp. 761–768.

[5] D. Bedard, M. Haberl, R. May, and M. Brennan (1987). "Evidence for novel mechanisms of polychlorinated biphenyl metabolism in *alcaligenes eutrophus* H850," *Applied and Environmental Microbiology*, vol. 53, no. 5, pp. 1103–1112.

[6] J. O. Berger (1985). *Statistical Decision Theory and Bayesian Analysis.* New York, NY: Springer-Verlag, 2nd ed.

[7] R. F. Bopp (1979). *Geochemistry of polychlorinated biphenyls in the Hudson River.* PhD thesis, Columbia University, New York.

[8] J. Brown, D. Bedard, M. B. J. Carnahan, H. Feng, and R. Wagner (1987a). "Polychlorinated biphenyl dechlorination in aquatic sediments," *Science*, vol. 236, pp. 709–712.

[9] J. Brown, R. Wagner, H. Feng, D. Bedard, M. Brennan, J. Carnahan, and R. May, (1987b). "Environmental dechlorination of PCBs," *Environmental Toxicology and Chemistry*, vol. 6, no. 8, pp. 579–593.

[10] B. R. Colby (1964). *Scour and Fill in Sand-Bed Streams.* Geological Survey Professional Paper 462-D.

[11] J. K. Culbertson and D. R. Dawdy (1964). *A Study of Fluvial Characteristics and Hydraulic Variables Middle Rio Grande New Mexico.* Geological Survey Water Supply Paper 1498-F.

[12] P. Doskey and A. Andren (1981). "Modelling the flux of atmospheric polychlorinated biphenyls across the air/water interface," *Environmental Science and Technology*, vol. 15, no. 15, pp. 705–711.

[13] F. M. Dunnivant (1981). *Congener-Specific PCB Chemical and Physical Parameters for Evaluation of Environmental Weathering of Aroclors*. PhD thesis, Clemson University.

[14] H. Fischer, E. List, R. Koh, and N. Brooks (1979). *Mixing in Inland and Coastal Waters*. New York: Academic Press.

[15] K. Furukawa, K. Tonomura, and A. Kamibayashi (1978). "Effect of chlorine substitution on the biodegradability of polychlorinated biphenyls," *Applied and Environmental Microbiology*, vol. 35, no. 2, pp. 223–227.

[16] M. Harkness, J. McDermot, D. Abramowicz, J. Salvo, W. Flanagan, M. Stephens, R. Mondello, R. May, J. Lobos, K. Carroll, M. Brennan, A. Bracco, K. F. adn G.L. Warner, P. Wilson, D. Dietrich, D. Lin, C. Morgan, and W. Gately (1993). "In situ stimulation of aerobic PCB biodegradation in Hudson River sediments," *Science*, vol. 259, pp. 503–507.

[17] D. Hawker and D. Connell (1988). "Octanol-water partition coefficients of polychlorinated biphenyl congeners," *Environmental Science and Technology*, vol. 22, no. 1, pp. 382–387.

[18] IMSL, Inc. 1985. MATH/LIBRARY Optimization, DUMPOL.

[19] S. W. Karickhoff, D. S. Brown, and T. A. Scott (1979). "Sorption of hydrophobic pollutants on natural sediments," *Water Research*, vol. 13, no. 3, pp. 241–248.

[20] E. W. Lane and W. Borland (1954). "River bed scour during floods," *American Society of Civil Engineers Transactions*, vol. 119, no. 2712, pp. 1069–1080.

[21] P. Larsson, L. Okla, S.-O. Ryding, and B. Westoo (1990). "Contaminated sediment as a source of PCBs in a river system," *Canadian Journal of Fisheries and Aquatic Science*, vol. 47, no.4, pp. 746–754.

[22] J. Quensen, S. Boyd, and J. Tiedje (1990). "Dechlorination of four commercial polychlorinated biphenyl mixtures (Aroclors) by anaerobic microorganisms from sediments," *Applied and Environmental Microbiology*, vol. 56, no. 8, pp. 2360–2369.

[23] G.-Y. Rhee, B. Bush, M. Brown, M. Kane, and L. Shane (1989). "Anaerobic biodegradation of polychlorinated biphenyls in Hudson River sediments and dredged sediments in clay encapsulation," *Water Research*, vol. 23, no. 8, pp. 957–964.

[24] R. D. Richtmyer and K. W. Morton (1967).*Difference Methods for Initial Value Problems*. 2nd ed. New York: John Wiley and Sons.

[25] J. Schnoor, C. Sato, D.McKechnie, and D. Sahoo (1987) *Processes, Coefficients, and Models for Simulating Toxic Organics and Heavy Metals in Surface Waters.* EPA/600/3-87/015, Washington, D.C.: USEPA.

[26] R. Schroeder and C. Barnes (1983) *Trends in polychlorinated biphenyl concentrations in Hudson River water five years after elimination of point sources.* Albany, NY: U.S. Geological Survey and the New York State Department of Environmental Conservation.

[27] G. D. Smith (1978).*Numerical Solution of Partial DIfferential Equations: Finite Difference Methods.* 2nd ed. New York: Oxford University Press.

[28] L. S. Steinberg (1993) *Bayesian Parameter Estimation for Surface Water Quality Models.* Ph.D. thesis, Duke University, Durham, NC.

[29] L. S. Steinberg, K. H. Reckhow, and R. L. Wolpert (1994). "Characterization of Parameters in Mechanistic Models: A Case Study of PCB Fate and Transport in Surface Waters," submitted to *Ecological Modeling*.

[30] R. Thomann and J. Mueller, *Principles of Surface Water Quality Modeling and Control.* New York: Harper and Row, 1987.

[31] R. V. Thomann, J. A. Mueller, R. P. Winfield, and C.-R. Huang (1990). "Model of fate and accumulation of PCB homologues in Hudson estuary," *Journal of Environmental Engineering*, vol. 117, no. 2, pp. 161–178.

[32] T. Tofflemire, L. Hetling, and S. Quinn (1979). *PCB in the Upper Hudson - Sediment distributions, water interactions and dredging*. Technical Paper 55, Albany, NY: New York State Department of Environmental Conservation.

[33] S. Warren, R. Bopp, and H. Simpson (1987). "Volatilization of PCBs from contaminated sediments and water." Submitted to the N.Y.S. Department of Environmental Conservation.

List of Figures

List of Tables

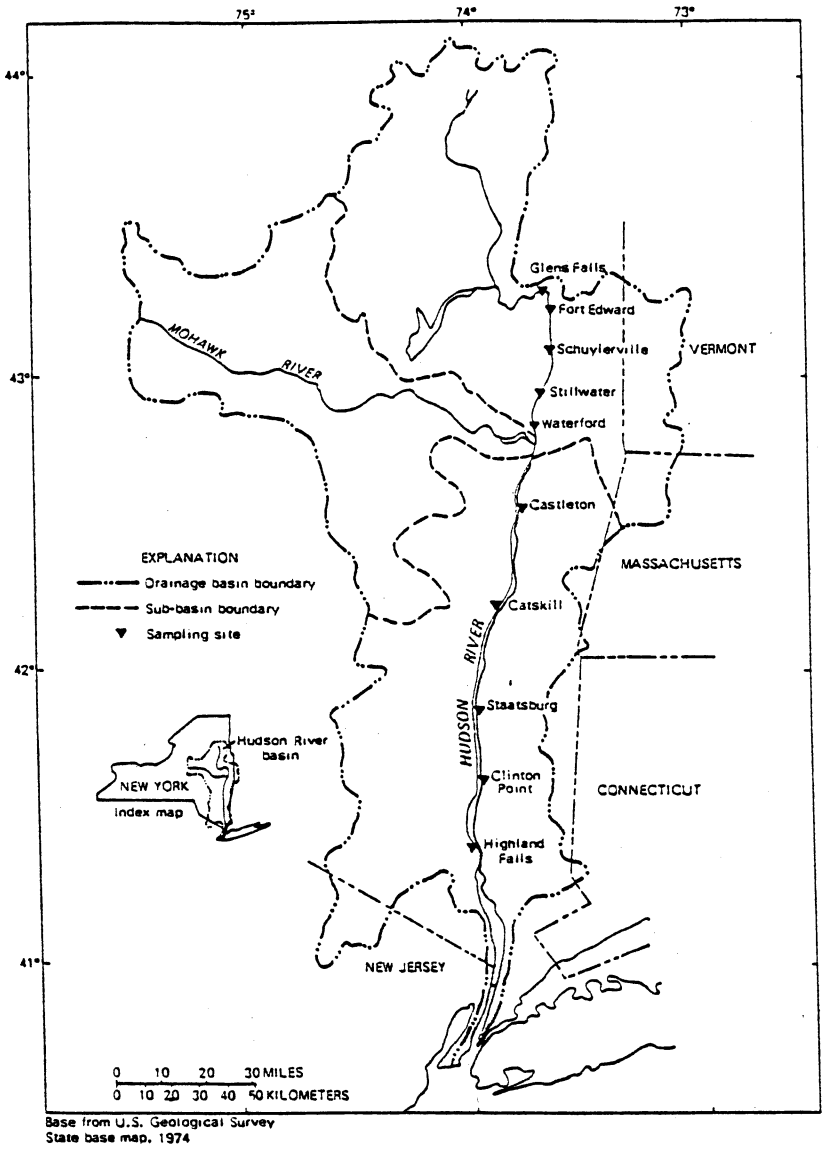

Figure 1: The Hudson River from Glens Falls to New York Harbor (adapted from Schroeder and Barnes 1983)

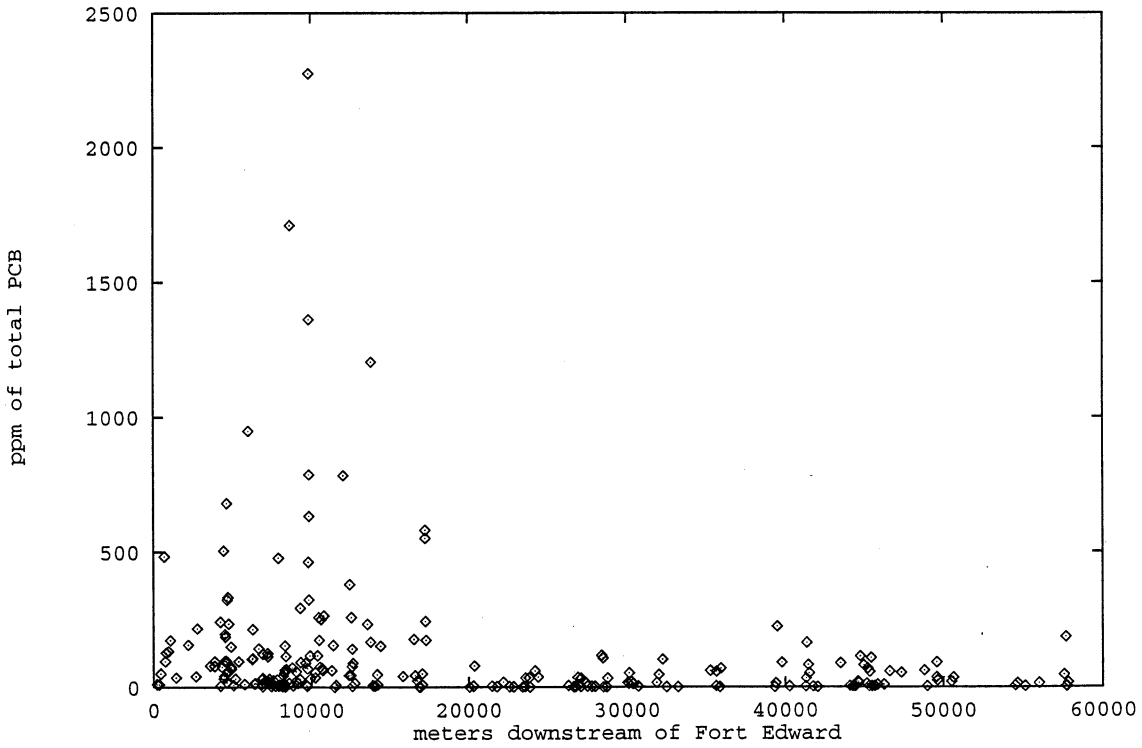

Figure 2: Maximum 1976-78 PCB concentration in each sediment core

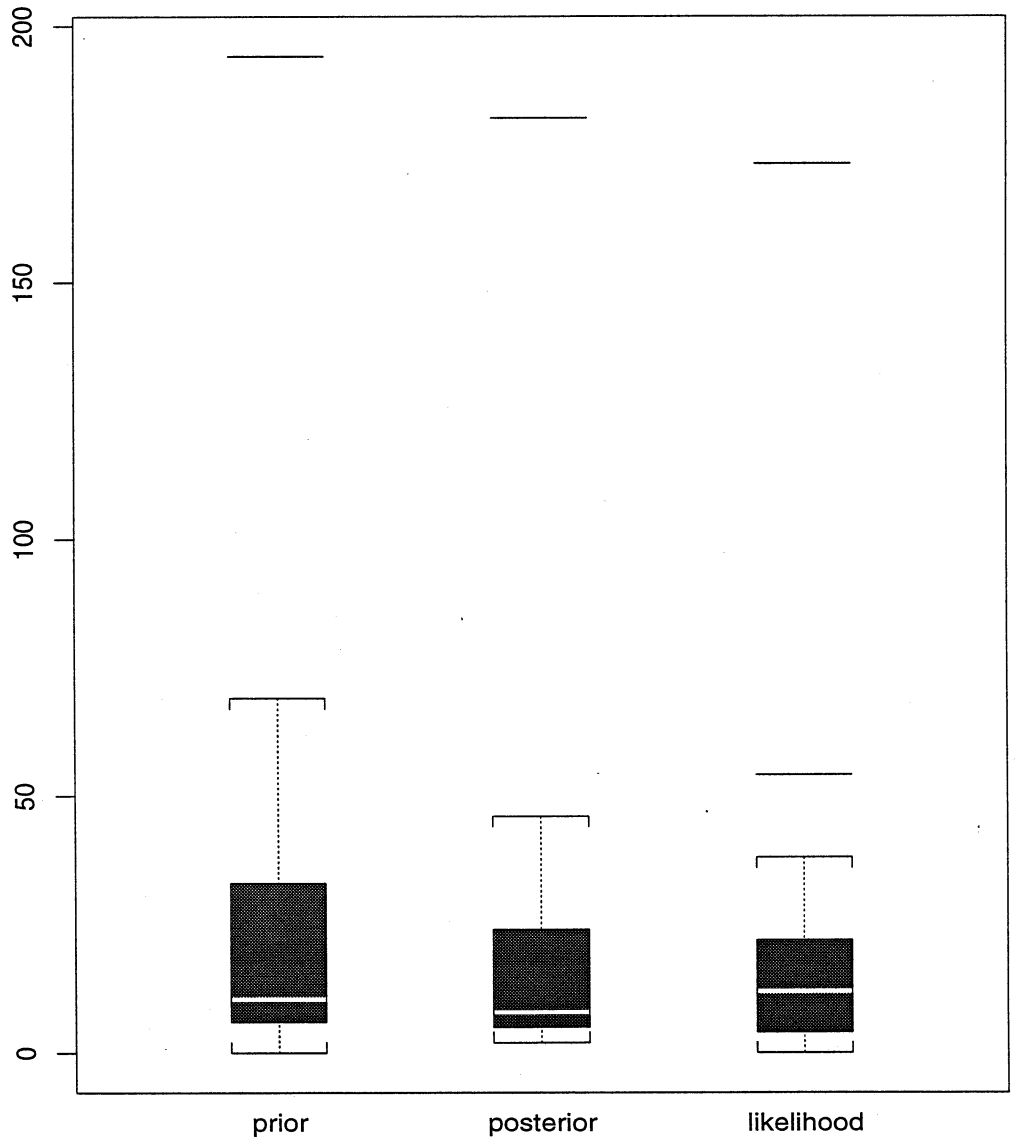

Figure 3: Boxplot showing the absolute value of the residuals (ppm) for each of the three parameter estimation techniques.

Table 1: Parameter values at the mode of the prior distribution

θ	description	parameter value
σ	measurement error	28 ppm
t_d	volatilization + aerobic biodegradation halflife	3.1 days
t_a	anaerobic biodegradation halflife	1400 days
ζ	contaminated sediment depth	0.07 m
v_r	resuspension velocity	0.09 m/yr

Table 2: Optimal parameter values for the posterior distribution

θ	description	parameter value
σ	measurement error	37 ppm
t_d	volatilization + aerobic biodegradation halflife	3.2 days
t_a	anaerobic biodegradation half-life	4400 days
ζ	contaminated sediment depth	0.32 m
v_r	resuspension velocity	0.02 m/yr

Table 3: Increase in the negative log of the posterior density as a function of changes in parameter values in the neighborhood of posterior mode

parameter	% change in parameter value					
varied	+50	+10	+5	-5	-10	-50
σ	6.20	0.55	0.22	0.03	0.28	27.44
t_d	0.23	0.04	0.04	0.03	0.03	0.36
t_a	0.49	0.14	0.11	0.12	0.14	0.54
ζ	0.27	0.09	0.09	0.12	0.14	0.54
v_r	0.35	0.16	0.14	0.09	0.08	0.50

Table 4: Change in the negative log of the likelihood function as a function of changes in parameter values in the neighborhood of the posterior mode

parameter varied	% change in parameter value					
	+50	+10	+5	-5	-10	-50
σ	2.05	-0.25	-0.25	0.16	0.64	26.66
t_d	-0.08	-0.12	-0.12	-0.13	-0.13	-0.13
t_a	-0.45	-0.19	-0.13	0.04	0.15	1.19
ζ	-0.60	-0.21	-0.13	0.04	0.15	1.23
v_r	0.65	0.12	0.04	-0.13	-0.23	-0.80

Table 5: Two locally optimal points of the likelihood function

Point 1	Point 2
$\sigma = 38$ ppm	$\sigma = 39$ ppm
$t_d = 5.0$ days	$t_d = 5.1$ days
$t_a = 7600$ days	$t_a = 8600$ days
$\zeta = 0.77$ m	$\zeta = 0.92$ m
$v_r = 0.004$ m/yr	$v_r = 0.003$ m/yr

Appendix. Notation

$C(x,t)$	concentration at location x at time t
c_a	water column PCB concentration
c_b	sediment PCB concentration
D	longitudinal dispersion coefficient
F_w	fraction of PCB which is dissolved
f_{oc}	fraction of organic carbon
g	gravitational acceleration
H	height of water over the weir
J	weir length
K	weir equation constant
K_p	partition coefficient
k_a	anaerobic dechlorination rate constant
k_d	combined rate constant for aerobic biodegradation and volatilization
k_o	aerobic biodegradation rate constant
k_{ow}	octanol-water partition coefficient
k_{vol}	volatilization rate constant
$\mathcal{L}$	likelihood function
n	number of observations
r	first order rate constant
S	slope of the water surface
s	suspended solids concentration
t_a	anaerobic dechlorination half-life
t_d	combine aerobic biodegradation and volatilization half-life
u	river velocity
v_r	sediment resuspension velocity
v_s	sedimentation velocity
w	river width
$\boldsymbol{y}$	vector of observations
z	water column depth
$\boldsymbol{\theta_p}$	vector of unknown parameters
μ	mean
π	probability density function, either prior or posterior
σ	variance
ζ	thickness of the contaminated sediment layer