Science of the Total Environment 557–558 (2016) 489–501

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Re-visiting projections of PCBs in Lower Hudson River fish using model emulation

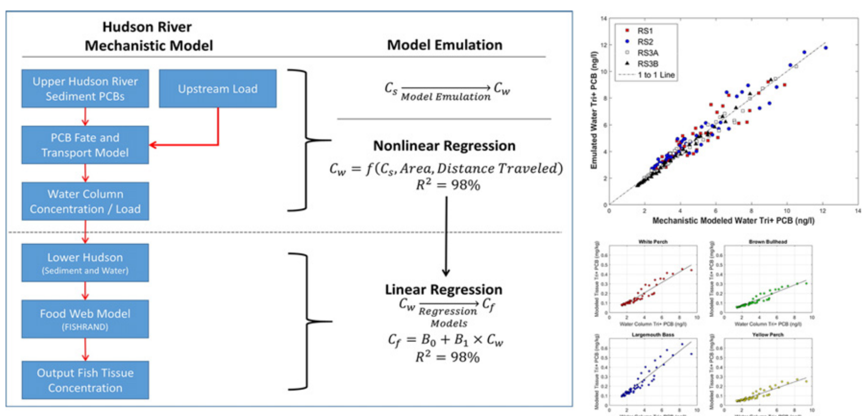
L. Jay Field [a,*], John W. Kern [b], Lisa B. Rosman [c]

[a] Assessment and Restoration Division, NOAA Office of Response and Restoration, 7600 Sand Point Way NE, Seattle, WA 98125, United States
[b] Kern Statistical Services, Inc., 5175 NE River Road, Sauk Rapids, MN, 56379, United States
[c] Assessment and Restoration Division, NOAA Office of Response and Restoration, 290 Broadway, New York, NY 10007, United States

## HIGHLIGHTS

- We emulated mechanistic model projections of fish PCBs in the lower Hudson River.
- Emulated models used updated sediment PCBs and recovery rate to revisit original predictions.
- Revised forecasts imply much longer time to recovery in lower Hudson River fish PCBs.
- Overestimating sediment recovery rates minimizes differences in remedial scenarios.
- Model emulation provides a mechanism to evaluate both bias and precision of models.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 8 October 2015
Received in revised form 10 February 2016
Accepted 10 February 2016
Available online 25 March 2016

Editor: Adrian Covaci

*Keywords:*
Model emulation
PCBs
Risk management
Hudson River
Natural recovery

## ABSTRACT

Remedial decision making at large contaminated sediment sites with bioaccumulative contaminants often relies on complex mechanistic models to forecast future concentrations and compare remedial alternatives. Remedial decision-making for the Hudson River PCBs Superfund site involved predictions of future levels of PCBs in Upper Hudson River (UHR) and Lower Hudson River (LHR) fish. This study applied model emulation to evaluate the impact of updated sediment concentrations on the original mechanistic model projections of time to reach risk-based target thresholds in fish in the LHR under Monitored Natural Attenuation (MNA) and the selected dredging remedy.

The model emulation approach used a combination of nonlinear and linear regression models to estimate UHR water PCBs as a function of UHR sediment PCBs and to estimate fish concentrations in the LHR as a function of UHR water PCBs, respectively. Model emulation captured temporal changes in sediment, water, and fish PCBs predicted by the mechanistic model over the emulation period. The emulated model, using updated sediment concentrations and a revised estimate of recovery rate, matched the trend in annual monitoring data for white perch and largemouth bass in the LHR between 1997 and 2014.

Our best predictions based on the emulated model indicate that the projected time to reach fish tissue risk-based thresholds in the LHR will take decades longer than the original mechanistic model projections.

Published by Elsevier B.V.

* Corresponding author.
*E-mail address:* jay.field@noaa.gov (L.J. Field).

http://dx.doi.org/10.1016/j.scitotenv.2016.02.072
0048-9697/Published by Elsevier B.V.

## 1. Introduction

Remediation decisions at large contaminated sediment sites with bioaccumulative contaminants often rely on highly parameterized mechanistic models to make long-range temporal projections comparing natural recovery and active remedial alternatives. At the Hudson River PCBs Superfund site in New York (Fig. 1), the U.S. Environmental Protection Agency (USEPA) used mechanistic contaminant fate and transport models linked to bioaccumulation models to predict future concentrations in fish (USEPA, 2000a, 2002). Model projections of temporal changes in fish concentrations played an important role in the comparative evaluation of remedial alternatives (USEPA, 2000b).

After USEPA's Record of Decision (ROD) (USEPA, 2002), extensive remedial design sediment sampling revealed that concentrations of PCBs in surface sediments were higher and more widespread than the models had predicted (Field et al., 2009; USEPA, 2010, 2012). Additionally, USEPA observed that PCB loads from the Upper Hudson River (UHR) to the Lower Hudson River (LHR) prior to the start of dredging in 2009 were substantially greater than predicted by the models and showed little evidence of decline (USEPA, 2010). Because modeled fish tissue PCB concentrations in the LHR are a function of PCB loads from the UHR, these findings imply that time to reach target thresholds for human consumption in fish in the LHR was underestimated by the original mechanistic model projections.

In this study, we used statistical model emulation to condense relationships between inputs and outputs of USEPA's linked mechanistic models to investigate sensitivity of model predictions to this new information. Model emulation reduces complex mechanistic models into computationally-efficient equations, dramatically reducing computational demands and time and effort to recalibrate and rerun the mechanistic models, while also maintaining a relevant and consistent representation of the underlying relationships within them (Logemann et al., 2004). The model emulator developed in this study was used to estimate new outputs associated with modified and updated inputs defining a range of remedial scenarios. The model emulator was also used to evaluate the sensitivity of model predictions to variation and uncertainty in initial sediment concentrations and different rates of natural recovery of surface sediment concentrations.

## 2. Methods

### 2.1. Study area

The Hudson River PCBs Superfund site extends approximately 321 km (200 miles) downstream from two General Electric (GE) capacitor manufacturing plants adjacent to the UHR to New York Harbor (Fig. 1). USEPA's ROD in 2002 (USEPA, 2002) called for dredging and monitored natural recovery (MNA) of PCB contaminated UHR sediments extending 64 km upstream from the Federal Dam at Troy. This area was divided into three main sections, River Sections (RS) 1 (Thompson Island Pool), RS2 (Schuylerville), and RS3. Because of its overall length, RS3 was subdivided into three modeling subsections RS3A (Stillwater), RS3B (Waterford) and RS3C (Troy). USEPA did not evaluate or select a remedy for the LHR tidal estuary (245 km between the Federal Dam and the Battery in New York City).

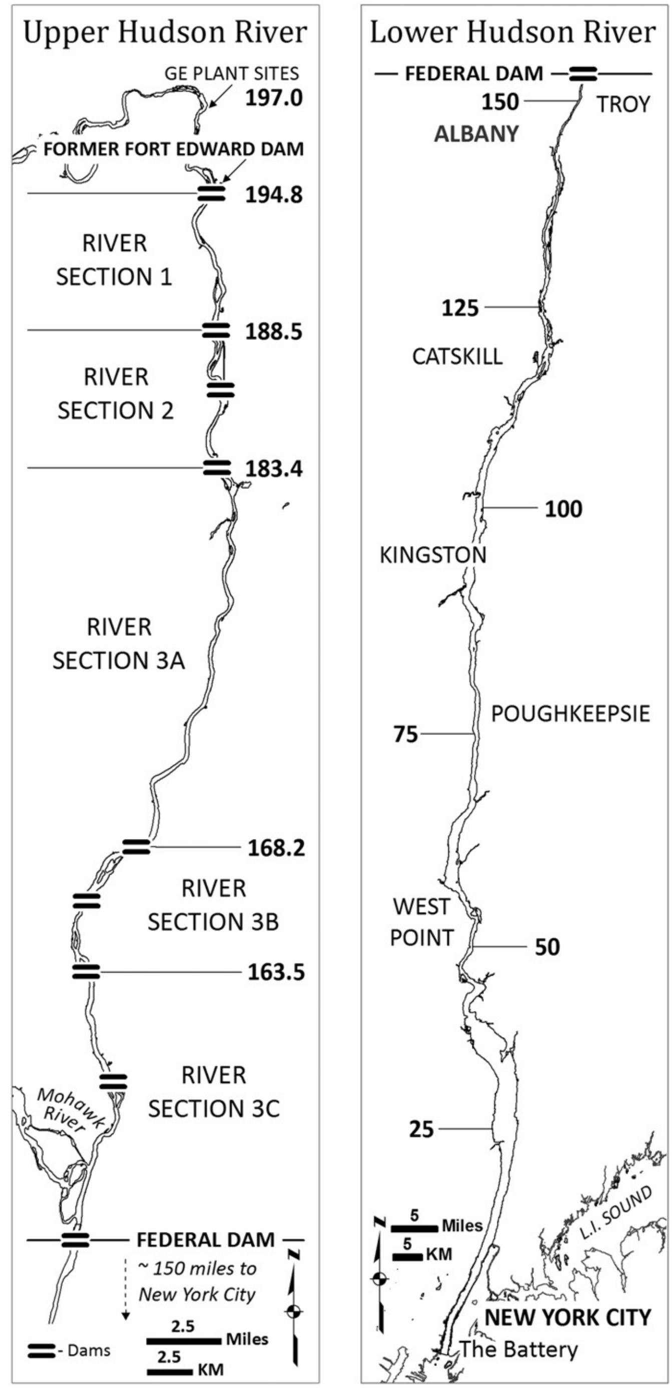



**Fig. 1.** Map showing the 321 km (200 mile) extent of the Hudson River PCBs Superfund site from Hudson Falls (above the GE plant sites) to The Battery in New York City. The left panel for the Upper Hudson shows the River Sections (RS) for the approximate 64 km (40 mile) remedial action area. The right panel for the Lower Hudson shows the 241 km (150 mile) tidal estuary with the fish model locations.

### 2.2. Sample sediment data

Sediment samples collected for PCB analysis between 1976 and 1999 by USEPA, GE and New York State were used during the Remedial Investigation and Feasibility Study (RI/FS) to assess risk and to predict future concentrations under various remedial scenarios (USEPA, 2000b, 2000c). Surface sediments were generally collected from the top 5 cm, although some penetrated as deep as 15 cm. Tri+ PCBs in water (average annual whole water concentrations), sediment and fish, the sum of trichlorobiphenyl and higher chlorinated homologues, were used for modeling because historic total PCB data did not effectively quantify mono- and di-chlorobiphenyl PCBs (USEPA, 2000a; Connolly et al., 2000). PCBs in fish tissue are primarily composed of Tri+ PCBs (USEPA, 2000a, 2002).

Subsequent to USEPA's ROD, GE collected sediment samples (mostly cores with some grab samples) from over 8000 locations throughout the UHR supporting design and implementation of the selected remedy. In RS1, most cores were collected on a triangular 24-meter (80-foot) grid from the entire pool. In RS2 and RS3, cores were collected almost exclusively within fine-grained sediments on triangular 24- or 50-

meter grids (QEA, 2002, 2005, 2007). This sampling design is considered approximately unbiased in RS1 and unbiased to fine-grained sediments in RS2 and RS3.

#### 2.2.1. Estimated pre-dredge concentrations

We averaged surface Tri+ PCB concentrations from design sampling conducted from 2002 through 2005 representing pre-dredge surface sediment concentrations in 2003. We used these averages for initial conditions comparing updated MNA and remedial (REM) scenarios.

Most (94%) of the samples represented the top 5 cm and the remainder were from the top 15 cm or less. Average concentrations from samples including intervals up to 15 cm in depth differed inconsequentially from samples composed of the 0–5 cm interval. The USEPA mechanistic model simulated PCB fate and transport in the top 4 cm.

#### 2.2.2. Estimated post-dredge surface sediment concentrations

Evaluating the change in surface sediment concentration following remediation required an estimate of expected post-dredging Tri+ PCB concentrations in sediment. Samples within the remedial design dredge footprints (Arcadis, 2013) were assigned a post-dredge surface sediment Tri+ PCB concentration of 0.25 mg/kg (USEPA, 2002) and arithmetic averages for each river subsection were recalculated to represent the concentration in 2003, the year USEPA expected dredging to commence.

#### 2.2.3. Estimated surface sediment concentration decay rate

Field et al. (2009) found that the exponential temporal decrease in sediment PCBs (exponential decay rates) estimated from USEPA's mechanistic models overstated the rate of natural recovery of surface sediments. GE conducted large-scale sediment surveys throughout the UHR in 1991 (O'Brien and Gere Engineers, Inc., 1993) and in 2002 through 2005 as part of remedial design (QEA, 2005, 2007). We compared average surface concentrations from these two surveys and calculated an exponential decay rate for each river section (Table 1). The average surface sediment Tri+ PCB concentration representing 2003 in each modeled river subsection was calculated, using only samples from the top 5 cm matching the top 5 cm sampling interval collected in 1991. In RS2 and RS3, these samples from 2003 can be considered representative of cohesive sediment deposits and directly comparable to samples from the cohesive sediment transects from 1991. By necessity, decay rate estimates for RS1 were based on comparison of remedial design samples, representing both cohesive and non-cohesive sediments, with samples representing cohesive sediments collected in 1991. Because cohesive sediments tended to have higher than average Tri+ PCB concentrations, the estimated decay rate is likely to overstate the actual rate. The overall average decay rate and confidence interval (CI) was used to guide selection of model emulation scenarios.

**Table 1**
Average surface (top 5 cm) sediment Tri+ PCB concentration (mg/kg) in 1991 and 2003 and estimated exponential decay rate.

| Model subsection | Cohesive sediment 1991[a] | Updated sediment 2003[b] | Exponential decay |
|---|---|---|---|
| 1 | 20<br>(227)[c] | 16.9<br>(3414)[c] | 1.4% |
| 2 | 18<br>(33) | 14.7<br>(1539) | 1.7% |
| 3A | 4.3<br>(103) | 3.4<br>(2129) | 2.0% |
| 3B | 5.7<br>(30) | 5.6<br>(682) | 0.1% |
| Average<br>95% confidence interval (CI) | | | 1.3%<br>(−0.1% to 2.6%) |

[a] O'Brien and Gere Engineers, Inc. (1993).
[b] Includes cohesive and non-cohesive sediments in River Section 1 and cohesive only in River Sections 2 and 3.
[c] Number of samples.

### 2.3. Selected remedy

The selected remedy, initiated in 2009, included both MNA and active remediation (dredging and backfill or capping followed by MNA) in the UHR. Sediment remediation areas were defined primarily on two criteria: surface concentrations (defined by USEPA as the top 30 cm) and mass-per-unit area (MPA), a measure of PCB inventory. Remediation areas were defined as follows: for RS1, a surface concentration of 10 mg/kg Tri+ PCBs in the surface or an MPA of 3 g/m$^2$ Tri+ PCBs; for RS2 and RS3, a surface concentration of 30 mg/kg Tri+ PCBs or an MPA of 10 g/m$^2$ Tri+ PCBs. Source control near GE plant sites, approximately 3 km upstream of the modeled area, was assumed under both MNA and active remediation scenarios.

### 2.4. Mechanistic model framework

The mechanistic numerical models developed by USEPA predicted sediment, water and fish Tri+ PCB concentrations in the RS1, RS2, RS3A, and RS3B reaches of the UHR (USEPA, 2000a). GE also developed similar mechanistic models that were generally consistent with those developed by USEPA (QEA, 1999a). USEPA used the projections of PCB load from the UHR (RS3B) to the LHR from the Upper Hudson River Toxic Chemical Model (HUDTOX) as input to the Farley model (Farley, 1999; USEPA, 1999) to calculate sediment and water concentrations in the LHR. Output from the Farley model was then used as input to USEPA's FISHRAND model, a mechanistic food web model, to predict Tri+ PCB concentrations in four species of fish (white perch, brown bullhead, largemouth bass, and yellow perch) at four LHR locations downstream of the Federal Dam at Troy (RM152 (Albany/Troy) (river kilometer [RK] 245), RM113 (Catskill) (RK 182), RM90 (Kingston) (RK 145), and RM50 (West Point) (RK 80) USEPA, 2002). While PCB-contaminated sediment in the UHR was the primary focus for remedial alternatives, reduction in PCB load to the LHR was a major remedial action objective and was expected to result in a reduction of PCB concentrations in lower river fish. Because initial PCB concentrations in LHR fish were lower than UHR fish, model projections indicated that LHR fish would reach human health risk management objectives (thresholds) much sooner than UHR fish.

We captured mechanistic model output by digitizing Tri+ PCB time series from the USEPA mechanistic model output for MNA and the selected remedy, including sediment (USEPA, 2000b: Figures 6-24, 6-26, 6-28, and 6-30; USEPA, 2002: Figures 363150-1, 3, 5, and 7) and water (USEPA, 2002: Figures 363150-10, -11, -12, and -13) for four model subsections in the UHR and fish at four locations in the LHR (USEPA, 2002: Figures 313787-2, 3, 4, and 5). Digitizing was accomplished using Plot Digitizer, a shareware Java program used to digitize scanned plots. Digitized sediment, water, and fish Tri+ PCBs time series were interpolated to equally-spaced annual time steps so that modeled values for each media could be paired temporally. Interpolation was conducted using linear interpolation using MATLAB© software (MATLAB 8.6, Release 2015b, The MathWorks Inc., Natick, MA, 2000). These time series simulated scenarios assumed dredging would begin in 2003 or 2004 and end by 2010 for the selected remedy.

### 2.5. Model emulation

Digitized input and output from mechanistic model projections provided a basis for using nonlinear optimization to fit a simplified mathematical model of water concentrations ($C_W$) in each UHR subsection as a function of 1) original and updated sediment Tri+ PCBs ($C_s$), 2) upstream source input (2 ng/L or 0 ng/L), 3) area of subsection, and 4) distance from the downstream dam in each subsection (see Supplementary Fig. 1). The emulated model structure is a simplified parameter version of the USEPA mechanistic model including four one-dimensional model compartments representing each river subsection.

#### 2.5.1. Model emulator

The model emulator represented each of the four river subsections with one model compartment composed of three terms representing PCB transfer to or from the water column: 1) upstream source minus deposition; 2) release/resuspension minus deposition of a fraction of these resuspended solids; and 3) post-dredge resuspension of disturbed residuals. The general form of the emulator within the *i*th subsection is:

$$Water\ Column\ Load_i = (Water\ Column\ Load_{i-1} - Deposition_i) + (Resuspension_i - Deposition_i) + Post\ Dredge\ Resuspension_i. \quad (1)$$

Each model compartment (i.e. river subsection) represents an impounded pool within which flows are generally laminar. Deposition of PCBs from the water column to the sediment bed was assumed proportional to distance traveled within each subsection with constant deposition rate per unit distance ($g_i,\ \ i = 1,2,3,4$) within river segments.

Release/resuspension of sediment PCBs to the water column was assumed to be directly proportional to average PCB concentration and area of PCB-containing cohesive sediments per river subsection with net sediment to water transfer coefficients ($\gamma_i;\ \ i = 1,2,3,4$) assumed constant through time.

Post-dredging sediment residuals were assumed to be more susceptible to resuspension with sediment to water transfer coefficients ($\beta_i;\ \ i = 1,2,3,4$) proportional to pre-dredge PCB concentrations and area dredged. These lower density disturbed residuals were assumed to decline with time at an 8% rate as they either flushed downstream, or became more consolidated and less susceptible to erosion.

Lower Hudson River fish Tri+ PCBs ($C_f$) were predicted from modeled water column Tri+ PCB concentrations ($C_w$) from the mechanistic model output for RS3B using linear regression.

#### 2.5.2. Emulator calibration

Net contaminant transfer coefficients were estimated by minimizing root mean squared error between temporally paired emulated and mechanistic modeled Tri+ PCB concentrations in water. The paired sediment and water time series for each of the 4 river sections spanned 30 years (2005–2034) for MNA and 25 years (2010–2034) for REM1 (the selected remedy) and each remedial scenario was modeled assuming: 1) partial source control with Tri+ PCB load decreasing from 0.16 kg/d to 0.0256 kg/d by the year 2005; and 2) complete source control, assuming upstream Tri+ PCB load would decrease from 0.16 kg/d to 0.0 kg/d (USEPA, 2000b). These 55 time steps and 4 river sections and 2 upstream load scenarios resulted in a system of 440 simultaneous nonlinear equations with 12 unknown net transfer coefficients which were solved using nonlinear optimization using MATLAB© scientific software (The MathWorks 2015). Full mathematical detail is provided in Appendix A. The estimated coefficients are summarized in Table S-1. Mechanistic water column Tri+ PCB concentrations from RS3B were treated as predictors of LHR fish Tri+ PCB concentrations and were calibrated by linear regression. Projections of LHR fish tissue Tri+ PCBs were calculated by applying this regression model to emulated water Tri+ PCB concentrations at the downstream end of RS3B.

Although we calibrated the model emulation to both upstream load scenarios, we found only small differences in future model projections of primary interest, so we focused on scenarios with average upstream source concentrations of 0.0256 kg/d (approximately 2 ng/L Tri+ PCB). This is reasonable because measured water column Tri+ PCB concentrations upstream of RS1 have been approximately 2 ng/L Tri+ PCB since 2004 (Farrar, 2011; USEPA, 2010). For the calibration step, we selected 2005 as the initial year for MNA because mechanistic model projections reached baseline concentrations of 2 ng/L Tri+ PCBs in that year. Initial year 2010 was selected for REM1 because dredging was anticipated to be completed by that time.

#### 2.5.3. Uncertainty

Analytical statistical theory for mechanistic simulation models is generally intractable due to their complexity, so statistical inference to model predictions is often limited. In situations where computer run-time for simulation models is relatively short, statistical inference may be available through Monte Carlo simulation or Bayesian Markov Chain Monte Carlo Methods (Raftery et al., 1995; Smith, 1994; USEPA, 1994). These examples have the commonality that mechanistic model equations are relatively simple and can be run repeatedly, a necessity for both Bayesian and Monte Carlo methods. Because linked fate and transport models often require extremely long run-times (Glaser and Bridges, 2007), Monte Carlo or Bayesian simulation is not directly applicable. Model emulation provides a solution to this computational problem by providing a surrogate model that can be run repeatedly within a reasonable period of time, while maintaining essential elements of the physical processes embodied in the mechanistic model. This advancement provides a mechanism to evaluate both bias and precision of models, providing risk managers with a more complete description of the reliability of predictions.

*2.5.3.1. Bias.* Our primary objective was to apply model emulation deterministically to evaluate bias in modeled forecasts associated with change in initial sediment bed Tri+ PCB concentrations. Future Tri+ PCB concentrations in sediment, water, and fish tissue were estimated using updated sediment Tri+ PCB concentrations reflecting averages from comprehensive remedial design sampling. Changes in these values associated with updated estimates of temporal decay rates in sediments were also considered. Using these modified model inputs, future Tri+ PCB concentrations in LHR fish were re-calculated and compared to human health total PCB risk thresholds of 0.05 mg/kg, 0.2 mg/kg and 0.4 mg/kg, representing levels protective of fish consumers eating one meal per week, one meal per month, and one meal every two months respectively (USEPA, 2002). USEPA considered Tri+ PCB and total PCB concentrations interchangeable in fish (USEPA, 2002).

These estimates representing central tendency or best estimates updated for new sediment surface and decay rates were compared with the original mechanistic model estimates.

*2.5.3.2. Precision.* We also estimated precision of model forecasts using parametric Monte Carlo simulation for auto-correlated time series of sediment Tri+ PCB concentrations. Synthetic sediment time series were generated that reproduced temporal autocorrelation patterns and between river section cross correlations similar to those in original EPA mechanistic modeled sediment time series. Each sediment Tri+ PCB concentration time series was simulated from a lognormal distribution with mean concentration

$$C_i(t) = C_{0i}e^{-kt+\varepsilon_i(t)}$$

where $C_{0i}$ is the initial sediment Tri+ PCB concentration in the *i*th subsection, and *k* is the PCB concentration decay rate. Because the sediment decay rate was estimated from just two points in time (1991 and 2003), we viewed this as a relatively uncertain parameter and as such investigated a relatively wide range of plausible decay rates uniformly distributed on the interval from 0.02 to 0.05. The residual time series $\varepsilon_i(t)$ was simulated as a normally distributed mean zero correlated random variable with autocorrelation and variance estimated from the residuals of an exponential fit to the mechanistic model time series. [The mathematical details of this probability model are summarized in Appendix B.]

This Monte Carlo simulation procedure involved four steps; 1) simulating four normally distributed auto-correlated sediment time series ($\varepsilon_i(t),\ \ i = 1,2,3,4$), 2) randomly selecting a uniformly distributed decay coefficient between 0.02 and 0.05, 3) calculating $C_i(t)$ and 4) applying the model emulator, to these four sediment time series, producing four corresponding Tri+ PCB time series for water and finally a synthetic fish tissue Tri+ PCB time series. These four steps were repeated 1000 times, and the fish Tri+ PCB time series were plotted, and the time to reach risk thresholds was calculated for each of the 1000 synthetic time series.

### 2.6. Remedial scenarios evaluated

Model emulation was used to evaluate the following remedial scenarios: (1) Mechanistic model projections for sediment PCB concentrations under Monitored Natural Attenuation (MNA1) and the selected remedy (REM1); (2) MNA (MNA2) and the selected remedy with updated sediment PCBs (REM2); and (3) An alternative remedial scenario (REM3), not considered in the ROD, that applies the RS1 cleanup target levels to RS2 and RS3 with updated sediment PCBs. For each of these scenarios, we applied both the original (8%) and the updated (3%) rate of exponential decrease in surface sediment PCBs.

## 3. Results

### 3.1. Model emulator

#### 3.1.1. UHR sediment to water

Fitting a set of nonlinear and linear regression models using inputs and outputs from the original mechanistic models provided a computationally simple means to reproduce the USEPA water column model Tri+ PCB results under MNA and selected remedy scenarios. The mechanistic model developed by USEPA predicted sediment and water Tri+ PCB concentrations in RS1, RS2, RS3A and RS3B that were used to compare remedial alternatives.

The four-compartment nonlinear model emulator with twelve parameters linking PCB transfer from sediment to water explained 98% ($R^2 = 0.98$) of the variation in mechanistic modeled water column concentration over the 30 year projection for MNA and the 25 year projection for REM1 (Fig. 2). This demonstrates that the model emulator successfully captures the changes in sediment and water concentrations predicted by the mechanistic model for MNA and for the selected remedy in the UHR model sections over the emulation period.

#### 3.1.2. UHR water to LHR fish

The mechanistic model predicted Tri+ PCB concentrations in four species of fish (white perch, brown bullhead, largemouth bass, and yellow perch) at four locations in the LHR (USEPA, 2002). Fish tissue Tri+ PCB concentrations in the LHR below the Federal Dam (RM152) had a strong linear relationship to water column Tri+ PCB at Waterford

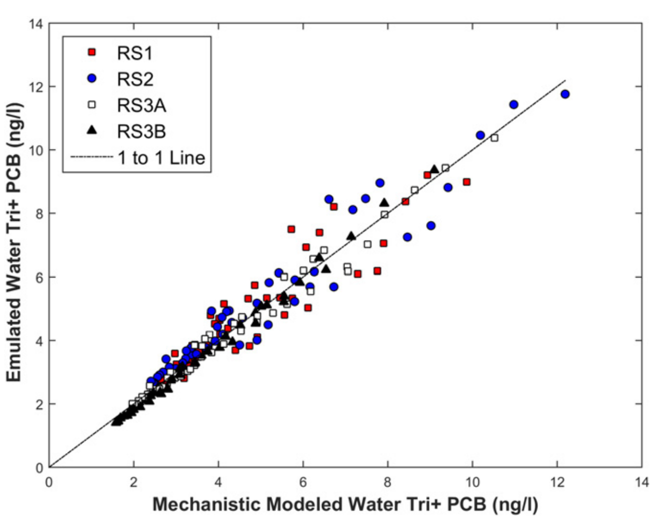



**Fig. 2.** Emulated vs original mechanistic model projected Tri+ PCB (ng/l) water concentrations by river subsections on the Upper Hudson River for MNA and the selected remedy.

(RS3B) in the UHR for all four modeled species ($R^2 \geq 0.90$, Fig. 3). This linear relationship between water Tri+ PCB at RS3B and LHR fish concentrations in the mechanistic model output provided the basis for the model emulation of fish PCBs.

Modeled fish tissue Tri+ PCBs for all four species at the other three LHR locations (RM113, RM90 and RM50) were also strongly linearly related to Tri+ PCB concentrations at Waterford, showing that the mechanistic model linking water to fish was effectively linear [Supplementary Table S-1 lists the regression coefficients and standard errors for white perch, brown bullhead, largemouth bass, and yellow perch at all four LHR locations].

Mechanistic food web model predictions of fish tissue concentrations for all four species at RM152 are strongly linearly related ($R^2 > 0.99$; Supplementary Fig. 2). Largemouth bass are predicted to have higher PCB concentrations than white perch, while brown bullhead and yellow perch are predicted to have lower concentrations.

Mechanistic model projections of white perch Tri+ PCB concentrations at RM113, 90, and 50 are also proportional to white perch Tri+ PCB concentrations at RM152 ($R^2 > 0.96$) and decrease with distance from the Federal Dam (Supplementary Fig. 3). The other three species had similar proportional relationships (not shown).

Emulation equations, with estimated coefficients, were applied to new model inputs such as new average PCB concentrations and decay rates in sediment (see Table A.2 for nonlinear regression coefficients).

The model emulation combined the nonlinear regression model between sediment and water with these linear regressions linking fish tissue and water column Tri+ PCBs to predict fish tissue Tri+ PCBs in the LHR from sediment Tri+ PCB concentrations in the four upper river sections. A comparison between the mechanistic model projections of Tri+ PCBs for all four species at RM152 and the emulation results are shown in Fig. 4 ($R^2 = 0.92$). Emulated model concentrations for largemouth bass and white perch tended to underestimate the mechanistic model at the higher concentrations (early in the time period).

#### 3.1.3. Updated surface sediment concentrations

The average Tri+ PCB concentration in sediment samples from the top 5 cm in 2003, exceeded the upper bound of the mechanistic model predictions (representing the top 4 cm) under MNA (MNA1) and were more than twice the mean concentration predicted for cohesive sediments in all four model subsections of the UHR (Table 2; Fig. 5). The GE mechanistic model for RS1 similarly understated average measured sediment PCBs in 2003 (QEA, 1999a).

The projected Tri+ PCBs concentrations in surface sediment under USEPA's natural recovery scenarios declined with an approximate 8% annualized exponential decay rate (USEPA, 2000a). Using the cohesive sediment data from the 1991 transect survey and the sediment data collected in 2003, we estimated the decay rate over the twelve year period to be 2% or lower in all four model sections (Table 1) with an average decay rate of 1.3% (95% CI = −0.1% to 2.6%). The 3% rate selected for simulated scenarios was a round number representing a reasonable upper bound for calculated decay rates shown in Table 1.

Dredging was expected to begin in 2003 and require 6 years to complete (USEPA, 2000b). In the emulation, we treated 2010 as the first post-dredging year. We assumed that natural recovery would continue outside the dredging footprint while dredging occurred. To estimate surface sediment concentrations in the initial post-dredging year, needed for simulating post-dredging scenarios, exponential decay rates of 8% and 3% were applied to the average surface concentration estimated from pre-design sampling in 2003. Post-dredging river-subsection averages were then calculated accounting for reduced concentrations due to dredging and backfilling (Table 2).

The post-dredging surface Tri+ PCB concentrations estimated for 2010 were also considerably higher than predicted by the USEPA models. In RS2 and RS3, where the target cleanup levels were at least a factor of 3 higher than for RS1, estimated post-dredging surface

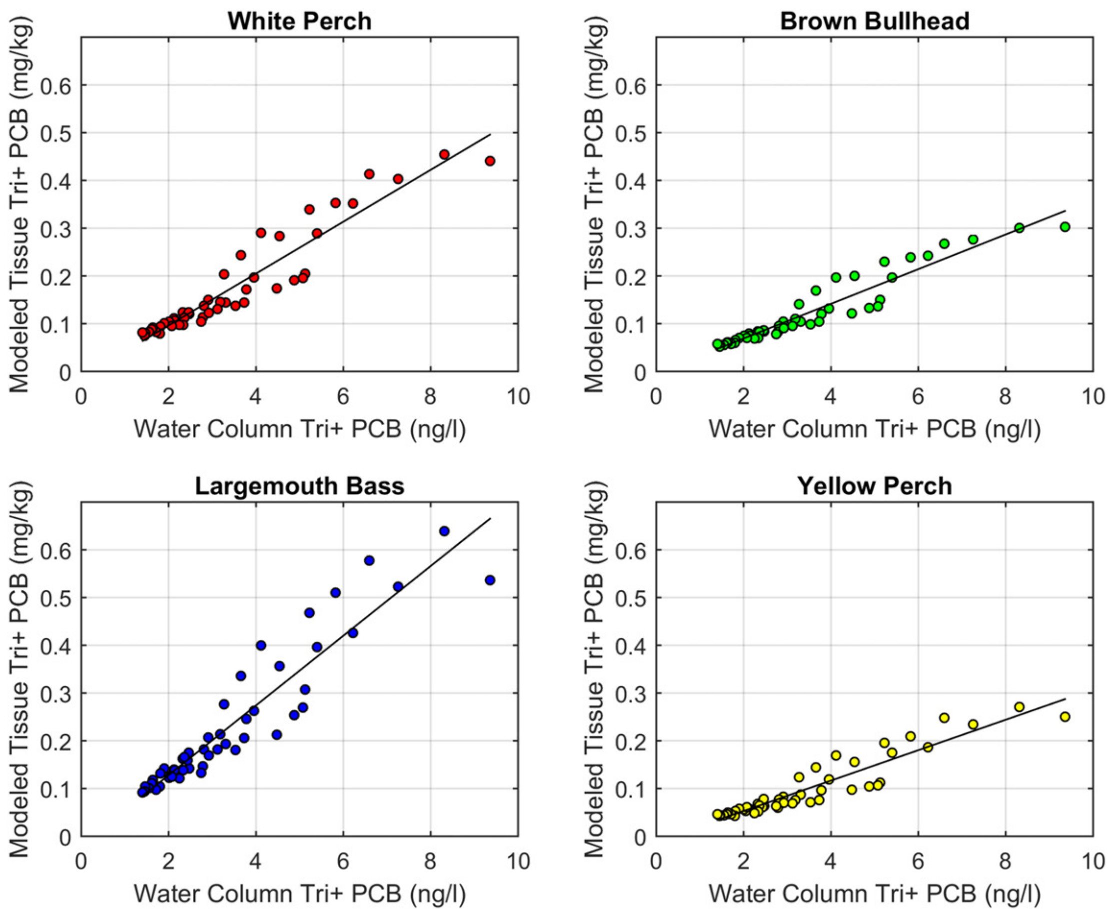



**Fig. 3.** Mechanistic model Tri+ PCB water concentrations (ng/l) at Waterford (RS3B) vs tissue concentrations (mg/kg) for white perch, brown bullhead, largemouth bass, and yellow perch from RM152 for MNA and the selected remedy.

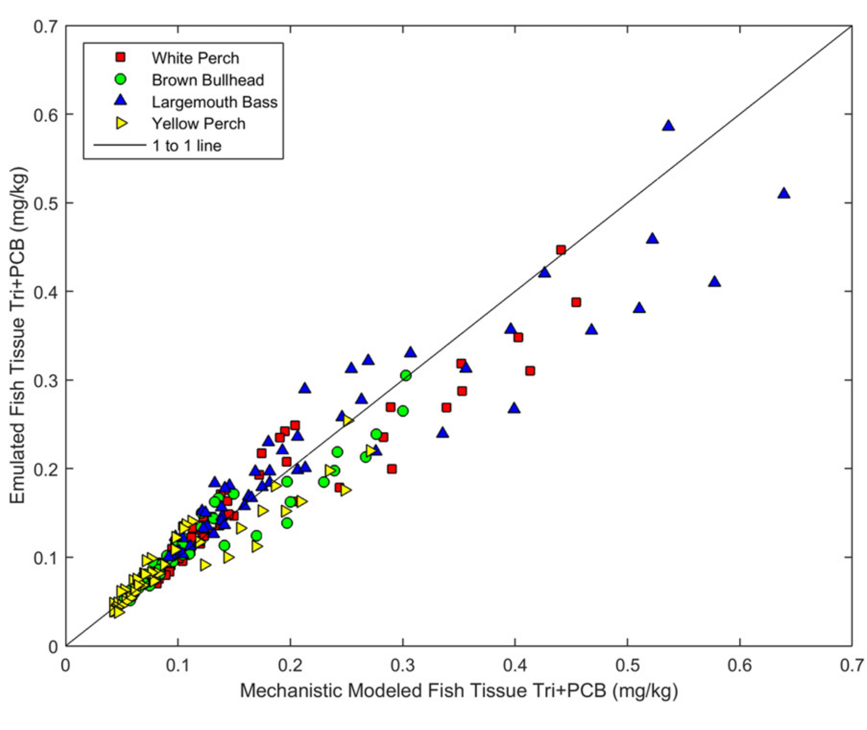



**Fig. 4.** Emulated vs original mechanistic model projected Tri+ PCB (mg/kg) fish concentrations for white perch, brown bullhead, largemouth bass, and yellow perch from RM152 for MNA and the selected remedy.

**Table 2**
Average Tri+ PCB concentrations (mg/kg) in surface sediment by river subsection under different remedial scenarios and rate of exponential decay in concentration between 2003 and 2010.

| River subsection | | Remedial scenario | | | | | |
|---|---|---|---|---|---|---|---|
| | Reach | MNA1[a] | MNA2[b] | REM1[c] | REM2[d] | REM2[e] | REM3[f] |
| | Year | 2003 | 2003 | 2010 | 2010 | 2010 | 2010 |
| | Decay | | | | 8% | 3% | 3% |
| RS1 | Thompson Island Pool | 8.5 | 16.9 | 0.5 | 0.8 | 1.1 | 1.1 |
| RS2 | Schuylerville | 6.5 | 14.7 | 1.0 | 2.8 | 3.9 | 1.0 |
| RS3A | Stillwater | 1.3 | 3.7 | 0.5 | 1.4 | 2.0 | 1.0 |
| RS3B | Waterford | 1.0 | 6.0 | 0.4 | 1.9 | 2.7 | 0.9 |

[a] MNA1: Mechanistic model predictions for Monitored Natural Attenuation for sediment concentrations in 2003.
[b] MNA2: Measured sediment concentrations in 2003 based on updated data.
[c] REM1: Mechanistic model predictions for the selected remedy for sediment concentrations post-remediation (2010).
[d] REM2: Estimated concentrations for the selected remedy post-remediation (2010) based on updated data, assuming 8% exponential decay since 2003.
[e] REM2: Estimated concentrations for selected remedy post-remediation (2010) based on updated data, assuming 3% exponential decay since 2003.
[f] REM3: Estimated post-remediation (2010) concentrations for hypothetical remedial scenario that applies RS1 cleanup levels to RS2 and RS3, based on updated data and assuming 3% exponential decay since 2003.

concentrations, based on updated data, are about 5 times higher than previously predicted based on the mechanistic model.

### 3.1.4. Emulated models with updated surface sediment concentrations pre- and post-removal

The effect of a lower natural recovery rate (3%) in sediment was also evaluated in combination with updated sediment surface Tri+ PCBs concentration. This updated decay rate is more consistent with the observed changes in surface concentrations during the 12 year period between the 1991 transect survey and the remedial design data collected in 2003, while not being overly conservative with respect to anticipated decay rates. The mechanistic model profile using USEPA's original projections of sediment concentrations under MNA (MNA1) and the selected remedy (REM1) was compared to the emulated model projections using an exponential decay rate of 8%. The computed exponential decay function closely matches the original model projections (Fig. 6), supporting the use of an exponential decay model for emulated results representing other decay rates (e.g., 3%) for surface sediment concentrations under MNA2, REM2 and REM3.

The emulated models projected LHR fish Tri+ PCBs using updated surface sediment concentrations (i.e., based on the 2003 pre-design sampling) as input. Estimates of pre- and post-removal surface sediment concentrations derived from the extensive remedial design sediment dataset (Table 2) provided more accurate characterization of surface Tri+ PCB concentrations prior to initiation of remediation.

Fig. 7 illustrates the difference between USEPA's original scenarios (MNA1 and REM1 with 8% decay rates) and updated scenarios (MNA2, REM2 and REM3 with updated sediment and 3% decay rates) for Tri+ PCB concentrations in white perch at RM152. The emulated LHR fish Tri+ PCB concentrations (MNA2, REM2, REM3) were substantively higher than USEPA's original mechanistic model predictions for MNA1 and REM1 and remain elevated over a much longer period. The updated sediment surface and decay rates for MNA2, REM2, and REM3 provide greater discrimination between remedial alternatives than in the evaluation of remedial alternatives prior to remedy selection.

The model emulator was used to estimate the number of years necessary to reach USEPA risk thresholds in white perch at RM152 under original modeled scenarios (MNA1, REM1) with the number of years to reach thresholds based on updated scenarios (MNA2, REM2, REM3) using two sediment exponential decay rates: 8% (mechanistic model) and 3% (upper bound of empirical estimate). Fig. 8 displays the number of years predicted to attain the 0.4 and 0.2 mg/kg Tri+ PCB thresholds for white perch at RM152 under remedial scenarios REM1, REM2 and REM3, each with 3% and 8% exponential decay rates. For all scenarios, using the updated sediment concentrations the time for fish tissue Tri+ PCB concentrations to reach remedial action objectives of 0.4 and 0.2 mg/kg is estimated to be substantively longer than originally predicted. For the original selected remedy (REM1) under either 8% or 3% decay assumptions, white perch at RM152 were projected to reach the 0.4 mg/kg threshold before or immediately after dredging was completed. With updated sediment concentrations (REM2) and 3% decay, white perch at RM152 were estimated to reach 0.2 mg/kg more than six decades longer than the original mechanistic model projections. The REM3 scenario greatly reduced the time to thresholds compared to REM2, but still longer than the original model predictions (REM1) [see Supplementary Tables S-2 and S-3 for time to 0.2 and 0.4 mg/kg thresholds for all scenarios, species, and locations].

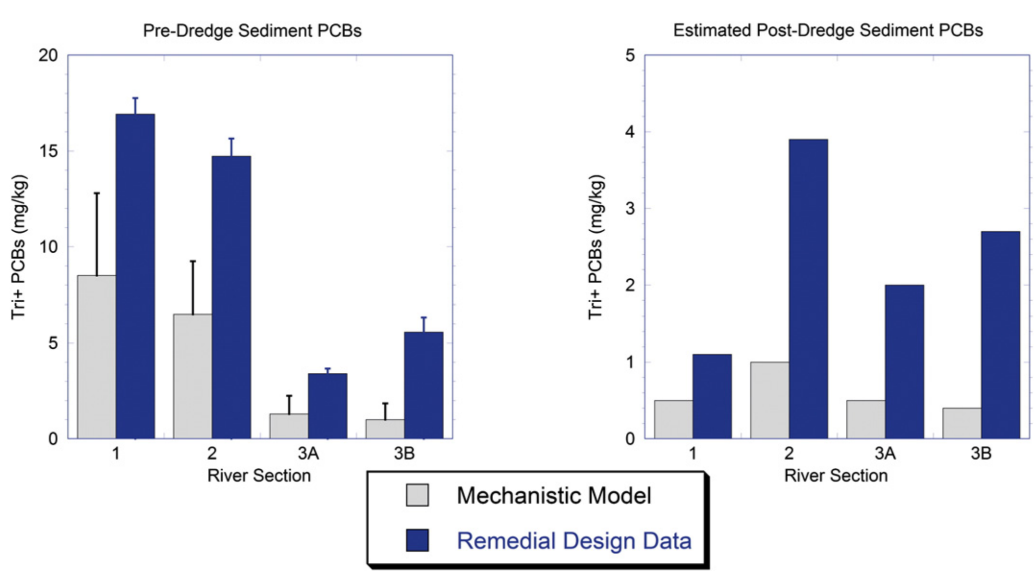



**Fig. 5.** Mechanistic model predictions of average and upper bound (error bars) surface sediment (top 4 cm) Tri+ PCB concentrations for 2003 pre-dredging (left panel) and post-dredge concentrations (right panel) compared to estimated river subsection average pre- and post-dredge sediment (top 5 cm) concentrations from remedial design sampling between 2002 and 2005 (approximately 2003).

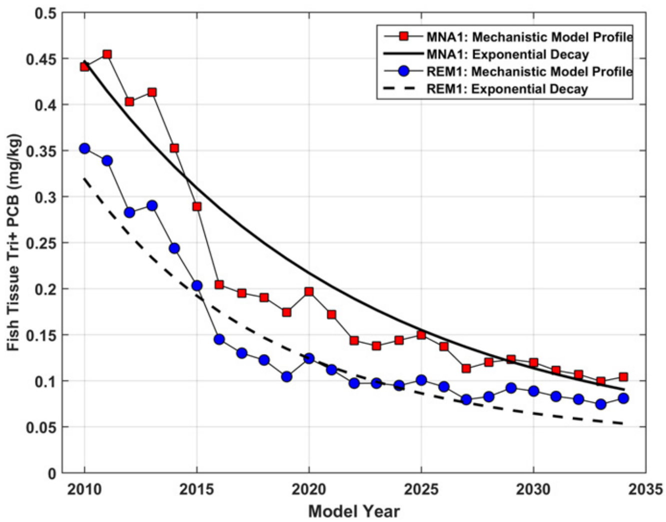



**Fig. 6.** Emulated model projections for white perch Tri+ PCB concentrations (mg/kg) from RM152 under MNA (MNA1) and the selected remedy (REM1) comparing the original mechanistic model (square and circle) results with simulated exponential decay rate of 8% (solid and dashed line).

## 3.2. Precision

Precision of emulator-based Tri+ PCB concentration in fish tissue was estimated using Monte Carlo simulation of equally likely sediment time-series with a range of decay rates (2% to 5%) and with statistical properties matching original mechanistic model sediment time series. The emulator was applied to these time-series, propagating uncertainty in sediment Tri+ PCB concentrations through to corresponding uncertainty in output Tri+ PCB concentrations in white perch at RM152. Fig. 9 shows the Monte Carlo distribution of future trajectories of fish tissue Tri+ PCB concentration, illustrating the uncertainty in estimates of the number of years needed to reach risk thresholds. The estimated number of years to thresholds were estimated to be 27 (95% CI: 19, 43), 49 (95% CI: 35, 77) and 102 (95% CI: 73, 162) for the 0.4 mg/kg, 0.2 mg/kg and 0.05 mg/kg risk based thresholds respectively.

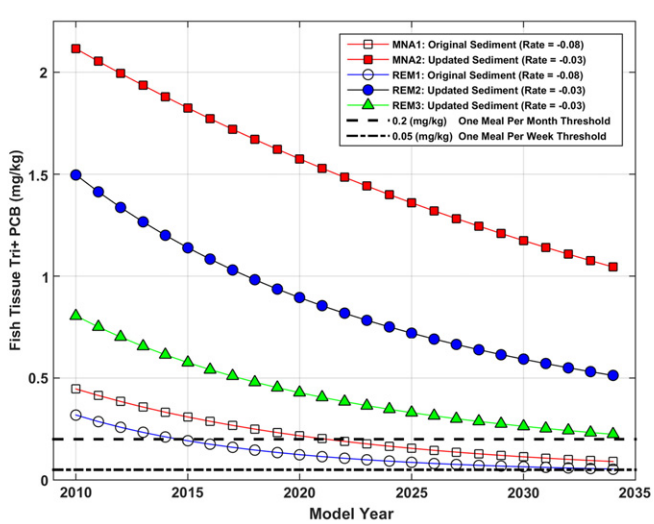



**Fig. 7.** Emulated model projections for white perch Tri+ PCB concentrations (mg/kg, wet weight) from RM152 for MNA (squares) and the selected remedy (REM) (circles) comparing the time to reach risk thresholds of 0.2 and 0.05 mg/kg at 8% (open symbols) and 3% (filled symbols) exponential decay rates for original mechanistic model concentrations (MNA1, REM1), updated sediment concentrations from remedial design sampling (MNA2, REM2), and hypothetical scenario that applies the RS1 target cleanup levels to RS2 and RS3 using updated sediment concentrations (REM3) (triangles).

# 4. Discussion

## 4.1. Interpretation of key findings

### 4.1.1. Model emulation

Model emulation provides a fast and inexpensive way to efficiently calculate outputs from inputs for complex mechanistic models, while retaining underlying physics-based properties. The method of model emulation is relatively new, with recent developments in global climate modeling stimulating the need to quantify uncertainty in complex mechanistic simulation models (Castruccio et al., 2014). An approach similar to ours was proposed by Margvelashvili et al. (2010) emulating a linked one-dimensional sediment/contaminant and three dimensional sediment transport model in the South-East Tasmanian coast of Australia.

For the Hudson River, sediment fate and transport model emulation successfully reproduced mechanistic model projections of sediment and water Tri+ PCB concentrations in the UHR and fish Tri+ PCB concentrations in the LHR. These results demonstrate that essential elements of the mechanistic mass balance model were captured by the emulator and support its validity for re-visiting temporal projections of fish tissue concentrations in the LHR with updated model inputs. Use of the emulator allowed us to update original predictions without necessitating access to computer codes that are often not readily available to third party investigators. Model emulation may also reduce the time to update complicated simulation models, because recalibration procedures may also entail re-evaluation of the physical mechanisms of the model itself. We believe these features of model emulation could enhance the transparency and accountability of the comparisons of alternative remedial scenarios.

### 4.1.2. Surface sediment concentrations and natural recovery

Extensive systematic remedial design sampling of surface sediment conducted to delineate dredge areas showed that the mechanistic model predictions of surface sediment concentrations underestimated surface PCBs under MNA and post-remediation scenarios and overestimated the rate of decrease in surface sediment PCBs. The higher than predicted post-remediation concentrations primarily resulted from high concentrations of PCBs in surface sediment adjacent to the planned dredge areas (Field et al., 2011).

Multiple reasons are possible for the mechanistic model underestimating surface sediment Tri+ PCBs, but processes that resulted in an overstated effective recovery rate (8%, MNA1 scenario) (as compared to our empirical estimate of <3% from data only available after the original model was developed) should be considered. Overestimated natural recovery rates are not unique to this model or this situation. For example, models developed by GE for the UHR had a similar effective decay rate (QEA, 1999a). Rates of recovery derived from data collected in the 1970s to mid-1980s have also led to overly optimistic estimates of rates of decline. Consistent with our findings, PCB concentrations in Great Lakes salmonids declined at high double digit rates in the 1970s and 1980s, but the inclusion of more recent data showed that declines have slowed to the low single digits in the 1990s and later (Rasmussen et al., 2014). Examination of PCB data from the 1970s to 2000s in several species of Great Lakes fish suggest that the estimates of contaminant decline were overly optimistic and responses to mitigation weaker than anticipated (Carlson et al., 2010; Sadraddini et al., 2011).

### 4.1.3. Estimated rate of recovery and fish concentrations

Monitoring data for adult white perch collected annually at RM152 in the late spring between 1997 and 2014 (NOAA, 2015) were normalized to 3% lipid for consistency with the USEPA FISHRAND model and overlaid on updated emulated model predictions for MNA (MNA2) at 3% and MNA1 at 8% decay. The original mechanistic model understates the measured tissue concentrations, whereas the updated predictions using 3% decay are more consistent with the measured data (Fig. 10).

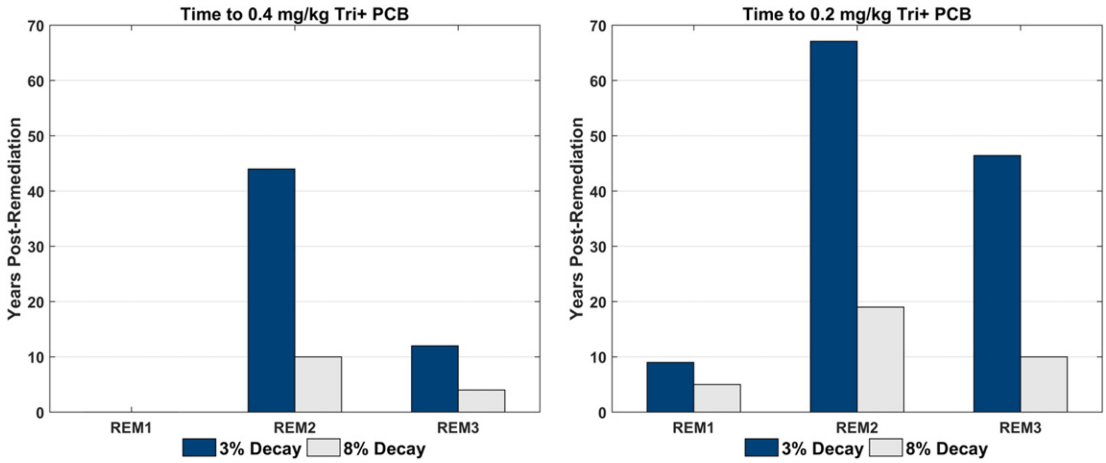



**Fig. 8.** Emulated model projections of the number of years to reach 0.4 and 0.2 mg/kg Tri+ PCB thresholds for white perch at RM152 under three remedial scenarios and two exponential decay rates, 3% and 8%: the selected remedy with original initial sediment concentrations (REM1), the selected remedy with updated initial sediment concentrations (REM2), and a hypothetical scenario that applies the RS1 target cleanup levels to RS2 and RS3 using updated sediment concentrations (REM3).

It could be argued that this apparently lower than expected decay rate in LHR white perch tissue concentrations is an artifact of Tri+ PCB releases from UHR dredging which began in 2009. However, the updated predictions equally describe trends in monitoring data collected between 1997 and 2009 (Fig. 10), supporting the lower than anticipated 3% recovery rate. It should also be noted that, due to a change in fish processing protocol between 2004 and 2013 (USEPA, 2015), lipid-adjusted Tri+ PCBs shown in Fig. 10 may understate actual concentrations during that time period. Adjusting these data for this change in protocol would shift Tri+ PCBs upward, suggesting even slower recovery rates, again supporting our finding that recovery rates are <8%. Similar results were observed for largemouth bass (Supplementary Fig. S-4). The monitoring data do not definitively identify the correct decay rate, but 3% is a demonstrably better fit to the data than 8%.

### *4.2. Use of model emulation to evaluate uncertainty*

Resource managers need to account for uncertainty in modeled forecasts to avoid selecting overly optimistic, or pessimistic, remedial options. For relatively simple measurement endpoints, statistical analyses are regularly used to quantify uncertainty. For example, uncertainty in exposure estimates is generally quantified using 95% confidence limits. When more complicated functions of the data are involved, the statistical methods of bootstrapping (Efron, 1979) and Monte Carlo simulation (Manly, 1991; USEPA, 1997) are used to describe uncertainty distributions. Bootstrap and Monte Carlo methods involve selecting equation inputs from statistical distributions to which model equations are applied, producing distributions of model outputs. Traditional metrics of uncertainty, such as confidence intervals or percentiles, are calculated directly from the output distributions. The time required to run linked sediment fate and transport models precludes direct application of bootstrap and Monte Carlo methods, because the model runs must be repeated many times to develop statistical distributions of output parameters.

Our model emulation provides a novel approach to extend the utility of complex linked sediment transport, contaminant fate and transport, and bioaccumulation models for the Hudson River by creating a computational shortcut that reliably predicts mechanistic model outputs from imperfectly known model inputs. By varying inputs to the model emulator (i.e. sediment concentrations and decay rates) within reasonably constrained ranges, the uncertainty distributions of emulated outputs were developed, simulating the uncertainty distributions of the mechanistic model. Importantly, because the mechanistic model is based on linked physical processes thought to be predictive, the model emulator can also be considered to be similarly predictive. The use of model emulation allowed for the investigation of the sensitivity of model outputs to uncertainty in model inputs, including both bias and precision.

#### *4.2.1. Bias*

The emulator was used in a deterministic way by modifying model inputs. The resulting mechanistic model forecasts were highly sensitive

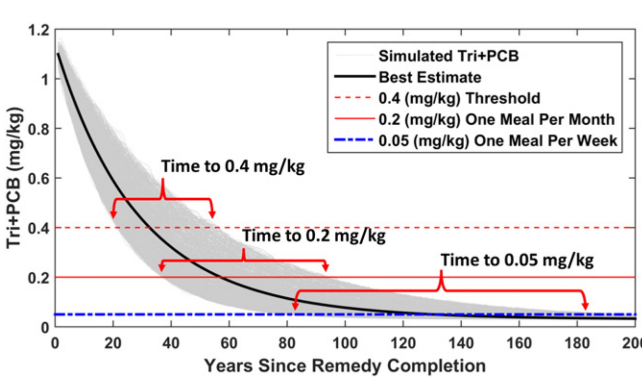



**Fig. 9.** Monte Carlo distribution of Tri+ PCB concentrations (mg/kg) in white perch at RM152 using the emulated model for the selected remedy with updated sediment concentrations and exponential decay rates in sediment Tri+ PCBs between 2 and 5%.

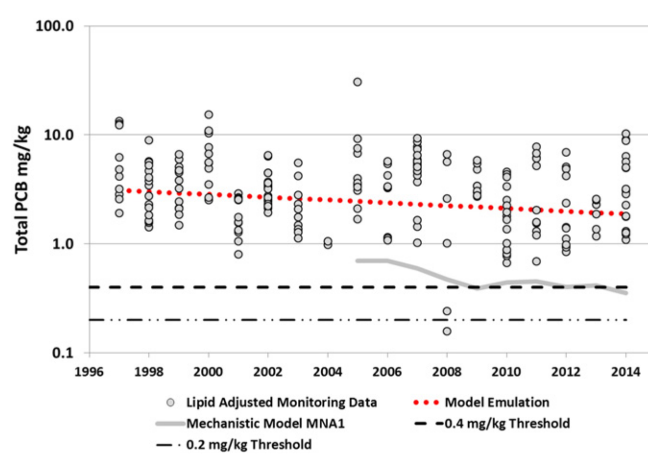



**Fig. 10.** Emulated model (dotted line) for white perch Tri+ PCB (mg/kg; normalized to 3.0% lipid) from RM152 with 3% exponential decay compared to monitoring data for white perch between 1997 and 2014 (circles) and risk thresholds (0.2 and 0.4 mg/kg PCBs) (horizontal dashed lines). Dredging began in 2009 and was completed in 2015.

to changes (e.g., bias) in initial sediment bed Tri+ PCB concentrations and temporal trend rates, but less so to variation in loads from upstream sources. This paper focuses on the scenario with upstream input concentration decaying to 2 ng/L Tri+ PCBs by 2005, which is consistent with recent monitoring data (USEPA, 2010). The mechanistic models indicated that recovery eventually would be limited with a 2 ng/L upstream baseline load compared to complete source control (upstream load = 0 ng/L). However, the emulated model for 0 upstream load (results not shown) did not differ much from the 2 ng/L model during the emulation period, possibly because initial higher than expected sediment concentrations and lower than expected decay rates mask relatively small differences due to upstream loads.

Updating input sediment Tri+ PCB concentrations to reflect more comprehensive, recent sampling led to the realization that concentrations observed in 2003 sample data exceeded the deterministic upper bound developed from the mechanistic model. Updating input bed sediment concentrations with this new information led to longer estimated recovery times for LHR fish, indicating reduced apparent benefit forecasted for the selected remedy.

The rate of natural recovery is more uncertain than surface sediment concentrations in 2003 because recovery estimates require comparisons with data from older sampling programs, which were based on subjective sampling designs and much smaller sample sizes. Although no completely unbiased sediment sampling program had been conducted prior to 2003, the 1991 UHR transect survey (O'Brien and Gere Engineers, Inc., 1993) was closest to an unbiased systematic sampling study with spatially extensive coverage and many sampling locations distributed throughout the UHR. Lack of unbiased estimates of mean surface concentration at multiple points in time limit the potential to accurately estimate the natural recovery rate. For our study of bias in the decay rates, we used 3% because, while we believe that our sediment decay rate estimate is the best available, the fact that it is based on just two time steps and because only one time step is based on a completely unbiased sampling design, the estimate of 1.3% exponential decay is highly uncertain. Therefore, for evaluating bias, we used 3% as a value that is meaningfully <8%, yet not overly pessimistic. Such subjectivity about sediment recovery rates, at one of the most heavily studied Superfund sites in the United States, is disconcerting and should stimulate a focus on improving the estimate of the rate of recovery at other contaminated sites where remedial alternatives are being evaluated.

#### 4.2.2. Precision

The precision of model forecasts was estimated using a parametric Monte Carlo approach to simulate autocorrelated time series of bed sediment Tri+ PCB concentrations. Sediment concentration inputs were modeled as a first order (i.e. exponential) decay function with temporally correlated residual errors. Application of the model emulator to the 1000 sets of simulated sediment time series resulted in corresponding ensembles of water and fish tissue time series. As discussed above, temporal recovery rates at the Hudson River site are highly uncertain, so the effects of this uncertainty were incorporated into this analysis by simulating first order decay rates as a range of values uniformly distributed from 2% to 5%. This range was chosen subjectively, but nonetheless the analysis illustrated that even modest uncertainty in decay rates can translate into a wide range of estimated times to recovery (Fig. 8). This result indicates that reliable estimates of exponential decay rates in contaminated media are required for reliable remedial alternatives comparisons.

Each of the 1000 simulated time series varies through time around its selected exponential decay rate. When data are strongly correlated temporally, concentration time series may wander far from the exponential decay curve for significant periods of time, leading to greater uncertainty in estimates of time to threshold values. Although results were not shown, the Monte-Carlo procedure was used to evaluate effects of temporal autocorrelation by holding the exponential decay rate fixed across all 1000 simulations. This analysis showed that times to reach threshold concentrations were insensitive to these types of excursions of sediment concentrations due to autocorrelation.

If large linked contaminant fate and transport models are to be used for remedial alternatives evaluation, supporting sediment data appropriate for estimating temporal decay rates are necessary. Frequently, high resolution geochronology sediment cores are used to deduce sedimentation rates and indirectly extrapolate natural recovery rates that are often extrapolated over large spatial regions. However, exposures to biotic receptors are generally assumed proportional to spatial averages, which may not be adequately represented by a small number of high resolution cores. This problem is likely exacerbated by the tendency for investigators to rely on high resolution cores with interpretable geochronology, which typically are collected in low energy areas with continuous deposition and greater than average sedimentation rates that are not representative of site conditions (USEPA, 1998; QEA, 1999b). Those rates, which could be considered to represent an upper bound on sedimentation rates, are then extrapolated over large areas with varying energy regimes and less interpretable geochronologies.

The model emulation approach was useful for quantifying bias and precision of mechanistic model forecasts of fish tissue Tri+ PCB concentrations at the Hudson River. Further application of the method is recommended at contaminated sediment sites where large contaminant fate and transport models have been developed for use in remedial decision-making. Model emulation at other large sites should provide further support for utilizing this approach when additional site data become available to evaluate model projections.

### 4.3. Improving model calibration and validation

Following the approach used by Castruccio et al. (2014), model emulation can also improve the objectivity and efficiency of model calibration and validation by using a mechanistic model to "pre-calculate" a relatively wide range of model input and output combinations from which a model emulator can be developed. The emulator is then used to iterate on model inputs until optimal combinations of input parameters minimizing error between outputs and sample data are obtained. The emulator provides a mechanism to efficiently calculate combinations of inputs and outputs, allowing many more combinations of model parameters to be evaluated than would otherwise be possible using the mechanistic model directly.

This approach would provide an understanding of the full range of inputs calibrating to the sample data. Combinations of model parameters resulting in similar model fit to data would be considered to represent similarly likely scenarios. If only a small range of model parameters fit the data well, one would conclude that the available data are adequate to uniquely identify the most likely model. In this situation, one could be confident in model projections, whereas a broad range of model parameter combinations resulting in similar model fit to data, would suggest that the sample data are inadequate to uniquely identify a likely model. In this situation, one would not ascribe a great deal of confidence in modeled projections.

### 4.4. Implications for remedy selection

The model emulation results demonstrate the importance of generating an accurate estimation of both surficial sediment concentrations and the rate of natural recovery of the sediment surface in order for mechanistic models to provide useful information for decision-makers on the relative comparisons among remedial alternatives. If the model-predicted rate of natural recovery is too high, the magnitude of the difference between MNA and an active remedy or between various active remedies, such as the selected remedy for the Hudson River site and a more comprehensive alternative, will be underestimated. USEPA considered two alternative dredging scenarios: the selected remedy (REM1) and a full section removal. The full section removal scenario essentially doubled the area to be dredged (additional 190 ha). According to USEPA's review of alternatives, full section removal would have been more protective, but the projected difference in fish concentrations (and risk) between the two remedial scenarios was considered too small to warrant the increased cost (USEPA, 2002). The difference between those two dredging alternatives was understated because of the overly optimistic rate of recovery of the surface sediment considered. This is illustrated in Fig. 7, which clearly discriminates between the different alternatives and shows the large difference in time to reach risk thresholds for emulated fish concentrations for the selected dredging remedy (REM1) and for the updated scenario with a more aggressive (but less than full section removal) remedy (REM3). This hypothetical remedy, which maintained the same target cleanup levels for surface sediment throughout the UHR, would involve removing an estimated additional 71 ha, <50% of the area under the full section removal scenario.

While we estimate risk thresholds would be reached meaningfully sooner under this hypothetical and more aggressive remedy (REM3) than under the selected remedy with updated sediment surface concentrations and decay rate (REM2), the estimated time to thresholds would still be longer than the original mechanistic model projections (REM1). Our analysis suggests that achievement of LHR fish PCB threshold concentrations targeted as remedial action objectives to protect human health will be delayed for up to several decades. Our analysis also implies that the remedial action objectives will not be met in the time frame identified in the 2002 ROD for the Hudson River (USEPA, 2002) without implementing a more comprehensive remedy.

Models are often considered to be most useful for evaluating uncertainty in predictions of the relative, as opposed to absolute, benefits for alternative remedial options (Glaser and Bridges, 2007). In such situations management teams may rationalize potential inaccuracies in model forecasts by assuming that relative comparison of forecast remedial effectiveness is possible even when absolute forecasts may be inaccurate or highly uncertain. Our analyses suggest that when models are biased or imprecise the relative differences between remedial alternatives can be significantly under- or over-estimated. In addition, the model emulation approach can serve to improve precision and reduce bias in model output, therefore more reliably discriminating among remedial alternatives. Box and Draper (1987) stated "All models are wrong, some are useful". The models discussed in this paper rely on accurate surface sediment concentrations and the rate of change to make reliable projections of concentrations in sediment, water, and biota. The best way for resource managers and decision-makers to know if the models used for comparing remedial options are useful is to collect systematic, unbiased data on surface sediment concentrations that can be used to estimate the rate of natural recovery and to regularly monitor fish tissues for bioaccumulative contaminants.

## 5. Conclusions

Our analyses demonstrate that pre-remedial surface sediment Tri+ PCBs in the Upper Hudson River were two to three times higher and estimated post-remediation Tri+ PCBs averaged about four times higher than predicted by the original mechanistic models used by USEPA in the Hudson River 2002 ROD. The rate of recovery, as measured by the exponential decay rate of Tri+ PCBs in surface sediment, was overestimated by the original mechanistic models. We estimated a mean of 1.3% and a 95% upper CI of ~3% compared to the ~8% derived from the original EPA and GE mechanistic models.

The emulated models successfully reproduced the mechanistic model projections for sediment and water in the UHR and fish in the LHR. The emulated models were used to incorporate the updated information on higher surface sediment concentrations and reduced rate of sediment recovery. Our model projections suggest that the original mechanistic model projections greatly underestimated the time to reach risk thresholds in the LHR fish, thereby extending by decades the time period for the project to reach its fish PCB-based remedial action objectives in the LHR.

The results also demonstrated the adverse impact of over-estimation of the rate of sediment recovery on the potential ability of risk managers to discriminate among alternative remedial scenarios.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2016.02.072.

## Acknowledgments

Funding for this work was provided by the National Oceanic and Atmospheric Administration (NOAA) Office of Response and Restoration. This manuscript benefited from the comments of internal and external peer reviewers. Data management and digitization were provided by EXA Data and Mapping Services. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of NOAA.

## Appendix A. Mathematical formulation for emulator

Table A.1 summarizes the locations of the four dams (*River Mile* = $d_i$), acres of cohesive sediments ($A_i$), distances between dams ($\delta_i = d_i - d_{i-1}$), area remediated and average distance between deposits and downstream dams ($\overline{d}_i$). Table 2 lists the Tri+ PCB concentrations ($c_{si}$) in surface sediment in 2003 and 2010 for each of the five scenarios evaluated in this study. The load at the $i$th dam is represented by $L_i$ and the transfer coefficients from water to sediment and sediment to water are represented by $\gamma_i$ and $g_i$ respectively. With this notation, the processes for deposition and resuspension at each model annual timestep were described mathematically in the following set of four equations which are nonlinear in the transfer coefficients

$$L_i = L_{i-1} \times (1 - g_i \times \delta_i) + \left\{\gamma_i \times (c_{si} \times A_i) \times \left(1 - g_i \times \overline{d}_i\right) + \beta_i \times (R_i \times c_{si} \times A_i)\right\} \times Q_i \quad \text{(A.1)}$$

where $i = 1,2,3,4$ indexes each of the four modeled sections of the river, $\beta_i$ represents the sediment to water net transfer coefficient for dredged residuals and $R_i$ represents the 8% decay of post-dredge residual concentrations. If discharge at successive dams is similar ($Q_i = Q_{i-1}$), Eq. (A.1) can also be expressed in terms of water column concentrations as opposed to loads by dividing both sides of Eq. (A.1) by $Q_i$ giving the following Eq. (A.2).

$$c_{wi} = c_{wi-1} \times (1 - g_i \times \delta_i) + \left\{\gamma_i \times (c_{si} \times A_i) \times \left(1 - g_i \times \overline{d}_i\right) + \beta_i \times (R_i \times c_{si} \times A_i)\right\} \quad \text{(A.2)}$$

For the Hudson River, results were similar for Eqs. (A.1) and (A.2) so the simpler Eq. (A.2) was used for these analyses.

Each of the 25 years from 2010 through 2034 provides a different set of modeled sediment bed and water column Tri+ PCB concentrations from which the best estimates of emulator net transfer coefficients ($g_i, \gamma_i$ and $\beta_i$, $i = 1,2,3,4$) can be estimated using constrained nonlinear least squares. These paired inputs and outputs from the EPA mechanistic model were available for two remedial scenarios; 1) natural recovery (MNA1), and 2) the selected remedy (REM1A). Each of these scenarios was also simulated with the assumptions of 0 and 2 ng/l PCBs entering from upstream of RS1. Modeled time series spanning 30 (2005–2034) and 25 (2010–2034) year time frames for MNA and active remediation respectively, under two sets of upstream input assumptions and four river sections provided 440 ($2 \times 25 \times 4 + 2 \times 30 \times 4$) nonlinear equations in 12 unknown net transfer coefficients (i.e., $g_i, \gamma_i$ and $\beta_i$, $i = 1,2,3,4$). The transfer coefficients were estimated

**Table A.1**
Summary of input parameters and initial conditions for calibrating model emulator.

| Reach | River section | Downstream river kilometer | River section length (km) | Area (ha) Cohesive sediment area (ha)[a] | Alternative REM1 remediated area[b] | Alternative REM2 remediated area[c] | Alternative REM3 remediated area[d] |
|---|---|---|---|---|---|---|---|
| Thompson Island Pool | RS1 | 303.4 | 10.1 | 42 | 114 | 124 | 124 |
| Schuylerville | RS2 | 295.2 | 8.2 | 54 | 31 | 35 | 56 |
| Stillwater | RS3A | 270.7 | 24.5 | 93 | 38 | 29 | 64 |
| Waterford | RS3B | 263.1 | 7.6 | 52 | 17 | 13 | 28 |
| Total | | | | | 200 | 201 | 272 |

[a] Cohesive sediment area from Tables 5.2a–5.2b in USEPA (2000b).
[b] Area for alternative REM1 from Tables 8–9 in USEPA (2000a).
[c] Alternative REM2 area calculated based on delineated dredge area.
[d] Alternative REM3 area based on delineated dredge area for the selected remedy and additional area estimated from number of cores exceeding RS1 target cleanup levels.

**Table A.2**
Estimated model emulation nonlinear regression coefficients.

| Model coefficients | River section RS1 | RS2 | RS3A | RS3B |
|---|---|---|---|---|
| Water to sed | 0.0000 | 0.0350 | 0.0157 | 0.0641 |
| Sed to water | 0.0160 | 0.0095 | 0.0078 | 0.0451 |
| Post dredge resuspension | 0.0251 | 0.0143 | 0.0283 | 0.0357 |

by constrained nonlinear least squares with MATLAB© Release 2011a (The MathWorks 2011).

## Appendix B. Probability model for synthetic sediment time series

The residual process $C_i(t) = C_{0i}e^{-kt+\varepsilon_i(t)}$ was simulated by randomly drawing an exponential decay rate ($k$) from a uniform probability distribution on the interval 0.02–0.05, followed by simulation of $\varepsilon_i(t)$ as a mean zero normally distributed random variable with covariance matrix $\boldsymbol{C}$ with the entries $c_{ij}$ defined as $cov(\varepsilon_i(t),\varepsilon_i(t+h)) = e^{-a_i h2}$, and covariance between subsections $i$ and $j$ given by $cov(\varepsilon_i(t),\varepsilon_j(t)) = c_{ij}$ for $i \neq j$. The constants $a_i$ and $c_{ij}$ were estimated from the four mechanistic modeled sediment Tri+ PCB concentration time series. The expected mean of the simulated sediment series for the $i$th subsection is $C_{0i}e^{-kt}$. The simulated series are distributed log-normally because $\varepsilon_i(t)$ is a normally distributed random variable.

The estimated coefficient $a_i$ defining the rate of decline in temporal auto correlation was 0.1. The resulting correlation matrix C was a real symmetric banded matrix with diagonal entries $C_{ii} = 1.0$ and with 5 non-zero off diagonal with values $C_{i,i\,\pm\,j} = 1$, 0.90, 0.67, 0.41, 0.20, and 0.08; for $j = 1, 2, \ldots, 5$ respectively and $i = 1, 2, 3, \ldots, 200$ years. The remaining values $C_{i,i\,\pm\,j} = 0$; for $j > 5$.

## References

Arcadis of New York, Inc. (Arcadis), August 2013. Phase 2 Final Design Report for 2013, Addendum No. 3, Hudson River PCBs Superfund Site. Prepared for General Electric Company.
Box, G.E.P., Draper, N.R., 1987. Empirical Model Building and Response Surfaces. John Wiley & Sons, New York, NY.
Carlson, D.L., Vault, D.S., Swackhamer, D.L., 2010. On the rate of decline of persistent organic contaminants in lake trout (*Salvelinus namaycush*) from the Great Lakes, 1970–2003. Environ. Sci. Technol. 44, 2004–2010.
Castruccio, S., McInerney, D.J., Stein, M.L., Liu Crouch, F., Jacob, R.L., Moyer, E.J., 2014. Statistical emulation of climate model projections based on precomputed GCM runs*. J. Clim. 27, 1829–1844.
Connolly, J.P., Zahakos, H.A., Benaman, J., Ziegler, C.K., Rhea, J.R., Russell, K., 2000. A model of PCB fate in the Upper Hudson River. Environ. Sci. Technol. 34, 4076–4087.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.
Farley, K.J., 1999. An Integrated Model of Organic Chemical Fate and Bioaccumulation in the Hudson River Estuary. Environmental Engineering Department, Manhattan College.
Farrar, K., 2011. Updates on remedial program at the GE Hudson Falls plant site. Presentation to the USEPA Community Advisory Group Meeting, June 30 (http://www.hudsoncag.ene.com/files/NYSDEC%20presentation%20on%20GE%20Hudson%20Falls%20Remediation.pdf).
Field, L.J., Kern, J.W., Rosman, L.B., 2009. Evaluation of Natural Recovery Models for Sediment in the Upper Hudson River. (February, http://www.darrp.noaa.gov/northeast/hudson/pdf/Battelle09_Field_NatRecovery_508.pdf).
Field, L.J., Rosman, L.B., Brosnan, T., Foley, R., 2011. Hudson River remedy: unremediated PCBs and the implications for restoration. http://www.darrp.noaa.gov/northeast/hudson/pdf/Battelle1_Field.final1.pdf (February).
Glaser, D., Bridges, T.S., 2007. Separating the wheat from the chaff: the effective use of mathematical models as decision tools. Integr. Environ. Assess. Manag. 3, 442–449.
Logemann, K., Backhaus, J.O., Harms, I.H., 2004. SNAC: a statistical emulator of the north-east Atlantic circulation. Ocean Model. 7, 97–110.
Manly, B., 1991. Randomization and Monte Carlo Methods in Biology. Chapman and Hall, New York.
Margvelashvili, N., Parslow, J., Herzfeld, M., Wild-Allen, K., Andrewartha, J., Rizwi, F., et al., 2010. Development of operational data-assimilating water quality modelling system for South-East Tasmania. OCEANS 2010 IEEE-Sydney. IEEE, pp. 1–5.
NOAA, 2015. Hudson River watershed database and mapping project. http://response.restoration.noaa.gov/environmental-restoration/environmental-assessment-tools/query-manager-marplot-data-maps.html (Last accessed: 03/20/2015).
O'Brien & Gere Engineers, Inc., 1993. Data summary report, Hudson River project. 1991 Sediment Sampling and Analysis Program. Prepared for General Electric Company Corporate Environmental Programs. 1993. O'Brien and Gere Engineers, Inc., Syracuse, New York.
Quantitative Environmental Analysis (QEA), 1999a. PCBs in the Upper Hudson River PCBs. Volume 3 Predictions of Natural Recovery and the Effectiveness of Active Remediation. Prepared for General Electric Company, Albany, New York (May 1999 Amended July 1999).
Quantitative Environmental Analysis (QEA), 1999b. PCBs in the Upper Hudson River PCBs. Volume 1 Historical Perspective and Model Overview. Prepared for General Electric Company, Albany, New York (May 1999).
Quantitative Environmental Analysis (QEA), 2002. Hudson River PCBs Site Design Support Sediment Sampling and Analysis Program Field Sampling Plan. Prepared for General Electric Company, Corporate Environmental Programs, Albany, NY (July, http://www.epa.gov/hudson/Design_Support_FSP_10_final.pdf).
Quantitative Environmental Analysis (QEA), 2005. Hudson River PCBs Site Phase 1 Dredge Area Delineation Report. Prepared for General Electric Company, Albany, NY (February 28, http://www.epa.gov/hudson/pdf/2005_2_28_phase1_DAD_text_tables_appendices.pdf).
Quantitative Environmental Analysis (QEA), 2007. Hudson River PCBs Site Phase 2 Dredge Area Delineation Report. Prepared by for General Electric Company, Albany, NY (December 17, http://www.epa.gov/hudson/pdf/phase2_dad_report.pdf).
Raftery, A.E., Givens, G.H., Zeh, J.E., 1995. Inference from a deterministic population dynamics model for bowhead whales. J. Am. Stat. Assoc. 90, 402–416.
Rasmussen, P.W., Schrank, C., Williams, M.C.W., 2014. Trends in Lake Michigan coho and chinook salmon, 1975-2010. J. Great Lakes Res. 40, 748–754.
Sadraddini, S., Azim, M.E., Shimoda, Y., Mahmood, M., Bhavsar, S.P., Backus, S.M., Ahronditsis, G.B., 2011. Temporal PCB and mercury trends in Lake Erie fish communities: a dynamic linear modeling analysis. Ecotoxicol. Environ. Saf. 74 (2011), 2203–2214.
Smith, R.L., 1994. Use of Monte Carlo simulation for human exposure assessment at a superfund site. Risk Anal. 14, 433–439.
USEPA, 1994. Technical Guidance Manual. Use of Monte Carlo Simulation in Risk Assessments. EPA903-F-94-001 — February 1994. http://www.epa.gov/reg3hwmd/risk/human/info/guide1.htm (Last accessed: 02/10/2015).
USEPA, 1997. Guiding principles for Monte Carlo Analysis. United States Environmental Protection Agency. EPA/630/R-97/001. http://www.epa.gov/raf/publications/pdfs/montecar.pdf (Last accessed 5/22/2014).
USEPA, 1998. Phase 2 Report — Review Copy Further Site Characterization and Analysis Volume 2C-A Low Resolution Sediment Coring Report Addendum to the Data Evaluation and Interpretation Report Hudson River PCBs Reassessment RI/FS July 1998, Prepared for the For U.S. Environmental Protection Agency Region II and the U.S. Army Corps of Engineers Kansas City District by Tams Consultants, Inc. Gradient Corporation, and Tetra Tech, Inc. (Last accessed: 02/10/2015).
USEPA, 1999. Phase 2: Further Site Characterization and Analysis, Volume 2E-A Baseline Ecological Risk Assessment for Future Risks in the Lower Hudson

River, Hudson River PCBs Reassessment RI/FS. December 1999. Prepared for the For U.S. Environmental Protection Agency Region II and the U.S. Army Corps of Engineers Kansas City District by Tams Consultants, Inc. and Menzie-Cura & Associates, Inc. (http://www.epa.gov/hudson/low-hud-future-era.pdf (Last accessed: 02/10/2015)).

USEPA, 2000a. Revised Baseline Modeling Report (RBMR), Hudson River PCBs Reassessment RI/FS Volume 2D.Prepared Prepared for USEPA Region 2 and USACE, Kansas City District by TAMS Consultants, Inc., Limno-Tech, Inc., Menzie-Cura & Associates, Inc., and Tetra Tech, Inc. (January 2000. http://www.epa.gov/hudson/reports.htm (Last accessed: 02/10/2015)).

USEPA, 2000b. Hudson River PCBs Reassessment RI/FS Phase 3 Report: Feasibility Study. TAMS Consultants, Inc. (December 2000. http://www.epa.gov/hudson/study.htm (Last accessed: 02/10/2015)).

USEPA, 2000c. Phase 2 Report, Further Site Characterization and Analysis. Volume 2E Revised Baseline Ecological Risk Assessment, Hudson River PCBs Reassessment RI/FS. Prepared for USEPA Region 2 and the USACE, Kansas City District by TAMS Consultants, Inc. and Menzie-Cura Associates, Inc. (November. http://www.epa.gov/hudson/revisedbera-text.pdf. (Last accessed: 02/10/2015)).

USEPA, 2002. Hudson River PCBs Site Record of Decision and Responsiveness Summary. January 2002. Prepared for the For U.S. Environmental Protection Agency Region 2 and the U.S. Army Corps of Engineers Kansas City District by Tams Consultants, Inc. (http://www.epa.gov/hudson/d_rod.htm#record (Last accessed: 02/10/2015)).

USEPA, 2010. Hudson River PCBs Site EPA Phase 1 Evaluation Report, Prepared for: US Environmental Protection Agency, Region 2 and US Army Corps of Engineers, Kansas City District, Prepared by:The Louis Berger Group, Inc., March 2010. (http://www.epa.gov/hudson/pdf/2010-03-15_Phase_1_Evaluation_Report_Text.pdf (Last accessed 2/10/2015)).

USEPA, 2012. First Five-year Review For Hudson River PCBs Superfund Site. U.S. Environmental Protection Agency, Region 2, New York, NY (June 2012, http://www.epa.gov/hudson/pdf/Hudson-River-FYR-6-2012.pdf (Last accessed 2/10/2015)).

USEPA, 2015. Letter from Judith Enck (Regional 2 Regional Administrator) to Ned Sullivan (President, Scenic Hudson), September 10.