Comparison of Measured and Predicted Environmental PCB Concentrations Using Simple Compartmental Models

ALISON CULLEN *

The Daniel J. Evans School of Public Affairs, Mailbox 353055, University of Washington, Seattle, Washington 98195-3055

The use of models to represent biological, chemical, and physical processes that govern the fate and transport of environmental contaminants is an enduring feature of risk assessments. Data collection is costly and time-consuming. Measuring future conditions is impossible regardless of the resources available. For these reasons, rarely do analysts have sufficient empirical data for estimating risks in a population of interest across the desired dimensions of space and time. The appropriate level of complexity, detail, and resource investment in a modeling exercise should be established by the intended use of the results generated and by the expected performance of the available modeling options. At one end of the spectrum, the concentration of a contaminant in an environmental medium may be estimated quickly and inexpensively using an intermedia partition coefficient or simple steady state compartmental model. In contrast, a complex dynamic model requiring vast stores of input data, computer power, and run time may be used to estimate concentration. However, because models are generally used when data are scarce or nonexistent, our ability to assess the accuracy and precision of various model options is often limited. The research presented below exploits a relatively unusual opportunity to compare concentration measurements of polychlorinated biphenyls (PCBs) for multiple environmental media with predictions from simple compartmental fate and transport models using two-dimensional Monte Carlo analysis. Simple compartmental models are found to predict measurements quite well overall, although decisions about the treatment of variability and autocorrelation in the airborne load of contaminant with season, weather system, and location influence their performance. The models are assessed at two sites near New Bedford Harbor in Massachusetts, one characterized by higher and more variable contaminant concentrations, while the other is a comparison or “background” site. The difference between model performance in the two locations illustrates some characteristics of situations in which simple models are most appropriate. Under background conditions of relatively low and consistent contaminant levels, a two compartment model generates excellent predictions of the sum of PCB congener concentration in soil based on air concentration. Under more contaminated or more variable conditions, model results are less predictive; however, most fall within an order of magnitude of the data. A comparison of model performance for predicting concentration of the sum of PCB congeners vs predictions of individual PCB congeners is also pursued. Individual congener concentrations in soil tend to be slightly overpredicted for lighter weight congeners, i.e., those more characteristic of the New Bedford Harbor region, while levels of heavier congeners tend to be underpredicted in circumstances where air concentrations are relatively low. A three compartment model representing air, soil, and plant matter is found to predict levels of PCBs in edible produce within about an order of magnitude of those measured, under the conditions considered. This work demonstrates the relevance and usefulness of results from simple and easily implemented models for fate and transport predictions.

Introduction

The use of models to represent biological, chemical, and physical processes that govern the fate and transport of environmental contaminants is an enduring feature of risk assessments. Data collection is costly and time-consuming. Measuring future conditions is impossible regardless of the resources available. For these reasons, rarely do analysts have sufficient data about the concentration of contaminants in environmental media for estimating risks in some specified scenarios. Contaminant concentration may be estimated quickly, inexpensively, and with a modest amount of input data using simple steady state compartmental models or, in contrast, using complex dynamic models requiring vast stores of input data, computer power, and run time. The former is frequently selected as a result of the advantages stated (1–4); yet, a crucial step in opting to use any model is believing that it will give relevant results. In light of the magnitude of resource expenditures to address human and ecological health risks, it is especially important to understand the quality of the predictions. However, because models are generally used when data are scarce or nonexistent, our ability to assess the accuracy and precision of model predictions is often limited.

A multipathway exposure assessment between 1993 and 1995 in the community surrounding the New Bedford Harbor Superfund site in Massachusetts provides an unusual opportunity to consider the performance of simple compartmental models to represent intermedia transport (5–7). This data set contains measurements of 58 polychlorinated biphenyl (PCB) congeners in outdoor air, soil, and backyard produce at several locations near the site (5). Ongoing work includes a comparison of the performance of the simple models assessed in this work with the performance of more complex models under the same conditions and also an exploration of Bayesian melding to improve model performance (6).

Approach

Compartmental models are used to predict concentrations of contaminants in soil and plants based on measured concentration in air and other model inputs. These models are developed elsewhere (2, 3); however, brief descriptions are included here. Two-dimensional probabilistic simulation in which variability and uncertainty are kept distinct is used to generate distributions of predicted concentration. In this context, variability refers to temporal, spatial, or interindi-
individual differences (heterogeneity) in the value of a model input. Uncertainty is a measure of the incompleteness of knowledge or information about an unknown quantity whose true value could be established if a perfect measuring device were available. The model predictions are compared to measured concentrations of PCBs in soil and plant tissue (7–9). The measurements are plotted alongside the predictions in log-probability space as distributions of concentration representing variability. Model predictions are generated for two sites in Massachusetts, Acushnet and Dartmouth. Acushnet is a town directly adjacent to New Bedford Harbor and downstream of the most contaminated portions. Dartmouth is a rural town 5 miles upwind of the harbor contamination. The analysis is carried out for total PCBs (the sum of 58 individual congeners), as well as for three individual congeners, which span the range of molecular weight.

Simple Compartmental Model for Soil Concentration. The long-term average concentration of a semivolatile contaminant in soil, $C_s$, reflects a dynamic equilibrium between multiple processes. For this class of chemicals, a balance between deposition of airborne contaminant to soil and degradation of the contaminant in the soil dominates the determination of concentration. Although a fraction of the deposited material is intercepted en route to the ground by plants and other structures, weathering processes ensure that most particles ultimately reach the ground. Likewise, the fraction of contaminant removed from the soil by plants via root uptake is small in relation to the reservoir that accumulates over the course of years; thus, it can be ignored in the calculation of soil concentration.

A differential equation describing the change in soil contaminant concentration with time, $C_s(t)$, according to these processes is

$$\frac{dC_s}{dt} = -vC_s + \frac{V}{D_p}C_a$$

where the subscript $j$ allows an accounting for PCB in various states including occurring as vapor, bound to particles of diameter $d$, and bound to particles of diameter $>2\mu m$ and $V_j =$ deposition velocity of particle or vapor class $j$ (m/day); $C_a =$ annual average concentration in air, in vapor, or particle class $j$, in soil (1/day); $t =$ duration of soil contact with contaminated air (day); $D =$ depth of mixed layer of soil in which crop is grown (m); $p =$ bulk density of soil (g/m$^3$).

$C_s$, the steady state solution for concentration of PCBs in soil (in mg/g) summed across vapor and particle categories, is given by

$$C_s = \sum_j \frac{VC_a}{bD_p}$$

Simple Compartmental Model for Concentration in Plant Tissue. Contaminant accumulates in plant tissue via two pathways: (i) deposition onto foliar plant surfaces and subsequent incorporation into tissues and (ii) uptake through the roots and translocation via the transpiration stream (3). The increase in chemical concentration in plant tissues due to foliar contact with airborne contaminant reflects a dynamic equilibrium between two processes: weathering of contaminant mass from the plant surfaces and deposition of airborne contaminant. The steady state solution is represented by the first term in eq 3. The efficiency of root uptake has been estimated with experimental measurements and is represented by the second term in eq 3. The concentration in the tissues of an initially uncontaminated plant, after a growing season of $t$ days, is estimated as

$$C_p(t) = \sum_j \left[ C_{aq}V_j \left(1 - \frac{\exp(-Wt)}{W}\right) + UC_s \right]$$

where $W =$ weathering rate constant of particulate adhering to crop (1/day); $Y =$ yield of crop per cultivated area (g/m$^2$); $I =$ fraction of contaminant intercepted by crop plants during deposition (%); $t =$ duration of the growing season (days); and $U =$ root uptake factor, i.e., the experimentally determined ratio of the concentration in plant tissues to the concentration in soil (%).

Input Distributions. Probabilistic assessment requires an analyst to assign distributions to represent variability and uncertainty in model inputs and/or about model form. In many analyses, model form is a subject of intense debate since a lack of knowledge about correct model form contributes significantly to overall uncertainty. Because the purpose of this analysis is to explore the performance of simple models, we take their form as given and develop distributions to represent the variability and uncertainty in model inputs. It is assumed that the comparison to measurements carried out here contributes insight about model uncertainty. Several of the required inputs are not expected to vary greatly from place to place, thus distributions developed for other purposes are used to represent them. These inputs include mixing depth of soil, soil bulk density, length of growing season, breakdown rate of PCBs in soil, interception effectiveness of plants, weathering rate of PCBs adhering to plants, root uptake rates, and yield of plant per area (3–4). Distributions to represent contaminant concentrations in air and soil and the deposition velocity of particles and vapor are developed for the specific location and conditions of interest as outlined below (10).

Concentration of PCBs in Ambient Air. There are multiple sources of variability and uncertainty to be considered during distribution development for concentration of PCBs in air from measured values. A set of measured values for concentration of a chemical in air is assumed to represent random draws from an underlying log-normal distribution since the concentration in air is generated by dilution processes (11, 12). First, one must consider temporal variability in measurements at a single location. Also, there is measurement uncertainty, which has been estimated at ±10% in these samples (8, 9). Next, random sampling error is introduced given the inherent reliance on a relatively small set of measurements. We account for random sampling error by assigning distributions to the parameters of the concentration distribution—the Student’s $t$-test and Chi square to the mean and standard deviation, respectively. This second-order assignment of distributions is the heart of two-dimensional probabilistic analysis. Finally, the representativeness of the measurements is unknown; however, given our research design, it is not possible to quantify the resulting uncertainty.

An often overlooked complexity in the development of distributions to represent concentration of a contaminant in air is autocorrelation. Autocorrelation arises from the fact that the magnitude of concentration on a given day is likely related to that on the previous day. This occurs because a weather pattern, and its accompanying evaporative conditions, tends to settle in an area for several days at a time. Accounting for this dependency reduces the overall variance in concentration. In the following two step method, annually averaged concentration is distributed assuming that the number of independent concentrations across the time period of interest is smaller than the number of days in that time period, resulting in a reduction in variance (10).

First, the mean and variance of a log-normal distribution are themselves distributed, Student’s $t$-test and Chi square,
TABLE 1. Summary Statistics for Air Concentration Measurements

<table>
<thead>
<tr>
<th>PCB congener</th>
<th>Acushnet</th>
<th>Dartmouth</th>
</tr>
</thead>
<tbody>
<tr>
<td>congener sum</td>
<td>ln(Ca) annual average (C in ng/m³)</td>
<td>μln(Ca)</td>
</tr>
<tr>
<td>28</td>
<td>1.6</td>
<td>-0.8</td>
</tr>
<tr>
<td>101</td>
<td>-0.6</td>
<td>-3.6</td>
</tr>
<tr>
<td>153</td>
<td>-2.2</td>
<td>-3.8</td>
</tr>
</tbody>
</table>

a Source, ref 8.

respective (13). The population mean is distributed

\[ \mu_{\ln(x)} \sim \ln(x) + \frac{s_{\ln(x)}}{n^{1/2}} t_{n-1} \]  

where \( \bar{x} \) is the mean of the logged air sample measurements, \( s_{\ln(x)} \) is the sample standard deviation of the logged air measurements, \( n \) is the number of samples, and \( t_{n-1} \) is the Student’s t distribution with \( n - 1 \) degrees of freedom.

Second, the average concentration is distributed according to a lognormal distribution (7). The mean of the average annual concentration in air accounting for dependency among daily values, \( \mu_{\ln}(x) \), is distributed

\[ \mu_{\ln}\langle x \rangle = \ln\left( \frac{\mu_x}{1 + \frac{\sigma_x^2 M}{\mu_x^2}} \right) \]  

where \( \mu_x \) is the mean of the air concentration measurements and \( M \) is the number of weather cycles contained in 1 year. Approximately 73 cycles of 5 days duration are assumed for this case (14).

The standard deviation of the average annual concentration in air accounting for dependency among daily values, \( \sigma_{\ln}(x) \), is distributed

\[ \sigma_{\ln}\langle x \rangle = \frac{\ln(1 + \frac{\sigma_x^2 M}{\mu_x^2})^{1/2}}{} \]  

Summary statistics for measured concentration of PCBs in air appear in Table 1.

Deposition Velocity. The velocity of deposition of a mass of airborne contaminant varies depending on whether the mass is bound to particulate matter or exists as a vapor. Large particles (roughly > 2 μm) deposit primarily by gravitational Stokes settling, and vapor deposition is predominantly by Brownian diffusion (15, 16). Small to medium size particles (≤2 μm) fall less quickly than either large particles or vapors. Small to medium size particles are too big to diffuse effectively and have insufficient mass for significant gravitational pull. A description of deposition velocity must therefore include information about the degree to which contaminant is bound to particles and the size distribution of those particles. For this analysis, we use vapor–particle partitioning findings (17, 18) and the EPA’s assumptions about deposition velocity with distributions developed as in Cullen (10); see Table 2.

Several sets of results are used in this comparison. First, measured concentration of PCBs in soil is compared to predictions from eq 2 for congeners 28, 101, and 153 as well as the sum of 58 of the most prevalent congeners for both the background site, Dartmouth, and the downwind site, Acushnet. Second, measured concentration of PCBs in tomatoes and potatoes grown at the Dartmouth site is compared to predictions from eq 3 for the sum of 58 congeners.

Results

Measured concentrations of PCBs are plotted on a log-probability scale alongside model predictions in Figures 1–3. Results for soil appear in Figures 1 and 2, for Dartmouth and Acushnet, respectively. Figure 3 contains illustrative results for tomatoes from the John George Farm in Dartmouth, MA. Figures 1–3 plot the percentiles of variability in model predictions of concentration (2.5th, 50th, and 97.5th percentile). For all figures, the band between the 2.5th and 97.5th percentile of variability is the 95% confidence interval on that percentile, i.e., uncertainty about the true concentration representing that percentile of variability. Measured concentrations of PCB are plotted as points on all figures for clarity of comparison, although it should be noted that measurement error bars of ±10% apply (8, 9). The sparseness of the measurements challenges statistical approaches to comparison. However, the figures allow a direct comparison of each distribution of predictions with the variability distribution of measurements, its median, and extreme percentiles.

Discussion

Reliance on fate and transport models to predict concentration in environmental media produces estimates that incorporate several sources of variability and uncertainty. Model uncertainty is incorporated in these estimates as the approach assumes that the model is correct for the purpose of exploring its performance. Temporal and spatial variability, random sampling error, and measurement error are represented in the form and parametrization of the input distributions to the models.

Upwind of the harbor at the Dartmouth site, the simple compartmental model for soil concentration based on

TABLE 2. Model Input Distributions

<table>
<thead>
<tr>
<th>category (j)</th>
<th>deposition velocity (m/day)</th>
<th>ln(V)</th>
<th>fraction of PCB mass in each vapor or particle category (j)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>sum of PCBs</td>
</tr>
<tr>
<td>vapor</td>
<td>U(1,3)</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>particles ≤ 2 μm</td>
<td>U(0,2)</td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>particles &gt; 2 μm</td>
<td>U(1,3)</td>
<td></td>
<td>0.42</td>
</tr>
</tbody>
</table>

a V, deposition velocity, is represented with a log uniform distribution. ln(V) thus follows a uniform distribution parametrized as indicated.

b Fraction of PCB mass existing as a vapor, adhering to small particles, or adhering to large particles is based on work of Whitby (17) and Falconer and Bidleman (18). c Sum of 58 PCB congeners measured.
measurements in air performs well. For the lightest congener considered, number 28 characterized by relatively high vapor pressure and low molecular weight, the model overpredicts slightly at the low end of the concentration distribution; however, the predicted median is within a factor of 3 of that measured (Figure 1a). The predictions for congener 101, a more chlorinated and consequently heavier compound, are in the range of the data and represent both the median and the variability in the measurements well (Figure 1b). For the most chlorinated and heavy congener, number 153, the model underpredicts the median and upper percentiles of variability (Figure 1c). At the upper percentiles of variability, the measured concentrations are about an order of magnitude above any value predicted by the model. Predictions of the sum of PCB concentrations (Figure 1d) are in the range of the data throughout and are quite reliable overall, similar to the congener 101 results. Variance is accurately predicted and fairly modest. Overall, the model predicts the range of the measured values very well with slight overprediction of the central tendency in some cases. Measured concentrations as high or higher than those predicted are observed. The overall strong performance of the model is consistent with the characteristics of this “clean” background location where concentrations are neither very high nor temporally variable. As congeners of increasingly large molecular weight are tackled by the model, it is observed to underpredict concentration. This finding should be interpreted in light of the particular physical chemistry of PCB. Specifically, heavier congeners adhere to fine particles rather than existing in the vapor phase. Particles in this range of aerodynamic diameter deposit more slowly than vapor (which moves by Brownian motion). The result is a potentially exaggerated reduction in modeled mass flux from air to soil. A striking consequence of these facts is that the unknown degree of partitioning of PCBs between vapor and solids in air, a traditional measurement challenge for chemists, contributes significantly to the overall uncertainty in this analysis. This important source of uncertainty continues to challenge modeling attempts to describe PCB fate and transport.

Downwind at the Acushnet site, the simple compartmental model performs less well than upwind. For congener 28, the median is predicted within a factor of 3 of that measured (Figure 2a). The concentrations of congeners 101 and 153 are overpredicted (Figure 2b,c) throughout the range of the data. The low end of the predicted 95% variability band closely bounds the measurements of concentration of the sum of 58 PCB congeners. Assuming that additional local sources are not contributing to the overall load of contaminant in the environment, the model's decreased performance may be explained by a combination of two factors: the violation of steady state assumptions and an incomplete understanding of the physical chemistry of PCBs. With respect to the former, variability in the source strength undermines the model assumptions about long-term partitioning behavior of PCBs between the air and the soil. Specifically, dredging of the contaminated harbor caused a release of PCBs, resulting in elevated air concentrations shortly before the field sampling; thus, the soil has not yet reached a new equilibrium with the air. With respect to the latter, vapor–particle partitioning of airborne PCBs in the environment is sufficiently poorly understood to interfere with model use under certain conditions. More specifically, assumptions about the vapor–particle partitioning of PCBs influence the model's treatment of contaminant deposition, which in turn governs the magnitude of the modeled mass flux from air to soil.

The model predictions of concentration in crops overestimate the measured values. The distribution of measured concentrations in vine crops (tomatoes in this case, Figure 3) falls slightly below the 2.5th percentile of variability for the sum of 58 PCB congeners. One possible explanation for
overestimation by the model is that the uptake factors for produce from soil may overestimate the mass flux of PCB into plants from the roots. This effect is more pronounced for the less contaminated soil and plants in Dartmouth than in Acushnet. The models again performed better for light congeners than for heavy ones. Similar results were obtained in model runs predicting concentration in a root crop, i.e., potatoes.

The use of models is a necessary component of most risk assessments. This research exploits an opportunity presented by an unusual data set—a multimedia exposure assessment carried out in the communities surrounding New Bedford. As a result of the range of environmental media for which PCB concentration measurements were made, it is possible to examine the performance of simple compartmental models. Given the simplicity of these steady state models, relative to the time and cost of field-based data collection, their performance is striking.

Not surprisingly, the models perform best in settings that are not influenced by short term variation in climatic conditions or contaminant source emissions. Upwind of the harbor, at the Dartmouth site, the steady state model is most appropriate since the concentrations in air are relatively constant over time and the air and soil are assumed to have reached mutual equilibrium. The profile of PCB congeners at this site is weighted toward the middle and is reflective of the profile in background air in the U.S. generally, with congeners such as number 101 figuring prominently. Downwind model performance is not as strong. The tendency of the model to underestimate concentration, under conditions where the steady state assumptions upon which the models rest are violated, is of obvious concern. Underestimation of concentration may lead to underestimation of risk levels and a potential lack of health protection. Analysts should take heed and exercise caution about misapplication. In general, the consequences of both over- and underprediction of exposure concentrations in health risk assessment cannot be dismissed lightly. However, traditionally, overprediction of contaminant concentration has been better tolerated by some decision makers seeking a margin of health protection, although overprediction certainly garners criticism for potentially wasting cleanup resources. Overall, the results presented here argue that appropriate application of the models is an efficient choice relative to measurement.

Acknowledgments

The author gratefully acknowledges statistical contributions from Christopher Bretherton and technical support from Kole

FIGURE 2. Model predictions vs measured concentrations of PCBs in soil in Acushnet, MA. (a) PCB congener 28. (b) PCB congener 101. (c) PCB congener 153. (d) Sum of 58 PCB congeners.

FIGURE 3. Model predictions vs measured concentrations of PCBs in vine vegetables (tomatoes) grown in Dartmouth, MA.
Kanter. Funding provided by University of Washington, through the U.S. EPA Agreement CR825173-01-0 for the National Research Center for Statistics and the Environment, supported a portion of this research.

**Literature Cited**


Received for review May 18, 2001. Revised manuscript received February 19, 2002. Accepted February 21, 2002.

ES0109895