Integrated Environmental Assessment and Management — Volume 10, Number 1—pp. 102–113

Case Study

From Sediment to Tissue and Tissue to Sediment: An Evaluation of Statistical Bioaccumulation Models

Nancy Judd,† Lucinda Tear,† and John Toll†*
†Windward Environmental LLC, Seattle, Washington, USA

(Submitted 3 December 2012; Returned for Revision 25 March 2013; Accepted 27 August 2013)

ABSTRACT

Biota–sediment accumulation factors (BSAFs) and biota–sediment accumulation regressions (BSARs) are statistical models that may be used to estimate tissue chemical concentrations from sediment chemical concentrations or vice versa. Biota–sediment accumulation factors and BSARs are used to fill tissue concentration data gaps, set sediment preliminary remediation goals (PRGs), and make projections about the effectiveness of potential sediment cleanup projects in reducing tissue chemical concentrations. We explored field-based, benthic invertebrate biota–sediment chemical concentration relationships using data from the US Environmental Protection Agency (USEPA) Mid-Continent Ecology Division (MED) BSAF database. Approximately two thirds of the 262 relationships investigated were very poor ($r^2 < 0.3$ or p-value ≥ 0.05); for some of the biota–sediment relationships that did have a significant nonzero slope (p-value < 0.05), lipid-normalized tissue concentrations tended to decrease as the colocated organic carbon (OC)-normalized sediment concentration increased. Biota–sediment relationships were further evaluated for 3 of the 262 datasets. Biota–sediment accumulation factors, linear regressions, model II regressions, illustrative sediment PRGs, and confidence intervals (CIs) were calculated for each of the three examples. These examples illustrate some basic but important statistical practices that should be followed before selecting a BSAR or BSAF or relying on these simple models of biota–sediment relationships to support consequential management decisions. These practices include the following: one should not assume that the relationship between chemical concentrations in tissue and sediment is necessarily linear, one should not assume the model intercept to be zero, and one should not place too much stock on models that are heavily influenced by one or a few high chemical concentration data points. People will continue to use statistical models of field-based biota–sediment chemical concentration relationships to support sediment investigations and remedial action decisions. However, it should not be assumed that the models will be reliable. In developing and applying BSAFs and BSARs, it is essential that best practices are followed and model limitations and uncertainties are understood, acknowledged, and quantified as much as possible. *Integr Environ Assess Manag* 2014;10:102–113. © 2013 SETAC

Keywords: Bioaccumulation Sediment Regression BSAF Risk management

INTRODUCTION

Biota–sediment accumulation factors (BSAFs) are ratios of paired sediment and tissue chemical concentrations, whereas biota–sediment accumulation regressions (BSARs) are regression relationships for paired sediment and tissue chemicals concentrations. Both are statistical models that are used to:

- Estimate tissue chemical concentrations based on sediment chemical concentrations when empirical tissue data are lacking
- Estimate future tissue chemical concentrations under alternative conditions (e.g., remedial actions, routine navigation dredging, open-water disposal)
- Estimate sediment preliminary remediation goals (PRGs) from risk-based tissue chemical concentrations

These 3 applications are common in risk assessments and feasibility studies at contaminated sediment sites as well as for dredge material sites.

All Supplemental Data may be found in the online version of this article.

* To whom correspondence may be addressed: NancyJ@windwardenv.com

Published online 16 September 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/ieam.1484

Both the US Environmental Protection Agency (USEPA) and the US Army Corps of Engineers (USACE) provide online databases for BSAFs (e.g., USEPA 2008; USACE 2009), and the USEPA has provided a methodology and guidelines for developing BSAFs (Burkhard 2009; USEPA 1994, 2000).

Early applications of BSAFs were based on equilibrium partitioning theory and the assumption that lipid-normalized tissue concentrations could be reasonably predicted as a multiple of organic carbon (OC)–normalized sediment concentrations (Lake et al. 1990; Ankley et al. 1992; Lee 1992; McFarland 1995; Pruell et al. 1990). Although BSAFs are relatively easy to calculate and use, the conditions for which they were originally intended—estimating chemical concentrations in benthic infaunal organisms exposed to contaminated sediment—generally do not drive sediment management decisions. Those decisions tend to be driven by tissue chemical concentrations in the fish and shellfish consumed by wildlife and people. Using BSAFs to predict tissue chemical concentrations in demersal or water column organisms at sites with multiple sources of chemicals (e.g., sediment, groundwater and surface water discharges, atmospheric deposition) and varied environmental conditions (e.g., hydrogeomorphology, habitats, food webs) deviates significantly from the conditions for which they were originally intended (i.e., conditions under which sediment chemical concentrations and sorptive phenomena govern tissue chemical concentrations).

The use of BSARs is consistent with USEPA guidance and published research. The USEPA's guidance for the risk assessment of metals (USEPA 2007) suggests using regressions to accommodate the mechanistic dependency (e.g., DeForest et al. 2007) of a bioconcentration factor or bioaccumulation factor on an exposure concentration. Linear regression also has been used to characterize the bioaccumulation of metals in soil (Sample et al. 1999; USEPA 2003). For many dioxins and furans (tetrachlorodibenzo-*p*-dioxins/tetrachlorodibenzofurans), Yunker and Cretney (2000) found BSAFs to be related to sediment concentrations and, consequently, recommended the use of a direct regression of sediment and tissue (i.e., a BSAR) instead of a BSAF for these chemicals (In that example, other factors such as soot carbon were thought to be influencing bioavailability in samples with higher tetrachlorodibenzo-*p*-dioxins/tetrachlorodibenzofurans concentrations. The effects of the OC sources or other factors on bioaccumulation are also present in other systems. Wong et al. (2001) identified numerous limitations with regard to the field application of BSAFs, possibly attributable in part to the biotransformation of chemicals by organisms.

The utility of regression-based bioaccumulation models (i.e., BSARs) is discussed in the USEPA's most recent guidance on BSAFs (Burkhard 2009). The guidance advises the careful inspection of relationships between BSAFs and other site factors to assess the linearity of the biota–sediment relationship and determine whether multiple BSAF "populations" exist at a site. In practice, BSAFs are often used regardless of evidence that the relationship between tissue and sediment is either not constant or not linear. In addition, although field and laboratory bioaccumulation testing methodologies continue to improve (Burkhard, Cowan-Ellsberry et al. 2012; Melwani et al. 2009; ORNL 1998; Parkerton et al. 2008), uncertainty associated with the calculation approach for BSAFs is rarely explored or discussed (e.g., confidence intervals [CIs] around BSAFs are not reported, BSAFs are used as constants with no variance).

Regression modeling practices are described in many statistical textbooks (Box and Draper 1987; Neter et al. 1990; Zar 1996; Helsel and Hirsch 2002). They can, and ideally should, involve an iterative process of trying different models and examining the error structures (i.e., differences between a model's predictions and empirical data) to select a model that best describes the relationship between the model's inputs and responses (in this case, sediment or tissue chemical concentrations). This evaluation can also take the form of a simple hypothesis test to determine whether the data fit a preselected model.

In this paper, we are not focused on the process of model fitting but rather on the consequences for managers. We are interested in understanding how the CIs around predictions from the site mean BSAF compare with those from a BSAR and how those differences might affect a sediment remediation decision. If a transformation of the data is required to make the model linear, how does that affect the prediction of the PRG and its CI? What strength of biota–sediment relationship is needed to support risk or cleanup decision making? The answers to these questions may influence cleanup decisions or how dredged material is managed. Basing contaminated sediment management decisions on flawed assumptions about biota–sediment relationships could lead to expenditures in the tens or hundreds of millions of dollars on actions that provide little or no marginal environmental or public health benefits.

METHODS

Hundreds of publicly available datasets of paired sediment and tissue chemistry data were screened using regression analyses (i.e., BSARs) and predetermined performance criteria to evaluate the strength and significance of the relationship. After this initial screening step, BSAFs and regressions were developed for 3 example paired sediment and tissue datasets. These 3 datasets were chosen to explore the influence of model selection and confidence intervals on model predictions. The different models and their confidence intervals were compared, and example sediment PRGs were calculated. Details on the data sources, selection of datasets, and the models applied (BSAF, BSAR, and model II) are provided below. The methods for PRG calculation were fairly straightforward and generally involved a reinterpretation of the biota–sediment relationships so that the sediment, rather than the tissue, concentration would be predicted.

Data sources

The EPA Mid-Continent Ecology Division (MED) BSAF database (USEPA 2008) was searched for example biota–sediment relationships. This dataset includes approximately 20 000 BSAFs from 20 locations (mostly Superfund sites, all field-collected sediment and tissue samples) for non-ionic organic chemicals, including polychlorinated biphenyls (PCBs), polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans, polycyclic aromatic hydrocarbons, pesticides, and other compounds. Freshwater, tidal, and marine species are included in the dataset. Only data with paired, detected sediment and detected tissue concentrations are included in the database. If either or both were not detected, the data pair is not included in the database.

The number of nondetects and detection limits in either tissue or sediment concentrations provides important information about the uncertainty in the data and any sediment–tissue relationship. Although the purpose of this paper is not to discuss the methods for dealing with nondetects, and the database used (i.e., the USEPA MED BSAF database) excludes nondetect data, the exclusion of nondetects from a dataset is not recommended. When nondetects are present in a dataset, methods appropriate for censored data (Helsel 2005, 2006) should be used.

In the USEPA MED BSAF database, 5 sedentary benthic and epibenthic species that are relevant to ecological or human health risk assessment were identified. Although some of these species may feed largely from the water column, they were selected because they are often used for the generation of BSAFs. Only sedentary species were selected because their relatively limited movement (as compared with that of fish or crab) increased the likelihood of finding a biota–sediment relationship. For these species, 344 of the datasets had more than one sample, and 262 datasets had a sample size of 4 or greater (Table 1). The datasets were from 5 different sites (i.e., each species included in the table was from a different site). Most of the datasets ($n = 164$) for all organisms and species groups were for individual PCB congeners (or co-elutions). Of the remaining 98 datasets, 55 involved individual polycyclic aromatic hydrocarbons, 23 involved pesticides, 16 involved dibenzo-*p*-dioxins or polychlorinated dibenzofurans congeners), 3 involved total PCBs, and 1 involved tributyltin.

Selection of datasets for analysis

The 262 datasets were tested for a significant, linear relationship between untransformed (i.e., arithmetic) tissue

Table 1. MED BSAF database datasets investigated

Organism/species group/location	Nr of datasets (i.e., nr of chemicals)[a]	Nr of datasets with $n \geq 4$
Alewife floater (*Anodonta implicata*), Charles River, Massachusetts	6	6
Blue mussel (*Mytilus edulis*), Narragansett Bay, Rhode Island	42	33
Hard clam (*Mercenaria mercenaria*), New Bedford Harbor, Massachusetts	56	52
Hard clam (*Pitar morrhuana*), Coddington Cove, Rhode Island	52	38
Crayfish (unidentified), Willamette River, Oregon	188	133
Total	**344**	**262**

[a]Data were available from only 1 study for each organism/species group/location. Thus, the number of datasets is equal to the number chemicals reported for each organism/species group/location.
BSAF = biota-sediment accumulation factor; MED = Mid-Continent Ecology Division.

and sediment concentrations, untransformed tissue and log (10) transformed sediment concentrations, and log (10) transformed tissue and log (10) transformed sediment concentrations. Because all of the chemicals in the database were hydrophobic organic chemicals, sediment concentrations were normalized based on OC content, and tissue concentrations were normalized based on lipid content before BSAR regressions were performed per Burkhard (2009) recommendations. The same evaluations in this paper may also be performed without normalizing tissue or sediment concentrations, which would allow more possibilities for identifying a relationship.

A regression model for any species–chemical dataset was screened in for further evaluation as a BSAR if the linear slope coefficient was significantly different than zero ($p < 0.05$) and the coefficient of determination (adjusted r^2) was 0.30 or greater (i.e., at a minimum, a weak relationship was established). The strengths of the relationships were evaluated as follows:

- No relationship: $0.0 \leq r^2 < 0.3$
- Weak relationship: $0.3 \leq r^2 < 0.5$
- Moderate relationship: $0.5 \leq r^2 < 0.7$
- Strong relationship: $0.7 \leq r^2 < 1.0$

The weak, moderate, and strong relationship designations were assigned for descriptive purposes and were not based on any convention, and 3 datasets with a statistically significant slope and a moderate or stronger relationship (i.e., $r^2 \geq 0.5$) were carried forward for more-detailed analysis and discussion.

The evaluation of the significance and coefficient of determination should be considered as an initial step in the evaluation of a BSAR. In addition, linear regression models should always be evaluated for goodness of linear fit, the distribution of residuals around the model, evidence of high-influence data pairs—especially those with high concentrations—and any issues related to data quality and applicability to the question at hand. For example, are the ranges of sediment and tissue concentrations sufficiently wide that a linear relationship would be expected (or are the ranges within the realm of measurement error), are the data of uniform quality and appropriately pooled so that the population to which the inference will be drawn from the model is clear, do the data include any nondetects or "greater than" data that require methods that explicitly consider censored data? All of these issues have been written about extensively in many introductory statistical texts and so are not discussed in depth in this paper. This additional model analysis was not a part of the screening to select datasets for further evaluation.

BSAF calculations

For each dataset selected for more-detailed analysis, BSAFs were calculated using Equation 1.

$$\mathrm{BSAF} = \frac{(C_{\mathrm{tiss,LN}})}{(C_{\mathrm{sed,OC}})} \quad (1)$$

Where

BSAF = site-specific biota-sediment accumulation factor
$C_{\mathrm{tiss,LN}}$ = organism tissue concentration, lipid-normalized (μg/kg lipid dry weight [dw])
$C_{\mathrm{sed,OC}}$ = surface sediment concentration, OC-normalized (μg/kg OC dw)

A BSAF was computed for each biota–sediment pair in a dataset, and the mean BSAF with its upper and lower 95% CIs was calculated (Equation 2).

$$\mathrm{mean}_{\mathrm{BSAF}} \pm t_{(.05,n-1)} \times \mathrm{SE}_{\mathrm{BSAF}} \quad (2)$$

Where

$\mathrm{SE}_{\mathrm{BSAF}}$ = standard error of the BSAF
t = t-statistic associated with the specified CI and sample size (n)

The mean tissue concentration predicted for each measured sediment concentration was calculated by rearranging Equation 1 to create Equation 3:

$$C_{\mathrm{tiss,LN}} = \mathrm{mean(BSAF)} \times C_{\mathrm{sed,OC}} \quad (3)$$

The CIs for each predicted tissue concentration were calculated using the upper and lower confidence limits on the BSAF in place of the mean BSAF. Likewise, sediment concentrations (hypothetical PRGs or $C_{\mathrm{sed,OC}}$) for each specified tissue concentration were "back calculated" (Equation 4):

$$C_{sed,OC} = \frac{(C_{tiss,LN})}{mean(BSAF)} \quad (4)$$

Again, confidence limits for predicted PRGs were calculated by replacing the mean BSAF with the upper and lower confidence limits around the mean BSAF in Equation 4.

Linear BSARs (Model I regressions)

Linear regressions developed for each data transformation of each dataset that screened in were further investigated to select the form of the relationship that was most linear and had the most homoscedastic residuals. The selected BSARs were then used for the calculation of tissue concentrations from sediment as well as for the calculation of sediment concentrations from tissue. Confidence intervals for both predicted sediment and predicted tissue were also computed.

For 2 of the 3 datasets evaluated in greater detail, the form that best met the assumptions of a linear model involved log transformations of both tissue and sediment. The best-fitting linear regression was used to predict mean tissue concentrations (Equation 5) along the range of measured sediment concentrations (CIs for the linear model). These are displayed graphically in the *Results and Discussion* section. In addition, the empirical mean sediment concentration that had been used to predict a given tissue concentration was back calculated for each empirical tissue concentration and for the mean empirical tissue concentration (e.g., toxicity reference value or other risk based-tissue level) (Equation 6). All calculations were conducted in the programming language R (Version 2.12).

$$C_{tiss,LN} = a + b \times C_{sed,OC} \quad (5)$$

$$C_{sed,OC} = \frac{(C_{tiss,LN} - a)}{b} \quad (6)$$

The CIs around predicted mean tissue and back-calculated sediment concentrations were calculated according to Neter, Wasserman, and Kutner (1990), as well as by bootstrapping (Efron and Tibshirani 1986). The back calculations have the same limitations as forward regression calculations. If model assumptions (e.g., linearity, homogeneous residuals) have not been met, the predictions and confidence intervals may not be correct.

Each bootstrap iteration involved sampling *n* biota–sediment pairs (with replacement) from the *n* biota–sediment pairs in the dataset, fitting a linear regression to the bootstrap sample, and saving the predicted mean tissue and back-calculated sediment concentrations from Equations 4 and 5. One thousand bootstrap iterations were used to create a bootstrap distribution of predicted tissue and back-calculated sediment concentrations. The mean predicted tissue and mean back-calculated sediment concentrations for each bootstrap distribution were calculated, along with the 2.5th and 97.5th percentiles (i.e., 95% CI).

When regressions are developed using log-transformed tissue and log-transformed sediment concentrations, back-transforming predictions of mean (log)tissue or (log)sediment concentrations to the original arithmetic units results in an estimate of the geometric mean (geomean) or median rather than the actual mean. If measures of mean exposure or effect are required, a correction factor could be applied to convert the geomean to the mean (see "Smearing Factor for Log-Transformed Data" in the Supplemental Data for discussion). However, those methods and results are not discussed in this paper.

To avoid the nuances of interpreting corrected predictions, in this paper, only bootstrap sampling was used to estimate the mean and CIs around mean predictions of tissue and sediment in the original units for regressions involving log-transformed variables. Each bootstrap prediction of (log)tissue and (log) sediment was first exponentiated before the calculation of the mean of the bootstrap distribution. This produced an estimate of the true mean in arithmetic units. In addition, the 2.5th and 97.5th percentiles of the 1000 bootstrap predictions of (log) tissue and (log)sediment were calculated and exponentiated to generate CI estimates (in arithmetic units).

Model II approach

A form of Model II regression called the geometric mean regression was also conducted in the programming language R (Version 2.12). As with the selected BSARs, the Model II regressions were then used for the calculation of tissue concentrations from sediment as well as the calculation of sediment concentrations from tissue. Again, confidence intervals were computed for both predicted tissue and predicted sediment.

Model II–type regressions, as discussed in Burkhard (2009) and in many other sources (e.g., Sokal and Rohlf 1969, 2012), are appropriate in cases in which the dependent variable has not been measured without error (as is assumed in Model I regressions, such as those discussed previously). The slope of a geometric mean regression is a function of the slopes of the regressions of X versus Y and Y versus X. As explained in Equations 16 through 19 in Burkhard (2009), a geometric mean regression was calculated for each pairing of (log)tissue and (log)sediment as Equation 7:

$$C_{tiss,LN} = a + b \times C_{sed,OC} \quad (7)$$

Where

$a = \text{mean}(y) - b \times \text{mean}(x)$
$b = \text{sqrt}(d/f$, where d and f are regression slopes from the equations below)
$\text{tissue} = c + d \times \text{sediment}$
$\text{sediment} = e + f \times \text{tissue}$

Because closed-form CIs, r^2 values, and significance are not readily available for Model II regressions, CIs for Model II regressions were constructed from bootstrapping (again, the 2.5th and 97.5th predictions from 1000 bootstrap iterations defined the 95% CI).

RESULTS AND DISCUSSION

Tissue versus sediment regressions

Of the datasets and relationships tested, approximately one third of the relationships met the screening criteria of statistical significance and $r^2 \geq 0.3$ (Table 2). Approximately one third of the relationships that met the screening criteria were negative (as shown in the total under the Summary of Coefficient of Determination in Table 2). A negative relationship indicates that as sediment concentrations increase, tissue concentrations decrease.

Only a limited suite of chemicals in hard clam and crayfish had relationships that passed the regression screening criteria

Table 2. Number of datasets and types of linear relationship between tissue and sediment

Type of relationship	Number of datasets		
	Tissue (μg/kg lipid): Sediment (μg/kg OC)	Tissue (μg/kg lipid): Log(sediment (μg/kg OC))	Log(Tissue (μg/kg lipid)): Log(sediment (μg/kg OC))
Summary of coefficient of determination			
$r < 0.0$ (negative slope)	79	66	80
$r^2 < 0.3$ (not even a weak relationship)	55	83	77
$r > 0$ and $r^2 \geq 0.3$ (some relationship)	128	113	105
Total	**262**	**262**	**262**
Summary of coefficient of determination and significance[a]			
Negative/significant[b]	7	15[c]	13[d]
Weak[c]/significant	1	3	6
Moderate[d]/significant	25	18	23
Strong[e]/significant	56	29	24
Total	**89**	**65**	**66**

[a]Significant relationship defined as p-value < 0.05.
[b]All negative and significant relationships were for individual PCBs or PAHs.
[c]$0.3 \leq r^2 < 0.5$.
[d]$0.5 \leq r^2 < 0.7$.
[e]$0.7 \leq r^2 < 1.0$.
OC = organic carbon; PAH = polycyclic aromatic hydrocarbon; PCB = polychlorinated biphenyl.

(i.e., $r^2 \geq 0.3$, $p < 0.05$). For hard clams, PCB congeners and total PCBs (for *Mercenaria mercenaria*) and chrysene (for *Mercenaria mercenaria* and *Pitar morrhuana*) had passing relationships. Dibenzo-*p*-dioxins/polychlorinated dibenzofurans congeners, dieldrin, and PCB congeners had passing relationships for crayfish (unidentified species). Many of the relationships, including those described in Table 2, were for individual PCB congeners. The dataset included only detect–detect data pairs. Nondetect data (for sediment or tissue concentrations) were not included in the database.

Example regressions/BSAFs

Three species–chemical combinations from the passing datasets were selected for further analysis. The 3 examples (PCBs in hard clam, chrysene in hard clam, and p,p′-dichlorodiphenyldichloroe–thylene [DDE] in crayfish) were selected because they illustrate different issues for BSAR development, had a moderate or stronger relationship (i.e., $r^2 \geq 0.5$), included at least 8 or more sediment–tissue data pairs, and were more important from a risk perspective than some other passing relationships (and therefore more relevant for PRG development). Of the datasets that passed the screening criteria, most would not be considered "good" linear relationships based on the distribution of the residuals. This was also true of the 3 BSARs selected for further evaluation. Residual plots are presented in the Supplemental Data (Figures S1 through S3); ideally residuals in such plots would be homogeneously distributed about the regression line. The following 3 examples illustrate key findings from the development of regressions and BSAFs.

Example 1. Importance of an intercept. Figure 1 presents total PCBs in untransformed hard clam tissue versus untransformed sediment collected in New Bedford Harbor, Massachusetts, USA. This example illustrates the importance of considering the intercept when characterizing the biota–sediment relationship. Both BSAFs and BSARs with the intercept forced through the origin tend to be biased to underpredict tissue concentrations at low sediment concentrations and overpredict tissue concentrations at high sediment concentrations because they fail to account for nonsediment sources (e.g., water) and, consequently, have steeper slopes. Thus, PRGs developed using BSAFs or by forcing a BSAR through the origin are liable to overestimate the reduction in tissue chemical concentration possible through sediment remediation. The 95% CI for the BSAF is conical, with very high uncertainty around the predicted tissue concentration at higher sediment concentrations. In this case, the linear regression BSAR is strong ($r^2 = 0.92$ and $p = 0.0002$) and therefore similar to the Model II relationship. This relationship was one of the best found in the database screened; using untransformed data, the relationship is quite linear across 2 orders of magnitude and also has fairly homogeneously distributed residuals (see the Supplemental Data, Figure S1).

Example 2. Nature of the relationship. Figures 2 and 3 present chrysene in hard clam tissue versus sediment from Coddington Cove, Rhode Island, USA. This example illustrates the importance of considering the scale of the relationship being investigated (e.g., linear, log linear, or nonlinear). Figure 2 shows the linear regressions for log-transformed (both sediment and tissue are log-transformed) and untransformed data (neither sediment nor tissue is transformed) plotted on a log scale. The relationship using untransformed data does not meet the screening criteria based on strength of relationship and significance, but the relationship using log-transformed data is

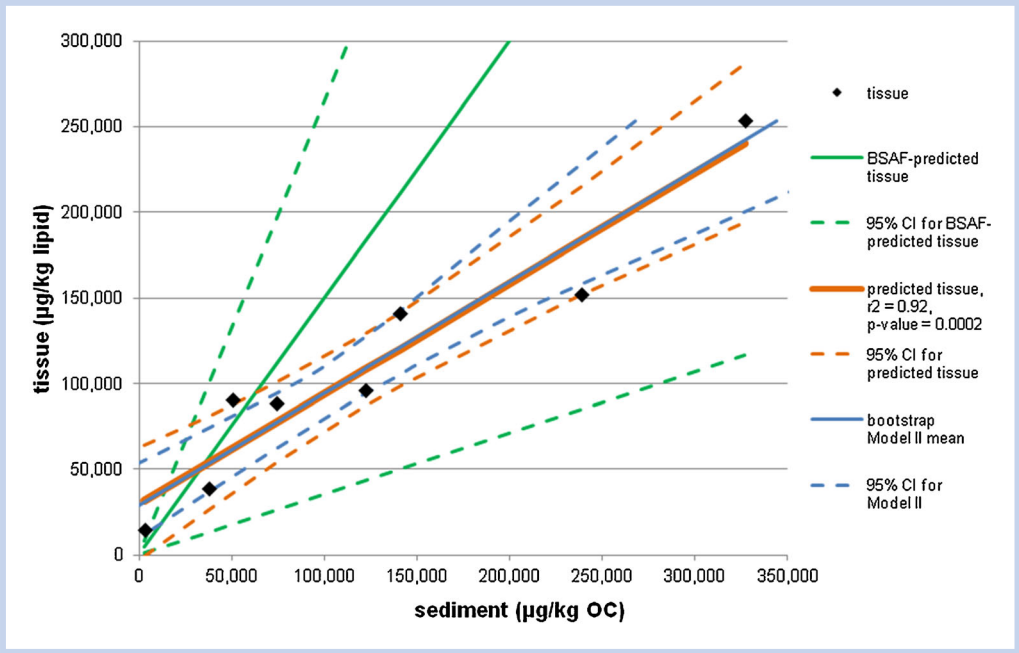

Figure 1. Hard clam (*Mercenaria mercenaria*) tissue versus sediment—total PCBs. New Bedford Harbor, Massachusetts, USA.

moderate and significant ($r^2 = 0.6$, $p = 0.0146$). In this example, predictions of tissue from sediment at the lower ranges of sediment concentration or vice versa would be strongly affected by the choice of the biota–sediment model. Figure 3 shows the BSAF, Model I linear (i.e., predicted [log] tissue regression [also shown in Figure 2]), and Model II regressions for the same data, also on a log scale. In this example, the moderate (but not strong) relationship between log-transformed tissue and log-transformed sediment causes the linear regression (Model I) and Model II relationships (Figure 3) to differ more than that for the PCBs in the hard clam tissue dataset represented in Figure 1.

The chrysene in hard clam tissue example also illustrates several considerations that arise when using log-transformed data in a linear model. First, the linear relationship described using log-transformed data is linear over orders of magnitude, and variance around predictions in the original units can be high. In addition, backtransformations of predictions of mean (log) tissue are geometric means and, as such, can underestimate mean risk. As mentioned, conversion factors must be applied to correct the backtransformation from the geomean to the mean, and these correction factors are accurate when the data are truly lognormal. The difference between the geometric mean and the mean can be large for lognormal distributions, especially those with high variance. This difference can be so large that the true mean exceeds the backtransformed CIs around the geometric mean. Given that CIs are generally expected to contain central tendencies, this scenario (i.e., the mean is outside the CIs of the geometric mean) is difficult to interpret.

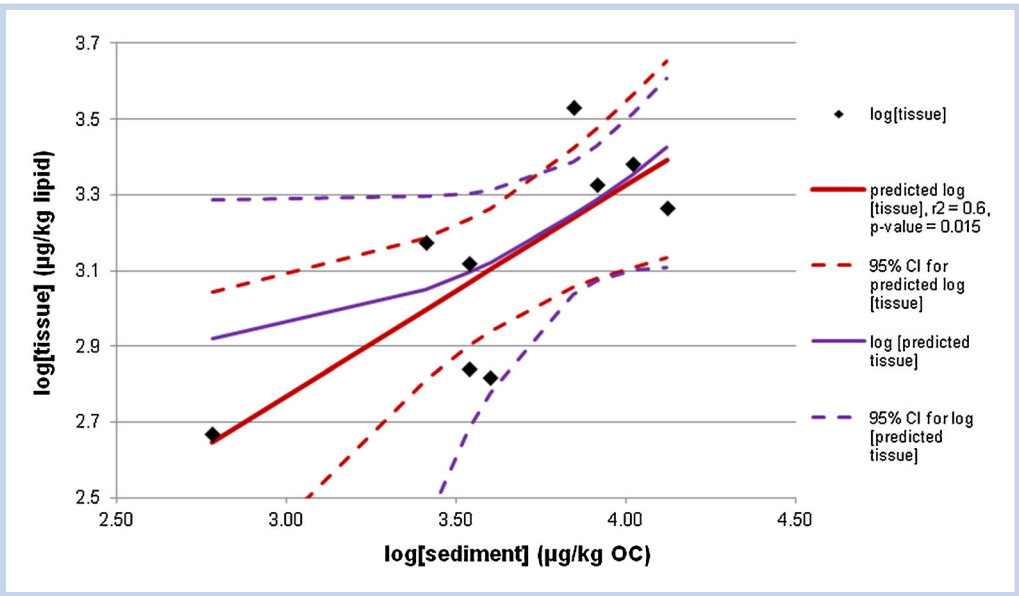

Figure 2. Hard clam (*Pitar morrhuana*) tissue versus sediment—chrysene. Coddington Cove, Rhode Island, USA (linear regressions).

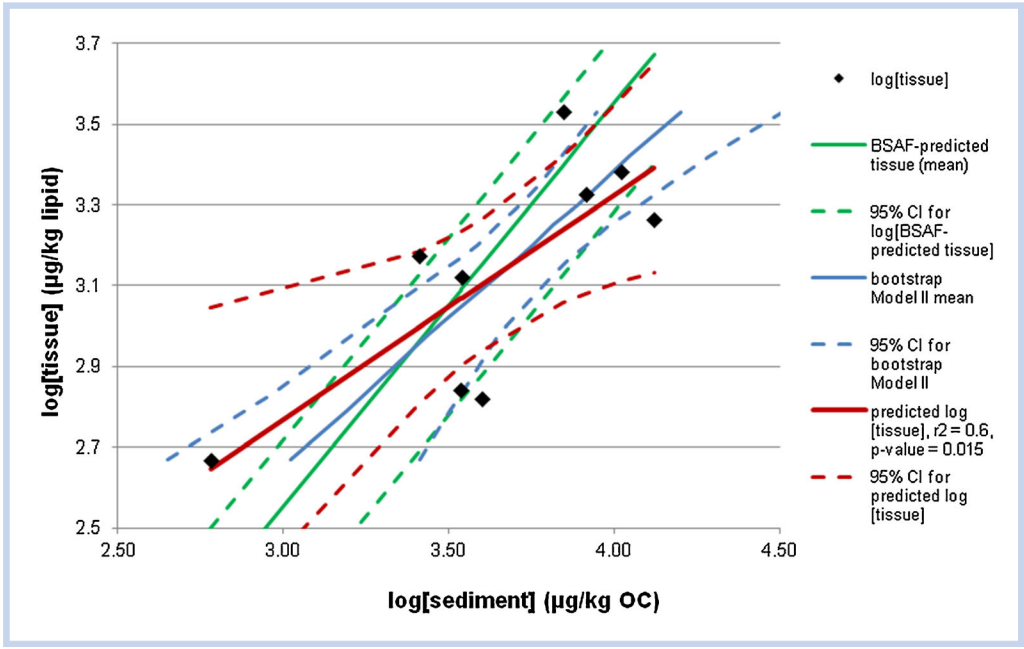

Figure 3. Hard clam (*Pitar morrhuana*) tissue versus sediment—chrysene. Coddington Cove, Rhode Island, USA (BSAF, Model I, and Model II regressions).

Example 3. Influence of a single high point, model choice, and model fit. Figures 4 and 5 present p,p′-DDE in crayfish tissue versus sediment collected in the Willamette River, Oregon, USA. In Figure 4, 2 linear regressions are included: 1 that uses all data and 1 that excludes the highest data pair. Removing the highest pair changes the relationship from strong and significant to insignificant. This is an example of a relationship whose significance would be highly affected by a nondetect in tissue or sediment when the other variable was high.

The highest pair in this dataset has a much larger effect on the regression (as measured by outlier statistics including Leverage, Cook's Distance, and DFBetas) than any of the other data pairs; and when the highest pair is included in the model, the distribution of residuals indicates that a linear model is not a good fit (see Figure S3 in the Supplemental Data). Explanations for the highest pair should be explored and discussed; for example, it could indicate the presence of more than one population of biota–sediment relationships in the dataset. Multiple populations could be created by fine or large-scale heterogeneity in physical or biological factors. If the highest pair is not part of the same population as the rest of the data but is retained in the dataset, tissue concentrations at lower sediment concentrations will likely be underpredicted. If the model (including the highest pair) is then used to back-calculate a sediment PRG, the reduction in the tissue chemical concentration that could be achieved in most of the population through

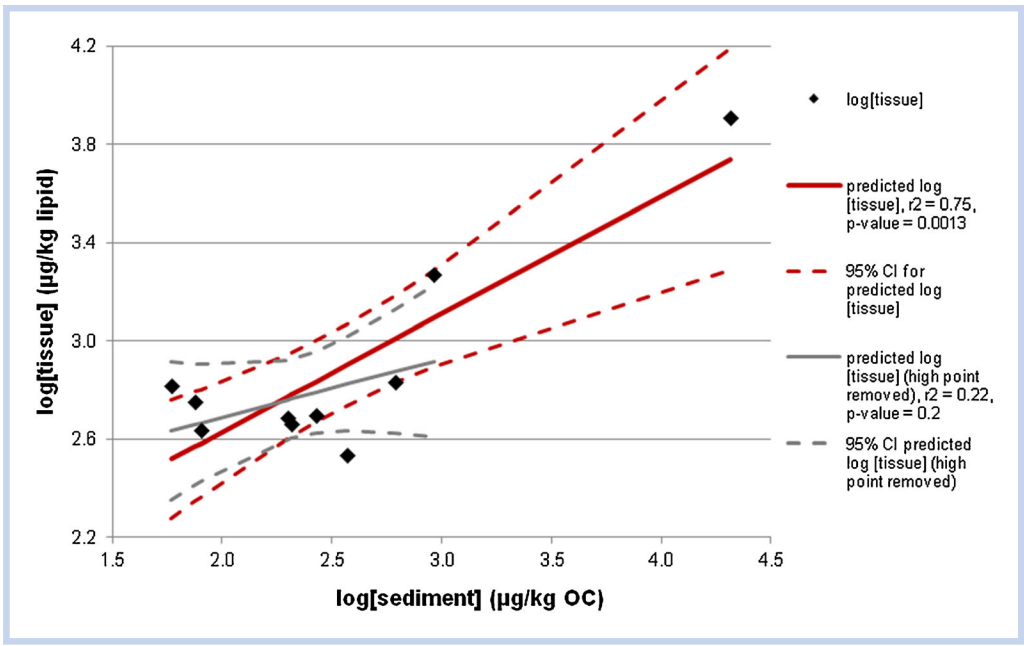

Figure 4. Unidentified crayfish tissue versus sediment p,p′-DDE—Willamette River, Oregon, USA (regression of all data, regression excluding highest point).

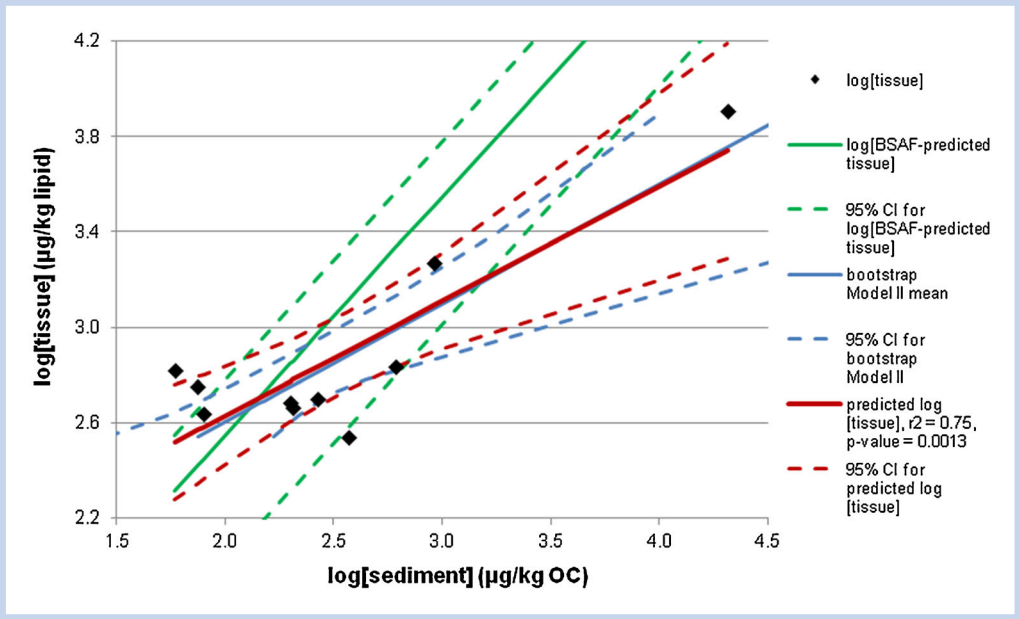

Figure 5. Unidentified crayfish tissue versus sediment p,p′-DDE—Willamette River, Oregon, USA (BSAF, Model I, and Model II regressions).

sediment remediation will likely be overestimated (i.e., the desired lower sediment concentration would not be achieved). If the highest pair appears to be valid and is the sole cause of a significant relationship, data transformations or other types of models should be considered.

With the inclusion of the highest data pair, the linear regression (Model I) and the Model II relationships (Figure 5) are fairly similar. If the highest pair were to be removed, the 2 relationships would be less similar (as with Example 2, chrysene in hard clam tissue). If the highest pair is considered to be valid, a hockey stick regression could be considered if it is plausible that 2 populations are present in the data or that a threshold response is justified (see Figure S4 in the Supplemental Data).

As with the chrysene example, this example (i.e., p,p′-DDE in crayfish tissue) demonstrates the importance of considering the scale of the relationship and, consequently, the scale at which predictions are valid. If the sediment and tissue concentrations are not logged, the regression is weaker and not significant ($r^2 = 0.39$, $p = 0.07$). Because the BSAF is computed in the original units, it could be considered insignificant as well. In addition, as with Example 2, because the BSAF does not include an intercept, the BSAF slope is steeper than both the linear (Model I) and Model II relationships and will result in underpredictions of tissue concentrations at low sediment concentrations and overpredictions of tissue concentrations at high sediment concentrations. The low data pair in Example 2 is quite influential, perhaps in a way that is analogous to the high data pair in Example 3.

Calculation of PRGs from Example Regressions and BSAFs. Table 3 presents PRGs calculated for all 3 of these examples using various models. The PRGs are usually defined as the sediment concentrations associated with specific target tissue concentrations, which may be equal to ecological toxicity reference value or human health risk–based tissue concentrations. In these examples, the target tissue concentrations selected were equal to the mean tissue concentrations as plotted on the y axis (raw or log-transformed). Thus, the CIs on regressions are at their narrowest. In reality, a target tissue concentration is likely to be lower than the existing mean tissue concentration, which will generally result in predicted PRGs with wider CIs that extend beyond the range of measured concentrations. When the relationship is on a log scale, the uncertainty in the PRGs may span orders of magnitude.

To avoid the complication of converting between mean target tissue concentrations and geometric mean values in log relationships (as discussed in the methods for linear BSAFs), a bootstrap approach was used for estimating the PRGs and 95% CI for the BSARs and Model II regressions of logged relationships (i.e., chrysene and p,p′-DDE). For total PCBs (a raw data BSAR relationship), the PRG and CI were calculated using the statistical formulas for reverse predictions as well as the bootstrap method. The results of the 2 approaches were similar. The minimum and maximum empirical sediment concentrations are presented for reference.

Another issue associated with the calculation of sediment PRGs for hydrophobic organics is that toxicity reference values and target tissue concentrations are often not expressed on a lipid-normalized basis. Total tissue concentrations must be converted to lipid-normalized concentrations, often using average lipid concentrations. The models then generate PRGs in terms of sediment OC concentrations.

In the case of p,p′-DDE (as discussed earlier in Example 3), the relationship was driven by the single highest data pair. A hockey stick regression was applied (see Figure S4 in the Supplemental Data) and provided a reasonable fit. The PRGs and CI generated from the hockey stick regression are also reported in Table 3. The mean PRGs from the hockey stick regression are higher than the BSAF estimate but lower than the estimates provided by the BSAR and Model II regressions. However, the CI for the hockey stick PRG estimate is narrower than those for the other models.

The use of reverse regressions (i.e., sediment vs. tissue) to develop PRGs was also considered. Although linear regression is generally used to predict a "dependent" variable (y) determined by an "independent" variable (x), a direct, ecological, mechanistic dependence of y on x (e.g., tissue on sediment) is not necessarily required if the ability to make

Table 3. PRGs calculated using various approaches

Chemical and target tissue concentration	Model	Median PRG (μg/kg OC)	Mean PRG (μg/kg OC)	2.5th Percentile (μg/kg OC)	97.5th Percentile (μg/kg OC)	Minimum sediment (μg/kg OC)	Maximum sediment (μg/kg OC)
Total PCBs (93 519 μg/kg lipid)	BSAR	99 311	99 311	67 488	131 134	3152	326 755
	BSAR bootstrap	99 351	98 584	71 768	121 341		
	BSAF	72 907	72 907	41 355	307 544		
	Model II bootstrap	100 434	99 441	74 279	121 729		
Chrysene (1585 μg/kg lipid)	BSAR (log-log) bootstrap	6173	6545[a]	3650	10 782	605	13 240
	BSAF (log-log)	4491	4491	3066	8388		
	Model II (log-log) bootstrap	5766	5913[a]	3831	9010		
p,p′-DDE (1413 μg/kg-lipid)	BSAR (log-log) bootstrap	1193	2704[b] 5.51E+284[a]	0	5.63E+07	57.8	20 625
	BSAF (log-log)	401	401	235	1373		
	Model II (log-log) bootstrap	1038	2164[a]	657	11 257		
	Hockey stick (log-log)	1426	1372[b] 1396[a]	1170	1718		

Table uses mean empirical tissue concentration as target tissue concentration.
[a]Mean(10 ^ log(PRG)).
[b]10 ^ mean(log(PRG)).
BSAF = biota-sediment accumulation factor; BSAR = biota-sediment accumulation regression; DDE = dichlorodiphenyldichloroethylene; OC = organic carbon; PRG = preliminary remediation goal.

protective management decisions does not require an exact quantification of partial correlations and their drivers (as might be attempted through path analysis). For example, for a given contaminant, concentration cycling may be occurring at a site such that benthic tissue concentrations do not depend strictly or solely on sediment concentrations. Both tissue and sediment concentrations might be dependent on the rate at which the chemical is entering the study area, and the decomposition of benthic tissue might be contributing chemical mass back to the sediment. This type of conceptual site model might support the use of a Model II–type regression; however, if the primary management objective is the determination of a sediment PRG with minimum variance from a tissue–residue threshold, the model could be framed so that the sediment PRG is the dependent variable in a Model I regression, which would minimize the residual variance in the dependent variable.

The use of a sediment–biota model as opposed to a biota–sediment model does not relieve the analyst or policy maker of any of the issues discussed in this paper. These include statistical or inferential issues related to inappropriate sampling design, inaccuracy or imprecision in the measurement of sediment and/or tissue and their pairings, the need for an intercept in the model, or the transformation of the data to produce a linear model. After all of these issues have been considered, the use of sediment as the dependent variable in a Model I regression could potentially increase the precision of PRG estimates.

SUMMARY AND RECOMMENDATIONS

Our search of the USEPA database and the examples presented in this paper show that for many datasets, not even a weak statistical relationship can be found between measured sediment concentrations and tissue concentrations. As noted by Burkhard, Arnot, et al. (2012b), "The key to measuring meaningful bioaccumulation metrics with accuracy is that the samples for the exposure medium must be representative of the actual exposure of the organisms collected. Furthermore, adequate field collection designs must also account for variability in the exposure concentrations and tissue residues." Many, if not most, study designs do not meet this standard; and when they do not, it is not possible to accurately quantify the uncertainty in biota–sediment relationships and any predictions based on these relationships.

Even when data are collected using a valid design, likely predicted tissue concentrations or PRGs and their associated CIs will differ depending on the model used and the sediment or tissue concentration used to predict the other variable. Consequently, selecting a model that best characterizes the relationship between tissue and sediment, both statistically and in terms of other relevant information, is important. The examples presented in this paper highlight some key considerations when developing biota–sediment relationships. These considerations include the following:

- Treatment of nondetects: What is the detection frequency and do the models explicitly account for nondetects?
- Model selection criteria: How good does a model need to be to be useful in a given management context?
- Shape of the relationships: Should the model include an intercept? What is the appropriate form of the model (linear, nonlinear, curvilinear)? Should the data be transformed?

- Validity of the data: Were the data collected according to a valid study design in which exposure areas and target organisms were truly matched?
- Presence of more than one population: Do the data indicate different biota–sediment relationships across the range of sediment concentrations measured?
- Use of BSAFs versus BSARs: What are the advantages of BSARs over BSAFs?

Treatment of nondetects

The analyses presented here rely on data from the MED BSAF database (USEPA 2008), which provides data for over 20 000 potential biota–sediment relationships (based on different chemical, species, and location data). This database does not include data for which either the sediment or tissue concentration of a data pair (or both) were nondetects. This is consistent with USEPA guidance on the use of only detect–detect pairs for BSAF development (Burkhard 2009). However, some datasets may have both high tissue concentrations and nondetected tissue concentrations at the highest sediment concentrations. If nondetects are not included in a dataset (as with the USEPA MED BSAF database [USEPA 2008]), one might overestimate the strength of a relationship. Therefore one must be explicit about whether nondetects have or have not been excluded and consider the impacts of their inclusion or exclusion on relationships and model predictions.

Assuming that detection limits are low, 2 different situations involving nondetected concentrations can be anticipated when evaluating the relationship between tissue and sediment concentrations at a site. In the first situation, the dataset includes one or more pairs with both high tissue and high sediment concentrations, as well as one or more pairs with a high concentration in one medium (either tissue or sediment) and a low or nondetected concentration in the other. This situation indicates uncertainty about the relationship between tissue and sediment. The reasons for this uncertainty should be explored; these might include different physical conditions at the locations where the 2 types of pairs were collected, which would imply the presence of different populations within the site and raise questions about how data should be pooled at the site, or the use of individual organisms that reveal differential uptake among individuals.

The second situation arises when neither tissue nor sediment concentrations are detected (i.e., a nondetected pair). In this situation, justifying removing the nondetected pairs if the detection limits for sediment is well below the range of sediment concentrations within which a PRG would be set or the detection limit for tissue is well below the range for which tissue concentrations would be predicted might be possible. As long as the predictions (for sediment or tissue) always fall within the range of detected concentrations, stratifying the population of concentrations so that only detected concentrations are modeled is valid, inasmuch as these concentrations constitute the population of interest, and the relationship between tissue and sediment below the detection limits does not affect the relationship above detection limits.

In all situations, the proper approach for dealing with nondetects can only be determined after careful consideration of (1) the frequency of nondetects, (2) the number of detection levels in each variable and the number of nondetects at each detection level, (3) any temporal or spatial factors correlated with nondetects that might help refine the definition of the population of interest and determine which data should be included in the final dataset, (4) the number and position of nondetected pairs in the dataset, (5) the position of target tissue concentrations within the range of detected and nondetected empirical tissue concentrations used to develop the model, and (6) the range of detected sediment concentrations used to develop the model. Nondetected data provide useful, and in some cases critical, information about the statistical, spatial, and temporal distribution of tissue and sediment concentrations at a site, as well as the relationship between tissue and sediment concentrations. The removal of nondetected data should be predicated on thoughtful consideration of how their removal would change the population being modeled and how the regression will be applied. If, after consideration of all these questions, the removal of nondetected data seems to be "valid" from a site management or policy perspective, evaluating the effects of the removal of any nondetected data by conducting analyses, with the nondetects included, using methods appropriate for censored data (e.g., nondetects and data analysis in the programming language R) as well as without the nondetects, may still be useful. A comparison of the results of both approaches could be used to validate assumptions and uncover any nuances that may have been overlooked during a priori considerations.

Model Selection Criteria

Based on the data compiled by USEPA in the MED BSAF database (USEPA 2008), few of the benthic invertebrate biota–sediment relationships met the assumptions required to fit a regression and calculate valid CIs. Of the 262 datasets evaluated, approximately two thirds of the biota–sediment relationships were very poor and did not meet minimum screening criteria (i.e., $r^2 > 0.3$, $p < 0.05$). Better guidance on sampling design should be developed to improve the quality of BSAFs and BSARs by adding to our knowledge of chemical bioavailability and exposure processes. Furthermore, if the goal of developing biota–sediment relationships is for decision-making, minimum criteria should be established for acceptable data quality and quantity and biota–sediment relationships. What quality and quantity of data are required to determine whether a relationship exists? What strength of relationship between tissue and sediment would be needed to make a prediction meaningful? Using a regression approach allows for the evaluation of the strength (r^2), significance (p-value), and shape of the relationship to be evaluated.

Given the profound differences in tissue predictions and estimated PRGs that can result from different statistical approaches, one must be extremely clear regarding

- The strength of the relationship between tissue and sediment that would be needed to make a prediction meaningful
- The extent to which modeling the uncertainty in both variables is considered necessary for the predictions to be meaningful
- Informational goals (i.e., is tissue or sediment being predicted?) and how informational needs change through the risk evaluation and remediation process
- Whether the relationship is linear or, if it is not, how the risk assessment process should accommodate the nonlinearity
- Whether the data are of sufficient quality to provide meaningful information (e.g., they do not have substantial

analytical errors, they are spatially colocated in a meaningful way)

Shape of the relationships

Considerations of the shapes of the relationships have been illustrated by the three examples discussed in detail. These considerations include whether the model should include an intercept, the appropriate form of the model (i.e., linear, nonlinear, or curvilinear), and whether the data should be transformed. Decisions about model shape affect the validity of the model, as well as the tissue and sediment predictions.

Validity of the data

When evaluating biota–sediment relationships, one must be aware of the study design and any issues associated with its execution. Is there any reason to expect that the sediment samples might not reflect the exposure of the organisms whose tissue is being modeled? Were any sampling constraints associated with any or all of the sediment samples collected? Are there abnormalities with the analytical data (e.g., unexpected number of nondetects)? These questions should be considered in the evaluation of biota–sediment relationships.

Presence of more than one population

The possibility that more than one population exists should also be considered. Are multiple species being used to develop a single biota–sediment relationship? Are there physiological reasons to expect that organisms might process exposures differently beyond a certain threshold? Are there differences in the ages of the organisms sampled? Numerous reasons exist for why multiple populations might exist in a dataset, and these should be considered when evaluating biota–sediment relationships.

Use of BSAFs versus BSARs

Further evaluation of 3 of the stronger biota–sediment relationships from the MED BSAF database (USEPA 2008) identified several ways in which BSARs may be preferable to BSAFs when quantifying the nature of the biota–sediment relationship at a site, especially when predictions of tissue or sediment concentrations are to be made for risk or remedial decision making. BSARs have an advantage over BSAFs in that they allow for the consideration of nonsediment contributions to tissue burdens via the model intercept, the exploration of the correct scale of the biota–sediment relationship (e.g., linear, log linear, or nonlinear), and the exploration of any temporal or spatial explanations for statistical outliers or nonlinearities within the relationship. With BSARs, it is also possible to develop and apply criteria for the exclusion of data (e.g., a specified leverage value). These types of evaluations can provide more information about the level of confidence in the relationship, help identify the best model, help identify whether multiple or subpopulations exist in the dataset, and aid in understanding whether other factors are affecting the biota–sediment relationship (e.g., other biological or sediment chemistry factors).

Although they have limitations, BSARs are better suited than BSAFs for answering practical sediment management questions. BSAR derivation does not presume that the relationship between tissue and sediment chemical concentrations is of a particular form, or even that such a relationship exists. BSARs tend to be based on a statistical analysis of site-specific data, and so they are governed by what the data say. Generally speaking, BSARs describe data trends, how well the underlying data fit the trend, and whether any particular data are unduly influencing BSAR selection. This promotes an understanding of model limitations and allows for uncertainties to be acknowledged, understood, and quantified.

In summary, BSAFs and BSARs, including Model I and Model II regressions, are often relied on to make important decisions with significant financial consequences, based on the presumption that those decisions will yield public health and environmental benefits. These decisions, committing many millions of dollars, are being made based on models that include large, unacknowledged uncertainties and, at least in some cases, are simply invalid. When biota–sediment accumulation relationships and their uncertainties are better understood and modeled, better public health and environmental cleanup decisions should follow.

Acknowledgment—The work presented here was supported solely by Windward Environmental. We thank our reviewers for their time and constructive comments.

SUPPLEMENTAL DATA

Figures S1–S4. Smearing Factor for Log Transformed Data

REFERENCES

Ankley GT, Lodge K, Call DJ, Balcer MD, Brooke LT, Cook PM, Kreis RG Jr, Carlson AR, Johnson RD, Niemi GJ., et al. 1992. Integrated assessment of contaminated sediments in the lower Fox River and Green Bay, Wisconsin. *Ecotox Environ Saf* 23:46–63.

Box GEP, Draper NR. 1987. Empirical model-building and response surfaces Wiley Series in Probability and Mathematical Statistics. New York (NY): John Wiley & Sons. 669 p.

Burkhard LP. 2009. Estimation of biota sediment accumulation factor (BSAF) from paired observations of chemical concentrations in biota and sediment. EPA/600/R-06/047; ERASC-013F. Ecological Risk Assessment Support Center. Cincinnati (OH): US Environmental Protection Agency.

Burkhard LP, Cowan-Ellsberry C, Embry MR, Hoke RA, Kidd KA. 2012a. Bioaccumulation data from laboratory and field studies: are they comparable? *Integr Environ Assess Manag* 8:13–16.

Burkhard LP, Arnot JA, Embry MR, Farley KJ, Hoke RA, Kitano M, Leslie HA, Lotufo GR, Parkerton TF, Sappington KG., et al. 2012b. Comparing laboratory and field-measured bioaccumulation endpoints. *Integr Environ Assess Manag* 8:17–31.

DeForest DK, Brix KV, Adams WJ. 2007. Assessing metal bioaccumulation in aquatic environments: The inverse relationship between bioaccumulation factors, trophic transfer factors and exposure concentration. *Aquat Toxicol* 84: 236–246.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist Sci* 1:54–77.

Helsel DR. 2005. Nondetects and data analysis: Statistics for censored environmental data. Hoboken (NJ): John Wiley & Sons. 250 p.

Helsel DR. 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65:2434–2439.

Helsel DR, Hirsch RM. 2002. Statistical methods in water resources. Chapter A3, Book 4, Hydrologic analysis and interpretation, Techniques of water-resources investigations of the United States Geological Survey [online]. Washington (DC): US Geological Survey. Updated 2002. Available from: http://pubs.usgs.gov/twri/twri4a3

Lake JL, Rubinstein NI, Lee H, Lake CA, Heltshe J, Pavignano S. 1990. Equilibrium partitioning and bioaccumulation of sediment-associated contaminants by infaunal organisms. *Environ Toxicol Chem* 9:1095–1106.

Lee H. 1992. Models, muddles, and mud: Predicting bioaccumulation of sediment-associated pollutants. In: Burton GA Jr, editor. Sediment toxicity assessment. Boca Raton (FL): Lewis Publishers. p. 267–293.

McFarland VA. 1995. Evaluation of field-generated accumulation factors for predicting the bioaccumulation potential of sediment-associated PAH

compounds. Long-term effects of dredging operations program technical report D- 95-2. Vicksburg (MS): Waterways Experiment Station, US Army Corps of Engineers.

Melwani AR, Greenfield BK, Byron ER. 2009. Empirical estimation of biota exposure range for calculation of bioaccumulation parameters. *Integr Environ Assess Manag* 5:138–149.

Neter J, Wasserman W, Kutner MH. 1990. Applied linear statistical models. 3rd ed. IRWIN. Homewood, IL. 1184 p.

[ORNL] Oak Ridge National Laboratory. 1998. Biota sediment accumulation factors for invertebrates: review and recommendations for the Oak Ridge Reservation. BJC/OR-112. Oak Ridge (TN): ORNL.

Parkerton TF, Arnot JA, Weisbrod AV, Russom CL, Hoke RA, Woodburn K, Traas T, Bonnell M, Burkhard LP, Lampi MA. 2008. Guidance for evaluating in vivo fish bioaccumulation data. *Integr Environ Assess Manag* 4:139–155.

Pruell RJ, Rubinstein NI, Taplin BK, LiVolsi JA, Norwood CB. 1990. 2,3,7,8-TCDD, 2, 3,7,8-TCDF and PCBs in marine sediments and biota: laboratory and field studies. EPA/600/8-90/068. Narragansett (RI): Environmental Research Laboratory, US Environmental Protection Agency.

Sample BE, Suter GW, Beauchamp JJ, Efroymson RA. 1999. Literature-derived bioaccumulation models for earthworms: development and validation. *Environ Toxicol Chem* 18:2110–2120.

Sokal RR, Rohlf FJ. 1969. Introduction to biostatistics.: 2nd ed. New York (NY): W.H. Freeman and Co. 363 p.

Sokal RR, Rohlf FJ. 2012. Biometry: The principles and practice of statistics in biological research. 4th ed. New York (NY): W.H. Freeman and Co. 937 p.

[USACE] US Army Corps of Engineers. 2009. BSAF database [online database]. Vicksburg (MS): USACE Environmental Laboratory. Updated 11/24/09. [cited 2012 March 7]. Available from: http://el.erdc.usace.army.mil/bsaf/bsaf.html

[USEPA] US Environmental Protection Agency. 1994. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. First Edition. EPA/600/R-99/024. Duluth (MN): USEPA.

[USEPA] US Environmental Protection Agency. EPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. Second Edition. EPA/600/R-99/064. Washington DC: USEPA Environmental Protection Agency

[USEPA] US Environmental Protection Agency. 2003. Guidance for developing ecological soil screening levels (Eco-SSLs). OSWER Directive 9285. 7-55 [online]. Office of Solid Waste and Emergency Response, Washington (DC): USEPA. Updated November 2003. Available from: http://rais.ornl.gov/documents/ecossl.pdf

[USEPA] US Environmental Protection Agency. 2007. Framework for metals risk assessmen. EPA 120/R-07/001. Office of the Science Advisor, Risk Assessment Forum, Washington DC: USEPA.

[USEPA] US Environmental Protection Agency. 2008. BSAF (biota-sediment accumulation factor) data set [online]. Mid-Continent Ecology Division, Duluth (MN): USEPA. Updated 1/25/08. [cited 2009 July 15]. Available from: http://www.epa.gov/med/Prods_Pubs/bsaf.htm

Wong CS, Capel PD, Nowell LH. 2001. National-scale, field-based evaluation of the biota-sediment accumulation factor model. *Environ Sci Technol* 35: 1709–1715.

Yunker MB, Cretney WJ. 2000. Bioavailability of chlorinated dibenzo-p-dioxins and dibenzofurans to Dungeness crab (*Cancer magister*) at marine pulp mill sites in British Columbia, Canada. *Environ Toxicol Chem* 19:2997–3011.

Zar JH. 1996. Biostatistical analysis. 3rd ed. Upper Saddle River (NJ): Prentice Hall. 662 p.