SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 20, No. 11, pp. 2561–2571, 2001

Printed in the USA
0730-7268/01 $9.00 + .00

UPTAKE OF INORGANIC CHEMICALS FROM SOIL BY PLANT LEAVES: REGRESSIONS OF FIELD DATA

Rebecca A. Efroymson,* Bradley E. Sample, and Glenn W. Suter II
Environmental Sciences Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831-6036, USA

(*Received* 11 *July* 2000; *Accepted* 9 *April* 2001)

Abstract—The estimation of chemical concentrations in wildlife foods, such as plant foliage, is often performed for risk assessments at contaminated sites. Regression models and uptake factors for use in estimating the uptake of inorganic elements from soil by above-ground plant tissues were derived in this study. These included models for arsenic, cadmium, copper, lead, mercury, nickel, selenium, and zinc. Models were developed using published data from soil contaminated in the field and were validated using measured concentrations from two contaminated sites. Single-variable regression models of log-transformed concentrations in plants versus log-transformed concentrations in soil are generally recommended over simple uptake factors for use in estimating plant uptake of inorganic contaminants in ecological risk assessments. Multiple regression models with soil concentration and pH as the variables are also recommended for estimating the uptake of four chemicals (cadmium, mercury, selenium, and zinc) by plants. Models for use in screening risk assessments, i.e., the upper 95% prediction limits on the regressions, are recommended to provide conservative estimates of uptake of inorganic chemicals by plants.

Keywords—Bioaccumulation Plant uptake Exposure Metal Ecological risk assessment

INTRODUCTION

The major pathway of exposure of terrestrial wildlife to chemical contaminants in soil at hazardous waste sites is through ingestion. Uptake through dermal and inhalation pathways is usually negligible [1]. Thus, the prediction or estimation of risks to wildlife requires knowledge of concentrations of contaminants in all ingested media (food, soil, and water) as well as the fractionation of their diets, ingestion rates, body weights, and habitats. In addition, the calculation of regulatory soil screening levels that are protective of ecological receptors at generic waste sites requires models to estimate chemical concentrations in wildlife foods based on soil concentrations [2]. A primary dietary medium for many terrestrial vertebrates is vegetation.

The collection of site-specific measurements of the uptake of contaminants in vegetation is often limited by the funding for or timing of an ecological risk assessment. In addition, if the particular chemical of interest is not viewed as a likely driver for the remedial decision, the relevant data will probably not be collected. Thus, models are needed to estimate concentrations of chemicals in plant tissues from concentrations in soil for a wide variety of soil types. The simplest linear model for estimating the concentrations of chemicals in vascular plants is the soil–plant uptake factor, the ratio of the concentration of a chemical in vegetation or a portion of vegetation such as foliage to that in soil [3]. Thus, the concentration of a contaminant in plants at a particular location is estimated by multiplying the measured concentration in soil by the soil–plant uptake factor. The usefulness of uptake factors lies in the ease by which distributions of these values can be developed and conservative (e.g., 90th percentile) values chosen for use in wildlife dietary uptake models for screening-level risk assessments. However, the evidence below suggests that uncertainty in uptake model predictions may be minimized if nonlinear models are employed and environmental factors and other sources of variability are incorporated in the model.

Although a constant soil–plant uptake factor is probably valid for narrow ranges of chemical concentration in the relatively nontoxic range [4,5], uptake factors have been demonstrated to be dependent on the chemical concentration in soil. Baes et al. [3], who developed soil–vegetative tissue uptake factors that are often used in human health and ecological risk assessments, found that the uptake factors for copper and zinc were inversely correlated with soil concentration. These metal contaminants are also nutrients, and it is not surprising that their accumulation would be regulated by plants. Alsop et al. [6] showed that the use of Baes factors underpredicted the uptake of zinc by oats and deer mice at concentrations below 75 mg/kg in soil and overpredicted metal concentrations in the plants and mice at concentrations exceeding 75 mg/kg in soil. Similarly, these results suggest that nonlinear models could be more useful for risk assessments than the Baes factors. The power model, which represents the rate of uptake as diminishing with increasing concentration, appears appropriate, and log-log transformations linearize that model. The reasonableness of this model is also suggested by its use by Neuhauser et al. [7] and Sample et al. [8] to obtain significant regressions for the uptake of inorganic elements by earthworms.

Although the primary objective of this study is to quantify the relationships between total concentrations of inorganic elements in soils and those in plants, it is recognized that environmental and physiological factors affect the extent of accumulation. For example, the amount of zinc in soil water and

---

* To whom correspondence may be addressed (efroymsonra@ornl.gov).

Publication 5067, Environmental Sciences Division, Oak Ridge National Laboratory. This article has been authorized by a contractor of the U.S. government. Accordingly, the U.S. government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution or allow others to do so for U.S. government purposes.

plant tissues is generally observed to increase with decreasing pH and cation exchange capacity [9]. Cadmium uptake by plants has been shown in numerous studies to decrease with increasing pH [10,11]. Sims and Kline [12] found significant multiple regression models between nickel, copper, and zinc in wheat and soybean and soil metal concentrations and pH, but not with soil metal concentrations alone. In addition to the total concentration of a chemical in soil, pH has been selected as an independent variable in this study.

In this article, we present single-variable regressions using natural log-transformed, above-ground plant and soil concentrations; multiple regressions of natural log-transformed plant concentration on natural log-transformed soil concentration and pH; and summary statistics for and distributions of soil–plant uptake factors for eight inorganic elements, including arsenic, cadmium, copper, lead, mercury, nickel, selenium, and zinc. Measurements of chemicals in plants and collocated soils from Bartlesville, Oklahoma, USA [13], and the Clark Fork River floodplain in Montana, USA [14], are used to validate the uptake models.

METHODS

Database development

Field and greenhouse studies in which concentrations of arsenic, cadmium, copper, lead, mercury, nickel, selenium, or zinc in which both surface soil and collocated, above-ground plant tissue were analyzed were identified. Information regarding soil and plant concentrations, soil parameters, exposure time, chemical form, dry or wet weight, extraction method, plant species, and plant part was compiled in a spreadsheet. Families and genera in the database are listed in the Appendix. The database included the growth forms (number of observations in parentheses) graminoid (525), forb/herb (544), forb/herb or vine (4), forb/herb or shrub (69), shrub (16), tree or shrub (17), tree (49), and unknown or composited samples (107). Approximately 30% of the data represented chemical concentrations in plant leaves, excluding stems, fruits, and seeds; the remaining above-ground samples included clippings, unspecified above-ground parts, or shoots. Samples of fruits or seeds alone were excluded from the database. The complete list of plant species in the database has been posted as a table at www.esd.ornl.gov/programs/ecorisk/guidance.html.

Tests in which salts (e.g., cadmium chloride, copper sulfate, sodium selenate) were added in solution to soil were excluded because of preliminary results that suggested regressions of concentrations in plants on concentrations in soil were different for field and salt chemical forms.

Only studies in which concentrations were expressed on an air- or oven-dry weight basis were used. Although most studies reported that plant material was washed, studies were not excluded if the extent of washing was not stated in the paper. Studies were used even if the individual investigators observed no correlation between concentrations of contaminants in soils and plants (e.g., arsenic in Norway spruce [15]; or copper in radish foliage [16]). Concentrations of chemicals in soil or plants were sometimes estimated from graphs in journal articles but only if estimates could be made within about 10%. Data for species that are known to hyperaccumulate metals were excluded. Data for which measured concentrations were below detection thresholds were excluded.

Each plant species or variety, soil type, location, and concentration of the test element in soil represented an independent observation in the dataset. Differences in exposure duration or above-ground plant part did not constitute separate observations; concentrations in soils or plants that differed on the basis of one of these two variables were averaged. The number of observations in these means, which ranged between one and six, was not retained in the subsequent statistical analysis.

Concentrations of contaminants in soil at the time of plant sampling were used if known. If these concentrations were not measured (as was often the case in pot studies), the initial concentration of the element measured in or added to soil was assumed to be equivalent to the final concentration. In field experiments, the change in soil concentration of an element over time was assumed to be minimal.

Observations were included in the database if the total chemical concentration in soil was measured, either by extraction with strong acid or by extraction with moderately strong acid (e.g., 4 N sulfuric acid) sometimes accompanied by heat. In one study, it was shown that extraction of arsenic with 6 M HCl for 2 h under constant rotation gave the same recovery as digestion in aqua regia, a mixture of concentrated nitric and hydrochloric acids [17]. Studies in which concentrations of contaminants in soil were determined by a partial extraction with diethylene triamine pentaacetic acid, weak acids, or water were excluded from analysis. Although concentrations of diethylene triamine pentaacetic acid-extracted contaminants from soils sometimes correlate with those taken up by plants [18], this method produces invalid estimates of bioavailability for some metals [18–20] and invalid comparisons of soils of varying pH [21].

For studies in which contaminant concentrations at multiple depths were measured, the concentration at the 0- to 10-, 0- to 15-, or 0- to 20-cm-depth interval was recorded. Where only a single soil depth was measured, it ranged from 5 to 70 cm. An exception was the 1- to 2-cm-depth samples in Severson et al. [22], which represented A horizons of the Frisian Islands, which have no B horizon development.

Studies included contamination from mine wastes (ores, tailings), smelter deposits, other industrial sources, vehicle and other urban emissions, wastewater effluents, composts, fertilizers, dredged materials, sewage sludges, fly ashes, flue dusts, nuclear waste, and arsenical pesticide residues [23, http://www.esd.ornl.gov/programs/ecorisk/guidance/html]. Where materials such as fertilizers were added to soil, data were excluded if mixing with soil did not occur. In addition, some measurements were taken from background locations. For example, chemical data for arsenic included as sources mine waste (24 observations), smelter operations (23 observations), fly ash disposal (18 observations), pesticide use (19 observations), nuclear waste (4 observations), unidentified urban sources (3 observations), background or no apparent anthropogenic source (13 observations), and unknown source (18 observations). Field studies in which a current, local atmospheric source of contaminants was present were excluded from the database. The initial database spreadsheet with references is included in Bechtel Jacobs Company [23]; however, additional data on plant taxa and plant parts are available in the supplemental dataset for this article at www.esd.ornl.gov/programs/ecorisk/guidance.html.

Model development and validation

Chemical uptake data were assigned to two groups, the original dataset used for initial model development and the validation dataset, employed to test the accuracy and predictive utility of the models. The two studies from Bartlesville [13] and the Clark Fork River watershed [14] were obtained for the purpose of model validation because data from these studies were used in ecological risk assessments, the expected primary use of the models developed in this article. Because sampling and analytical variability and environmental characteristics are likely to be correlated among data from the same study, it was assumed that data from wholly independent studies (i.e., studies from which no data were included in the initial model development) would be unbiased and would provide a better test of the uptake models than would randomly selected observations extracted from the combined dataset.

Soil–plant uptake factors were calculated for each observation in the original dataset. Summary statistics were generated for each chemical. To facilitate the use of the uptake factors in probabilistic risk evaluations, the distribution of the calculated factors for each chemical in the database was evaluated. The Shapiro–Wilk test (PROC UNIVARIATE; [24]) was applied to the untransformed and natural-log-transformed uptake factors for each chemical to determine whether the distribution of uptake factors differed significantly from normal or log normal, respectively.

To further evaluate the relationship between the contaminant concentration in soil and plants, single-variable and multiple regressions were performed using SAS PROC REG [24]. Contaminant concentrations in both soil and plants were natural-log transformed prior to regression analyses. Because data concerning the number of individuals and samples included in composites or means were not available for all observations, a weighting of observations was not applied. Linear regression models of natural-log-transformed plant concentration on natural-log-transformed soil concentration were developed for each chemical. Multiple regression models incorporating soil pH were also developed for each chemical, although pH was not available for all observations.

To evaluate the appropriateness and accuracy of various models for generating estimates for general application, estimated concentrations in above-ground plant tissues were generated using best-estimate and conservative models. The best-estimate models were the median uptake factor and single-variable and multiple regressions developed in this study. The conservative models, which are needed for screening assessments, were also generated using the 90th percentile uptake factor and the upper 95% prediction limit for the single-variable regression model from this study. The 95% upper prediction limit was calculated according to a method from Dowdy and Wearden [25].

For each chemical and model, differences between estimated and measured concentrations in validation observations were evaluated using Wilcoxon signed-rank tests (PROC UNIVARIATE; [24]). These nonparametric tests were used because no assumptions of normality were necessary and data were not generally normally distributed. Differences were considered significant if $p(H_0 = 0) \leq 0.05$. Relative accuracy and quality of different estimations were evaluated by calculating the proportional deviation of the estimate from the measured value [8] as

$$PD = (M_i - E_i)/M_i$$

where PD = proportional deviation, M_i = measured concentration for chemical in plant at soil concentration i, and E_i = estimated concentration for chemical in plant at soil concentration i.

The percentage of estimated values that exceeded their corresponding measured value was also tabulated by each chemical and estimation method. The relative quality of general estimation methods was evaluated by the following criteria from Sample et al. [8], with these criteria indicating increasing quality: median PD closest to zero (indicates that estimates center around measured values); PD with narrowest range (indicates relative precision of method); percentage overestimation closest to 50% (indicates that estimates center around measured values); and difference between estimated and measured values not significantly different, as determined by Wilcoxon signed-rank tests. Estimation methods were evaluated using these criteria in a weight-of-evidence approach. The fourth criterion was weighted somewhat less than the other three if all criteria did not indicate the same, preferred estimation method; the Wilcoxon test can be influenced by sample size.

Indicators of quality of conservative estimation methods were smallest, negative median PD value (indicates that method overestimates while minimizing the degree of overestimation), and PD with narrowest range (to minimize the degree of overestimation).

Linear regressions of the natural-log-transformed concentrations in the plant and soil validation dataset were performed and compared with single-variable regression models (i.e., soil concentration only) developed from the original observations using the F-test procedure for comparing regression lines outlined in Draper and Smith [26]. Differences were considered significant if $p \leq 0.05$.

Following validation analyses, the original and validation datasets were pooled, even if they differed significantly, and uptake factors and single and multiple regression models were recalculated. These results were reported as the final uptake factor or regression model.

RESULTS

Modeling results

Soil–plant regression models and uptake factors were developed for eight inorganic chemicals, including arsenic, cadmium, copper, lead, mercury, nickel, selenium, and zinc (Fig. 1). In the original dataset with salts data excluded, the number of observations ranged from 99 for arsenic to 164 for zinc (Table 1). The number of studies incorporated in the models ranged from 7 for nickel to 20 for zinc. Six of eight distributions of uptake factors fit a log-normal distribution more closely than a normal distribution, although only the distribution of uptake factors for arsenic, lead, selenium, and zinc fit the log-normal form well (Table 1). Median uptake factors for all chemicals were less than one; however, the maximum uptake factor for all chemicals exceeded one. The distributions of uptake factors for all eight chemicals spanned at least two orders of magnitude; for copper, the range of uptake factors was less than three orders of magnitude, and for arsenic, the range was greater than five orders of magnitude. As an example, the cumulative distribution of uptake factors for selenium is presented in Figure 2.

Regression of the natural log of chemical concentrations in plants versus the natural log of those in soil produced significant model fits for seven of eight chemicals using the original dataset (Table 2). The exception was arsenic ($p = 0.06$). Slopes of all significant regression models were positive. The r^2 values for the significant models ranged from 0.12 for nickel to 0.68 for mercury. The slopes of all regression models were positive.

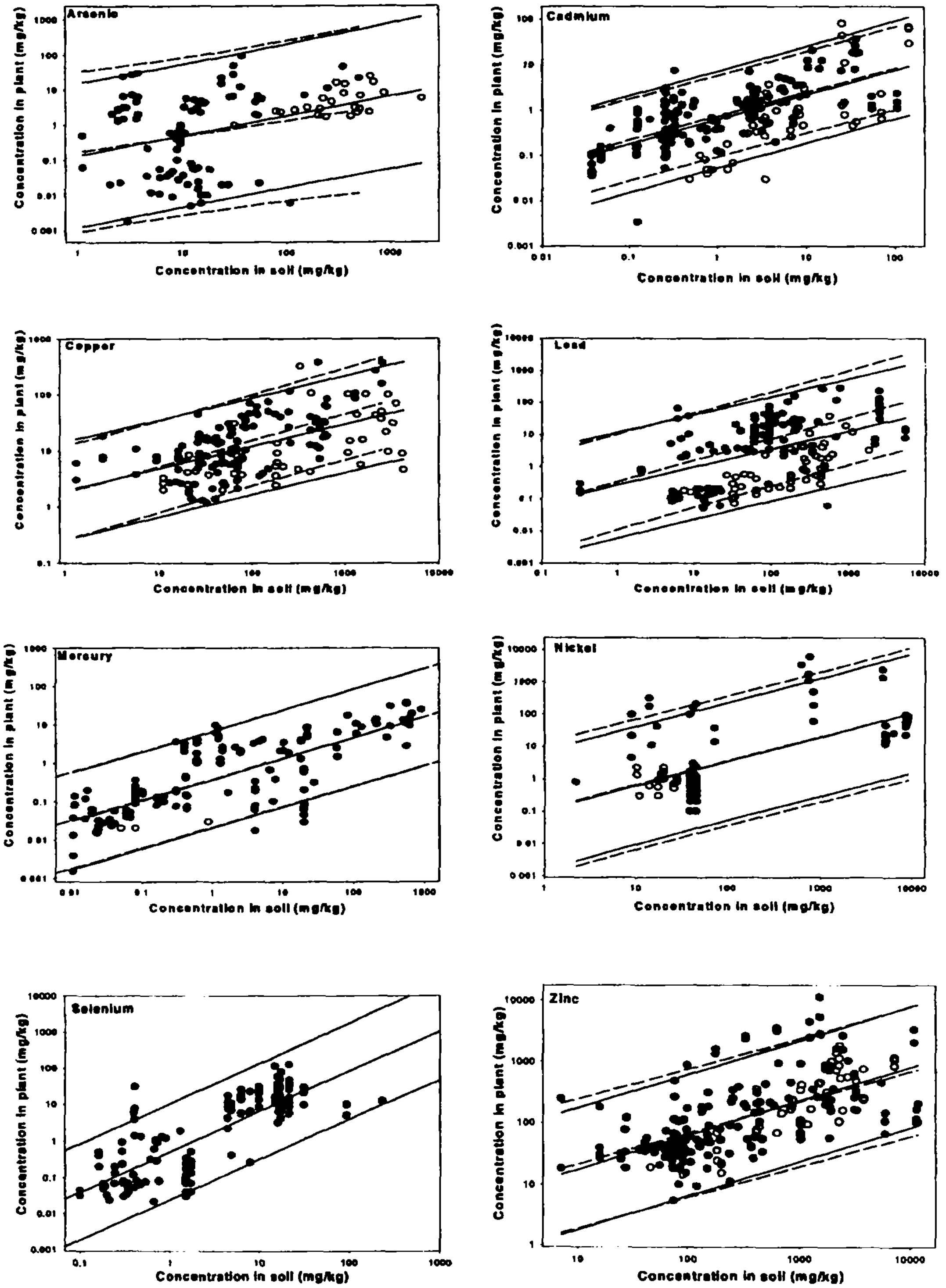

Fig. 1. Scatterplots of chemical concentrations in above-ground vegetation versus soil. Lines represent single-variable linear regression models and 95% prediction intervals for natural-log-transformed data for two datasets, i.e., the original, not including validation data (---), and original plus validation data sets (—). ● represents observations from the original dataset and ○ represents those from the validation dataset. No validation data were available for Se.

Intercepts differed significantly from zero for all eight chemicals.

The soil pH was not available for many observations in the database; thus, the inclusion of this variable in the regression models resulted in decreases in sample size (Table 3). Consequently, the single-variable and multiple regression models are not directly comparable. The addition of soil pH in the regression model resulted in significant model fits for all chemicals except nickel. The pH contributed significantly to the model fit for copper, lead, mercury, and selenium (Table 3).

Table 1. Summary statistics for soil-to-plant uptake factors (dry wt basis)

Chemical	n (studies)	n (observations)	Mean	Standard deviation	Minimum	Median	90th percentile	Maximum	Distribution
Based on original dataset only									
Arsenic	9	99	0.56	1.5	0.000056	0.047	1.2	9.1	Log normal
Cadmium	17	155	1.7	2.7	0.011	0.83	3.8	23	Neither[a]
Copper	17	125	0.44	0.91	0.014	0.20	0.71	7.4	Neither[a]
Lead	19	133	0.34	1.1	0.00011	0.12	0.56	11	Log normal
Mercury	12	142	1.5	2.5	0.0015	0.66	5.0	12	Neither
Nickel	7	90	0.91	3.1	0.0022	0.014	2.4	22	Neither
Selenium	14	156	2.3	8.7	0.020	0.70	3.0	77	Log normal
Zinc	20	164	1.3	3.2	0.0086	0.43	2.6	34	Log normal
Based on original and validation datasets combined									
Arsenic	11	122	0.45	1.4	0.000060	0.038	1.1	9.1	Log normal
Cadmium	19	207	1.4	2.4	0.0087	0.59	3.3	23	Neither[a]
Copper	19	180	0.34	0.78	0.0011	0.12	0.63	7.4	Log normal
Lead	21	189	0.25	0.92	0.00011	0.039	0.47	11	Neither[a]
Mercury	14	145	1.5	2.5	0.0015	0.65	5.0	12	Neither
Nickel	9	111	0.75	2.7	0.0022	0.018	1.4	22	Neither
Selenium	16	158	2.25	8.7	0.020	0.67	3.0	77	Log normal
Zinc	22	220	1.02	2.8	0.0086	0.37	1.8	34	Log normal

[a] Log normal closest.

Validation results

Data for model validation were available for all chemicals, but the two observations of selenium in soils and plants, which were identical concentrations, were insufficient for the construction of a regression model. A comparison of single-variable regression models for the log-transformed contaminant concentrations from original and validation datasets indicated that the models were statistically significantly different for cadmium (p = 1E−8), copper (p = 1E−5), lead (p = 2E−16), and zinc (p = 0.02).

The predictive utility of soil–plant uptake factors and regression models was determined in part by evaluating the statistical significance of differences between measured and estimated values. Using the validation dataset, significant differences between measured and estimated concentrations were observed for six of seven chemicals using the median uptake factor; such a difference was not observed in the case of mercury (Table 4). Significant differences in concentrations measured and those estimated using the single-variable regression model were observed for arsenic, copper, and lead but not for cadmium, copper, nickel, or zinc (Table 4). Significant differences between concentrations measured and those estimated using the multiple regression model with pH were found for lead and nickel only. All three general estimation methods overestimated measured plant concentrations for over 50% of validation soil concentrations for all chemicals except for the uptake factor for nickel (10% overestimation) and the single-variable regression model for arsenic (17% overestimation). Median proportional deviations of estimated values from measured values ranged from −0.19 for the multiple regression model for Zn to a maximum of −48.61 for the multiple regression model for nickel (Table 4).

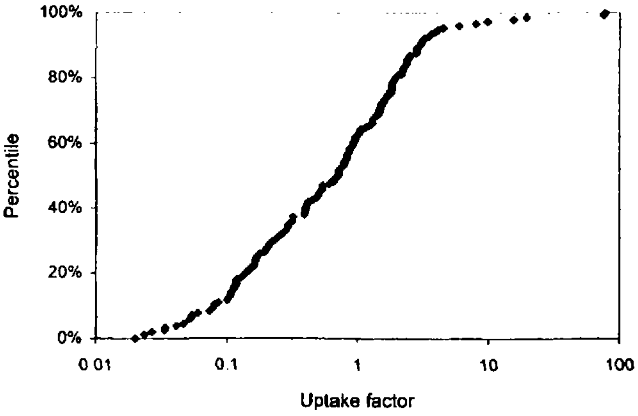

Fig. 2. Cumulative distribution of soil–plant uptake factors for selenium. Chemical concentrations were measured in contaminated soils and plants from many sites.

Table 2. Results of regression of ln(concentration in plant) on ln(concentration in soil)[a]

Chemical	n	$B_0 \pm$ SE	$B_1 \pm$ SE	r^2	p Model fit
Based on original dataset only					
Arsenic	99	-1.75 ± 0.60^{b}	0.44 ± 0.23^{c}	0.77	0.0573
Cadmium	155	-0.30 ± 0.08^{d}	0.53 ± 0.05^{d}	0.45	0.0001
Copper	125	0.57 ± 0.25^{e}	0.47 ± 0.05^{d}	0.41	0.0001
Lead	133	-1.09 ± 0.33^{b}	0.67 ± 0.07^{d}	0.44	0.0001
Mercury	142	-0.96 ± 0.12^{d}	0.53 ± 0.04^{d}	0.68	0.0001
Nickel	90	-2.12 ± 0.60^{d}	0.74 ± 0.11^{d}	0.12	0.0001
Selenium	156	-0.68 ± 0.14^{d}	1.11 ± 0.07^{d}	0.63	0.0001
Zinc	164	1.89 ± 0.33^{d}	0.50 ± 0.06^{d}	0.32	0.0001
Based on original and validation datasets combined					
Arsenic	122	-1.99 ± 0.43^{d}	$0.56 \pm 0.13^{d,f}$	0.15	0.0001
Cadmium	207	-0.48 ± 0.09^{d}	$0.55 \pm 0.04^{d,f}$	0.45	0.0001
Copper	180	0.67 ± 0.21^{d}	$0.39 \pm 0.04^{d,f}$	0.31	0.0001
Lead	189	-1.33 ± 0.35^{d}	$0.56 \pm 0.07^{d,f}$	0.24	0.0001
Mercury	145	-1.00 ± 0.12^{d}	$0.54 \pm 0.04^{d,f}$	0.60	0.0001
Nickel	111	-2.22 ± 0.47^{d}	$0.75 \pm 0.09^{d,f}$	0.37	0.0001
Selenium	158	-0.68 ± 0.14^{d}	1.10 ± 0.07^{d}	0.63	0.0001
Zinc	220	1.58 ± 0.28^{d}	$0.56 \pm 0.05^{d,f}$	0.40	0.0001

[a] Model: ln(concentration in above-ground plant) = B_0 + B_1 (ln[concentration in soil]), where concentrations (mg/kg) are expressed on a dry weight basis and B_0 = intercept of regression and B_1 slope of regression. SE = standard error.

[b] $0.001 < p \leq 0.01$.

[c] $p > 0.05$.

[d] $p \leq 0.001$.

[e] $0.01 < p \leq 0.05$.

[f] Slope different from one, indicative that plant uptake factor is not best model ($p \leq 0.05$); only combined dataset tested.

Table 3. Results of regression of ln(concentration in plant) on ln(concentration in soil) and pH[a]

Chemical	n	B_0 ± SE	B_1 ± SE	B_2 ± SE	r^2	p Model fit
Based on original dataset only						
Arsenic	24	−4.85 ± 2.74[b]	0.68 ± 0.31[c]	0.40 ± 0.56[b]	0.77	0.0001
Cadmium	117	0.70 ± 0.64[b]	0.54 ± 0.06[d]	−0.17 ± 0.10[b]	0.45	0.0001
Copper	85	−0.67 ± 0.47[b]	0.29 ± 0.07[d]	0.27 ± 0.08[e]	0.41	0.0001
Lead	103	−3.40 ± 1.05[e]	0.68 ± 0.08[d]	0.31 ± 0.15[c]	0.44	0.0001
Mercury	79	−4.12 ± 1.16[d]	0.64 ± 0.06[d]	0.42 ± 0.19[c]	0.68	0.0001
Nickel	36	−0.43 ± 2.33[b]	0.13 ± 0.12[b]	0.62 ± 0.35[b]	0.12	0.1235
Selenium	146	−8.94 ± 0.73[d]	0.98 ± 0.05[d]	1.18 ± 0.11[d]	0.85	0.0001
Zinc	167	2.28 ± 0.51[d]	0.57 ± 0.08[d]	−0.13 ± 0.10[b]	0.85	0.0001
Based on original and validation dataset combined						
Arsenic	47	−2.56 ± 0.76[e]	0.69 ± 0.06[d]	0.02 ± 0.12[b]	0.78	0.0001
Cadmium	170	1.15 ± 0.64[b]	0.56 ± 0.05[d]	−0.27 ± 0.10[e]	0.46	0.0001
Copper	140	0.51 ± 0.49[b]	0.36 ± 0.05[d]	0.01 ± 0.08[b]	0.33	0.0001
Lead	159	−1.93 ± 1.03[b]	0.56 ± 0.08[d]	0.04 ± 0.14[b]	0.27	0.0001
Mercury	82	−4.19 ± 1.14[d]	0.64 ± 0.06[d]	0.42 ± 0.19[c]	0.68	0.0001
Nickel	57	−2.06 ± 2.53[b]	0.57 ± 0.10[d]	0.26 ± 0.39[b]	0.36	0.0001
Selenium	148	−8.83 ± 0.72[d]	0.99 ± 0.05[d]	1.17 ± 0.11[d]	0.85	0.0001
Zinc	193	2.36 ± 0.44[d]	0.64 ± 0.06[d]	−0.21 ± 0.08[e]	0.41	0.0001

[a] Model: ln(concentration in above-ground plant) = $B_0 + B_1(\ln[\text{soil}]) + B_2(\text{pH})$, where concentrations (mg/kg) are expressed on a dry weight basis, B_0 = intercept of regression, B_1 = soil concentration coefficient, and B_2 = pH coefficient. SE = standard error.

[b] $p > 0.05$.

[c] $0.01 < p \le 0.05$.

[d] $p \le 0.001$.

[e] $0.001 < p \le 0.01$.

Using the selection criteria above, the best estimates for the uptake of arsenic and nickel by plants are provided by the single-variable regression model, and the best estimates for copper and mercury are provided by the multiple regression with pH. Results for cadmium, lead, and zinc are similar for the single-variable and multiple regressions. (These results do not constitute a recommendation of these models. A test of significance of pH as a variable was also required.)

Among conservative estimation methods, both the 90th percentile uptake factor and the 95% upper prediction limit for the single-variable regression model significantly overestimated measured concentrations in plants for all chemicals except mercury. The 95% upper prediction limit produced the best conservative estimate (i.e., smallest negative median and smallest range proportional deviation) of chemical concentrations in above-ground plant tissue for arsenic, copper, and zinc, with percent overestimates ranging from 96 to 100% (Table 5). The best conservative estimates for cadmium, lead, and mercury were obtained using the 90th percentile uptake factor. Conservative estimation methods were approximately equivalent for nickel.

Models incorporating data from validation

Final regression models and soil–plant uptake factors that incorporate data from both the original and validation datasets were calculated for all chemicals. Uptake factors based on the combined dataset were generally similar to those based on only the original dataset, though the median uptake factors of the combined dataset were always lower than those from the original dataset except in the case of nickel (Table 1). Distributions of uptake factors for most chemicals more closely resembled a log-normal than a normal shape. No median uptake factor was greater than one, though for six of eight chemicals, the 90th percentile uptake factor was >1 (Table 1).

In general, results of simple regression analyses differed little between the original and combined datasets. The model fit for arsenic was improved (and became significant) after the addition of the validation data to the dataset (Table 2). Slopes and intercepts of single-variable regressions for all chemicals became highly significant ($p \le 0.001$), even if they were not significant prior to the inclusion of the validation data.

After the inclusion of the validation data, the multiple re-

Table 4. Comparison of quality of general estimation methods as determined by the proportional deviation (PD) of the estimated values from measured validation values; PD = (measured − estimated)/measured

		Median uptake factor		Single-variable regression model		Regression model with pH	
Chemical	n	Median PD (range)	% Overestimated	Median PD (range)	% Overestimated	Median PD (range)	% Overestimated
Arsenic	23	−1.85[a] (−14.83 to 0.11)	96	0.48[a] (−0.38 to 0.88)	17	−3.46[b] (−43.90 to 0.95)	65
Cadmium	52	−4.64[a] (−94.83 to 0.7417)	90	−1.82[b] (−46.33 to 0.95)	67	−1.65[b] (−43.90 to 0.95)	69
Copper	55	−2.28[a] (−180.37 to 0.80)	85	−1.25[a] (−18.17 to 0.92)	85	−0.65[b] (−7.17 to 0.97)	67
Lead	56	−24.16[a] (−179.97 to −0.02)	100	−11.42[a] (−67.58 to 0.40)	95	−11.01[a] (−60.75 to 0.70)	91
Mercury	3	−1.65[b] (−18.00 to −0.66)	100	−3.92[b] (−10.79 to 2.82)	100	−1.02[b] (−6.19 to −0.92)	100
Nickel	21	0.656[a] (−0.36 to 0.94)	10	−0.27[b] (−3.54 to 0.70)	67	−46.81[a] (−224.39 to −16.90)	100
Zinc	56	−1.20[a] (−8.58 to 0.77)	82	−0.21[b] (−5.29 to 0.91)	55	−0.19[b] (−4.60 to 0.90)	59

[a] Estimate significantly different from measured as determined by Wilcoxon signed-rank test ($p \le 0.001$).

[b] Estimate not significantly different from measured ($p < 0.05$).

Table 5. Comparison of quality of conservative estimation methods as determined by the proportional deviation (PD) of the estimated values from measured validation values; PD = (measured − estimated)/measured

		90th percentile uptake factor		Upper 95% prediction limit for simple regression model	
Chemical	*n*	Median PD (range)	% Overestimated	Median PD (range)	% Overestimated
Arsenic	23	−70.64[a] (−396.75 to 21.31)	100	−38.11[a] (−99.27 to −7.647)	100
Cadmium	52	−24.78[a] (−436.65 to −0.18)	100	−85.13[a] (−12,886.84 to −6.04)	100
Copper	55	−10.66[a] (−643.88 to 0.28)	98	−9.82[a] (−90.51 to 0.60)	98
Lead	56	−119.94[a] (−838.87 to −3.92)	100	−209.36[a] (−1,159.20 to −9.09)	100
Mercury	3	−19.00[b] (−142.33 to 11.50)	100	−217.53[b] (−252.21 to −135.30)	100
Nickel	21	−58.88[a] (−234.78 to −9.95)	100	−59.58[a] (−216.27 to −13.37)	100
Zinc	56	−12.16[a] (−56.23 to −0.39)	100	−8.04[a] (−46.16 to 0.31)	96

[a] Estimate significantly different from measured as determined by Wilcoxon signed-rank test ($p \leq 0.001$).

[b] Estimate not significantly different from measured ($p < 0.05$).

gression model for nickel that included pH became significant (Table 3). Thus, multiple regression models for all eight chemicals were statistically significant. pH was no longer a significant contributor to the regression for copper and lead but became a significant contributor to the regressions for cadmium and zinc (Table 3). The slope of the regression for nickel, which was not significantly different from zero prior to the inclusion of validation data, became significant after their inclusion. With the inclusion of the validation data in the multiple regression with pH, the intercept for arsenic became significant and the intercept for lead became insignificant.

DISCUSSION

The measurement of chemicals in vegetation at specific hazardous waste sites is recommended for remedial investigations, but such samples are often not obtained. To estimate concentrations of inorganic chemicals in above-ground plant tissue from concentrations in soil, risk assessors must use empirical models. Such models are usually soil–plant uptake factors. In this study, uptake factors, single-variable regressions of log-transformed chemical concentrations in plants on log-transformed concentrations in soil, and multiple regressions incorporating pH as well as chemical concentration in soil were generated for eight common inorganic contaminants of soil using published measurements of chemicals in soil and above-ground vegetation. All single-variable and multiple regressions in which validation data were incorporated were statistically significant, even if the chemicals of concern were nutrients (copper, zinc) or if a large fraction of the chemical may volatilize from soil and enter the plant through the leaves (mercury [9,27]).

In contrast, in some of the individual studies from which data were drawn, a significant relationship between chemical concentrations in soil and plants was not observed, but this lack of a relationship could be due to the narrow range of chemical concentrations in soil or few observations in an individual study, soil characteristics or taxonomic differences that were dominant contributors to variability, plant regulatory control over the uptake of essential elements, or inappropriate measurement of exposure concentrations (e.g., wrong soil depth or lateral sampling area for tree roots).

In the multiple regressions, pH was a statistically significant variable for accumulation of cadmium, mercury, selenium, and zinc, and the multiple regression incorporating soil concentration and pH predicted the plant concentration of copper and lead in the validation dataset better than other models. In general, these findings were consistent with the literature, where pH has been observed to affect uptake of cadmium [10,11], selenium [28,29], and zinc [30]. Davies [16] found that lead uptake by radish was best predicted by total lead in soil, and the regression of plant lead on soil lead concentration in that study (like this) was not improved by adding other soil characteristics. The relationship between soil pH and plant uptake of mercury has not been investigated. In contrast with the results of this study, an association of nickel [12,31] and copper [12] concentrations in plants and pH has previously been observed.

Because soil–plant uptake factors are commonly used by risk assessors and are required by many multimedia models, distributions of factors were developed to provide median uptake factors for use as nonconservative estimates of chemical concentrations in plants and 90th percentile soil–plant uptake factors for use as conservative estimates of chemical concentrations in plants. It is not surprising that uptake factors did not lead to the best estimates of plant tissue concentrations in the validation dataset. Uptake factors are a specialized case of the natural-log-transformed, single-variable regression model, where the slope is one. Differences between calculated slopes and one were significant for all chemicals except selenium. For other elements, the calculated slope was less than one, suggesting that the uptake factor should generally decrease with higher concentrations of the chemical in soil. Moreover, for the four chemicals with distributions of uptake factors that were not well fit by either normal or log-normal distributions, empirical distributions (sometimes called custom distributions) are recommended for Monte Carlo analysis of exposure models. In Table 6, median soil–plant uptake factors are compared to Baes factors for vegetative components of plants (foliage and stems) [3], which are used widely in risk assessments.

Table 6. Comparison of geometric mean uptake factors from Baes [3] and median factors from the present study

Chemical	Present study	Baes [3]
Arsenic	0.038	0.04
Cadmium	0.59	0.55
Copper	0.12	0.40
Lead	0.039	0.045
Mercury	0.65	0.90
Nickel	0.018	0.06
Selenium	0.67	0.025
Zinc	0.37	1.5

The source of any discrepancy between factors is unknown because the data used in the derivation of the Baes factors are not published. The use of the two sets of uptake factors could lead to substantial differences in the estimation of risks associated with chemicals such as selenium and zinc.

The amount of variability explained by the soil concentration and pH ranged from r^2 values of 0.27 to 0.85. It should be possible to reduce the high scatter around the regression lines and the high variability in uptake factors for single chemicals by accounting for the other factors that influence uptake (e.g., soil parameters, chemical forms, cocontaminants, nutrients, plant taxa, exposure time, extent of tilling, potential volatility [e.g., Hg], and other biases of the data from which the models were derived). For example, numerous multi-crop studies found differential uptake of inorganic contaminants among the crops. It has also been observed that deciduous trees may accumulate greater concentrations of heavy metals in foliage from soil than do conifers [32]. The physiological differences that explain variability in accumulation of different inorganic chemicals by different plant species are largely unknown [33], though in a study of radiocesium, the rooting depth of plants was most important [34]. Additionally, temperature is expected to affect the uptake of all contaminants. For example, ryegrass grown at 25°C accumulated more cadmium and lead than that grown at 15°C [20].

A large potential source of measurement error for soil concentrations used in all uptake models is the depth to which soil concentrations were measured. The depth interval at which various plants in different environments obtain water and nutrients and the relative biomass of feeder roots at different depths are unknown. Concentrations of a chemical in the top 5 cm of soil versus that in the top 15 cm of soil may vary as much as an order of magnitude, particularly if the source of soil contamination is aerial deposition.

Measurements of accumulation of chemicals by plants are usually taken at a single time without knowledge of whether or not vegetation may be in equilibrium with the soil with respect to chemical transport. However, longer exposure does not necessarily lead to higher plant concentrations. Both the age of the plant and seasonal processes apparently affect uptake. For example, for all leafy and root crops grown in a muck soil, heavy metal concentrations were greater in young crops in the early summer than in mature crops [35].

The variance in the regressions reported here would probably decrease if additional factors were incorporated. Concentrations of soil nutrients [5,17] and cocontaminants [4,36] should aid in the estimation of uptake. The form of the element, such as selenium or mercury, could be represented in a model of uptake by plants [37,38], as could plant species [17,38,39].

RECOMMENDATIONS

Single-variable regressions of the natural log-transformed chemical concentration in the plant on the log-transformed concentration in soil are recommended as good tools for estimating concentrations of contaminants in plant tissues for all eight chemicals tested. The regressions based on combined original and validation datasets should be used rather than those based on the original data alone. Multiple regressions with chemical concentration in soil and pH provide equally good estimates of cadmium, selenium, and zinc in plants and estimates of mercury in plants that are probably better than those obtained from the single-variable regression. Soil–plant uptake factors should be used only if required by risk managers or by a multimedia model that is optimal for other reasons. In addition, users of these regressions should determine that the taxa of concern at a particular site are represented in these regressions. It should be noted, e.g., that the vast majority of observations were from nonwoody plants.

The regression models could underestimate concentrations of chemicals that are consumed by herbivores, given that many of the leaf concentrations in the database did not include particles attached to leaves that wildlife would consume. The reader should note that models developed in this study are not recommended for use in estimating contaminant concentrations in fruits, seeds, or roots because these plant parts were not part of the database unless fruits and seeds were ancillary components of above-ground clippings. Plants typically bioaccumulate inorganic elements in these structures to a different extent than in foliage or stems [5,31,32,40,41].

The regression models are not recommended for contaminated sites where concentrations of chemicals in soil are above those from which the regressions were derived. Even at moderate chemical concentrations in soil, these models are not recommended for estimating chemical concentrations in hyperaccumulating plant species or in plants growing on soils to which chelating agents have been added [42]. Hyperaccumulation by some plants has been demonstrated to occur at levels outside of the prediction intervals presented here, e.g., cadmium in shoots of some populations of *Thlaspi caerulescens* grown in minespoil-contaminated or sewage sludge-contaminated soil [43], cadmium [44] and zinc [45] in leaves of *Cardaminopsis halleri* grown in smelter-contaminated soil, and nickel in leaves of several plant species in serpentine soils [46]. However, the phenomenon of hyperaccumulation is sometimes observed in roots rather than shoots [44]. Whether or not the hyperaccumulation of chemicals could lead to unacceptable risks to wildlife populations would depend on the size of the contaminated area planted with hyperaccumulating species, the suitability of this land as habitat, the extent of hyperaccumulation in the plant parts that constitute food for wildlife (e.g., foliage), and the relative preference of hyperaccumulating plant species by consumers.

The 95% upper prediction limit for the single-variable regression is recommended over the 90th percentile uptake factor for providing conservative estimates of plant uptake of contaminants. The method provided the best conservative estimate for four of eight chemicals, and the log-transformed regression models consistently proved to be better than uptake factors for estimating chemical concentrations in plants.

Measurements of contaminant concentrations in plants at a specific waste site should be taken if possible. They are always more precise than estimates of these concentrations for assessing risks to herbivorous or omnivorous wildlife. Even a small number of samples (e.g., 10 or 20) from which site-specific regressions or uptake factors can be developed would surely give more precise and accurate estimates of concentrations of chemicals in plants at the site than the use of the models recommended above.

Acknowledgement—Research sponsored by the Oak Ridge Operations Office, U.S. Department of Energy, under contract DE-AC05-00OR22725 with the UT-Battelle, LLC. We would like to thank John Beauchamp for statistical guidance on the study methodology. Michael Aplin and M. Elizabeth Will compiled some of the data for this study. We would also like to thank Daniel S. Jones, Mark Peterson, and Christopher Welsh from Oak Ridge National Laboratory for comments on an earlier draft of this manuscript.

REFERENCES

1. Suter GW II, Efroymson RA, Sample BE, Jones DS. 2000. *Ecological Risk Assessment for Contaminated Sites*. Lewis, Boca Raton, FL, USA.
2. U.S. Environmental Protection Agency. 2000. Ecological soil screening level guidance, Draft. Office of Emergency and Remedial Response, Washington, DC.
3. Baes CFI, Sharp RD, Sjoreen AL, Shor RW. 1984. A review and analysis of parameters for assessing transport of environmentally released radionuclides through agriculture. ORNL-5786. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
4. Carlson RW, Bazzaz FA. 1977. Growth reduction in American sycamore (*Plantanus occidentalis* L.) caused by Pb-Cd interaction. *Environ Pollut* 12:243–253.
5. Jiang QQ, Singh BR. 1994. Effect of different forms and sources of arsenic on crop yield and arsenic concentration. *Water Air Soil Pollut* 74:321–343.
6. Alsop WR, Hawkins ET, Stelljes ME, Collins W. 1996. Comparison of modeled and measured tissue concentrations for ecological receptors. *Hum Ecol Risk Assess* 2:539–557.
7. Neuhauser EF, Cukic ZV, Malecki MR, Loehr RC, Durkin PR. 1995. Bioconcentration and biokinetics of heavy metals in the earthworm. *Environ Pollut* 89:293–301.
8. Sample BE, Suter II GW, Beauchamp JJ, Efroymson RA. 1999. Literature-derived bioaccumulation models for earthworms: Development and validation. *Environ Toxicol Chem* 18:2110–2120.
9. Bysshe SE. 1988. Uptake by biota. In Bodek I, Luman WJ, Reehl WF, Rosenblatt DH, eds, *Environmental Inorganic Chemistry: Properties, Processes, and Estimation Method*. Pergamon, New York, NY, USA, pp 4-1–4.7-1.
10. He QB, Singh BR. 1994. Crop uptake of cadmium from phosphorus fertilizers: I. Yield and cadmium content. *Water Air Soil Pollut* 74:251–265.
11. Miller JE, Hassett JJ, Koeppe DE. 1976. Uptake of cadmium by soybeans as influenced by soil cation exchange capacity, pH, and available phosphorus. *J Environ Qual* 5:157–160.
12. Sims JT, Kline JS. 1991. Chemical fractionation and plant uptake of heavy metals in soils amended with co-composted sewage sludge. *J Environ Qual* 20:387–395.
13. PTI Environmental Services. 1995. National zinc site remedial investigation feasibility study. Vol 4. Ecological risk assessment—Operable unit 2. Cyprus Amax Minerals Company, Salomon, OK, USA. City of Bartlesville, OK, USA.
14. PTI Environmental Services. 1994. Regional ecorisk field investigation. Upper Clark Fork River Basin. ARCO, Bellevue, WA, USA.
15. Wyttenbach A, Bajo S, Furrer V, Langenauer M, Tobler L. 1997. The accumulation of arsenic, bromine, and iodine in needles of Norway spruce (*Picea abies* [L.] Karst.) at sites with low pollution. *Water Air Soil Pollut* 94:417–430.
16. Davies BE. 1992. Inter-relationships between soil properties and the uptake of cadmium, copper, lead, and zinc from contaminated soils by radish (*Raphanus sativus* L.). *Water Air Soil Pollut* 63: 331–342.
17. Otte ML, Rozema J, Beek BJ, Broekman RA. 1990. Uptake of arsenic by estuarine plants and interactions with phosphate, in the field (Rhine Estuary) and under outdoor experimental conditions. *Sci Total Environ* 97/98:839–854.
18. Sadiq M. 1985. Uptake of cadmium, lead and nickel by corn grown in contaminated soils. *Water Air Soil Pollut* 26:185–190.
19. Sadiq M. 1986. Solubility relationships of arsenic in calcareous soils and its uptake by corn. *Plant Soil* 91:241–248.
20. Hooda PS, Alloway BJ. 1993. Effects of time and temperature on the bioavailability of Cd and Pb from sludge-amended soils. *J Soil Sci* 44:97–110.
21. Miles LJ, Parker GR. 1979. Heavy metal interaction for *Andropogon scoparius* and *Rudbeckia hirta* grown on soil from urban and rural sites with heavy metals additions. *J Environ Qual* 8: 443–449.
22. Severson RC, Gough LP, Van den Boom G. 1992. Baseline element concentrations in soils and plants, Wattenmeer National Park, North and East Frisian Islands, Federal Republic of Germany. *Water Air Soil Pollut* 61:169–184.
23. Bechtel Jacobs Company. 1998. Empirical models for the uptake of chemicals from soil by plants. ES/ER/TM-198. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
24. SAS Institute. 1988. *SAS®/STAT User's Guide*, Release 6.03 ed. Cary, NC, USA.
25. Dowdy S, Wearden S. 1983. *Statistics for Research*. John Wiley, New York, NY, USA.
26. Draper N, Smith H. 1981. *Applied Regression Analysis*, 2nd ed. John Wiley, New York, NY, USA.
27. Lindberg SE, Jackson DR, Huckabee JW, Janzen SA, Levin MJ, Lund JR. 1979. Atmospheric emission and plant uptake of mercury from agricultural soils near the Almaden mercury mine. *J Environ Qual* 8:572–578.
28. Banuelos GS. 1996. Managing high levels of boron and selenium with trace element accumulator crops. *J Environ Sci Health A* 31:1179–1196.
29. Carlson CL, Adriano DC, Dixon PM. 1991. Effects of soil-applied selenium on the growth and selenium content of a forage species. *J Environ Qual* 20:363–68.
30. Lagerwerff JV. 1971. Uptake of cadmium, lead, and zinc by radish from soil and air. *Soil Sci* 111:129–133.
31. Sauerbeck DR, Hein A. 1991. The nickel uptake from different soils and its prediction by chemical extractions. *Water Air Soil Pollut* 57–58:861–871.
32. Greenleaf-Jenkins J, Zasoski RJ. 1986. Distribution, availability, and foliar accumulation of heavy metals from dewatered sludge applied to two acid forest soils. In West SD, Zasoski RJ, eds, *Nutritional and Toxic Effects of Sewage Sludge in Forest Ecosystems*. College of Forest Resources, University of Washington, Seattle, WA, USA.
33. Peterson PJ. 1983. Adaptation to toxic metals. In Robb DA, Pierpont WS, eds, *Metals and Micronutrients: Uptake and Utilization by Plants*, Vol 21. Academic, London, UK, pp 51–69.
34. Guillitte O, Melin J, Wallberg L. 1994. Biological pathways of radionuclides originating from the Chernobyl fallout in a boreal forest ecosystem. *Sci Total Environ* 157:207–215.
35. Hutchinson TC, Czuba M, Cunningham L. 1974. Lead, cadmium, zinc, copper, and nickel distributions in vegetables and soils of an intensely cultivated area and levels of copper, lead and zinc in the growers. *Trace Subst Environ Health* 8:81–93.
36. Miller JE, Hassett JJ, Koeppe DE. 1977. Interactions of lead and cadmium on metal uptake and growth of corn plants. *J Environ Qual* 6:18–20.
37. Gissel-Nielsen G, Bisbjerg B. 1970. The uptake of applied selenium by agricultural plants 2. The utilization of various selenium compounds. *Plant Soil* 32:382–96.
38. Bache CA, Gutenmann WH, St John LE Jr, Sweet RD, Hatfield HH, Lisk DJ. 1973. Mercury and methylmercury content of agricultural crops grown on soils treated with various mercury compounds. *J Agric Food Chem* 21:607–613.
39. Haghiri F. 1973. Cadmium uptake by plants. *J Environ Qual* 2: 93–95.
40. Sadana US, Singh B. 1987. Yield and uptake of cadmium, lead and zinc by wheat grown in a soil polluted with heavy metals. *J Plant Sci Res* 3:11–17.
41. Baker DA. 1983. Uptake of cations and their transport within the plant. In Robb DA, Pierpont WS, eds, *Metals and Micronutrients: Uptake and Utilization by Plants*, Vol 21. Academic, London, UK, pp 3–19.
42. Huang JW, Chen J, Berti WR, Cunningham SD. 1997. Phytoremediation of lead-contaminated soils: Role of synthetic chelates in lead phytoextraction. *Environ Sci Technol* 31:800–805.
43. Hutchinson JJ, Young SD, McGrath SP, West HM, Black CR, Baker AJM. 2000. Determining uptake of 'non-labile' soil cadmium by *Thlaspi caerulescens* using isotopic dilution techniques. *New Phytol* 146:453–460.
44. Dahmani-Muller H, van Oort F, Gélie B, Balabane M. 2000. Strategies of heavy metal uptake by three plant species growing near a metal smelter. *Environ Pollut* 109:231–238.
45. Bert V, MacNair MR, de Laguerie P, Saumitou-Laprade P, Petit D. 2000. Zinc tolerance and accumulation in metallicolous and nonmetallicolous populations of *Arabidopsis halleri* (Brassicaceae). *New Phytol* 146:225–233.
46. Reeves RD, Baker AJM, Borhidi A, Berazaín R. 1999. Nickel hyperaccumulation in the serpentine flora of Cuba. *Ann Bot* 83: 29–38.

APPENDIX

Plant taxa represented in the database

Family	Genus	Number of species	Dataset[b]	Chemical
Dicots				
Acanthaceae	*Justicia*	1	o	Hg
Aceraceae	*Acer*	2	o	Hg, Zn
Anacardiaceae	*Mangifera*	1	o	Ni
Annonaceae	*Asimina*	1	o	Ni
Apiaceae	*Daucus*	1	o	Cd, Cu, Hg, Ni, Pb, Zn
	Pastinaca	1	o	Cd, Cu, Pb
Asclepiadaceae	*Calotropis*	1	o	Hg
Asteraceae	*Ambrosia*	2	o, v	Cd, Cu, Ni, Pb, Se, Zn
	Artemisia	1	o	Cd, Hg, Pb
	Bidens	1	v	Cd, Cu, Hg, Ni, Pb, Zn
	Centaurea	1	o	Ni
	Cirsium	1	o	Se
	Conyza	1	o	Se
	Dittrichia	1	o	Ni
	Lactuca	2	o	Cd, Cu, Hg, Pb, Se, Zn
	Reichardia	1	o	Ni
	Rudbeckia	1	o	Cd, Cu, Pb, Zn
	Scorzonera	1	o	Ni
	Senecio	1	o	Ni
	Sonchus	1	o	Se
Boraginaceae	*Lithospermum*	1	o	Cu, Pb, Zn
Brassicaceae	*Alyssum*	2	o	Ni
	Brassica	6	o	As, Cd, Cu, Hg, Pb, Se, Zn
	Lepidium	1	o	Cd, Cu, Hg, Pb, Zn
	Raphanus	1	o	Cd, Cu, Hg, Ni, Pb, Zn
	Sisymbrium	1	o	Se
	Thlaspi	3	o	Pb, Ni, Zn
Campanulaceae	*Jasione*	1	o	Ni
Caryophyllaceae	*Cerastium*	1	o	Ni
	Dianthus	1	o	Ni
	Herniaria	1	o	Ni
	Minuartia	1	o	Ni
	Silene	1	o	Ni
	Spergularia	1	o	Se
Chenopodiaceae	*Atriplex*	1	o	Se
	Beta	1	o	Hg, Se
	Spinacia	1	o	Cd, Cu, Hg, Se, Zn
Cistaceae	*Cistus*	1	o	Ni
Convolvulaceae	*Evolvulus*	1	o	Hg
	Ipomoea	1	o	Hg
Crassulaceae	*Sedum*	1	o	Ni
Ericaceae	*Vaccinium*	1	o	Ni, Zn
Euphorbiaceae	*Croton*	2	o	Hg
	Euphorbia	3	o	Ni
	Jatropha	1	o	Hg
Fabaceae	*Astragalus*	1	o	Cu, Hg
	Genista	1	o	Cu, Hg
	Glycine	1	o	As, Cd, Cu, Hg, Pb, Zn
	Lotus	1	o	As, Se
	Medicago	1	o	As, Hg, Se
	Melilotus	1	o	Se
	Phaseolus	1	o	Ni, Se
	Pisum	1	o	Hg
	Trifolium	3	o	Cd, Cu, Pb, Se, Zn
	Quercus	1	o	Cu, Pb, Zn
Frankeniaceae	*Frankenia*	1	o	Se
Gentianaceae	*Centaurium*	1	o	Ni
Geraniaceae	*Erodium*	1	o	Se
Hydrophyllaceae	*Phacelia*	1	o	Cu, Zn
Lamiaceae	*Mentha*	1	o	Cd, Hg, Pb
	Ocimum	1	o	Cd, Hg, Pb
	Stachys	1	o	Ni
	Thymus	1	o	Ni
Linaceae	*Linum*	2	o	Cd, Ni
Malvaceae	*Hibiscus*	1	o	Se
Moraceae	*Morus*	1	o	Hg
Onagraceae	*Epilobium*	1	o	Se
Papaveraceae	*Argemone*	1	o	Hg
Plantaginaceae	*Plantago*	3	o	Hg, Ni
Polygonaceae	*Rumex*	1	o	Ni

APPENDIX
Continued

Family	Genus	Number of species	Dataset[b]	Chemical
Rosaceae	*Potentilla*	1	o	Ni
	Sanguisorba	1	o	Ni
Rubiaceae	*Galium*	1	o	Ni
Rutaceae	*Citrus*	1	o	Ni
Salicaceae	*Populus*	1	o	Cu, Pb, Zn
	Salix	1	o	As, Cu, Hg, Pb, Se, Zn
Solanaceae	*Lycopersicon*	1	o	Hg, Ni
	Nicotiana	1	o	Hg
	Solanum	2	o	Hg, Ni
Urticaceae	*Urtica*	1	o	Cu, Zn
Violaceae	*Hybanthus*	1	o	Ni
Zygophyllaceae	*Tribulus*	1	o	Hg
Monocots				
Cyperaceae	*Cyperus*	1	o	Hg
	Scirpus	1	o	Ni
Liliaceae	*Allium*	1	o	Cd, Cu, Pb
	Smilacina	1	o	Cu, Pb, Zn
Musaceae	*Musa*	1	o	Ni
Poaceae	*Agropyron*	1	v	As, Cd, Cu, Pb, Zn
	Agrostis	2	o	As, Cd, Cu, Pb, Zn
	Ammophilia	1	o	As, Cu, Hg, Pb, Se, Zn
	Andropogon	2	o, v	Cd, Cu, Ni, Pb, Zn
	Avena	1	o	Cd, Cu, Hg, Ni, Pb, Zn
	Bromus	5 or 6	o	As, Cd, Pb, Se, Zn
	Chloris	1	o	Hg
	Cynodon	1	o, v	Cd, Cu, Hg, Ni, Pb, Zn
	Dactylis	1	o	As, Ni, Se
	Deschampsia	1 or 2	o, v	As, Cd, Cu, Ni, Pb, Zn
	Echinochloa	1	o	As, Se
	Festuca	3	o	Cd, Ni, Pb, Se, Zn
	Hordeum	2	o	Cd, Hg, Pb, Se, Zn
	Lolium	2	o	Cd, Cu, Pb, Se
	Panicum	1	v	Cd, Cu, Ni, Pb, Zn
	Pennisetum	1	o	As, Hg
	Phleum	1	o	As, Hg, Se
	Phragmites	1	o	Cu, Pb, Se, Zn
	Poa	1 or 2	o, v	Cd, Cu, Pb, Zn
	Polypogon	1	o	Se
	Sorghastrum	1	v	Cd, Cu, Ni, Pb
	Sorghum	1	o	As, Se
	Stipa	1	o	Ni
	Triticum	2	o	Cd, Hg, Pb
	Vulpia	1	o	Se
	Zea	1	o	As, Cd, Cu, Pb, Se, Zn
	Various	Unknown	o	As, Cd, Cu, Zn
Ferns				
Aspleniaceae	*Asplenium*	1	o	Ni
	Ceterach	1	o	Ni
Parkeriaceae	*Ceratopteris*	1	o	As, Hg
Pteridaceae	*Cheilanthes*	1	o	Ni
Gymnosperms				
Pinaceae	*Picea*	1	o	As, Cd, Cu, Pb, Zn
	Pinus	2	o	As, Hg
Horsetails				
Equisetaceae	*Equisetum*	1	o	Hg

[a] Additional composited samples of tree foliage, ferns, forbs, and grasses were included in the dataset.

[b] o = original dataset; v = validation dataset.