Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors

Xiaoxia Lu, Shu Tao *, Hanyin Hu, R.W. Dawson

Department of Urban and Environmental Sciences, Peking University, Beijing 100871, People's Republic of China

Abstract

A bioconcentration factor (BCF) estimation model for a wide range of nonionic organic compounds was developed on the basis of molecular connectivity indices and polarity correction factors. The nonlinear topological modeling using polarity correction factors resulted in the best BCF estimation quality for all of the 239 compounds studied, with a mean absolute estimation error of 0.478 log units. Residual analysis indicated that the estimation errors came from many sources including BCF measurement, test species, and selection of descriptors. Statistical robustness of the developed model was validated by modified jackknifed tests where random deletion of a set of compounds and specific deletion of a class of compounds were both performed. Comparison between the MCI-based (molecular connectivity indices) model and a K_{ow}-based (octanol/water partition coefficient) model revealed that the BCF estimation based on topological parameters was as good as that achieved by K_{ow}. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Bioconcentration factor; Molecular connectivity indices; Polarity correction factor; Octanol/water partition coefficient

1. Introduction

Bioconcentration factor (BCF) is an important eco-toxicological parameter describing the accumulation of chemicals in organisms, primarily aquatic, that live in contaminated environments. The accumulation of organic pollutants in fish is especially concerned because fish serve as food for many species including humans. However, in both human and environmental risk assessments, the indispensable BCF data are usually not easily available. Since the experimental determination of BCF values is expensive and time consuming, many investigators tend to use estimation methods to supply the missing data.

Relating bioconcentration to measures of hydrophobicity such as octanol/water partition coefficient (K_{ow}) is the usual methodology for BCF estimation. The bioconcentration is assumed to be a thermodynamically driven partitioning process between water and the lipid of fish, and therefore is modeled using K_{ow} as a surrogate for biological lipids. Other traditional approaches include using water solubility (S) and soil adsorption coefficient (K_{oc}) as the descriptors. Though good results were achieved in some studies, the application of these methods is limited by the availability of parameter data. Frequently, data on the parameters, when not available in the literature, have to be either experimentally determined or estimated by other methods.

In the molecular connectivity method, the estimation runs directly from structure to activity. Molecular

---

* Corresponding author. Tel.: +86-10-6275-1938; fax: +86-10-6275-1938.

E-mail address: taos@urban.pku.edu.cn (S. Tao).

connectivity indices (MCI) are representation of molecular structure and have been demonstrated correlating to many physicochemical properties including K_{ow}, S and K_{oc} (Sabljic, 1987a,b, 1993; Nirmalakhandan and Speece, 1988). Sabljic pioneered the study of relationship between BCF and MCI in 1982. He found that the second-order valence chi index ($^2\chi^v$) in a parabolic form gave good correlation with the measured log BCF data of some halocarbons (Sabljic and Protic, 1982). The established model was further tested by a set of chemicals including polycyclic aromatic hydrocarbons, alkyl and alkenyl benzenes, substituted phenols and other structurally similar compounds (Sabljic, 1987a,b). In addition, Koch discovered from his study that the first-order valence chi index ($^1\chi^v$) of some organic compounds correlated linearly with the measured log BCF data (Koch, 1983). But the previous studies were mostly based on nonpolar organic compounds. A molecular connectivity model suitable for estimating the BCFs of polar organic compounds is not available at this time. One of the reasons might be that the bioconcentration process of polar compounds is much more complicated than that of nonpolar ones. Generally speaking, the bioconcentration process is controlled by polar and nonpolar interactions among chemical, water and fish. Polarity of a chemical complicates this process. Obviously, more efforts are needed before a satisfactory topological model for estimating the BCFs of both polar and nonpolar compounds can be established.

The objective of this study was to develop a screening-level BCF estimation model for a wide range of nonionic organic compounds by use of molecular connectivity indices and polarity correction factors. A comprehensive BCF database containing the bioconcentration features of both polar and nonpolar compounds in fish had been assembled for the purpose of modeling. Robustness of the developed model would be tested by internal validations. Estimation quality of the MCI-based model would be compared with that of a K_{ow}-based model.

2. Methodology

2.1. Collection of measured BCF data

Experimentally determined BCF values for 239 nonionic organic compounds were collected from an extensive literature review (Chiou et al., 1977; Veith et al., 1980; Donald, 1982; Oliver and Niimi, 1983, 1984, 1985; Davis and Dobbs, 1984; Freitag et al., 1985; Opperhuizen et al., 1985; McCarty, 1986; Connell and Hawker, 1988; Isnard and Lambert, 1988; Gobas et al., 1989, 1990; Smith et al., 1990; Voogt et al., 1990; Mackay et al., 1992; Nendza, 1993; Fox et al., 1994;

Watze de Wolf et al., 1994; Devillers et al., 1996; Tsuda et al., 1997). The studied compounds included halogenated benzenes, chlorinated biphenyls, chlorinated naphthalenes, chlorinated aliphatic hydrocarbons, monocyclic aromatic hydrocarbons, polycyclic aromatic hydrocarbons, chlorinated dibenzo-dioxins, chlorinated dibenzofurans, phenols, anilines, phthalates, carbonates, phosphates, esters, ethers and nitroaromatics. These chemicals are ubiquitous contaminants in the environment due to their wide use in industry and agriculture. BCF data were taken from studies where it appeared that steady states of bioconcentration had been reached. Both equilibrium and kinetic methods were included. Most of the collected BCF data were obtained under flow conditions, while a few were determined from semistatic conditions. Test species related to several fish such as rainbow trout, guppies, fathead minnows, bluegill sunfish, golden ide, etc. There were differences in the lipid content and weight of the fish and exposure concentration and duration. The BCFs used in this study were measured basing on the total weight of fish. Log BCF values ranged from −0.22 to 5.97. Further, multiple log BCF data were available for most of the compounds studied.

A BCF database was constructed using Excel and included 239 records. Each record contained fields for chemical name, molecular structure, log K_{ow}, molecular connectivity indices, measured log BCF data plus partial reference citations and key experimental details if available. Experimental details included fish species, exposure concentration of the test compound, lipid content of the test organism, test method, test condition and duration. The medium value of the available log BCF data for each compound was used for model development.

2.2. Calculation of molecular connectivity indices

Molecular connectivity indices are calculated from the hydrogen-suppressed molecule. A series of indices can be derived and each expresses specific structural information (Kier and Hall, 1986). The general calculation equations are as follows:

$$^m\chi = \sum_{j=1}^{n}\left(\prod_{i=1}^{m+1}\delta_i\right)^{-1/2}, \qquad ^m\chi^v = \sum_{j=1}^{n}\left(\prod_{i=1}^{m+1}\delta_i^v\right)^{-1/2},$$

where, δ is the atomic delta value, δ^v is the valence atomic delta value, i the nonhydrogen atom, m and n refer to the connectivity level and the subgraph number, respectively. All molecular connectivity indices in this study were calculated by computer using the Wintox software developed by Jorgensen et al. (1998).

2.3. Model development and robustness test

A previously established BCF estimation model by means of molecular connectivity indices was adopted to predict the log BCF values of the 239 compounds in this study. In view of the residuals of various chemicals, modifications involving two major steps were made to the estimation model: firstly, polarity correction factors were introduced to describe the influence of polar functional groups on the bioconcentration process of polar compounds; and secondly, nonlinear estimation was used to fit the bioconcentration data of compounds with large molecules.

The statistical robustness of the regression model was validated using modified Jackknifed tests (Dietrich et al., 1980). Deletion of a random set of compounds and deletion of a specific class of compounds were both conducted. By comparing the differences in the coefficients of determination (R^2) between the pre-deletion and post-deletion and among various deletions, "outliers" were identified making it possible to examine chance correlation within the model. All of the statistical analyses were performed using Statistica Software.

3. Results

3.1. Difference between polar and nonpolar compounds

According to chemical composition, the nonionic organic compounds can be generally grouped into two categories: polar compounds and nonpolar compounds. In this study, those containing only C, H, Cl and Br were defined as nonpolar compounds and they consisted of hydrocarbons and halogenated hydrocarbons. The total was 121 and the measured log BCF values ranged from 0.30 to 5.97. Others with at least a polar fragment in the molecular were referred to polar compounds, including anilines, carbonates, ethers, phenols, phthalates, phosphates, nitroaromatics, chlorinated dibenzo-dioxins and dibenzofurans. The total was 118 and the measured log BCF values were between −0.22 and 4.51. Classification as well as the measured log BCF of each substance is listed in Table 1.

A BCF estimation model by means of molecular connectivity indices was established in our previous study. It was based on the measured BCF data of 80 nonpolar organic compounds (Lu et al., 1999). Through stepwise regression, five molecular connectivity indices were selected as descriptors in the model as follows:

$$\log \text{BCF} = 0.757 {}^{0}\chi^{v} - 2.650 {}^{1}\chi^{v} + 3.372 {}^{2}\chi - 1.186 {}^{2}\chi^{v}$$
$$- 1.807 {}^{3}\chi_{c} + 0.770, \quad n = 80, \ R^2 = 0.907, \ s = 0.364.$$
$$(1)$$

Eq. (1) was adopted as a preliminary model to estimate BCF for the 239 compounds in this study (including the previously used 80 nonpolar chemicals). Relationship between the measured and calculated log BCF values of the 239 compounds is displayed in Fig. 1.

The results showed that most polar compounds were overestimated by Eq. (1). As can be seen from Fig. 1, significant difference exists between polar and nonpolar compounds. Most of the polar compounds are found scattered below the nonpolar compounds, and the nonpolar compounds are distributed along the diagonal line. This phenomena reflects the bioconcentration features of both polar and nonpolar compounds. It was reported that under conditions where molecules are of the same size, the bioconcentration factors of polar compounds are usually less than those of nonpolar ones. According to the results predicted by Eq. (1), it is evident that the structural properties associated with various compounds are not fully described by the molecular connectivity indices. Therefore, additional descriptors are needed especially for assessing the BCF of polar compounds.

Another feature in Fig. (1) is that both polar and nonpolar compounds appear parabolic distribution within the study range. For those falling on right side of the parabolic line, common characteristics can be identified among them, that is, at least two aromatic rings and six halogenated groups are contained in the molecules. These compounds tend to have high log K_{ow} values as well. Obviously, the linear model does not fit in these cases.

3.2. Linear modeling with polarity correction factors

Analysis of the residuals calculated from Eq. (1) revealed that compounds sharing certain polar functional groups tend to have residuals relatively consistent in sign and magnitude. For instance, the estimation error of ethers was about −0.90 log units, and of anilines, it was about minus 1.20 log units. On this basis we identified some polar groups and introduced polarity correction factors (PCFs) to describe the contribution of those groups to the process of bioconcentration. As both carbonyl (–CO–) and sulfinyl (–SO–) appeared only once in the studied molecules, no correction factors were used for them. Similarly, no specific correction factors were assigned to azo (–N=) and nitrile (–CN). But the influence of azo (–N=) and nitrile (–CN) was described by the amino correction factor. In all, eight polar correction factors were introduced for the following eight kinds of polar groups: hydroxy (–OH), amino (–NH₂, –NH–, –N–), nitro (–NO₂), aminocarbonyloxy (–NCOO–), aminocarbonylthio (–NCOS–), oxycarbonyl (–COO–), oxy (–O–), and phosphor (–PO₃–, –PO₂–). Different polarity correction factors were used to describe the effects of different groups, and for the same group,

Table 1
Measured and calculated log BCFs of the studied compounds[a]

Class	Chemical	Measured log K_{ow}	Measured log BCF	Calculated log BCF by Eq. (4)	Calculated log BCF by Eq. (5)
CAH	1,1,1-Trichloroethane	2.47	0.95	1.520	1.604
CAH	1,1,2,2-Tetrachloroethane	2.39	0.90	1.780	1.553
CAH	1,1,2,3,4,4-Hexachloro-1,3-butadiene	4.78	3.83	3.703	3.075
CAH	Trichloroethylene	2.42	1.59	1.778	1.572
CAH	1,2-Dichloroethane	1.45	0.30	0.574	0.955
CAH	Tetrachloromethane	2.73	1.48	1.542	1.770
CAH	Trichloromethane	1.90	0.78	1.226	1.241
CAH	Hexachloroethane	3.93	2.92	2.597	2.534
CAH	Pentachloroethane	2.89	1.83	2.184	1.872
CAH	Tetrachloroethylene	2.53	1.74	2.371	1.642
MAH	Benzene	2.13	0.64	1.195	1.388
MAH	m-Methylstyrene	3.37	1.55	2.351	2.177
MAH	Octachlorostyrene	6.29	4.52	4.826	4.036
MAH	p-Methylstyrene	3.37	1.50	2.343	2.177
MAH	Styrene	2.95	1.13	1.842	1.910
MAH	Toluene	2.73	1.12	1.693	1.770
MAH	Ethylbenzene	3.15	1.19	1.703	2.037
MAH	o-Xylene	3.12	1.24	2.185	2.018
MAH	m-Xylene	3.20	1.27	2.197	2.069
MAH	p-Xylene	3.15	1.27	2.190	2.037
MAH	Isopropylbenzene	3.72	1.55	2.261	2.400
MAH	2-Phenyldodecane	8.19	2.65	3.698	3.126
PAH	2-Chlorophenanthrene	4.07	3.63	3.812	2.987
PAH	2-Methylphenanthrene	4.86	3.48	3.783	3.490
PAH	9-Methylanthracene	5.07	3.66	3.780	3.259
PAH	Acenaphthalene	3.92	2.59	2.533	2.527
PAH	Anthracene	4.54	2.83	3.345	2.922
PAH	Benzo[a]anthracene	5.61	4.00	4.365	3.968
PAH	Biphenyl	3.88	2.64	2.865	2.502
PAH	Fluerene	4.38	3.23	2.969	2.820
PAH	Naphthalene	4.70	1.64	2.210	3.024
PAH	Phenanthrene	4.46	3.42	3.332	3.236
PAH	Pyrene	4.88	3.43	3.796	3.503
PAH	2-Methylnaphthalene	4.11	3.20	2.730	2.648
PAH	Acenaphthylene	3.97	2.58	2.686	2.559
PAH	Benzo[a]pyrene	5.97	3.42	4.756	4.197
CB	1,2,3,4-Tetrachlorobenzene	4.65	3.77	3.217	2.992
CB	1,2,3,5-Tetrachlorobenzene	4.65	3.36	3.235	2.992
CB	1,2,3-Trichlorobenzene	4.27	3.11	2.755	2.750
CB	1,2,4,5-Tetrachlorobenzene	4.67	3.76	3.238	3.005
CB	1,2,4-Trichlorobenzene	4.04	3.26	2.768	2.604
CB	1,2-Dichlorobenzene	3.71	2.48	2.255	2.393
CB	1,3,5-Trichlorobenzene	4.08	3.38	2.786	2.629
CB	1,3-Dichlorobenzene	3.44	2.65	2.266	2.222
CB	1,4-Dichlorobenzene	3.37	2.52	2.259	2.177
CB	Hexchlorobenzene	5.73	4.26	3.985	3.679
CB	2,4,5-Trichlorotoluene	4.80	3.87	3.211	3.087
CB	Chlorobenzene	2.84	1.85	1.729	1.840
CB	Pentachlorobenzene	5.18	3.86	3.638	3.329
BB	1,2,4,5-Tetrabromobenzene	5.13	3.79	3.673	3.297
BB	1,2,4-Tribromobenzene	4.54	3.66	3.259	2.922
BB	1,3,5-Tribromobenzene	4.51	3.85	3.276	2.903
BB	1,3-Dibromobenzene	3.78	2.80	2.677	2.438

Table 1 (*continued*)

Class	Chemical	Measured log K_{ow}	Measured log BCF	Calculated log BCF by Eq. (4)	Calculated log BCF by Eq. (5)
BB	1,4-Dibromobenzene	3.89	2.83	2.670	2.508
BB	Bromobenzene	2.99	1.70	1.957	1.935
BB	Hexabromobenzene	6.07	3.04	3.844	3.896
BB	1,2-Dibromobenzene	3.64	3.10	2.667	2.349
CN	1,2,3,4-Tetrachloronaphthalene	5.50	4.10	4.087	4.243
CN	1,3,5,7-Tetrachloronaphthalene	6.38	4.53	4.128	4.803
CN	1,3,5,8-Tetrachloronaphthalene	5.96	4.40	4.128	4.536
CN	1,3,7-Trichloronaphthalene	5.59	4.43	3.732	4.300
CN	1,4-Dichloronaphthalene	4.88	3.56	3.261	3.848
CN	1,8-Dichloronaphthalene	4.41	3.79	3.264	3.549
CN	2,3-Dichloronaphthalene	4.71	4.04	3.275	3.740
CN	2,7-Dichloronaphthalene	4.81	4.04	3.278	3.804
CN	2-Monochloronaphthalene	3.90	3.63	2.767	2.514
CN	Octachloronaphthalene	6.42	3.44	4.972	4.828
PCB	2,2′,4,5-Tetrachlorobiphenyl	5.85	5.00	4.820	4.466
PCB	2,2′,5,5′-Tetrachlorobiphenyl	6.10	4.87	4.556	4.625
PCB	2,2′,5-Trichlorobiphenyl	5.60	4.27	4.223	4.307
PCB	2,2′,4,4′-Tetrachlorobiphenyl	6.11	4.02	4.556	4.631
PCB	2,2′-Dichlorobiphenyl	5.00	3.26	3.824	3.925
PCB	2,3′,4′,5-Tetrachlorobiphenyl	5.60	4.77	4.554	4.307
PCB	2,3-Dichlorobiphenyl	5.20	4.25	3.821	4.052
PCB	2,4,4′-Trichlorobiphenyl	5.62	4.63	4.222	4.319
PCB	2,4′,5-Trichlorobiphenyl	5.67	3.75	4.222	4.351
PCB	2,4,5-Trichlorobiphenyl	5.90	4.02	4.228	4.498
PCB	2,4′-Dichlorobiphenyl	5.10	3.55	3.825	3.988
PCB	2,5-Dichlorobiphenyl	5.16	4.20	3.832	4.027
PCB	3,3′,4,4′-Tetrachlorobiphenyl	6.36	3.90	4.552	4.790
PCB	3,5-Dichlorobiphenyl	5.37	3.78	3.847	4.160
PCB	4-Chlorobiphenyl	4.63	2.69	3.373	2.979
PCB	4,4′-Dichlorobiphenyl	5.58	3.28	3.825	4.294
PCB	2,2′,3,3′-Tetrachlorobiphenyl	6.18	4.23	4.535	4.676
PCB	2,2′,4,4′,6-Pentachlorobiphenyl	6.23	3.37	4.824	4.708
PCB	2,2′,4,5,5′-Pentachlorobiphenyl	6.38	5.40	4.820	4.803
PCB	2,2′,4,4′,5,5′-Hexachlorobiphenyl	6.92	4.83	5.014	5.147
PCB	2,2′,4,4′,6,6′-Hexachlorobiphenyl	6.54	4.93	5.021	4.905
PCB	2,2′,3,3′,4,4′,5,5′-Octachlorobiphenyl	7.80	5.08	5.141	5.248
PCB	2,2′,3,3′,4,4′,5,5′,6,6′-decachlorobiphenyl	8.18	4.02	4.978	5.006
PCB	2,2′,3,5′-Tetrachlorobiphenyl	5.75	4.84	4.546	4.402
PCB	2,2′,4,5′-Tetrachlorobiphenyl	5.85	4.84	4.556	4.466
PCB	2,2′,6,6′-Tetrachlorobiphenyl	5.21	3.85	4.546	4.058
PCB	2,2′,3,4,5′-Pentachlorobiphenyl	6.29	5.38	4.802	4.746
PCB	2,2′,3′,4,5-Pentachlorobiphenyl	6.29	5.43	4.811	4.746
PCB	3,3′,4,4′,5-Pentachlorobiphenyl	6.89	5.81	4.815	5.128
PCB	2,2′,3,3′,4,4′-Hexachlorobiphenyl	6.74	5.77	4.976	5.032
PCB	2,2′,3,3′,6,6′-Hexachlorobiphenyl	6.22	5.43	4.987	4.701
PCB	2,2′,3,4,4′,5-Hexachlorobiphenyl	6.83	5.88	4.994	5.089
PCB	2,2′,3,4,5,5′-Hexachlorobiphenyl	6.82	5.81	4.994	5.083
PCB	2,2′,3,5,5′,6-Hexachlorobiphenyl	6.64	5.54	4.998	4.969
PCB	3,3′,4,4′,5,5′-Hexachlorobiphenyl	7.42	5.97	5.007	5.491
PCB	2,2′,3,4,4′,5′,6-Heptachlorobiphenyl	7.20	5.84	5.120	5.631
PCB	2,2′,3,4,5,5′,6-Heptachlorobiphenyl	7.11	5.93	5.099	5.689
PCB	2,2′,3,3′,4,4′,5,6-Octachlorobiphenyl	7.56	5.92	5.128	5.402
PCB	2,2′,3,3′,4,5,5′,6-Octachlorobiphenyl	7.62	5.88	5.145	5.363
PCB	2,2′,3,3′,5,5′,6,6′-Octachlorobiphenyl	7.24	5.82	5.152	5.606
PCB	2,2′,3,3′,4,4′,5,5′,6-Nonachlorobiphenyl	8.09	5.71	5.098	5.063

Table 1 (*continued*)

Class	Chemical	Measured log K_{ow}	Measured log BCF	Calculated log BCF by Eq. (4)	Calculated log BCF by Eq. (5)
PBB	2,2′,5,5′-Tetrabromobiphenyl	6.50	4.80	4.468	4.879
PBB	2,4,6-Tribromobiphenyl	6.03	3.93	4.390	4.580
PBB	4,4′-Dibromobiphenyl	5.72	4.19	4.069	4.383
PBB	2,2′,4,4′,6,6′-Hexabromobiphenyl	7.20	3.96	3.935	5.631
CDO	2,7-Dichlorodibenzo-*p*-dioxin	5.75	2.13	2.653	3.692
CDO	2,8-Dichlorodibenzo-*p*-dioxin	5.60	2.82	2.653	3.596
CDO	1,2,4-Trichlorodibenzo-*p*-dioxin	6.35	2.36	2.999	4.074
CDO	1,2,3,4-Tetrachlorodibenzo-*p*-dioxin	6.60	2.55	3.274	4.233
CDO	1,2,3,7-Tetrachlorodibenzo-*p*-dioxin	6.91	3.24	3.292	4.430
CDO	1,3,6,8-Tetrachlorodibenzo-*p*-dioxin	7.20	3.36	3.313	3.710
CDO	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin	7.02	4.06	3.310	3.824
CDO	1,2,3,4,7-Pentachlorodibenzo-*p*-dioxin	7.44	3.21	3.499	3.557
CDO	1,2,3,4,7,8-Hexachlorodibenzo-*p*-dioxin	7.79	3.54	3.651	3.333
CDO	1,2,3,4,6,7,8-Heptachlorodibenzo-*p*-dioxin	8.20	3.16	3.713	3.072
CDO	Octachlorodibenzo-*p*-dioxin	8.60	2.76	3.699	2.817
CDO	1,2,3,7,8-Pentachlorodibenzo-*p*-dioxin	6.64	4.50	3.516	4.258
CDO	Dibenzo(1,4)dioxan	4.19	3.85	1.738	2.699
CDF	Dibenzofuran	4.21	3.34	2.216	2.712
CDF	2,3,7,8-Tetrachlorodibenzofuran	6.53	3.53	3.892	4.188
CDF	2,3,4,7,8-Pentachlorodibenzofuran	6.92	4.03	4.127	4.437
CDF	Octachlorodibenzofuran	8.20	2.94	4.403	3.072
CDF	1,2,3,4,6,7,8-Heptachlorodibenzofuran	7.92	3.62	4.384	3.250
CDF	Benzo[b]furan	2.86	2.56	1.119	1.852
Phenol	2,4-Dichlorophenol	3.30	1.50	1.954	2.063
Phenol	Hydroquinone	0.55	1.60	0.609	0.500
Phenol	Pentachlorophenol	3.81	2.74	3.257	2.387
Phenol	2,4,6-Trichlorophenol	3.06	2.43	2.444	1.910
Phenol	2-Chlorophenol	2.16	2.33	1.429	1.407
Phenol	3-Chlorophenol	2.50	1.25	1.441	1.623
Phenol	4-Cyanophenol	1.60	0.91	0.670	1.050
Phenol	2,6-Dibromo-4-cyanophenol	2.61	1.67	2.025	1.623
Phenol	4,5-Dichloroguaiacol	2.86	2.03	1.947	1.782
Phenol	2-Methyl phenol	1.95	1.03	1.394	1.273
Phenol	Tetrachloroguaiacol	4.45	2.71	2.721	2.795
Phenol	2,3,5,6-Tetrachlorophenol	4.39	2.15	2.875	2.756
Phenol	2,4,6-Tribromophenol	4.23	2.71	2.891	2.654
Phenol	3,4,5-Trichloroguaiacol	4.11	2.41	2.361	2.578
Phenol	4,5,6-Trichloroguaiacol	3.74	1.97	2.356	2.343
Phenol	Phenol	1.46	1.24	0.904	0.961
Phenol	4-*t*-Butyl phenol	3.31	1.86	2.176	2.139
Phenol	4,6-Dichloroguaiacol	1.88	1.74	1.945	1.159
Phenol	2,4-Dimethylphenol	2.42	2.18	1.886	1.572
Phenol	4-Bromophenol	2.59	1.56	1.653	1.681
Phenol	*p*-sec-Butyl phenol	3.08	1.57	2.031	1.992
Phenol	*p*-Nonyl phenol	5.76	2.45	2.784	2.450
Phenol	*p*-Dodecyl phenol	7.91	3.78	2.931	3.304
Aniline	3,3′-Dichlorobenzidine	3.51	2.79	2.669	2.266
Aniline	2-Chloroaniline	1.93	0.57	1.223	1.261
Aniline	3-Chloroaniline	1.91	0.34	1.236	1.248
Aniline	Diphenylamine	3.50	1.48	2.338	2.260
Aniline	Pentachloroaniline	5.08	3.17	3.035	3.265
Aniline	2,3,4,5-Tetrachloroaniline	4.57	2.69	2.659	2.941
Aniline	2,3,5,6-Tetrachloroaniline	4.46	3.03	2.659	2.871
Aniline	2,3,4-Trichloroaniline	3.68	2.31	2.214	2.374

Table 1 (*continued*)

Class	Chemical	Measured log K_{ow}	Measured log BCF	Calculated log BCF by Eq. (4)	Calculated log BCF by Eq. (5)
Aniline	2,4,5-Trichloroaniline	3.69	2.61	2.235	2.381
Aniline	2,4,6-Trichloroaniline	3.69	2.73	2.233	2.381
Aniline	3,4,5-Trichloroaniline	3.32	2.70	2.233	2.145
Aniline	N-phenyl-2-naphthylamine	4.38	2.17	3.363	2.820
Aniline	4-Chloroaniline	1.88	0.23	1.228	1.229
Aniline	2,4-Dichloroaniline	2.91	1.98	1.747	1.884
Aniline	3,4-Dichloroaniline	2.78	1.48	1.747	1.802
Aniline	Aniline	0.90	0.41	0.698	0.500
Ester	Diethyl phthalate	1.40	2.07	1.963	0.923
Ester	Dimethyl phthalate	1.61	1.76	1.886	1.057
Ester	Cypermethrin	6.05	2.91	2.815	2.858
Ester	Deltamethrin	6.20	2.66	2.237	2.953
Ester	Permethrin	6.50	3.39	3.611	3.144
Ester	Fenvalerate	6.20	2.79	3.985	3.978
Ester	Benzyl butyl phthalate	4.05	2.89	3.038	2.610
Ester	Bis(2-ethylhexyl) phthalate	4.65	2.34	2.125	0.000
Ester	Ethyl acetate (acetic acid ethyl ester)	0.73	1.48	0.666	0.500
Ether	Bis(2-chloroethyl) ether	1.12	1.04	0.584	0.745
Ether	Methoxychlor	4.30	3.10	3.419	2.769
Ether	2,4,6-Trichloroanisole	4.48	2.94	2.728	2.884
Ether	2-t-Butoxy ethanol	0.39	−0.22	0.619	0.500
Ether	t-Butyl isopropyl ether	2.14	0.76	1.524	1.394
Ether	t-Butyl methyl ether	1.24	0.18	1.123	0.821
Ether	2,4,6-Tribromoanisole	4.48	2.94	3.055	2.884
Ether	2,4,5-Trichlorodiphenyl ether	5.44	4.18	3.679	3.495
Ether	3,3′,4,4′-Tetrachlorodiphenyl ether	5.78	4.51	3.976	3.711
Nitro	2-Nitrophenol	2.15	1.60	0.718	1.401
Nitro	2-Chloronitrobenene	2.52	2.10	1.550	1.636
Nitro	3-Chloronitrobenzene	2.50	1.89	1.559	1.623
Nitro	4-Chloronitrobenzene	2.39	2.00	1.552	1.553
Nitro	2,3-Dichloronitrobenzene	3.05	2.16	2.040	1.973
Nitro	2,4-Dichloronitrobenzene	3.05	2.07	2.053	1.973
Nitro	2,5-Dichloronitrobenzene	3.03	2.05	2.053	1.961
Nitro	3,4-Dichloronitrobenzene	3.04	2.07	2.052	1.967
Nitro	3,5-Dichloronitrobenzene	3.09	2.23	2.070	1.999
Nitro	2-Methyl-4,6-dinitrophenol	2.13	0.16	0.979	1.388
Nitro	3-Nitrophenol	2.00	1.40	0.738	1.305
Nitro	Pentachloronitrobenzene	4.77	2.40	3.249	3.068
Nitro	2,3,4,5-Tetrachloronitrobenzene	3.93	1.89	2.905	2.534
Nitro	2,3,5,6-Tetrachloronitrobenzene	4.38	3.20	2.910	2.820
Nitro	2,3,4-Trichloronitrobenzene	3.61	2.20	2.493	2.330
Nitro	2,4,5-Trichloronitrobenzene	3.40	1.84	2.513	2.196
Nitro	2,4,6-Trichloronitrobenzene	3.69	2.88	2.517	2.381
Nitro	Chlornitrofen	3.67	3.04	3.223	2.368
Nitro	4-Nitroaniline	1.39	0.64	0.524	0.917
Nitro	3-Nitroaniline	1.37	0.92	0.531	0.904
Nitro	2-Nitroaniline	1.85	0.91	0.521	1.210
Phos	Phenthoate	3.69	1.56	1.908	2.154
Phos	Diazinon	3.81	1.80	1.784	2.230
Phos	Disulfoton	3.95	2.37	1.233	2.319
Phos	IBP	3.21	0.97	2.748	1.848
Phos	Chloropyrifos	4.82	3.18	2.587	2.873
Phos	Fenthion	4.09	2.68	2.572	2.408
Phos	Fenitrothion	3.47	2.00	2.422	2.014

Table 1 (*continued*)

Class	Chemical	Measured log K_{ow}	Measured log BCF	Calculated log BCF by Eq. (4)	Calculated log BCF by Eq. (5)
Phos	Isofenphos	3.30	2.17	1.824	1.906
Phos	EPN	3.85	3.05	2.700	2.256
Phos	Leptophos	5.88	3.78	3.778	3.548
Other	a-BHC	3.89	2.95	3.049	2.508
Other	b-BHC	3.96	2.86	3.049	2.553
Other	Lindane	3.85	2.84	3.049	2.483
Other	Chlordane	5.00	4.58	3.827	3.215
Other	Heptachlor	5.44	4.14	3.953	3.495
Other	*o,p'*-DDT	5.75	4.57	4.680	4.402
Other	*p,p'*-DDE	5.69	4.71	4.931	4.364
Other	*p,p'*-DDT	5.98	4.84	4.673	4.548
Other	Hexachlorocyclopentadiene	5.04	3.09	3.530	3.240
Other	Acrolein	0.90	2.54	1.513	0.500
Other	Acrylonitrile	1.20	1.68	0.622	0.796
Other	Carbaryl	2.56	1.22	1.311	1.662
Other	Dieldrin	4.53	3.71	3.035	2.915
Other	Heptachlor epoxide	3.65	4.16	3.099	2.355
Other	5-Bromoindole	3.00	1.15	1.846	1.942
Other	Molinate	2.88	1.41	0.935	1.865
Other	Thiobencarb	3.40	2.03	2.500	2.196
Other	Acridine	3.45	2.61	2.511	2.228
Other	BPMC	3.18	1.41	1.314	2.056
Other	Xanthene	4.23	3.62	2.488	2.724

[a] CAH: chlorinated aliphatic hydrocarbon; MAH: monocyclic aromatic hydrocarbon; PAH: polycyclic aromatic hydrocarbon; CB: chlorinated benzene; BB: brominated benzene; CN: chlorinated naphthalene; PCB: polychlorinated biphenyl; PBB: polybrominated biphenyl; CDO: chlorinated dibenzo-dioxins; CDF: chlorinated dibenzofuran; Nitro: nitro aromatic compound; Phoso: phosphorate.
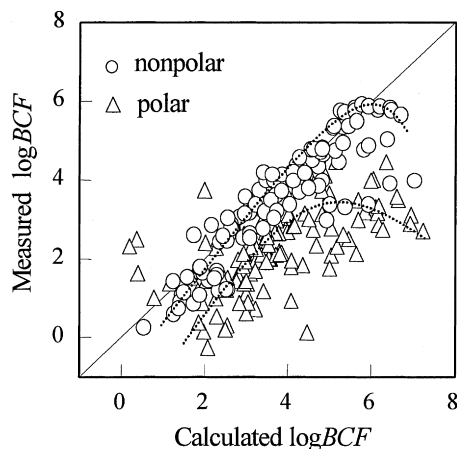
Fig. 1. Relationship between the calculated and measured log BCFs of 239 compounds (calculated by Eq. (1)).

$$\log \text{BCF} = \sum_i a_i \chi_i + \sum_j F_j n_j + c, \qquad (2)$$

where χ_i is the ith molecular connectivity index, n_j is the number of the jth polar functional group, a_i and F_j are regression coefficients for respective variables, and c is a regression constant (y-intercept). F_j is also referred to the ith polarity correction factor.

Values of five molecular connectivity indices ($^0\chi^v$, $^1\chi$, $^2\chi$, $^2\chi^v$, $^3\chi_c$) and numbers of eight polar groups (n_{OH}, n_{NH_2}, n_{NO_2}, n_{NCOO}, n_{NCOS}, n_{COO}, n_O, n_{POO}) were used as the independent variables in the regression. The coefficient of determination (R^2) of the above regression equation was 0.762. The mean absolute error of estimation for all the studied compounds was 0.525 log units. When the eight polar correction factors were deleted and only the five molecular connectivity indices were used in the regression, the obtained R^2 decreased to 0.587 and the mean absolute estimation error increased to 0.711 log units. Clearly, the MCI model with PCF had better estimation quality than the original MCI model, especially for polar compounds. The mean absolute estimation error with and without polarity corrections for the 118 polar compounds were 0.517 and 0.757 log units, respectively. Nevertheless, for some compounds with large molecular

whether it was connected with an aliphatic chain or an aromatic ring, the same PCF was applied. An assumption was made that the contributions of various PCFs were additive. Multivariate linear regression was performed in the following form:

sizes, large difference between the measured and calculated log BCFs still existed.

3.3. Nonlinear modeling with polarity correction factors

It is generally agreed that nonlinear modeling is more suitable considering the BCF prediction of large molecules. For instance, in Sabljic's study where several large molecules such as DDE and aldrin were included, a parabolic relation between log BCF and $^2\chi^v$ was identified (Sabljic, 1987). Nonlinear models have also been extensively derived from the relationship between log BCF and log K_{ow} (Jorgensen et al., 1998).

To provide a best fit for all the BCF data of the studied compounds, a nonlinear estimation analysis was conducted with the addition of three topological terms:

$$\log BCF = K_0(^0\chi^v)^{t_0} + K_1(^1\chi^v)^{t_1} + K_2(^2\chi^v)^{t_2}$$
$$+ \sum_i a_i\chi_i + \sum_j F_j n_j + c \tag{3}$$

$^0\chi^v$, $^1\chi^v$ and $^2\chi^v$ were selected in the additional terms because they expressed basic structure and electron information. The regression values of k_0, k_1, k_2, t_0, t_1 and t_2 were −0.191, 1.625, −0.0002, 2, 0.5 and 1.8, respectively. $(^2\chi^v)^{1.8}$ was deleted from the model due to its small regression coefficient of −0.0002. Ran the nonlinear estimation once again without $(^2\chi^v)^{1.8}$, the statistical summary indicated that $^1\chi$ and $^2\chi^v$ were not significant in the regression. When deleting $^1\chi$ and $^2\chi^v$, the estimation accuracy of the model was only slightly decreased. All of the other parameters were important in the modeling as follows:

$$\log BCF = -0.041(^0\chi^v)^2 - 5.809(^1\chi^v)^{0.5} + 0.615^2\chi$$
$$- 0.785^3\chi_c + 1.564^0\chi^v + \sum_j F_j n_j + 3.179,$$
$$n = 239, \quad R^2 = 0.810, \quad s = 0.615. \tag{4}$$

Deletion of any variable from Eq. (4) would obviously drop the estimation accuracy. Therefore, Eq. (4) was determined to be the final model developed in this study. Table 2 lists the polar functional groups along with the obtained values of polarity correction factors.

3.4. Residual analysis of the developed model

Calculated by Eq. (4), the mean absolute estimation error for all the 239 compounds was 0.478 log units. No significant deviation appeared for either polar compounds or nonpolar ones. The 118 polar compounds had an average absolute residual of 0.488 log unit, and the 121 nonpolar compounds had a similar value of 0.468 log unit. For compounds with large molecules,

Table 2
Polar functional groups and polarity correction factors

Group	Factor
–NCOO–	−2.698
–POO–	−0.982
–NO$_2$	−1.201
–NH$_2$	−0.884
–NCOS–	−1.151
–O–	−0.871
–COO–	−1.151
–OH	−0.644

calculated residuals were mostly within 1.0 log unit. The relationship between the measured and calculated log BCFs of all the studied compounds is plotted in Fig. 2.

In order to find out where the estimation error might be coming from, compounds with absolute residuals higher than 1.0 log unit have been enumerated in Table 3. Relevant information such as test species, condition and method are also given.

Three features can be summarized from Table 3. First, compounds tested in static conditions are underestimated, such as dibenzo(1,4)dioxan, benzo[b]furan and dibenzofuran; Second, compounds with very large or complex molecular structures are overestimated, for instance, octachloronaphthalene, octachlorodi-benzofuran, IBP and 2-phenyldodecane; and third, the relative estimation errors of polar compounds are generally larger than those of nonpolar ones. The three features also appeared for compounds with absolute residuals less than 1.0 log units. Since the measured BCF data came from many sources and involved various test conditions, it is likely that part of the estimation error might be caused by the observed data themselves.
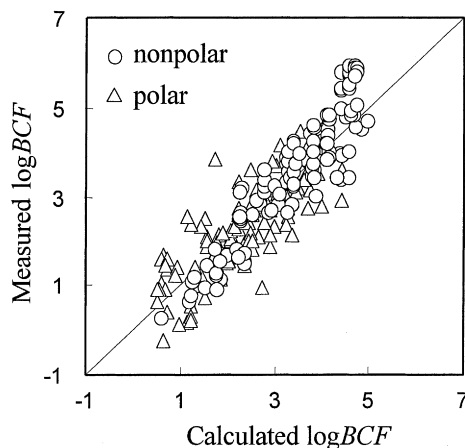

Fig. 2. Relationship between the calculated and measured log BCF of 239 compounds (calculated by Eq. (4)).

Table 3
List of compounds with residual over 1.0 log unit

Compound	Type[a]	Test[b] species	Test condition	Test[c] method	Measured[d] log BCF	Calculated[e] log BCF	Residual (log unit)	Relative error (%)
Dibenzo(1,4)dioxan	p	Gu	Static	C_f/C_w	3.85	1.738	2.112	54.9
IBP	p	Ki	Flow	C_f/C_w	0.97	2.748	−1.778	183.3
Octachloronaphtha-lene	np	Go	Flow	C_f/C_w	3.44	4.972	−1.532	44.5
Octa-chlorodibenzofuran	p	Gu	Flow	C_f/C_w, k_1/k_2	2.93	4.403	−1.468	50.0
2,2',4,4',6-Penta-chlorobiphenyl	np	Go	Flow	C_f/C_w	3.37	4.824	−1.454	43.2
Benzo[b]furan	p	Gu	Static	C_f/C_w	2.56	1.119	1.441	56.3
Benzo[a]pyrene	np	Bl	Flow	C_f/C_w	3.42	4.756	−1.336	39.1
Fenvalerate	p	Ra	Flow	C_f/C_w	2.79	3.985	−1.200	43.1
N-phenyl-2-naph-thylamine	p	Fa	Flow	C_f/C_w	2.17	3.363	−1.193	55.0
Xanthene	p	Gu	Static	C_f/C_w	3.62	2.488	1.132	31.3
Disulfoton	p		Flow	C_f/C_w	2.37	1.233	1.132	47.9
Dibenzofuran	p	Fa	Both	C_f/C_w	3.34	2.216	1.119	33.5
Heptachlor epoxide	p	Fa	Flow	C_f/C_w	4.16	3.099	1.061	25.5
Acrylonitrile	p	Bl	Flow	C_f/C_w	1.68	0.622	1.058	63.0
2-Phenyldodecane	np	Ra	Flow	C_f/C_w	2.65	3.698	−1.048	39.5
Acrolein	p	Bl	Flow	C_f/C_w	2.54	1.513	1.027	40.4
2,3,4,5-Tetrachloro-nitrobenzene	p	Ra	Flow	C_f/C_w	1.88	2.905	−1.020	54.1

[a] p: Polar compound; np: nonpolar compound.
[b] Gu: guppy; Ki: killifish; Go: golden ide; Bl: bluegill sunfish; Ra: rainbow trait; Fa: fathead minnow.
[c] C_f/C_w: plateau method; k_1/k_2: kinetic method.
[d] The median value of all the measured log BCF data for a compound.
[e] Calculated by Eq. (4).

Actually, for many compounds in this study, more than one measured BCF data was available. The maximum measure difference for a compound amounted to 2.5 log units.

3.5. Robustness test of the developed model

The robustness of the developed MCI-based model was validated using modified Jackknifed tests as suggested by Dietrich et al. (1980). Fifty compounds were randomly deleted from the regression and a Jackknifed model was developed in the same way as Eq. (4). This operation was repeated 20 times and 20 Jackknifed R^2 were obtained. The mean Jackknifed R^2 was 0.814, close to the original R^2 (0.810) of Eq. (4). The variation coefficient of the 20 Jackknifed R^2 was only 0.017, indicating the developed model was robust. Variations in the regression values of all the parameters were calculated and depicted in Fig. 3.

Except for –NCOS–, the Jackknifed regression values were all quite stable with variation coefficients less than 0.2. One of the reasons for the unstable regression coefficient of –NCOS– might be that the group was contained only by two compounds. The number of data relating to this variable was much less compared with
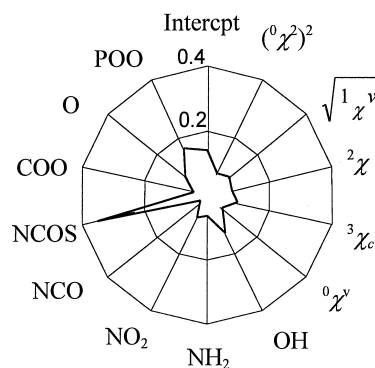

Fig. 3. Variation coefficients of the regression coefficients in the Jackknifed test.

other variables in the regression. Therefore, more fluctuation would be expected. For other parameters, especially the molecular connectivity indices, the regression values were rather robust with variation coefficients less than 0.1. Variations in the regression values of polarity correction factors were relatively large since these variables relied on the specific functional groups of the studied compounds more than molecular

connectivity indices did. In addition, the test also revealed that some compounds had significant influence on the model.

To further examine which kind of compounds have the most influence on the model, the studied compounds were divided into 17 classes according to various molecular structures. One class of compounds at a time was deleted and the regression was run in the same way as indicated in Eq. (4); with the process being repeated 17 times. The variations obtained from the Jackknifed R^2 and s are shown in Fig. 4. Among the 17 classes of compounds, polychlorinated biphenyls displayed the most significant influence on the model. By deleting them, the Jackknifed R^2 decreased to 0.747, while s also decreased to 0.585. The decline of the Jackknifed R^2 seems principally caused by the decrease of data used for the regression, since polychlorinated biphenyl was the biggest class including 41 compounds. The Jackknifed R^2 obtained after deleting other classes of compounds did not change much compared with the original R^2, and

increase of the Jackknifed R^2 was in accordance with decline of the s. Among them, chlorinated dibenzo-dioxins and chlorinated dibenzofurans exhibited relatively more effects on the model. The Jackknifed R^2 corresponding to these two classes were 0.8269 and 0.8238, while the Jackknifed s were 0.5999 and 0.5983, respectively. This indicted that shortcomings existed in the model concerning the BCF estimation of chlorinated dibenzo-dioxins and dibenzofurans.

3.6. Comparison of the MCI-based model with a K_{ow}-based model

For the purpose of comparison in BCF estimation accuracy between the novel MCI method and the usual K_{ow} method, a K_{ow}-based model with correction factors was selected from a recently published paper (Meylan et al., 1999). Based on the measured K_{ow} and BCF data of the 239 compounds in this study, the K_{ow}-based model was established in the same manner as being reported. When multiple measured log K_{ow} data available for a compound, the medium value was used. The obtained model was as follows:

$$
\begin{aligned}
&\log K_{ow} < 1 \\
&\quad \log \mathrm{BCF} = 0.50, \quad n = 5, \\
&1 < \log K_{ow} < 7 \\
&\quad \log \mathrm{BCF} = 0.032 + 0.636 \log K_{ow} + \sum F_i \\
&\quad n = 214, \quad R^2 = 0.781, \quad s = 0.614, \\
&\log K_{ow} > 7 \\
&\quad \log \mathrm{BCF} = 8.032 - 0.638 \log K_{ow} + \sum F_i, \\
&\quad n = 20, \quad R^2 = 0.795, \quad s = 0.617,
\end{aligned}
\tag{5}
$$

where F_i is the ith correction factor. In the K_{ow}-based model, correction factors were indicator variables for compounds with certain structural features. Each factor was counted only once no matter how many times the functional group appeared in the molecule, but more than one factor might apply to a given compound (Meylan et al., 1999). Since only the estimation accuracy was concerned here, the regression values of correction factors and their application rules were not presented in this paper. Comparison in estimation accuracy between the MCI-based model and the K_{ow}-based model was shown in Table 4 and Fig. 5.

Fig. 4. Distribution of Jackknifed R^2 and s during the deletions of each class of compounds. 1: Chlorinated aliphatic hydrocarbons; 2: monocyclic aromatic hydrocarbons; 3: polycyclic aromatic hydrcarbons; 4: chlorinated benzenes; 5: brominated benzenes; 6: chlorinated naphthalenes; 7: polychlorinated biphenyls; 8: polybrominated biphenyls; 9: chlorinated dibenzo-dioxins; 10: chlorinated dibenzfuran; 11: phenols; 12: anilines; 13: esters; 14: ethers; 15: nitro aromatic compounds; 16: phosphorates; 17: others).

Table 4
Comparison between the MCI-based model and the K_{ow}-based model

Model	Mean absolute estimation error (log unit)			Mean relative estimation error (%)		
	Total	Nonpolar	Polar	Total	Nonpolar	Polar
MCI-based	0.478	0.468	0.488	26.2	18.1	34.5
K_{ow}-based	0.494	0.493	0.494	27.4	20.5	34.5

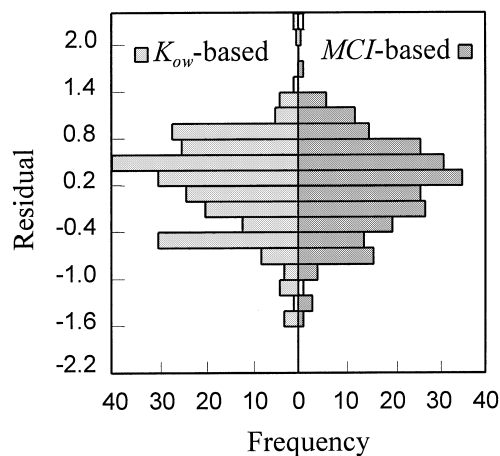Fig. 5. Comparison between the MCI-based model and the K_{ow}-based model.

Clearly, for the 239 compounds in this study, the MCI-based model had better estimation quality than the K_{ow}-based model did, particularly for nonpolar compounds. In our previous study, it was also found the BCFs of 80 nonpolar compounds could be predicted more accurately by MCIs than by K_{ow} (Lu et al., 1999). It seems that MCIs are better descriptors than K_{ow} concerning the BCF estimation of nonpoloar compounds. However, for the polar compounds studied, the estimation results were similar in both cases, though the mean absolute estimation error calculated from the MCI-based model was slightly lower that achieved by the K_{ow}-based model.

As can be seen from Fig. 5, the two residual distributions differed with each other in some aspects, but within certain ranges, the total frequencies were about equal. For instance, the amount of compounds with residuals less than 0.5 log units was 59% for the MCI-based model and 57% for the K_{ow}-based model. Moreover, among the total 27 compounds which had residuals above 1.0 log units, eight compounds were found appearing in both cases. Among the total 141 compounds which had residuals above 0.5 log units, 59 ones appeared in both cases.

4. Discussion

A topological model for the BCF estimation of nonionic organic compounds in fish was developed by using molecular connectivity indices and polarity correction factors. The physical significance of the developed model could be explainable. According to McFarland's probability model, in order to be absorbed in a biological system, chemical must penetrate a sequence of hydrophobic and hydrophilic barriers, and therefore the bioconcentration process is controlled by polar and nonpolar interactions among water and fish. In the topological model, the regression values of polarity correction factors are all negative, indicating that polar interactions are utilized to provide an attractive force between chemical and water, rather than between chemical and fish tissue. On the other hand, molecular connectivity indices have been known to represent chemical's structural attribute and hydrophobic character, and thus quantify the nonpolar interactions among various components. The nonlinear relation of molecular connectivity indices to log BCF reveals that super large molecules are expected to have low BCFs. These results correspond to many reports from bioconcentration studies. For example, Opperhuizen found that a chemical with molecular cross-sector over 9.5 A could not penetrate the membrane easily due to steric hindrance (Opperhuizen et al., 1985).

By using the methodology reported here, the BCF could be estimated to within 0.5 log units for three-fifths of the 239 compounds studied. This level of accuracy is quite good considering the many sources of error that may impact the model. Residual analysis indicated that the estimation error might come from BCF measurement, molecular structure, test species, etc. BCF measurement is a large source of error. Uncertainty associated with a given measured BCF may arise from the exposure concentration, test condition, duration of the experiment, and determination of the concentration in water and fish. In this study, multiple measured BCF values were available for some compounds, and the maximum measure difference for compounds amounted to 2.5 log units. Test species is also a major source of error. Differences between fish species are reflected in lipid content, blood low, and metabolism of compounds. Use of BCF basing on lipid content was suggested to eliminate some of the intra- and inter-species variability. In our previous study, a similar regression was performed using BCF_l (basing on lipid) instead of BCF (basing on whole weight) for the same set of compounds, however, no significant difference was found (Lu et al., 1999). BCF_l were not used in this study, because most data were not available. The third main source of error might be the selection of polarity correction variables. The polar functional groups were identified in accordance with how many times they appeared in the studied molecules. No correction was applied to those polar functional groups which appeared only once. For some groups with similar structures such as –NH_2, –NH–, –N– and –CN, one polarity correction factor was used.

The modified Jackknifed tests validated that the developed MCI-based model was statistically robust. Deleting 50 compounds randomly from the regression each time, variations in the coefficients of determination (R^2) between the pre-deletion and post-deletion and among various deletions were rather small. However, the

Jackknifed R^2 declined greatly when 41 polychlorinated biphenyls were deleted from the regression, indicating that the determined model was much influenced by the structures of the training compounds and their numbers. Relatively, the regression values of molecular connectivity indices were more stable than those of polarity correction factors. The reason might be that the two kinds of descriptors had different applications. Molecular connectivity indices were used in all of the 239 compounds studied, whereas polarity correction factors were just used by polar compounds.

Although shortcomings existed, the MCI-based model was superior to the K_{ow}-based model in many aspects, including estimation accuracy, model structure, and parameter availability. As for the studied 239 compounds, the average estimation error calculated by the MCI-based model was 0.016 log units less than that achieved by the K_{ow}-based model. The MCI-based model had a better quality especially for the BCF estimation of nonpolar compounds. Besides, the MCI-based model had a simpler structure than the K_{ow}-based model did. The K_{ow}-based model was developed according to an improved method for estimating BCF from K_{ow} (Meylan et al., 1999). Three equations consisted the K_{ow}-based model, and each equation was applied to compounds within a certain $\log K_{ow}$ range. Whereas in the MCI-based model, one equation was used for all the studied compounds. Furthermore, the most significant advantage of using topological variables as descriptors of BCF lies in that data on parameters are readily available. The topological values can be calculated quantitatively and directly from a given molecule. However, data on traditional descriptors such as K_{ow}, when not available in the literature, have to be experimentally determined or estimated by other methods.

5. Conclusion

Estimation of bioconcentration factors for a wide range of nonionic organic compounds on basis of molecular topology was investigated. Molecular connectivity indices were found to be good descriptors of BCF for nonpolar compounds but not for polar ones. When polarity correction factors were introduced into the linear molecular connectivity model, the BCF estimation for polar compounds was much increased. Nonlinear modeling with polarity correction factors further improved the predictive quality of the developed model because it provided a better fit for compounds with large molecular sizes. The BCF could be estimated to within 0.5 log units for 59% of the studied 239 compounds by using the methodology reported here. This level of accuracy was quite good considering the many sources of error that might impact the model. Modified Jackknifed tests showed that the statistical model was robust, yet

some compounds were discovered to have significant influences on the model. Comparison between the MCI-based model and a K_{ow}-based model indicated that the estimation accuracy based on MCI was not inferior to that achieved by K_{ow}.

References

Chiou, C.T., Freed, V.H., Schmedding, D.W., Kohnert, R.L., 1977. Partition coefficient and bioaccumulation of selected organic chemicals. Environ. Sci. Technol. 11, 475–478.

Connell, D.W., Hawker, D.W., 1988. Use of polynomial expressions to describe the bioconcentration of hydrophobic chemicals by fish. Ecotoxicol. Environ. Safety 16, 242–257.

Davis, R.P., Dobbs, A.J., 1984. The prediction of bioconcentration in fish. Water Res. 18, 1253–1262.

Devillers, J., Bintein, S., Domine, D., 1996. Comparison of BCF models based on log P. Chemosphere 33, 1047–1065.

Dietrich, D.W., Dreyer, N.D., Hansch, C.J., 1980. Confidence interval estimators for parameters associated with quantitative structure-activity relationships. J. Med. Chem. 23, 1201–1205.

Donald, M., 1982. Correlation of bioconcentration factors. Environ. Sci. Technol. 16, 274–278.

Fox, K., Zauke, G.P., Butte, W., 1994. Kinetics of bioconcentration and clearance of 28 polychlorinated biphenyl congeners in Zebrafish. Ecotoxicol. Environ. Safety 28, 99–109.

Freitag, D., Ballhorn, L., Geyer, H., 1985. Environmental hazard profile of organic chemicals: an experimental method for the assessment of the behavior of organic chemicals in the ecosphere by means of simple laboratory tests with [14]C Labelled Chemicals. Chemosphere 14, 1589–1616.

Gobas, F.A.P.C., 1990. Bioaccumulation of some polychlorinated dibenzo-p-dioxins and octachlorodibenzofuran in the guppy, poecilia reticulata. Chemosphere 20, 495–512.

Gobas, F.A.P.C., Clark, K.E., Shiu, W.Y., 1989. Bioconcentration of polybrominated benzenes and biphenyls and related superhydrophobic chemicals in fish: role of bioavailability and elimination into the feces. Environ. Toxicol. Chem. 8, 231–245.

Isnard, P., Lambert, S., 1988. Estimating bioconcentration factors from octanol–water partition coefficient and aqueous solubility. Chemosphere 17, 21–34.

Jorgensen, S.E., Sorensen, B.H., Mahler, H., 1998. Handbook of estimation methods in ecotoxicology and environmental chemistry. Lewis, Washington, DC.

Kier, L.B., Hall, L.H., 1986. Molecular Connectivity in Structure-Activity Analysis. Research Studies Press, England.

Koch, R., 1983. Molecular connectivity index for assessing ecotoxicological behaviour of organic compounds. Toxicol. Environ. Chem. 6, 87–96.

Lu, X.X., Tao, S., Cao, J., Dawson, R.W., 1999. Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices. Chemosphere 39, 987–999.

Mackay, D., Shiu, W.Y., Ma, K.C. (Eds.). 1992. Illustrated Handbook of Physical–Chemical Properties and Environmental Fate for Organic Chemicals. Lewis, Michigan.

McCarty, L.S., 1986. The relationship between aquatic toxicity QSARs and bioconcentration for some organic chemicals. Environ. Toxicol. Chem. 5, 1071–1080.

Meylan, W.M., Howard, P.H., Boethling, R.S., Aronson, D., Printup, H., Gouchie, S., 1999. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/ water partition coefficient. Environ. Toxicol. Chem. 18, 664–672.

Nendza, M., 1993. QSAR of bioconcentration: validity assessment of log Pow/log BCF correlations. In: Nagel, R., Loskill, R., (Eds.). Bioaccumulation in Aquatic Systems. VCH, Weinheim, pp. 43–66.

Nirmalakhandan, N.N., Speece, R.E., 1988. Prediction of aqueous solubility of organic chemicals based on molecular structure. Environ. Sci. Technol. 22, 328–338.

Oliver, B.G, Niimi, A.J., 1983. Bioconcentration of chlorobenzenes from water by rainbow trout: correlations with partition coefficients and environ mental residues. Environ. Sci. Technol. 17, 287–291.

Oliver, B.G, Niimi, A.J., 1984. Rainbow trout bioconcentration of some halogenated aromatics from water at environmental concentrations. Environ. Toxicol. Chem. 3, 271–277.

Oliver, B.G, Niimi, A.J., 1985. Bioconcentration factors of some halogenated organics for rainbow trout: limitations in their use for prediction of environ mental residues. Environ. Sci. Technol. 19, 842–849.

Opperhuizen, A., vanderVelde, E.W., Gobas, F.A.P.C., 1985. Relationship between bioconcentration in fish and steric factors of hydrophobic chemicals. Chemosphere 14, 1871–1896.

Sabljic, A., 1987a. On the prediction of soil sorption coefficients of organic pollutants from molecular structure: application of molecular topology model. Environ. Sci. Technol. 21, 358–366.

Sabljic, A., 1987b. The predication of fish bioconcentration factors of organic pollutants from the molecular connectivity model.. Z gesamte. Hyg. 33, 493–496.

Sabljic, A., Gusten, H., Hermens, J., Opperhuizen, A., 1993. Modelling octanol/water partition coefficients by molecular topology: chlorinated benzenes and biphenyls. Environ. Sci. Technol. 27, 1394–1402.

Sabljic, A., Protic, M., 1982. Molecular connectivity: a novel method for predition of bioconcentration factor of hazardous chemicals. Chem. Biol. Interact. 42, 301–310.

Smith, A.D., Bharath, A., Mallard, D., 1990. Bioconcentration kinetics of some chlorinated benzenes and chlorinated phenols in American flagfish, Jordanella floridae. Chemosphere 20, 379–386.

Tsuda, T., Kojima, M., Harada, H., 1997. Relationships of bioconcetration factors of organophosphate pesticides among species of fish. Comp. Biochem. Physiol. 116, 213–218.

Veith, G.D, Macek, K.J., Petrocelli, S.R., 1980. An evaluation of using partition coefficients and water solubility to estimate bioconcentration factors for organic chemicals in fish. In: Eaton, J.G., Parrish, P.R., Hendricks, A.C. (Eds.), Aquatic Toxicology, ASTM STP 707. American Society for Testing and Materials, Philadelphia, PA, pp. 116–129.

Voogt, P., Muir, D.C.G., Webster, G.R.B., 1990. Quantitative structure-activity relationships for the bioconcentration in fish of seven polychlorinated dibenzodioxins. Chemosphere 21, 1385–1396.

Watzede, W., SEYedema, E., Seinen, W., 1994. Bioconcentration kinetics of chlorinated anilines in guppy, poecilia reticulata. Chemosphere 28, 159–167.

Xiaoxia Lu obtained her Masters degree in Environmental Sciences from the Chinese University of Geosciences (1996), and since 1996, she has been a Ph.D student at the Department of Urban and Environmental Sciences, Peking University, China.

Dr. Shu Tao is a Professor of Environmental Sciences at the Department of Urban and Environmental Sciences, Peking University, China. He received his Ph.D in Environmental Sciences from the University of Kansas (1984), USA.