QSAR MODELLING OF SOIL SORPTION.
IMPROVEMENTS AND SYSTEMATICS OF log K_{OC} vs. log K_{OW} CORRELATIONS

Aleksandar Sabljić,[1,2,*] Hans Güsten,[1] Henk Verhaar,[3] and Joop Hermens[3]

[1]Institut für Meteorologie und Klimaforschung, Forschungszentrum Karlsruhe/Universität Karlsruhe
D-76021 Karlsruhe, Federal Republic of Germany

[2]Institute Ruđer Bošković, P.O.B. 1016, HR-10001 Zagreb, Republic of Croatia

[3]Research Institute of Toxicology, University of Utrecht, P.O. Box 80176, NL-3508 TD Utrecht,
The Netherlands

ABSTRACT

A systematic study was performed to evaluate the quality and reliability of the quantitative relationships between the soil sorption coefficients and the n-octanol/water partition coefficients (log K_{oc} vs. log K_{ow}). A system of QSAR models has been derived which is based on a reliable set of experimental log K_{ow} data (LOGPSTAR data) or reliable estimates calculated with the latest version of the ClogP algorithm. Particular emphasis has been made to clearly define the boundaries for application of developed models as well as the quality of estimates. Thus, for each developed model the application domain is clearly defined through the chemical (structural) domain, substituents domain, and X-variable domain. Finally, the QSAR model with the first-order molecular connectivity indices has been incorporated in the derived system of QSAR models since the soil sorption estimates of the predominantly hydrophobic chemicals based on the log K_{ow} data have large uncertainties, particularly in the log K_{ow} data range from 4 to 7.5.

1. INTRODUCTION

Background Information

The sorption of commercial chemicals by soil and sediment plays a very important role in their transport and mobility in the environment [1]. Furthermore, the sorption may significantly influence the chemical and biological transformation or degradation of chemicals in the aquatic environment. Thus, the measured or accurately estimated soil sorption coefficients of xenobiotic chemicals are of critical importance for evaluating their fate and potential exposure to chemicals in the environment and, consequently, for the whole process of environmental risk assessment [2]. A basic idea behind this systematic study is to develop a reliable user-friendly methodology for estimating soil sorption coefficients of a large number of commercial chemicals.

---

* Corresponding author: Phone (385-1) 456-1089; Fax (385-1) 272-648; E-mail:sabljic@olimp.irb.hr

Despite extensive experimental work carried out at numerous laboratories for more than 30 years, measured soil sorption coefficients are available for less than 500 chemicals [3-6] and reliable data on sorption coefficients are in many cases unavailable, not even for the High Production Volume (HPV) chemicals [7]. Thus, a considerable number of studies have been performed to develop quantitative models for estimating the soil sorption coefficients of organic chemicals [3-6, 8-10]. A thorough literature search has shown that nearly 50 relevant studies [3-6, 8-51] have been published during the last 25 years. Their analysis has shown that the majority of QSAR models [3-5, 12-14, 22-29, 31-33, 39-46] for estimating soil sorption coefficients (log K_{oc}'s) use either the n-octanol/water partition coefficient (log K_{ow}) or the aqueous solubility (log S) as molecular descriptors. These two physico-chemical properties are highly intercorrelated [3, 4, 8, 32] and, consequently, these two groups of QSAR models can be considered as identical or parallel. Unfortunately, these models are mainly class-specific, they are based on a small number of chemicals, and no external validation was performed. Furthermore, a large variety in slopes (from 0.40 to 1.04) and intercepts (from -0.88 to 1.79) of QSAR models describing linear relationships between soil sorption coefficients and n-octanol/water partition coefficients point to their large uncertainty and low reliability of estimates.

Additional research is needed in this area in order to develop reliable general as well as class-specific quantitative relationships between the soil sorption coefficients and the n-octanol/water partition coefficient (log K_{oc} vs. log K_{ow}). Thus, the primary objective of this study is to develop a reliable methodology for estimating soil sorption coefficients from n-octanol/water partition coefficients. Particular emphasis will be placed on clearly defining the boundaries for the application of developed models as well as on the quality of estimates. For each developed model its application domain will be uniquely defined by unambiguous description of its chemical (structural) domain, substituents domain, and X-variable domain. Finally, for predominantly hydrophobic chemicals, the QSAR model with the first-order molecular connectivity indices [11] will be redefined and reevaluated. Several attempts have already been made [52, 53] to evaluate quality and applicability of QSAR models important in exposure assessment. In this study, the applicability of developed models will be described, for the first time, in a very precise and uniform manner.

Developing Strategy

General, subgeneral, and class-specific QSAR models will be systematically developed for all available soil sorption data. This way we hope to develop the reliable general models that will have very wide range of applicability but relatively low accuracy as well as the reliable class-specific models with narrow application domains and high accuracy. The range of applicability and level of accuracy of subgeneral models will be somewhere in between. In this modelling approach, log K_{oc} vs. log K_{ow}, there will always be a trade-off between the range of model applicability and the accuracy of estimates. However, this way it will be possible to make reliable and accurate estimates for simple chemicals and reasonable estimates for complex chemicals that have many functional groups and which are large in size.

The following strategy was adopted to make this research effort more effective. It was decided to use the n-octanol/water partition coefficient (log K_{ow}) as structural descriptors. There is a large database (MedChem Master file) with evaluated and recommended log K_{ow} values for nearly 10,000 chemicals [54]. The recommended log K_{ow} values are known as LOGPSTAR data. Furthermore, there is also a standard method for estimating n-octanol/water partition coefficients, the ClogP program [55]. The program indicates if there is a problem or some uncertainty in calculating the n-octanol/water partition coefficients. Thus, in the modelling process we will use the LOGPSTAR data as first choice. If the LOGPSTAR data are not available, we will use the ClogP data, but only if they are reliable, i.e. all fragment values have been determined, no

approximation is used in the calculation, and/or no other type of warning is given. If LOGPSTAR or reliable ClogP data are not available, that chemical was not used in the modelling process.

The most common and today generally accepted quantitative measure of the sorption of organic pollutants by soil or sediment from aqueous solutions is the soil sorption coefficient (log K_{oc}). This chemical-specific parameter provides a relative measure of mobility in aqueous/soil systems. In general, compounds with higher log K_{oc} values will be less mobile than those with lower values. In the modelling process, we will use the log K_{oc} data from the most recent compilations of soil sorption coefficients [5, 6]. It is assumed that, in general, the most recent compilations will also contain the most reliable data presently available. In most instances, the median value is used when several measured K_{oc} values are available for a given chemical. All soil sorption data used in this study have been determined for nonionic species of respective chemicals. Thus, QSAR models, developed during the course of this study, will be applicable only for nonionized chemicals.

Table 1. List of predominantly hydrophobic chemicals with their soil sorption coefficients (log K_{oc}), the first-order molecular connectivity indices ($^1\chi$), and reliable, measured or calculated, n-octanol/water partition coefficients (log K_{ow}).

Chemicals	log K_{oc}	$^1\chi$	log K_{ow}
Benzene	1.96	3.000	2.19
Toluene	2.39	3.394	2.79
Ethylbenzene	2.22	3.931	3.15
1,2-Dimethylbenzene	2.41	3.805	3.12
1,3-Dimethylbenzene	2.34	3.788	3.20
1,4-Dimethylbenzene	2.52	3.788	3.15
Propylbenzene	2.86	4.432	3.72
1,3,5-Trimethylbenzene	2.82	4.182	3.42
1,2,3-Trimethylbenzene	2.80	4.218	3.59
1,2,4,5-Tetramethylbenzene	3.12	4.609	4.00
n-Buthylbenzene	3.39	4.932	4.38
Styrene	2.96	3.932	2.95
Naphthalene	3.11	4.966	3.30
Anthracene	4.27	6.933	4.45
Phenanthrene	4.28	6.949	4.56
Tetracene	5.81	8.899	5.90
Pyrene	4.66	7.933	4.88
Dibenz[1,2;5,6]anthracene	6.22	10.90	6.50
Fluorene	3.70	6.449	4.18
Fluoranthene	4.62	7.949	5.16
1-Methylnaphthalene	3.36	5.377	3.87
2-Methylnaphthalene	3.64	5.360	3.86
1-Ethylnaphthalene	3.78	5.915	4.39
2-Ethylnaphthalene	3.76	5.898	4.38
9-Methyanthracene	4.81	7.360	5.07
Benz(a)anthracene	5.30	8.916	5.79
7,12-Dimethylbenz(a)anthracene	5.37	9.771	5.80
3-Methylcholanthrene	6.10	10.33	6.42
Chlorobenzene	2.34	3.394	2.90
1,2-Dichlorobenzene	2.50	3.805	3.43
1,3-Dichlorobenzene	2.48	3.788	3.53
1,4-Dichlorobenzene	2.63	3.788	3.44
1,2,3-Trichlorobenzene	3.29	4.215	4.14

Table 1.   (continued)

Chemicals	log K_{oc}	$^1\chi$	log K_{ow}
1,2,4-Trichlorobenzene	3.15	4.198	4.05
1,3,5-Trichlorobenzene	2.85	4.182	4.19
1,2,3,5-Tetrachlorobenzene	3.20	4.609	4.66
Pentachlorobenzene	3.50	5.037	5.18
Hexachlorobenzene	3.99	5.464	5.73
Bromobenzene	2.49	3.394	2.99
Iodobenzene	3.10	3.394	3.25
Biphenyl	3.27	5.966	4.01
2-Chlorobiphenyl	3.47	6.377	4.53
2,2'-Dichlorobiphenyl	3.92	6.788	4.97
2,4'-Dichlorobiphenyl	4.14	6.771	5.10
2,4,4'-Trichlorobiphenyl	4.62	7.165	5.62
2,5,2'-Trichlorobiphenyl	4.23	7.182	5.60
2,4,2'-Trichlorobiphenyl	4.84	7.182	5.76
2,6,2',6'-Tetrachlorobiphenyl	5.01	7.609	5.94
2,5,3',4'-Tetrachlorobiphenyl	4.84	7.575	6.23
2,5,2',5'-Tetrachlorobiphenyl	4.73	7.575	6.09
2,3,4,,2',5'-Pentachlorobiphenyl	4.62	8.003	6.37
2,4,5,,2',5'-Pentachlorobiphenyl	4.63	7.986	6.50
2,3,4,,2',3',4'-Hexachlorobiphenyl	5.05	8.430	7.32
2,4,5,,2',4',5'-Hexachlorobiphenyl	5.65	8.397	7.16
p,p'-DDT	5.31	8.874	6.91
p,p'-DDE	4.82	8.575	6.96
Dichloromethane	1.44	1.414	1.25
Trichloromethane	1.65	1.732	1.97
Tetrachloromethane	1.85	2.000	2.83
Bromodichloromethane	1.78	1.732	2.09
Dibromochloromethane	1.92	1.732	2.23
Tribromomethane	2.06	1.732	2.67
1,1-Dichloroethane	1.48	1.732	1.79
1,2-Dichloroethane	1.52	1.914	1.48
1,2-Dibromoethane	1.80	1.914	1.96
1,1,1-Trichloroethane	1.82	2.000	2.49
1,1,2-Trichloroethane	1.87	2.270	1.89
1,1,2,2-Tetrachloroethane	1.90	2.643	2.39
1,2-Dichloropropane	1.67	2.270	1.99
1,2-Dibromo-3-chloropropane	2.11	2.808	2.26
1,1-Dichloroethene	1.81	1.732	2.13
1,2-Dibromoethene	1.65	1.914	1.95
Tetrachloroethene	2.38	2.643	3.40
Trichloroethene	2.00	2.270	2.61
1,3-Dichloropropene	1.53	2.414	1.76
α-HCH	3.25	5.464	3.80
β-HCH	3.36	5.464	3.78
γ-HCH	3.00	5.464	3.76
Aldrin	4.69	8.276	6.50
Chlordane	5.15	8.114	5.80
Mirex	6.00	9.500	5.28

2. CALCULATION OF DESCRIPTORS AND STATISTICAL METHODS

The n-octanol/water partition coefficients (log K_{ow}) used in this study were collected from the MedChem Master file [54] or calculated by the ClogP program [55]. Molecular connectivity indices were calculated on an Apple Macintosh SE/30 personal computer using the GRAPHIII computer program [56, 57]. It must be pointed out that the $^1\chi$ indices can be easily calculated by hand.

A regression analysis was carried out on an Apple Macintosh SE/30 personal computer using the SYSTAT, statistical analysis system, version 5.2.1 [58]. The following statistical parameters were used to test the quality of generated regression equations: the correlation coefficient (r), the coefficient of determination (r^2), the standard error of the estimate (s), a test of the null hypothesis (F-test) and the cross-validated explained variance (Q^2).
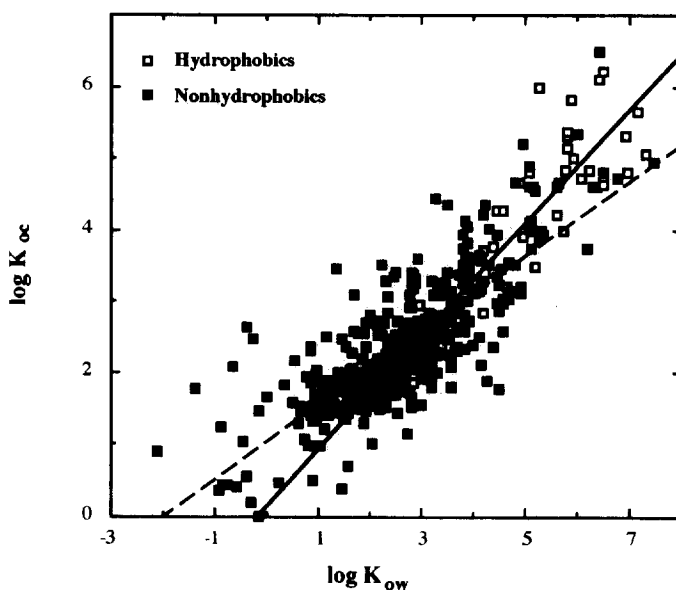

Figure 1. Plots of log K_{oc} vs. log K_{ow} for hydrophobic (open squares) and nonhydrophobic chemicals (closed squares). Different trends for these two classes of chemicals are clearly shown by different slopes of lines describing QSAR models for hydrophobic (solid line) and nonhydrophobic (dashed line) chemicals.

3. RESULTS

GENERAL SOIL SORPTION MODELS

Predominantly hydrophobic chemicals. We will uniquely define the predominantly hydrophobic chemicals as compounds that contain only carbon (C), hydrogen (H), and halogen (F, Cl, Br, I) atoms. A graphical analysis of the plots log K_{oc} vs. log K_{ow} (Fig. 1) has shown that predominantly hydrophobic chemicals display trends different from all other classes of chemicals. The soil sorption data are available for 81 predominantly hydrophobic chemicals (Tab. 1). There are two models that may be used to estimate soil sorption coefficients of predominantly hydrophobic chemicals. The more accurate model is the linear correlation of the soil sorption coefficients with the first-order molecular connectivity indices (Fig. 2).

$$\log K_{oc} = 0.70 + 0.52 * {}^1\chi \qquad\qquad (1)$$

$$N = 81 \quad r = 0.981 \quad r^2 = 0.961 \quad s = 0.264 \quad F^{1,79} = 1993 \quad p > 99.9\% \quad Q^2 = 0.959$$
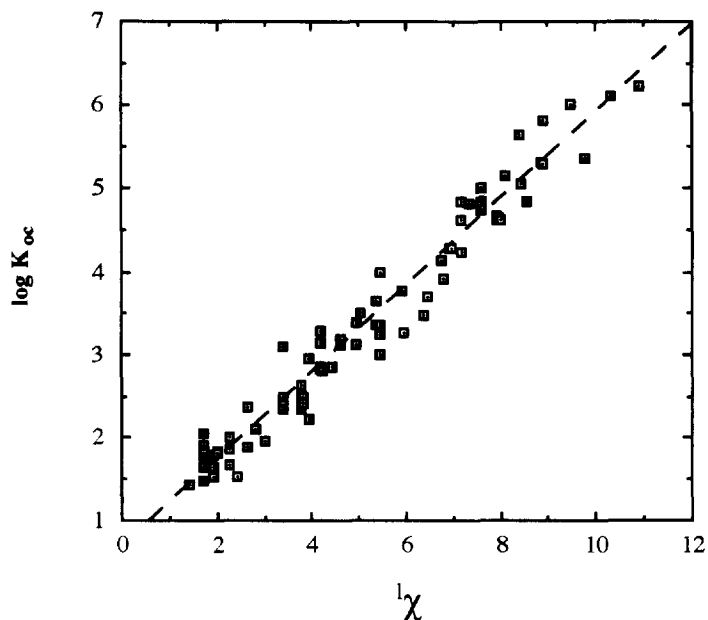
Figure 2. Plot of log K_{oc} vs. the first-order molecular connectivity indices for predominantly hydrophobic chemicals. The dashed line represents QSAR model described by Eq. 1.
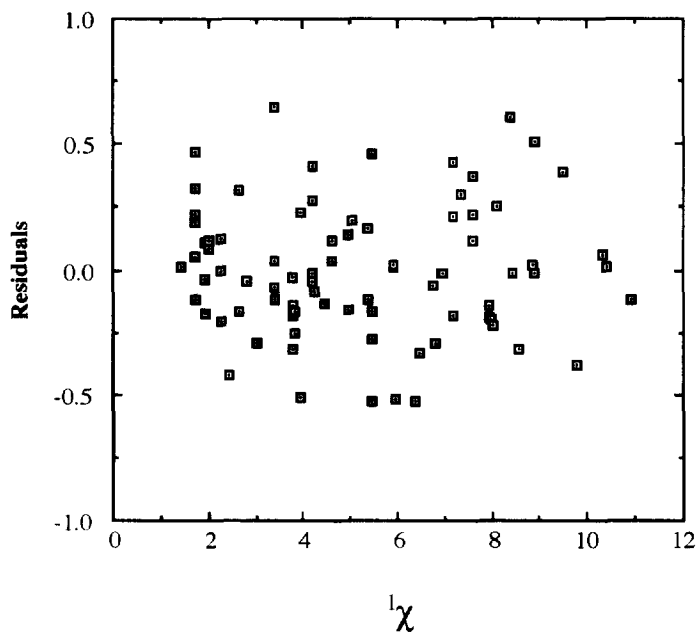

Figure 3. Plot of residuals vs. the first-order molecular connectivity indices for Eq. 1.

All estimated soil sorption coefficients are within the range of ± 0.64 log units from the measured values. In this model standard error of the estimate is very small and it can be applied to all predominantly hydrophobic chemicals. The plot of residuals vs. the $^{1}\chi$ index shows that the precision of estimates is uniform across the whole range of independent variable (see Fig. 3). The application domain of this model are all predominantly

hydrophobic chemicals with 3 to 22 carbon or halogen atoms and the K_{oc} estimates between 1 and 6.5 log units. Estimates outside this range cannot be considered reliable. This model was externally validated at several stages [9, 34, 59] during its development and it has always shown a very good performance.

The second model for the estimation of soil sorption coefficients of predominantly hydrophobic chemicals is based on the reliable log K_{ow} data described above.

$$\log K_{oc} = 0.10 + 0.81*\log K_{ow} \tag{2}$$

$$N = 81 \quad r = 0.943 \quad r^2 = 0.887 \quad s = 0.451 \quad F^{1,79} = 629 \quad p>99.9\% \quad Q^2 = 0.881$$

Mirex was a clear outlier in this model with residual of nearly 4 standard errors (1.65 log units). However, its influence on the regression model is negligible and it is not excluded from this model (Eq. 2). All other estimated soil sorption coefficients are within the range of ±1 log unit from the measured values. The plot of residuals vs. the log K_{ow} data shows that the precision of estimates is higher for the less hydrophobic chemicals and lower for the more hydrophobic chemicals (see Fig. 4). For the less hydrophobic chemicals (the log K_{ow} data from 1 to 4) the spread of residuals is from 0.2 to 0.5 log units. However, for the more hydrophobic chemicals (the log K_{ow} data from 4 to 7.5) the spread of residuals is from 0.6 to 1.0 log units. Nevertheless, this model has relatively low standard error of the estimate and it can be applied to predominantly hydrophobic chemicals that have log K_{ow} values in the range from 1 to 4.
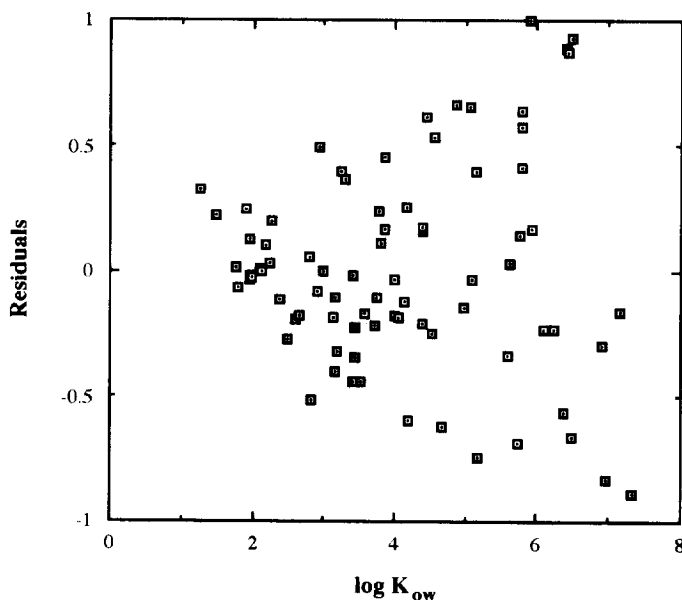

Figure 4. Plot of residuals vs. the n-octanol/water partition coefficients for Eq. 2.

Nonhydrophobic chemicals. All chemicals that cannot be classified as predominantly hydrophobic chemicals are uniquely defined as nonhydrophobic chemicals. This classification is based strictly on structural properties of those chemicals and does not imply anything about their lipophilicity. Appropriate log K_{oc} and log K_{ow} values were available for 390 chemicals (Tab. 2). The linear regression approach with this large set of data resulted in the following QSAR model:

$$\log K_{oc} = 1.02 + 0.52*\log K_{ow} \tag{3}$$

$$N = 390 \quad r = 0.795 \quad r^2 = 0.631 \quad s = 0.557 \quad F^{1,388} = 666 \quad p>99.9\% \quad Q^2 = 0.619.$$

Table 2. List of nonhydrophobic chemicals with their soil sorption coefficients (log K_{oc}) and reliable, measured or calculated, n-octanol-water partition coefficients (log K_{ow}).

Chemicals	log K_{oc}	log K_{ow}	Chemicals	log K_{oc}	log K_{ow}
Acetanilide	1.43	1.16	4-iso-Propyl-N-(1,1-dimethyl-2-propynyl)benzamide		
2-Chloroacetanilide	1.58	0.923		2.17	3.60
3-Methylacetanilide	1.45	1.68	Isoxaben	2.40	4.00
3-Fluoroacetanilide	1.57	1.61	Benzamide	1.46	0.64
3-Chloroacetanilide	1.86	2.15	2-Nitrobenzamide	1.45	-0.15
3-Bromoacetanilide	2.01	2.32	3-Nitrobenzamide	1.95	0.77
3-Trifluoromethylacetanilide	1.75	2.56	4-Methylbenzamide	1.78	1.18
3-Nitroacetanilide	1.94	1.47	4-Nitrobenzamide	1.93	0.82
4-Fluoroacetanilide	1.48	1.47	2-Chlorobenzamide	1.51	0.64
4-Bromoacetanilide	1.95	2.29	2,6-Dichlorobenzamide	0.53	0.77
4-Methoxyacetanilide	1.40	1.05	3,5-Dinitrobenzamide	2.31	0.83
Butyranilide	1.71	2.05	N-Methylbenzamide	1.42	0.86
Propachlor	2.42	2.18	Trichloroacetamide	0.99	1.04
3,4-Dichloroacetanilide	2.34	2.98	Diethylacetamide	1.84	0.34
3-Chloro-4-methoxyacetanilide	1.95	1.81	R-40244	2.55	3.36
Alachlor	2.28	3.52	Maleic hydrazide	0.45	-0.84
Butachlor	2.86	4.50	Bentazon	1.52	2.80
Norfluorazon	3.28	2.30	Captan	2.30	2.35
Acetchlor	2.32	2.39	Folpet	3.27	2.85
Metholachlor	2.46	3.13	Pyrazon	2.04	1.65
Metalaxyl	1.57	1.65	Captafol	3.32	3.83
Methanol	0.44	-0.77	Aniline	1.41	0.94
Ethanol	0.20	-0.31	3-Methylaniline	1.65	1.40
1-Propanol	0.48	0.25	4-Methylaniline	1.90	1.39
1-Butanol	0.50	0.88	4-Chloroaniline	1.96	1.88
1-Pentanol	0.70	1.56	4-Bromoaniline	1.96	2.26
1-Hexanol	1.01	2.03	3-Trifluoromethylaniline	2.36	2.29
1-Heptanol	1.14	2.72	3-Chloro-4-methoxyaniline	1.93	1.85
1-Octanol	1.56	3.00	3-Methyl-4-bromoaniline	2.26	2.53
1-Nonanol	1.89	4.26	2,4-Dichloroaniline	2.72	2.91
1-Decanol	2.59	4.57	2,6-Dichloroaniline	3.25	2.82
Propyleneglycol	0.36	-0.92	3,5-Dichloroaniline	2.49	2.90
sec-Phenylethylalcohol	1.50	1.36	3,4-Dichloroaniline	2.29	2.69
Diphenylmethanol	2.34	2.67	2,3,4-Trichloroaniline	2.60	3.68
Napronamide	2.62	3.36	2,3,4,5-Tetrachloroaniline	3.03	4.57
Pronamide	2.31	3.87	2,3,5,6-Tetrachloroaniline	3.94	4.46
4-Chloro-N-(1,1-dimethyl-2-propynyl)benzamide			Pentachloroaniline	4.62	5.08
	1.90	3.08	N-Methylaniline	2.28	1.66
N-(1,1-dimethyl-2-propynyl) benzamide			N,N-Dimethylaniline	2.26	2.31
	1.54	2.169	Diphenylamine	2.78	3.50
4-Methoxy-N-(1,1-dimethyl-2-propynyl)benzamide			Azobenzene	3.30	3.82
	1.83	2.37	Carbaryl	2.40	2.36
4-Methyl-N-(1,1-dimethyl-2-propynyl)benzamide			Chlorpropham	2.53	3.51
	1.76	2.67	Ethyl-N-phenylcarbamate	1.82	2.30
4-Bromo-N-(1,1-dimethyl-2-propynyl)benzamide			Propoxur	1.67	1.52
	2.01	3.23	Aldicarb	1.50	1.13
4-Fluoro-N-(1,1-dimethyl-2-propynyl)benzamide			Propham	1.83	2.60
	1.68	2.51	Oxamyl	1.00	-0.47

Table 2 continued

Chemicals	log K_{oc}	log K_{ow}	Chemicals	log K_{oc}	log K_{ow}
EPTC	2.38	3.21	Dinitramin	3.63	3.89
Pebulate	2.80	3.84	Fluchloralin	3.55	4.70
Cycloate	2.54	4.11	Butralin	3.98	5.16
Carbofuran	1.75	2.32	GS-39985	3.23	4.49
Aldicarb sulfoxide	0.56	-0.39	GS-38946	3.17	3.96
Aldicarb sulfone	0.42	-0.57	SD11830	2.37	0.86
Methyl-N-phenylcarbamate	1.73	1.76	SD12030	3.10	3.34
Propyl-N-phenylcarbamate	2.06	2.80	SD12346	3.30	3.87
Butyl-N-phenylcarbamate	2.26	3.30	SD12400	2.58	1.69
Pentyl-N-phenylcarbamate	2.61	3.80	SD12639	2.56	1.75
Methyl-N-(3-chlorophenyl)carbamate			SD13207	2.71	2.28
	2.15	2.58	3,5-Dinitroaniline	2.55	1.89
Methyl-N-(3,4-dichlorophenyl)carbamate			Diethyl phthalate	1.84	2.47
	2.74	3.54	Di-isobutyl phthalate	3.14	4.11
Chlorbufam	2.21	3.59	Di-n-butyl phthalate	3.14	4.72
Butylate	2.11	4.15	Di-n-hexyl phthalate	4.72	6.77
Methiocarb	2.32	2.92	n-Butylbenzyl phthalate	3.21	4.91
Triallate	3.35	4.53	Methyl benzoate	2.10	2.12
Meobal	1.71	2.09	Ethyl benzoate	2.30	2.64
Asulam	2.48	-0.27	Phenyl benzoate	3.16	3.59
Carbendazim	2.35	1.52	Ethyl 4-nitrobenzoate	2.48	2.33
Methomyl	1.30	0.60	Ethyl phenylacetate	1.89	2.28
Benomyl	2.71	2.12	Ethyl heptanoate	2.61	3.36
4-Chlorobenzaldoxime N-methylcarbamate			Ethyl octanoate	3.02	3.88
	2.19	2.27	Tetrachlorophthalate	3.30	3.20
Benzaldoxime N-methylcarbamate			Di-2-ethylhexyl phthalate	4.94	7.45
	1.80	1.49	Ethyl 4-hydroxybenzoate	2.21	2.47
3-Bromophenylcarbamate	1.89	2.10	Ethyl hexanoate	2.06	2.83
3-Ethylphenylcarbamate	1.66	2.12	Dimethyl phthalate	1.60	1.56
3-Methylphenylcarbamate	1.48	1.59	Ethyl pentanoate	1.97	2.30
4-Isopropylphenylcarbamate	1.94	2.52	Ethyl 3,5-dinitrobenzoate	2.74	2.13
3-Methoxyphenylcarbamate	1.44	1.18	Ethyl 4-methylbenzoate	2.59	3.14
4-Methoxyphenylcarbamate	1.40	1.18	Chloramben methyl	2.74	2.55
4-t-Butylphenylcarbamate	2.07	2.92	Fluazyfop butyl	1.76	4.50
Vernolate	2.33	3.84	Ethofumesate	2.53	2.16
Diallate	3.28	3.67	Fenvalerate	3.74	6.20
Pirimicarb	1.90	1.70	Permethrin	4.80	6.50
BPMC	1.71	2.78	Dinoseb	2.09	3.56
Molinate	1.92	3.21	Nitrobenzene	2.20	1.83
Thiobencarb	3.27	3.40	3-Aminonitrobenzene	1.73	1.37
2,6-Dinitro-trifluoro-p-toluidine			3,4-Dichloronitrobenzene	2.53	3.12
	2.56	2.65	4-Aminonitrobenzene	1.88	1.39
Trifluralin	3.93	5.34	4-Bromonitrobenzene	2.42	2.55
Benfluralin	3.99	5.29	3-Chloro-4-bromo-nitrobenzene		
2,6-Dinitro-N-n-propyl-trifluoro-p-toluidine				2.60	3.18
	3.61	3.96	2,3,4,5-Tetrachloro-nitrobenzene		
Nitralin	2.92	2.81		4.23	4.18
Oryzalin	3.40	2.79	2,3,5,6-Tetrachloro-nitrobenzene		
Chlornidine	3.94	3.82		4.05	3.89
Profluralin	4.01	5.08	Pentachloronitrobenzene	4.36	4.22

Table 2 continued

Chemicals	log K_{oc}	log K_{ow}	Chemicals	log K_{oc}	log K_{ow}
Dalapon	0.40	1.47	2-Chlorophenylurea	1.61	1.27
Fenac	1.80	3.20	Monuron	1.95	1.94
Silvex	3.28	3.80	3-(3-Fluorophenyl)-1,1-dimethylurea		
2,4,5-T acid	1.99	3.31		1.73	1.37
MCPA	1.73	2.68	Diuron	2.40	2.68
2,4-D acid	1.66	2.81	Linuron	2.70	3.20
Chloramben	1.48	1.90	3-(4-Fluorophenyl)-1,1-dimethylurea		
Anthracene-9-carboxylic acid	2.74	3.85		1.43	1.13
Dicamba	1.50	2.21	Neburon	3.40	3.80
Picloram	1.30	1.87	3-(3-Chlorophenyl)-1,1-dimethylurea		
Benazolin	1.52	1.34		1.79	2.00
6-Chloropicolinic acid	1.37	1.50	2-Fluorophenylurea	1.32	0.88
Triclorpyr	1.43	2.53	4-Fluorophenylurea	1.52	1.04
4-Hydroxybenzoic acid	1.43	1.58	3-Fluorophenylurea	1.77	1.29
4-Nitrobenzoic acid	1.54	1.89	3-Phenyl-1-cyclohexylurea	2.07	3.13
4-Methylbenzoic acid	1.77	2.27	3-Phenyl-1-methylurea	1.29	1.12
3,4-Dinitrobenzoic acid	1.53	1.64	Monolinuron	2.10	2.30
3,6-Dichlorosalicylic acid	2.30	2.79	4-Bromophenylurea	2.06	1.98
Acetic acid	0.00	-0.17	3-Chlorophenylurea	2.01	1.82
Benzoic acid	1.50	1.87	Chloroxuron	3.55	4.11
Hexanoic acid	1.46	1.92	Siduron	2.31	3.65
Phthalic acid	1.07	0.73	Fluometuron	2.00	2.42
Phenylacetic acid	1.45	1.41	3,4-Dichlorophenylurea	2.53	2.64
Phenol	1.43	1.46	3-Bromophenylurea	2.12	2.08
2,3-Dichlorophenol	2.65	2.84	Metobromuron	2.10	2.38
2,4-Dichlorophenol	2.75	3.06	3-(3,4-Dichlorophenyl)-1-methylurea		
2,4,6-Trichlorophenol	3.02	3.69		2.46	2.94
2,4,5-Trichlorophenol	3.36	3.72	3-(4-Methylphenyl)-1,1-dimethylurea		
3,4,5-Trichlorophenol	3.56	4.14		1.51	1.33
2,3,4,6-Tetrachlorophenol	3.35	4.45	3-(4-Methoxyphenyl)-1,1-dimethylurea		
Pentachlorophenol	3.73	5.12		1.40	0.83
4-Bromophenol	2.41	2.59	3-Trifluoromethylphenylurea	1.98	2.31
4-Nitrophenol	2.37	1.91	3-Phenyl-1-cyclopropylurea	1.74	1.65
2-Chlorophenol	2.60	2.15	Chlorbromuron	2.70	3.09
3-Chlorophenol	2.54	2.50	Chlortoluron	2.02	2.41
3,4-Dichlorophenol	3.09	3.33	3-Phenyl-1-cycloheptylurea	2.37	3.69
3,5-Dimethylphenol	2.83	2.35	Metoxuron	1.72	1.64
2,3,5-Trimethylphenol	3.61	2.92	3-(3-Chloro-4-methoxyphenyl)		
4-Methylphenol	2.70	1.94		1.84	1.76
2-Methoxyphenol	1.56	1.32	3-(3-Chloro-4-methylphenyl)-1-methylurea		
3-Methoxyphenol	1.50	1.58		2.10	2.61
3-Hydroxyphenol	0.98	0.80	3-Chloro-4-methoxyphenylurea		
4,5,6-Trichloroguaiacol	2.80	3.19		2.00	1.37
Tetrachloroguaiacol	2.85	3.63	3-(3-Methoxyphenyl)-1,1-dimethylurea		
Catechol	2.03	0.95		1.72	1.08
Chlorotalonil	3.26	2.90	Tebuthiuron	1.83	1.62
Dichlobenil	2.31	2.74	Diflubenzuron	3.83	3.88
3-Methylphenylurea	1.56	1.29	3-(3,5-Dimethylphenyl)-1,1-dimethylurea		
Phenylurea	1.50	0.83		1.73	1.90
Fenuron	1.40	0.98	VEL3510	2.51	1.15

Table 2 continued

Chemicals	log K_{oc}	log K_{ow}	Chemicals	log K_{oc}	log K_{ow}
Isouron	2.47	1.47	Demeton-S-methyl	1.49	1.02
Chlorsulfuron	2.19	2.14	Phosalone	2.63	4.38
Sulfometuron methyl	1.62	2.31	Dicrotophos	1.66	0.00
3-Methyl-4-fluorophenylurea	1.75	1.59	Simazine	2.10	2.18
3-Methyl-4-bromophenylurea	2.37	2.49	Propazine	2.40	2.93
3-Phenyl-1-cyclopentylurea	1.93	2.65	Ametryn	2.59	2.98
Methabenzthiazuron	2.80	1.98	Terbutryn	2.85	3.74
3-(3-Chlorophenyl)-1-methylurea			Prometon	2.60	2.99
	1.93	2.18	Atrazine	2.24	2.61
3-(3,5-Dimethyl-4-bromophenyl)-1,1-dimethylurea			Ipazine	2.91	2.87
	2.53	2.92	Trietazin	2.76	3.34
4-Phenoxyphenylurea	2.56	2.80	Dipropetryn	3.07	3.92
Azinophos methyl	2.28	2.75	Terbuthylazine	2.32	3.06
Carbophenothion	4.66	5.66	Prometryn	2.85	3.51
Carbophenothion methyl	4.67	4.82	Metribuzin	1.71	1.70
Chlorfenvinphos	2.47	3.10	Cyanazine	2.28	2.22
Chlorpyrifos	3.70	5.27	sec-Bumeton	2.78	3.20
Chlorpyrifos methyl	3.52	4.31	Metamitron	2.17	0.55
Crotoxyphos	2.00	3.30	NIA23486	1.70	1.96
Diazinon	2.75	3.81	Amitrole	1.25	-0.87
Dichlorvos	1.67	1.43	Thiabendazol	3.24	2.47
Dimthoate	1.20	0.78	Imazalil	3.73	3.82
Disulfoton	3.22	4.02	Tricyclazol	3.09	1.70
Dowco275	2.41	2.60	Triadimefon	2.71	2.77
Dyfonate	3.44	3.94	Propiconazole	3.39	3.50
EPN	3.12	4.57	Benzyltriazole	1.69	0.92
Ethion	4.06	5.07	3,4-Dichlorobenzyltriazole	2.33	2.54
Ethoprophos	1.80	3.59	4-Trifluoromethylbenzyltriazole		
Fenamiphos	2.51	3.23		1.77	1.87
Fenamiphos sulfone	1.64	0.92	4-n-Butylbenzyltriazole	2.16	2.81
Fenamiphos sulfoxide	1.57	0.95	4-Chlorobenzyltriazole	1.98	1.84
Fenitrothion	2.63	3.30	4-Fluorobenzyltriazole	1.87	1.10
Fensulfothion	2.52	2.23	4-Methoxybenzyltriazole	1.80	1.10
Fensulfothion sulfide	3.18	4.16	PP450	1.88	2.30
Fensulfothion sulfone	2.17	2.56	Oxadiazon	3.51	4.80
IBP	2.40	3.21	Bromacil	1.60	2.11
Isazophos	2.01	2.70	Isocil	2.11	2.44
Leptophos	4.50	6.31	Terbacil	1.66	1.89
Malathion	3.07	2.36	Bifenthrin	5.35	6.00
Methidathion	1.53	2.42	Methyl isothiocyanate	0.97	0.94
Mevinphos	2.30	1.20	Fluridone	2.85	3.16
Parathion	3.20	3.83	1-Butylamine	1.88	0.97
Parathion methyl	3.00	2.86	Dimethylamin	2.63	-0.38
Phorate	2.70	3.56	Biquinoline	4.02	4.31
Piperophos	3.44	4.92	Phenazine	3.37	2.84
Profenophos	3.03	4.68	2,3,7,8-Tetrachlorodibenzodioxine		
Terbufos	2.82	4.48		6.50	6.42
Terbufos sulfone	2.18	2.48	Anisole	1.54	2.11
Terbufos sulfoxide	2.18	2.21	Methoxychlor	4.90	5.08
Trichlorfon	1.90	0.51	1,2-Dimethoxybenzene	2.03	1.60

Table 2 continued

Chemicals	$\log K_{oc}$	$\log K_{ow}$	Chemicals	$\log K_{oc}$	$\log K_{ow}$
Diphenylether	3.29	4.21	3,5,6-Trichloropyridin-2-ol	2.11	3.21
1,4-Dimethoxy-2,5-dichlorobenzene			1-Naphthol	3.33	2.84
	3.10	3.34	5-Indanol	3.40	2.49
Bis(2-chloroisopropyl) ether	1.67	1.614	Dieldrin	4.55	5.20
Bis(2-chloroethyl) ether	1.88	1.29	Nitrapyrin	2.62	3.41
Benzophenone	2.63	3.18	Urea	0.90	-2.11
4-Acetylbiphenyl	3.22	3.47	Ethenethiourea	2.07	-0.66
2-Acetonaphthalene	2.93	2.75	Methylurea	1.76	-1.40
Acetophenone	1.54	1.58	Endosulfan	4.13	3.83
2-Aminoanthracene	4.45	3.26	Pentafluorophenyl methyl sulfone		
6-Aminochrysene	5.21	4.98		1.46	1.40
1-Aminonaphthalene	3.51	2.25	Cinmethylin	2.60	3.83
Benzidine	3.46	1.34	Dimethrimol	2.30	3.70
3,3'-Dichlorobenzidine	4.35	3.51			

Nine chemicals (asulam, fluazyfop butyl, dimethylamine, 2,3,7,8-tetrachlorodibenzodioxin, 2-aminoanthracene, 6-aminochrysene, benzidine, 3,3'-dichlorobenzidine, and methylurea) were outliers with residuals between 1.5 and 2.1 log units. Outliers constitute only 2% of the modelling set and their removal from the regression analysis resulted in a very similar model. Thus, the influence of outliers on the regression model is negligible and they are not excluded from this model (Eq. 3.). All other estimated soil sorption coefficients are within ± 1.4 log units of the measured values. The plot of residuals vs. the $\log K_{ow}$ data shows that the precision of estimates is uniform across the whole range of independent variable (see Fig. 5). This model exhibits a large standard error of the estimate but it has a clearly defined application domain. The only limitation in the application of this model is that the $\log K_{ow}$ values of chemicals must be in the range from -2 to 8.
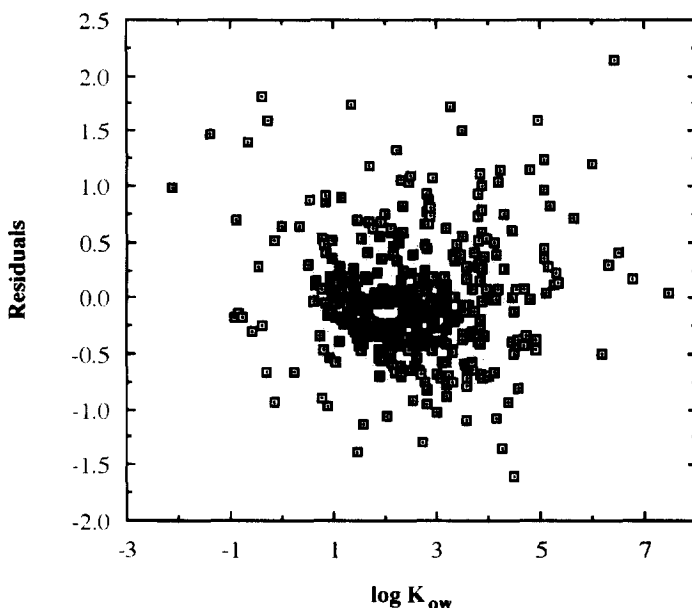

Figure 5. Plot of residuals vs. the n-octanol/water partition coefficients for Eq. 3.

For this model, the analysis of residuals has been performed on a chemical class by class basis. Results of this analysis are that the soil sorption coefficients are overestimated for n-alkyl alcohols (0.9 log unit) and organic acids (0.55 log units), while a systematic underestimation is found for the log K_{oc} data of amino-PAHs and aliphatic amines (1-2 log units) and alkyl ureas (1-1.5 log units). Thus, caution should be exercised when estimating the log K_{oc} values for these classes of chemicals.
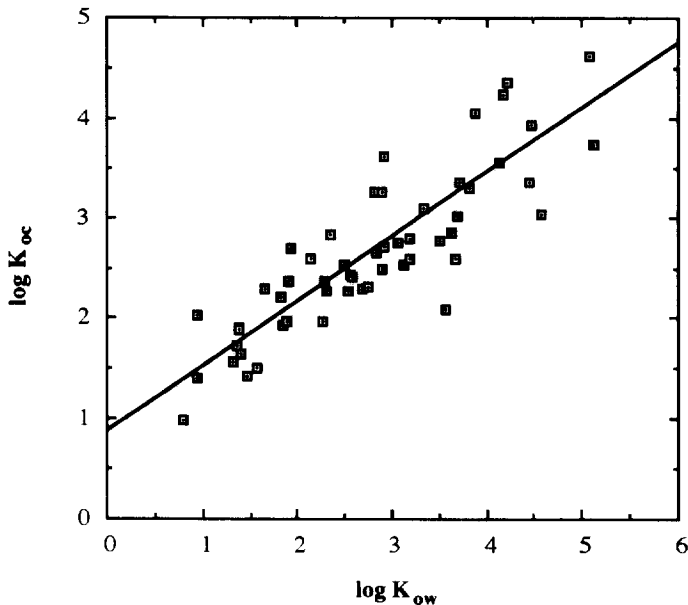

Figure 6. Plot of log K_{oc} vs. log K_{ow} for phenol type chemicals. The dashed line represents QSAR model described by Eq. 4.
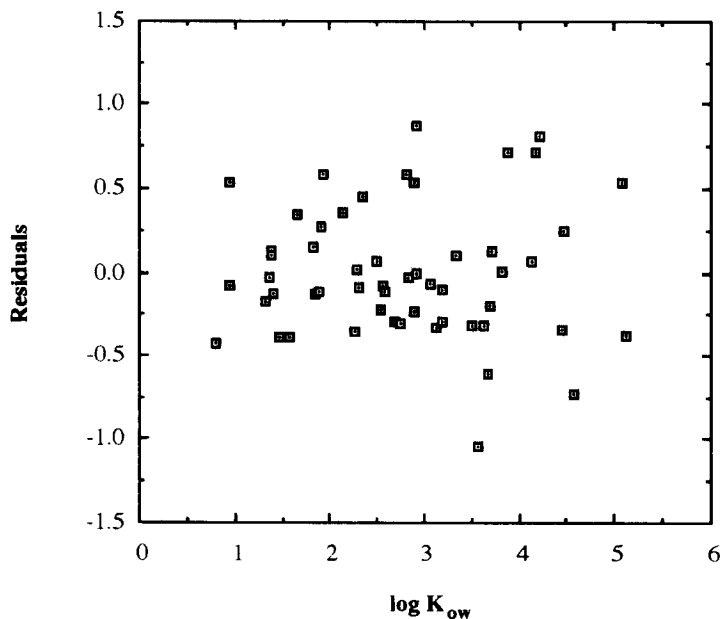
Figure 7. Plot of residuals vs. the n-octanol/water partition coefficients for Eq. 4.

SUBGENERAL SOIL SORPTION MODELS

Phenol type chemicals. A graphical analysis of the plots log K_{oc} vs. log K_{ow} has shown that substituted phenols and a similar type of chemicals display specific trends (see Fig. 6). Thus, the subgroup of chemicals, which will be called the phenol type chemicals, are substituted phenols (Cl, Br, CH_3, OH, NO_2, CH_3O), substituted anilines (Cl, Br, CH_3, CF_3, CH_3O, N-Me), chlorinated benzonitriles, and substituted nitrobenzenes (Cl, Br, NH_2). Appropriate log K_{oc} and log K_{ow} values were available for 54 phenol type chemicals. The linear regression approach with these data resulted in the following QSAR model:

$$\log K_{oc} = 0.90 + 0.63 * \log K_{ow} \qquad (4)$$
$$N = 54 \quad r = 0.865 \quad r^2 = 0.744 \quad s = 0.401 \quad F^{1,52} = 155 \quad p > 99.9\% \quad Q^2 = 0.736.$$

Dinoseb, a nitrobenzene derivative, was a marginal outlier with residual of -1.04 log units and it is not excluded from this model (Eq. 4). All other estimated soil sorption coefficients are within ±0.88 log units of the measured values. The plot of residuals vs. the log K_{ow} data shows that the precision of estimates is uniform across the whole range of independent variable (see Fig. 7). This model has a low standard error of the estimate and it can be applied to phenol type chemicals with log K_{ow} values in the range from 1 to 5.

Agricultural chemicals. A graphical analysis of the plots log K_{oc} vs. log K_{ow} has shown that the largest chemical cluster include the following chemical classes: acetanilides, carbamates, esters, phenylureas, phosphates, triazines, triazoles, and uracils (see Fig. 8). We will call these chemical classes agricultural chemicals. The log K_{oc} and log K_{ow} values are available for 216 agricultural chemicals. The linear regression approach using these data resulted in the following QSAR model:

$$\log K_{oc} = 1.09 + 0.47 * \log K_{ow} \qquad (5)$$
$$N = 216 \quad r = 0.826 \quad r^2 = 0.681 \quad s = 0.425 \quad F^{1,214} = 460 \quad p > 99.9\% \quad Q^2 = 0.670.$$
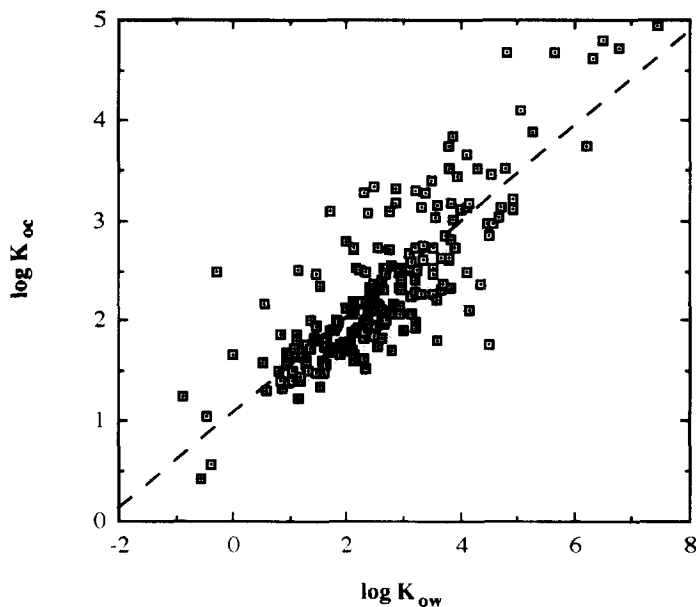

Figure 8. Plot of log K_{oc} vs. log K_{ow} for agricultural chemicals. The dashed line represents QSAR model described by Eq. 5.

Four chemicals (asulam, fluazyfop butyl, carbophenthion methyl, and tricyclazol) were outliers with residuals between 1.2 and 1.5 log units. Outliers constitute less than 2% of the modelling set and their removal from the

regression analysis resulted in a very similar model. Thus, the influence of outliers on the regression model is negligible and they are not excluded from this model (Eq. 5). All other estimated soil sorption coefficients are within ±1 log unit of the measured values. The plot of residuals vs. the log K_{ow} data shows that the precision of estimates is uniform across the whole range of independent variable (see Fig. 9).
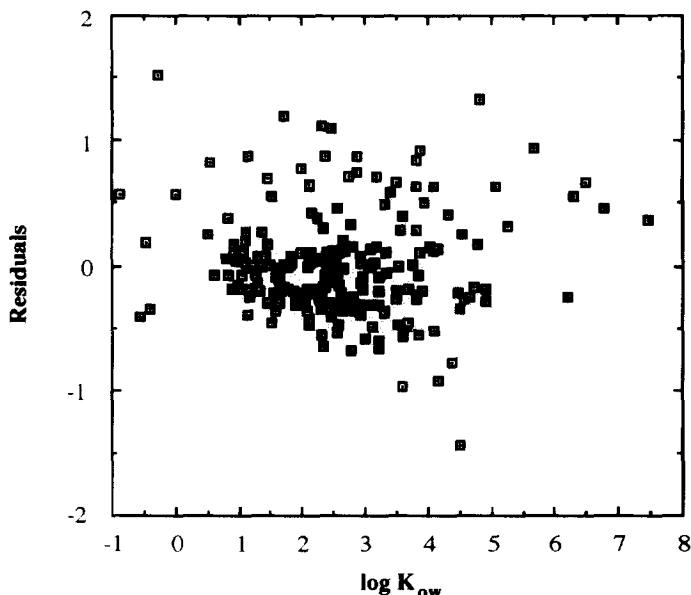

Figure 9. Plot of residuals vs. the n-octanol/water partition coefficients for Eq. 5.

An analysis of residuals was also performed for this model on a chemical class by class basis. Results of this analysis show that the soil sorption coefficients have not been systematically overestimation or underestimation of for any class of agricultural chemicals. This model exhibits a relatively low standard error of the estimate and it can be applied to the agricultural chemicals listed above. The only limitation in application of this model is that the log K_{ow} value of any particular chemical must be in the range from -1 to 8.

Alcohols and organic acids. A graphical analysis of the plots log K_{oc} vs. log K_{ow} has shown that alcohols and organic acids are located predominantly below all other chemical classes and on the average those chemicals have the smallest soil sorption coefficients. Thus, it seems to be worth to establish QSAR model for these two chemical classes alone (see Fig. 10). The linear regression with 36 alcohols and organic acids resulted in the following QSAR model:

$$\log K_{oc} = 0.50 + 0.47*\log K_{ow} \tag{6}$$
$$N = 36 \quad r = 0.850 \quad r^2 = 0.715 \quad s = 0.388 \quad F^{1,34} = 89 \quad p>99.9\% \quad Q^2 = 0.665.$$

Silvex, an organic acid, is a marginal outlier with a residual of 0.98 log units and it is not excluded from this model (Eq. 6). All other estimated soil sorption coefficients are within ±0.8 log units of the measured values and the precision of estimates is uniform across the whole range of independent variable (see Fig. 11). This model has a low standard error of the estimate and it can be applied to alcohols and organic acids with log K_{ow} values in the range from -1 to 5. It seems that this model is more appropriate for these two chemical classes than the general QSAR model for nonhydrophobic chemicals.
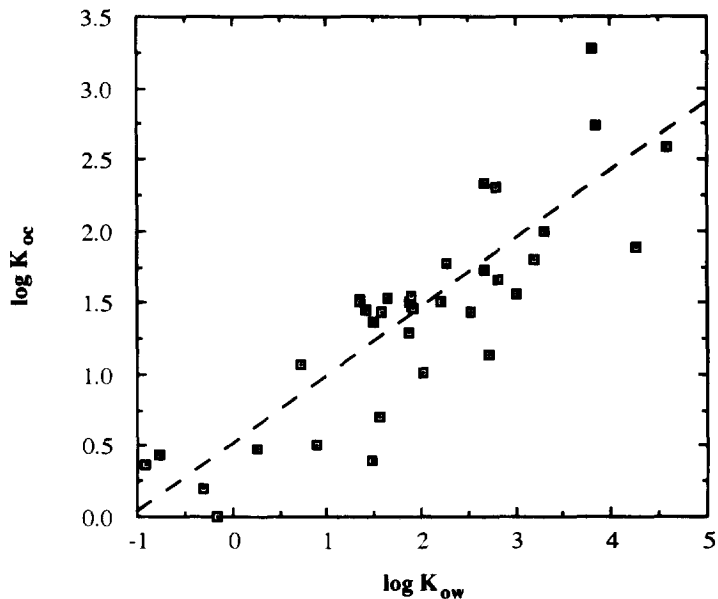
Figure 10. Plot of log K_{oc} vs. log K_{ow} for alcohols and organic acids. The dashed line represents QSAR model described by Eq. 6.
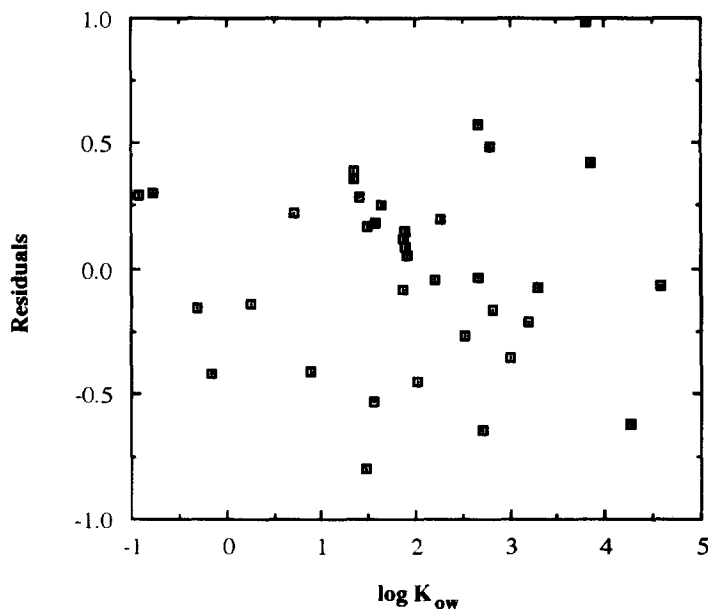

Figure 11. Plot of residuals vs. the n-octanol/water partition coefficients for Eq. 6.

CLASS-SPECIFIC SOIL SORPTION MODELS

Finally, an effort was made to calculate class-specific QSAR models for all nonhydrophobic classes of chemicals. Altogether, 14 class-specific QSAR models have been produced and they are all listed below. Some of these models will be discussed in the following section.

Acetanilides $\log K_{oc} = 1.12 + 0.40 * \log K_{ow}$ (7)
$$N = 21 \quad r = 0.719 \quad r^2 = 0.491 \quad s = 0.339 \quad F^{1,19} = 20 \quad p > 99.9\% \quad Q^2 = 0.436$$

Alcohols $\log K_{oc} = 0.50 + 0.39 * \log K_{ow}$ (8)
$N = 13$ $r = 0.876$ $r^2 = 0.747$ $s = 0.397$ $F^{1,11} = 36$ $p > 99.9\%$ $Q^2 = 0.718$

Amides $\log K_{oc} = 1.25 + 0.33 * \log K_{ow}$ (9)
$N = 28$ $r = 0.679$ $r^2 = 0.440$ $s = 0.491$ $F^{1,26} = 22$ $p > 99.9\%$ $Q^2 = 0.385$

Anilines $\log K_{oc} = 0.85 + 0.62 * \log K_{ow}$ (10)
$N = 20$ $r = 0.905$ $r^2 = 0.808$ $s = 0.341$ $F^{1,18} = 81$ $p > 99.9\%$ $Q^2 = 0.804$

Carbamates $\log K_{oc} = 1.14 + 0.365 * \log K_{ow}$ (11)
$N = 43$ $r = 0.760$ $r^2 = 0.568$ $s = 0.408$ $F^{1,41} = 56$ $p > 99.9\%$ $Q^2 = 0.492$

Dinitroanilines $\log K_{oc} = 1.92 + 0.38 * \log K_{ow}$ (12)
$N = 20$ $r = 0.909$ $r^2 = 0.817$ $s = 0.242$ $F^{1,18} = 86$ $p > 99.9\%$ $Q^2 = 0.787$

Esters $\log K_{oc} = 1.05 + 0.49 * \log K_{ow}$ (13)
$N = 25$ $r = 0.874$ $r^2 = 0.753$ $s = 0.463$ $F^{1,23} = 74$ $p > 99.9\%$ $Q^2 = 0.736$

Nitrobenzenes $\log K_{oc} = 0.55 + 0.77 * \log K_{ow}$ (14)
$N = 10$ $r = 0.839$ $r^2 = 0.666$ $s = 0.583$ $F^{1,8} = 19$ $p > 99\%$ $Q^2 = 0.567$

Organic acids $\log K_{oc} = 0.32 + 0.60 * \log K_{ow}$ (15)
$N = 23$ $r = 0.865$ $r^2 = 0.736$ $s = 0.336$ $F^{1,21} = 62$ $p > 99.9\%$ $Q^2 = 0.682$

Phenols & Benzonitriles $\log K_{oc} = 1.08 + 0.57 * \log K_{ow}$ (16)
$N = 24$ $r = 0.865$ $r^2 = 0.737$ $s = 0.373$ $F^{1,22} = 65$ $p > 99.9\%$ $Q^2 = 0.699$

Phenylureas $\log K_{oc} = 1.05 + 0.49 * \log K_{ow}$ (17)
$N = 52$ $r = 0.790$ $r^2 = 0.616$ $s = 0.335$ $F^{1,50} = 83$ $p > 99.9\%$ $Q^2 = 0.540$

Phosphates $\log K_{oc} = 1.17 + 0.49 * \log K_{ow}$ (18)
$N = 41$ $r = 0.856$ $r^2 = 0.726$ $s = 0.452$ $F^{1,39} = 107$ $p > 99.9\%$ $Q^2 = 0.700$

Triazines $\log K_{oc} = 1.50 + 0.30 * \log K_{ow}$ (19)
$N = 16$ $r = 0.567$ $r^2 = 0.273$ $s = 0.379$ $F^{1,14} = 6.6$ $p > 97\%$ $Q^2 = 0.045$

Triazoles $\log K_{oc} = 1.405 + 0.47 * \log K_{ow}$ (20)
$N = 15$ $r = 0.811$ $r^2 = 0.631$ $s = 0.482$ $F^{1,13} = 25$ $p > 99.9\%$ $Q^2 = 0.573$

4. DISCUSSION

First, we will address the question of outliers. The majority of outliers (dimethylamine, 2-aminoanthracene, 6-aminochrysene, benzidine, 3,3'-dichlorobenzidine, and methylurea) are members of chemical groups (alkyl ureas, alkyl amines, and amino-PAHs) whose $\log K_{oc}$ data are systematically underestimated by the general QSAR model for nonhydrophobic chemicals (Eq. 3). This subject will be discussed in the following paragraphs. Some outliers (asulam and fluazyfop butyl) appear in two models. Thus, it seems that there is a large uncertainty in their $\log K_{oc}$ data since we have applied only reliable $\log K_{ow}$ data. The possibility cannot be excluded that these chemicals can specifically interact with the various constituents of the soil [60].

There are several chemical groups (alkyl ureas, amino-PAHs, alkyl amines, and azo-PAHs) whose $\log K_{oc}$ data are systematically underestimated by the general QSAR model. This result indicates that these chemical groups establish specific bindings with appropriate soil constituents [60] which is demonstrated by the larger soil sorption coefficients than can be expected from their n-octanol/water partition coefficients.

The log K_{oc} data of alcohols and organic acids are systematically overestimated by the general QSAR model. This result supports the idea that these two chemical classes should form a separate model. The reason for this behaviour is not clear.

Two class-specific models, for amides and dinitroanilines, deserve special attention. The QSAR model for amides (Eq. 9) has a very large standard error of the estimates, and the plot of log K_{oc} vs. log K_{ow} data (Fig. 12) has shown that data points are very scattered in this diagram (up to 2 log units). Thus, we cannot expect that estimates of the soil sorption coefficients of amides, made either with the class-specific model or with the general QSAR model, will be very accurate. On the other hand, dinitroanilines, on the average, have larger soil sorption coefficients than any other chemical class evaluated. This is emphasised by the very large intercept (1.92) of this class-specific model. Such a specific behaviour can be explained by a very strong binding interaction of amino and/or nitro groups with appropriate soil constituents [60]. However, it must be pointed out that only 3 out of 20 log K_{ow} data have been measured and all the others are calculated values. Thus, it is also possible that log K_{ow} values calculated for dinitroanilines are systematically inaccurate.
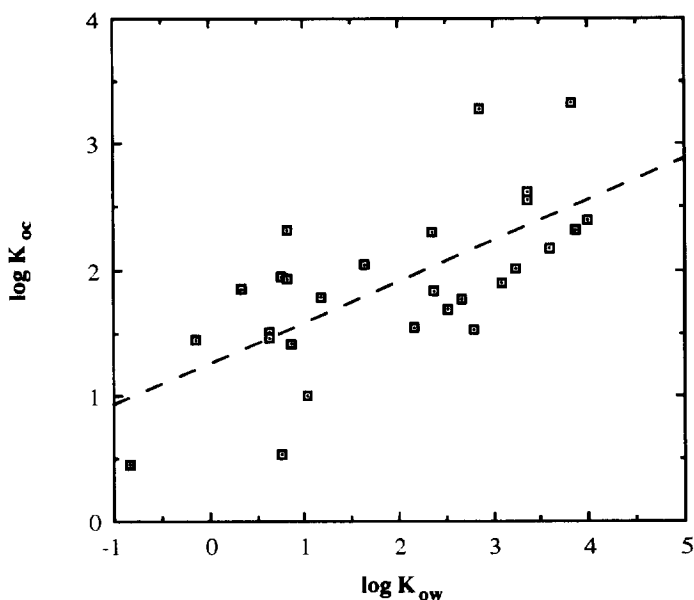

Figure 12. Plot of log K_{oc} vs. log K_{ow} for amides showing the large scatter in data. The dashed line represents the model for amides described by Eq. 9.

Specific interactions with soils and sediments (hydrogen bonding, dipole interactions, charge transfer, etc.) which are feasible for chemicals such as alkyl ureas, amines, alcohols, organic acids, amides and dinitroanilines cannot be adequately described by n-octanol/water partition coefficients alone since this global molecular descriptor accounts primarily for nonspecific interactions which are results of dispersive forces. Thus, in order to improve the quality of estimates for these chemicals classes, other molecular descriptors which reflect more specific interactions must be used in addition to n-octanol/water partition coefficients. However, the results of this systematic study are a solid basis for developing more precise models for specifically interacting chemicals and such effort is in progress.

Now, we will discuss the two most important points for application of derived QSAR models: (a) Which of these QSAR models can be recommended for application and reliable estimation of soil sorption

coefficients and (b) how to select the most appropriate model for some specific cases? There is no direct answer to or simple recipe for these questions. In principle, we can recommend for application all QSAR models above, but the selection of the most appropriate model will depend on the required accuracy of estimates and on the structure and complexity of chemical(s) for which estimations must be made.

Table 3. Decision making tree for the system of QSAR soil sorption models.

ORGANIC CHEMICAL ? — NO → No **QSAR** Available

YES

Salt, Surfactant, Organometallic, or Ionized Chemical ? — YES → No **QSAR** Available

NO

Hydrophobic Chemical with C, H, and Halogens? — YES → Hydrophobic Model Eq. 1

NO

Can a Chemical be Clearly Defined as a Member of Any of 14 Chemical Classes by
a) X-variable domain and
b) chemical domain and
c) substituent domain
— YES → Use Appropriate Class-Specific **QSAR** Model Eqs. 7-20

NO

Alcohol or Organic Acid? — YES → Alcohols and Organic Acids Model - Eq. 6

NO

Phenol, Aniline, Nitrobenzene or Benzonitrile? — YES → Phenols Model Eq. 4

NO

Agricultural Chemical ? — YES → Agricultural Model - Eq. 5

NO

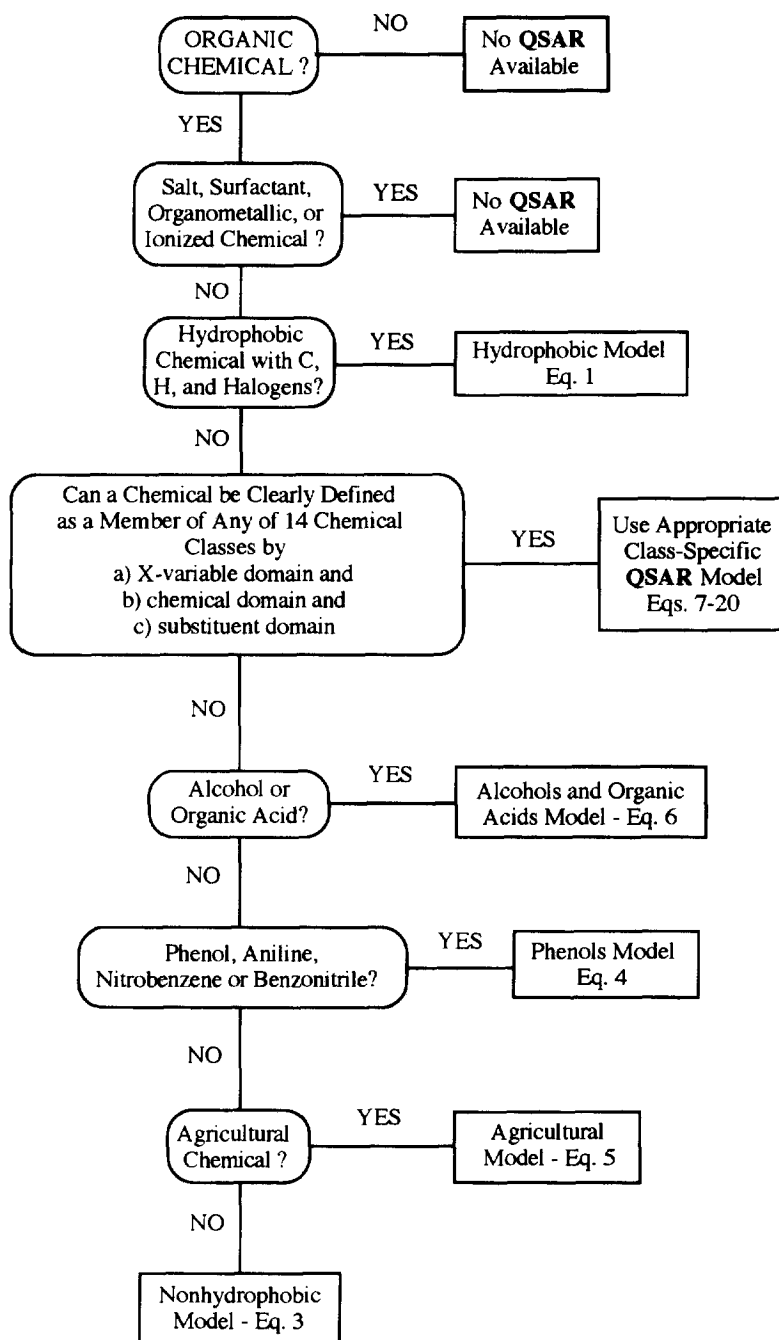Nonhydrophobic Model - Eq. 3

Table 4. List of all derived QSAR models for soil sorption and their main characteristics: the application domains and quality of estimates.

MODEL	ACCURACY ($\pm 2\sigma$ range = 95%)	X-VARIABLE DOMAIN (log K_{ow} in log units)	CHEMICAL DOMAIN	SUBSTITUENTS DOMAIN or WARNINGS
Hydrophobic (Eq. 1)	0.54 log units	3 - 22 Carbon or halogens	All chemicals with C, H, F, Cl, Br, and I atoms	
Hydrophobic (Eq. 2)	0.83 log units*	1 - 7.5	All chemicals with C, H, F, Cl, Br, and I atoms	
Nonhydrophobic (Eq. 3)	1.0 log units	(-2.0) - 8.0	All chemicals not classified as hydrophobics	**Overestimated** n-alkyl alcohols (0.9 log units), organic acids (0.55 log units) **Underestimated** amino-PAHs (1-2 log units), aliphatic amines (1-2 log units), alkyl ureas (1.0-1.5 log units)
Phenols (Eq. 4)	0.76 log units	1.0 - 5.0	Phenols Anilines Benzonitriles Nitrobenzenes	Cl, Br, CH_3, OH, NO_2, CH_3O Cl, Br, CH_3, CF_3, CH_3O, N-Me Cl Cl, Br, NH_2
Agricultural (Eq. 5)	0.77 log units	(-1.0) - 8.0	Acetanilides, carbamates, esters, phenylureas, phosphates, triazines, triazoles, uracils	
Alcohols & acids (Eq. 6)	0.78 log units	(-1.0) - 5.0	Alcohols Organic acids	Alkyl, phenalkyl, OH All
Acetanilides	0.35 log units	0.9 - 5.0	Anilides	CH_3O, Cl, Br, NO_2, CF_3, CH_3
Alcohols	0.80 log units	(-1.0) - 5.0	Alcohols	Alkyl, phenalkyl, OH
Amides	0.98 log units	(-1.0) - 4.0	Acetamides Benzamides	F, Cl, Br, CH_3O, Alkyl NO_2, N-Me

Table 4. (continued)

MODEL	ACCURACY (±2σ range = 95%)	X-VARIABLE DOMAIN (log K_{ow} in log units)	CHEMICAL DOMAIN	SUBSTITUENTS DOMAIN or WARNINGS
Anilines	0.68 log units	1.0 - 5.1	Anilines	Cl, Br, CF_3, CH_3, N-Me, N,N-di-Me
Carbamates	0.67 log units	(-1.0) - 5.0	Carbamates	Alkyl, Alkenyl, Cl, Br, N-Me, CH_3O
Dinitroanilines	0.48 log units	0.5 - 5.5	Dinitroanilines	CF_3, Alkyl-SO_2, NH_2SO_2, CH_3, t-Bu
Esters	0.68 log units	1.0 - 8.0	Phthalates Benzoates Phenylacetates Hexanoates, heptanoates Octanoates	Alkyl, phenyl, Cl Alkyl, phenyl, NO_2, OH, Cl, NH_2 Alkyl, phenalkyl Alkyl Alkyl
Nitrobenzenes	0.77 log units	1.0 - 4.5	Nitrobenzenes	Cl, Br, NH_2
Organic acids	0.67 log units	(-0.5) - 4.0	Organic acids	All
Phenols	0.75 log units	0.5 - 5.5	Phenols Benzonitriles	Cl, Br, NO_2, CH_3, CH_3O, OH Cl
Phenylureas	0.48 log units	0.5 - 4.2	Phenylureas	CH_3, CH_3O, F, Cl, Br, Cyc-alkyls, CF_3, PhO
Phosphates	0.76 log units	0.0 - 6.5	All phosphates	
Triazines	0.38 log units	1.5 - 4.0	Triazines	Cl, CH_3O, CH_3S, NH_2, N-Alkyl
Triazoles	0.96 log units	(-1.0) - 5.0	Triazoles	Alkyl, CH_3O, F, Cl, CF_3, NH_2

* The precision of the estimates is higher for the less hydrophobic chemicals and lower for the more hydrophobic chemicals. For chemicals with log K_{ow} data from 1 to 4 the spread of residuals is from 0.2 to 0.5 log units and for the chemicals with log K_{ow} data from 4 to 7.5 the spread of residuals is from 0.6 to 1.0 log units.

Table 5. List of all derived QSAR models for soil sorption with their chemical domains.

Model #	Regression Equation	Chemical Domain
1	$\log K_{oc} = 0.70 + 0.52*{}^{1}\chi$	Predominantly hydrophobics
2	$\log K_{oc} = 0.10 + 0.81*\log K_{ow}$	Predominantly hydrophobics
3	$\log K_{oc} = 1.02 + 0.52*\log K_{ow}$	Nonhydrophobics
4	$\log K_{oc} = 0.90 + 0.63*\log K_{ow}$	Phenols, anilines, benzonitriles, & nitrobenzenes
5	$\log K_{oc} = 1.09 + 0.47*\log K_{ow}$	Acetanilides, carbamates, esters, phenylureas, phosphates, triazines, triazoles, & uracils
6	$\log K_{oc} = 0.50 + 0.47*\log K_{ow}$	Alcohols & organic acids
7	$\log K_{oc} = 1.12 + 0.40*\log K_{ow}$	Acetanilides
8	$\log K_{oc} = 0.50 + 0.39*\log K_{ow}$	Alcohols
9	$\log K_{oc} = 1.25 + 0.33*\log K_{ow}$	Amides
10	$\log K_{oc} = 0.85 + 0.62*\log K_{ow}$	Anilines
11	$\log K_{oc} = 1.14 + 0.365*\log K_{ow}$	Carbamates
12	$\log K_{oc} = 1.92 + 0.38*\log K_{ow}$	Dinitroanilines
13	$\log K_{oc} = 1.05 + 0.49*\log K_{ow}$	Esters
14	$\log K_{oc} = 0.55 + 0.77*\log K_{ow}$	Nitrobenzenes
15	$\log K_{oc} = 0.32 + 0.60*\log K_{ow}$	Organic acids
16	$\log K_{oc} = 1.08 + 0.57*\log K_{ow}$	Phenols & benzonitriles
17	$\log K_{oc} = 1.05 + 0.49*\log K_{ow}$	Phenylureas
18	$\log K_{oc} = 1.17 + 0.49*\log K_{ow}$	Phosphates
19	$\log K_{oc} = 1.50 + 0.30*\log K_{ow}$	Triazines
20	$\log K_{oc} = 1.405 + 0.47*\log K_{ow}$	Triazoles

In order to help prospective users (QSAR practitioners) to successfully apply the derived system of QSAR models, we have designed a flowchart or decision tree (Tab. 3) to select the appropriate model. In addition, an appendix has been prepared (Tabs. 4 an 5) that summarises the complete results and information on derived QSAR models. Table 4 contains information on the accuracy, chemical domain, X-variable domain, and substituents domain for all 20 QSAR models. In addition, specific warnings are given in Table 4 for cases in which soil sorption coefficients are systematically overestimated or underestimated for certain chemical classes. Table 5 lists all 20 QSAR models developed in this study together with their chemical domains.

The strategy for selecting the most appropriate QSAR model is designed at two stages. At the first stage, an effort is made to pinpoint, within the strict rules, to the more specific model with high accuracy for a particular chemical. If this is not feasible, a selection is made at the second stage among more general models in order to select the most appropriate one. Each QSAR model is uniquely defined by its chemical domain, substituents domain, and X-variable domain. Thus, in order to make a reasonable estimate of the soil sorption coefficient, the characteristics of a chemical must comply with the application domain of a particular QSAR model. To obtain the best possible estimation of soil sorption coefficients from the derived system of QSAR models (i.e. the log K_{ow} data) as well as a correct information about its accuracy, these rules or structural and chemical requirements must be strictly obeyed.

As a general conclusion, it is worth noting that the more specific models are more accurate with a narrower application domain. On the other hand, the more general models are less accurate, but their application domains are much wider. Thus, for simple chemicals it should be possible to obtain accurate estimates while for the complex chemicals, those with many functional groups and large in size, we should not aim at achieving a very high accuracy, but we should rather make a reasonable estimate with a greater uncertainty than a very accurate result that may be wrong.

Acknowledgment

5. REFERENCES

1. Pesticides in the Soil Environment: Processes, Impact, and Modeling, H.H. Cheng, Editor, Soil Science Society of America, Inc., Madison, Wisconsin, 1990.
2. Commission of the European Communities, Council Regulation on the Evaluation and Control of the Environmental Risk of Existing Substances, European Communities, Brussels, 1992.
3. G.G. Briggs (1981). Theoretical and Experimental Relationships Between Soil Adsorption, Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors, and the Parachor. J. Agric. Food Chem. **29**, 1050-1059.
4. E.E. Kenaga, C.A.I. Goring (1980). Relationship Between Water Solubility, Soil Sorption, Octanol-Water Partitioning, and Concentration of Chemicals in Biota. In J.C. Eaton, P.R. Parrish, A.C. Hendricks, eds., Aquatic Toxicology. STP 707, (Third Symposium).American Society for Testing and Materials, Philadelphia (PA), pp. 78-115.

5. Z. Gerstl (1990). Estimation of organic chemical sorption by soils. J. Con. Hydrology **6**, 357-375.

6. W. Meylan, Ph.H. Howard, and R.S. Boethling (1992). Molecular topology/fragment contribution method for predicting soil sorption coefficients. Environ. Sci. Technol. **26**, 1560-1567.

7. EINECS - European Inventory of Existing Commercial Chemical Substances, Advanced edition, Commission of the EC, Luxemburg, Office of Official Publications of the European Communities, 1987 (8 volumes).

8. J.S. Means, S.G. Wood, J.J. Hassett, W.L. Banwart (1982). Sorption of Amino- and Carboxy-Substituted Polynuclear Aromatic Hydrocarbons by Sediments and Soils, Environ. Sci. Technol. **16**, 93-98.

9. A. Sabljić (1987). On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. Environ. Sci. Technol. **21**, 358-366.

10. J.H. Park, H.J. Lee (1993). Estimation of Bioconcentration Factor in Fish, Adsorption Coefficient for Soils and Sediments and Interfacial Tension with Water for Organic Nonelectrolytes Based on the Linear Solvation Energy Relationships. Chemosphere **26**, 1905-1916.

11. A. Sabljić (1989). Quantitative Modelling of Soil Sorption for Xenobiotic Chemicals. Environ. Health Perspect. **83**, 179-190.

12. A.S. Abdul, T.L. Gibson, and D.N. Rai (1987). Statistical Correlations for Predicting Partition Coefficient for Nonpolar Organic Contaminants Between Aquifer Organic Carbon and Water. Hazardous Waste and Hazardous Materials **4**, 211-222.

13. S.W. Karickhoff (1981). Semi-Empirical Estimation of Sorption of Hydrophobic Pollutants on Natural Sediments and Soils. Chemosphere, **10**, 833-846.

14. S.W. Karickhoff, D.S. Brown, T.A. Scott (1979). Sorption of Hydrophobic Pollutants on Natural Sediments. Water Res. **13**, 241-248.

15. S. Okouchi, H. Saegusa (1989). Prediction of Soil Sorption Coefficients of Hydrophobic Organic Pollutants by Adsorbability Index. Bull. Chem. Soc. Jpn. **62**, 922-924.

16. D.A. Bahnick, W.J. Doucette (1988). Use of Molecular Connectivity Indices to Estimate Soil Sorption Coefficients for Organic Chemicals. Chemosphere **17**, 1703-1715.

17. Y-P. Chin, C.S. Peven, W.J. Weber Jr (1988). Estimating Soil/Sediment Partition Coefficients for Organic Compounds by High Performance Reverse Phase Liquid Chromatography. Wat. Res. **22**, 873-881.

18. J. Hodson, N.A. Williams (1988). The Estimation of the Adsorption Coefficients (K_{oc}) for Soils by High Performance Liquid Chromatography. Chemosphere **17**, 66-77.

19. L. Pussemier, R. De Borger, P. Cloos, R. Van Bladel (1989). Relation Between the Molecular Structure and the Adsorption of Arylcarbamate, Phenylurea and Anilide Pesticides in Soil and Model Organic Adsorbent. Chemosphere **18**, 1871-1882.

20. G. Szabo, S.L. Prosser, R.A. Bulman (1990). Determination of the Adsorption Coefficient (K_{oc}) of Some Aromatics for Soil by RP-HPLC on Two Immobilized Humic Acid Phases. Chemosphere **21**, 777-788.

21. L. Pussemier, G. Szabo, R.A. Bulman (1990). Prediction of the Soil Adsorption Coefficients K_{oc} for Aromatic Pollutants. Chemosphere **21**, 1199-1212.

22. R.J. Hance (1969). An Empirical Relationship Between Chemical Structure and the Sorption of Some Herbicides by Soil. J. Agric. Food Chem. **17**, 667-668.

23. K. Kawamoto, K. Urano (1989). Parameters for Predicting Fate of Organochlorine Pesticides in the Environment. (II) Adsorption Constant to Soil. Chemosphere 19, 1223-1231.

24. J. Kanazawa (1989). Relationship Between the Soil Sorption Constants for Pesticides and Their Physicochemical Properties. Environ. Toxicol. Chem. 8, 477-484.

25. P.S.C. Rao, J.M. Davidson (1980). Estimation of Pesticide Retention and Transformation Parameters Required in Nonpoint Source Pollution Models. In: Environmental Nonpoint Source Pollution. Ann Arbor Sci. Publ. Inc., Mich., pp. 23-67.

26. J.J. Hasset , J.C. Means, W. L. Banwart, S. G. Wood (1980). Sorption Properties of Sediments and Energy-Related Pollutants. EPA-600/3-80-041, pp. 1-133.

27. P.J. McCall, R.L. Swann, D.A. Laskowski, S.M. Unger, S.A. Vrona, H.J. Dishburger (1980). Estimation of Chemical Mobility in Soil from Liquid Chromatographic Retention Times. Bull. Environ. Contam. Toxicol. 24, 190-195.

28. R.P. Schwarzenbach, J. Westall (1981). Transport of Nonpolar Organic Compounds from Surface Water to Groundwater. Laboratory Sorption Studies. Environ. Sci Technol. 15, 1360-1367.

29. C.T. Chiou, L.J. Peters, V.H. Freed (1979) A Physical Concept of Soil-Water Equilibria for Nonionic Organic Compounds. Science 206, 831-832.

30. A. Sabljić, M. Protić (1982). Relationships Between Molecular Connectivity Indices and Soil Sorption Coefficients of Polycyclic Aromatic Hydrocarbons. Bull. Environ. Contam. Toxicol. 28, 162-165.

31. U. Mingelgrin, Z. Gerstl (1983). Reevaluation of Partitioning as a Mechanism of Nonionic Chemicals Adsorption in Soils. J. Environ. Qual. 12, 1-11.

32. C.T. Chiou, P.E. Porter, D.W. Schmedding (1983). Partition Equilibria of Nonionic Organic Compounds Between Soil Organic Matter and Water. Environ. Sci. Technol. 17, 227-231.

33. K. Schellenberg, C. Leuenberger, R.P. Schwarzenbach (1984) Sorption of Chlorinated Phenols by Natural Sediments and Aquifer Materials. Environ. Sci. Technol. 18, 652-657.

34. A. Sabljić (1984). Predictions of the Nature and Strength of Soil Sorption of Organic Pollutants by Molecular Topology. J. Agric. Food Chem. 32, 243-246.

35. G. Szabo, G. Farkas, R.A. Bulman (1992). Evaluation of Silica-Humate and Alumina-Humate HPLC Stationary Phases for Estimation of the Adsorption Coefficient, K_{oc}, of Soil for Some Aromatics. Chemosphere 24, 403-412.

36. S.A. Boyd (1982). Adsorption of Substituted Phenols by Soil. Soil Sci. 134, 337-343.

37. Z. Gerstl, C.S. Helling (1987). Evaluation of Molecular Connectivity as a Predictive Method for the Adsorption of Pesticides by Soils. Environ. Sci. Health. B22, 55-69.

38. S. Sekušak, A. Sabljić (1992). Soil Sorption and Chemical Topology. J. Math. Chem. 11, 271-280.

39. R.L. Swann, D.A. Laskowski, P.J. McCall, K. Vander Kuy, H.J. Dishburger (1983). A Rapid Method for the Estimation of the Environmental Parameters Octanol/Water Partition Coefficient, Soil Sorption Constant, Water to Air Ratio, and Water Solubility. Residue Rev. 85, 17-28.

40. D.S. Brown, E.W. Flagg (1981). Empirical Prediction of Organic Pollutant Sorption in Natural Sediment, J. Environ. Qual. 10, 382-386.

41. G.G. Briggs (1981). Adsorption of Pesticides by Some Australian Soils, Aust. J. Soil Res. 19, 61-68.

42. P. Lagas (1988). Sorption of Chlorophenols in the Soil. Chemosphere 17, 205-216.

43. B.J. Eadie, N.R. Morehead, P.F. Landrum (1990). Three-phase Partitioning of Hydrophobic Organic Compounds in Great Lakes Waters. Chemosphere 20, 161-178.

44. E.E. Kenaga (1980). Correlation of Bioconcentration Factors of Chemicals in Aquatic and Terrestrial Organisms with Their Physical and Chemical Properties. Environ. Sci. Technol. **14**, 553-556.

45. B. von Oepen, W. Kordel, W. Klein, G. Schuurmann (1991). Predictive QSPR Models for Estimating Soil Sorption Coefficients: Potential and Limitations Based on Dominating Processes. Sci. Tot. Environ. **109/110**, 343-354.

46. C.A.M. van Gestel, W. Ma, C.E. Smit (1991). Development of QSARs in Terrestrial Ecotoxicology: Earthworm Toxicity and Soil Sorption of Chlorophenols, Chlorobenzenes and Dichloroanilines. Sci. Tot. Environ. **109/110**, 589-604.

47. R. Koch, M. Nagel (1988). Quantitative Structure Activity Relationships in Soil Ecotoxicology. Sci. Tot. Environ. **77**, 269-276.

48. R. Koch (1983). Molecular Connectivity Index for Assessing Ecotoxicological Behaviour of Organic Compounds. Toxicol. Environ. Chem. **6**, 87-96.

49. S. Bintein, J. Devillers (1994). QSAR for Organic Chemical Sorption in Soils and Sediments. Chemosphere **28**, 1171-1188.

50. H. Lohninger (1994). Estimation of Soil Partition Coefficients of Pesticides from Their Chemical Structure. Chemosphere **29**, 1611-1626.

51. W. Kördel, J. Stutte, G. Kotthoff (1993). HPLC-Screening Method for the Determination of the Adsorption-Coefficient on Soil - Comparison of Different Stationary Phases. Chemosphere **27**, 2341-2352.

52. OECD Environment Monograph No 67. Application of structure-activity relationships to the estimation of properties important in exposure assessment. OECD, Paris, France, 1993.

53. IUCT. 1992. Handbuch zum SAR-Programm Version 3.0. Fraunhofer-Institut für Umweltchemie und Ökotoxikologie, Schmallenberg, Germany.

54. D. Weininger (1989). MedChem User's Manual V 3.54, Daylight Chem. Info. Systems, Inc. Irvine, CA, USA.

55. A.J. Leo (1993). Calculating $logP_{oct}$ from structure. Chem. Rev. **93**, 1281-1306.

56. A. Sabljić, D. Horvatić (1993). GRAPH III - A Computer Program for Calculating Molecular Connectivity Indices on Microcomputers. J. Chem. Inf. Comput. Sci. **33**, 292-295.

57. GRAPH III, Version 1.2 for Macintosh, LPC d.o.o., Bijenicka 30, HR-41000 Zagreb, Croatia, Fax: (385-1) 455-5506

58. SYSTAT for Macintosh, Version 5.2 (1992). SYSTAT, Inc., Evanston, Illinois

59. A. Sabljić (1987). Nonempirical Modelling of Environmental Distribution and Toxicity of Major Organic Pollutants, In QSAR in Environmental Toxicology-II, K.L.E. Kaiser (ed.), D. Reidel Publishing Co., Dordrecht, Holland, pp. 309-332.

60. B. von Oepen, W. Kördel, W. Klein (1991). Sorption of Polar and Nonpolar Compounds to Soils. Processes, Measurements and Experience with the Applicability of the Modified OECD-Guideline . 106. Chemosphere **22**, 285-304.