Molecular Descriptors Family on Structure Activity Relationships 6. Octanol-Water Partition Coefficient of Polychlorinated Biphenyls

Lorentz Jäntschi and Sorana Bolboacă

Technical University of Cluj-Napoca, Romania, http://lori.academicdirect.org
"Iuliu Hațieganu" University of Medicine and Pharmacy, http://sorana.academicdirect.ro

Abstract

Octanol-water partition coefficient of two hundred and six polychlorinated biphenyls was model by the use of an original method based on complex information obtained from compounds structure. The regression analysis shows that best results are obtained in four-varied model (r^2 = 0.9168). The prediction ability of the model was studied through leave-one-out analysis ($r^2_{cv(loo)}$ = 0.9093) and in training and test sets analysis. Modeling the octanol-water partition coefficient of polychlorinated biphenyls by integration of complex structural information provide a stable and performing four-varied model, allowing us to make remarks about relationship between structure of polychlorinated biphenyls and associated octanol-water partition coefficients.

Keywords

PolyChlorinated Biphenyls (PCBs), Molecular Descriptors Family (MDF), Structure-Property Relationships (SAR), Octanol-water partition coefficient

Background

Polychlorinated biphenyls (PCBs), stable organic industrials chemicals widely used as insulating fluids, hydraulic and lubricating fluids, heat exchanger fluids and as additives in adhesive inks and paints [1] are persistent in the environment [2] as well as in the living tissue [3].

Quantitative structure-property relationships of PCBs were previous studied taking into consideration octanol-water partition coefficients and soil-water partition coefficients [4] and/or other physicochemical properties [5].

Based on the complex information offered by the structure of polychlorinated biphenyls congeners, octanol-water partition coefficients express as log K_{ow} was modeled by applying of an original methodology. Thus, the aim of the paper is to present the performances of the original methodology in estimation and prediction of octanol-water partition coefficients of polychlorinated biphenyls.

Materials and Methods

A set of two-hundred and six polychlorinated biphenyls congeners with measured octanol-water partition coefficients were included into analysis. The values for the octanol-water partition coefficients were take from a previous reported study [6]. There were included ten PCBs congener group: mono-, di-, tri-, terta-, penta-, hexa-, hepta-, octa-, nona-, decachlorobiphenyl. Table 1 contains the PBCs number, the structure (chlorine-filled) and associated octanol-water partition coefficients (express as logK_{ow}).

The original methodology is based on molecular descriptors family computed based on the structure of the PCBs. The steps used to model the activity of interest were presented in details on [7] and were:

- Step 1: Sketch of the three-dimensional structure of polychlorinated biphenyls congeners;
- Step 2: Create the file with the measured octanol-water partition coefficients of the polychlorinated biphenyls congeners;
- Step 3: Generating, computing and filtering the members of molecular descriptors family for polychlorinated biphenyls congeners;
- Step 4: Finding and identifying the SAR models for polychlorinated biphenyls congeners;
- Step 5: Validate the SAR model by a cross-validation analysis [8];
- Step 6: Analyze the selected SAR model.

Results and Discussions

Modeling of the octanol-water partition coefficients of the polychlorinated biphenyls congeners was run on mono-, bi-, and tetra-varied SARs. The model which obtained best performance was the four-varied model and is presented here. The equation of the four varied model is:

$$\hat{Y}_{logKow} = 3.039 - 0.421 \cdot IIDDKGg + 0.044 \cdot IHDRKEg + 0.070 \cdot aHMmjQti - 37.502 \cdot aSMMjQg$$

The abbreviation associated with the studied PCBs congener (PBC no.), the measured octanol-water partition coefficients (express as $\log K_{ow}$), the values of the descriptors used and estimated octanol-water partition coefficients by the model ($\hat{Y}_{logKow}$) and the absolute differences between estimated by the model and measured octanol-water partition coefficients ($|\hat{Y}-\log K_{ow}|$) are in table 1.

Table 1. Polychlorinated biphenyls abbreviation, $\log K_{ow}$, values for descriptors used by model, $\hat{Y}_{logKow}$, and $|\hat{Y}_{logKow} - \log K_{ow}|$

PCB no.	Structure (chlorine-filled)	$\log K_{ow}$	IIDDKGg	IHDRKEg	aHMmjQt	aSMMjQg	$\hat{Y}_{logKow}$	$\lvert\hat{Y}-\log K_{ow}\rvert$
1	2	4.6010	5.7503	91.1540	0.0244	$3.67 \cdot 10^{-5}$	4.6477	0.0467
2	3	4.4210	6.8329	100.870	0.0286	$5.60 \cdot 10^{-4}$	4.6022	0.1812
3	4	4.4010	7.1099	105.020	0.0303	$1.50 \cdot 10^{-4}$	4.6845	0.2835
4	2,2'	5.0230	5.6688	98.0270	0.0454	$2.17 \cdot 10^{-4}$	4.9804	0.0426
5	2,3'	5.0210	6.0092	104.370	0.1765	$4.61 \cdot 10^{-5}$	5.1330	0.1120
6	2,4	5.1500	7.0663	113.130	0.0205	$4.01 \cdot 10^{-5}$	5.0646	0.0854
7	2,4'	5.3010	7.2970	115.000	0.0079	$8.57 \cdot 10^{-5}$	5.0476	0.2534
8	2,5	5.1800	5.8788	102.720	0.1013	$4.82 \cdot 10^{-5}$	5.1096	0.0704
9	2,6	5.3110	5.3684	95.9580	0.1265	$4.87 \cdot 10^{-5}$	5.0274	0.2836
10	3,3'	5.3430	6.8183	115.510	0.0464	$2.30 \cdot 10^{-4}$	5.2688	0.0742
11	3,4	5.2950	7.2304	118.150	0.0067	$4.30 \cdot 10^{-5}$	5.2163	0.0787
12	3,5	5.4040	6.7261	115.520	0.0357	$5.34 \cdot 10^{-4}$	5.2959	0.1081
13	4,4'	5.3350	7.3646	124.560	0.0065	$1.00 \cdot 10^{-4}$	5.4409	0.1059
14	2,2',3	5.3110	6.5110	114.030	0.0668	$2.44 \cdot 10^{-4}$	5.3336	0.0226
15	2,2',4	5.7610	6.9032	119.960	0.0867	$2.52 \cdot 10^{-4}$	5.4317	0.3293
16	2,2',5	5.5510	6.8266	120.050	0.0579	$2.66 \cdot 10^{-4}$	5.4654	0.0856
17	2,2',6	5.4810	5.8973	106.490	0.0865	$3.80 \cdot 10^{-4}$	5.2550	0.2260
18	2,3,3'	5.5770	7.8504	128.320	0.7667	$8.00 \cdot 10^{-5}$	5.4564	0.1206
19	2,3,4	5.5170	7.4010	125.250	0.0235	$5.28 \cdot 10^{-5}$	5.4591	0.0579
20	2,3,4'	5.4210	8.1564	132.790	0.0508	$7.75 \cdot 10^{-5}$	5.4753	0.0543
21	2,3,5	5.5770	6.5446	122.510	0.7310	$1.40 \cdot 10^{-4}$	5.7444	0.1674
22	2,3,6	5.6710	6.2218	110.890	0.0325	$1.01 \cdot 10^{-4}$	5.3195	0.3515
23	2,3',4	5.6770	8.1210	134.330	0.0325	$5.58 \cdot 10^{-5}$	5.5578	0.1192
24	2,3',5	5.6670	7.3770	129.310	0.1674	$8.11 \cdot 10^{-5}$	5.6575	0.0095
25	2,3',6	5.4470	6.2046	113.400	0.0385	$1.01 \cdot 10^{-4}$	5.4381	0.0089
26	2,4,4'	5.6910	8.4470	139.210	0.0266	$6.66 \cdot 10^{-5}$	5.6355	0.0555

27	2,4,5	5.7430	7.1079	126.640	0.0245	$5.28 \cdot 10^{-5}$	5.6439	0.0991
28	2,4,6	5.5040	6.5149	114.430	0.0420	$1.50 \cdot 10^{-4}$	5.3515	0.1525
29	2,4',5	5.6770	7.5935	133.180	0.0270	$7.78 \cdot 10^{-5}$	5.7278	0.0508
30	2,4',6	5.7510	6.4853	116.740	0.0429	$1.60 \cdot 10^{-4}$	5.4657	0.2853
31	2',3,4	5.5720	7.7422	129.640	0.0889	$7.82 \cdot 10^{-5}$	5.5131	0.0589
32	2',3,5	5.6670	7.2479	126.270	0.0128	$7.99 \cdot 10^{-5}$	5.5668	0.1002
33	3,3',4	5.8270	8.6854	141.730	0.1163	$3.69 \cdot 10^{-4}$	5.6414	0.1856
34	3,3',5	4.1510	8.1095	137.500	0.3422	$3.53 \cdot 10^{-2}$	4.4021	0.2511
35	3,4,4'	4.9410	8.9668	146.520	0.3565	$1.86 \cdot 10^{-4}$	5.7583	0.8173
36	3,4,5	5.7670	7.0857	129.590	0.0297	$1.89 \cdot 10^{-3}$	5.7151	0.0519
37	3,4',5	5.8970	8.3823	142.120	0.0550	$3.49 \cdot 10^{-4}$	5.7827	0.1143
38	2,2',3,3'	5.5610	7.7406	134.940	0.1340	$2.94 \cdot 10^{-4}$	5.7430	0.1820
39	2,2',3,4	6.1110	7.8567	137.950	0.0258	$2.88 \cdot 10^{-4}$	5.8198	0.2912
40	2,2',3,4'	5.7670	8.0115	139.920	0.1247	$2.83 \cdot 10^{-4}$	5.8488	0.0818
41	2,2',3,5	5.7570	7.2531	133.770	0.0875	$2.68 \cdot 10^{-4}$	5.8942	0.1372
42	2,2',3,5'	5.8110	7.5381	135.540	0.0446	$3.09 \cdot 10^{-4}$	5.8479	0.0369
43	2,2',3,6	5.5370	6.5064	121.120	0.1066	$3.47 \cdot 10^{-4}$	5.6478	0.1108
44	2,2',3,6'	5.5370	6.6121	121.520	0.1112	$4.71 \cdot 10^{-4}$	5.6166	0.0796
45	2,2',4,4'	6.2910	8.1688	145.660	0.2853	$3.00 \cdot 10^{-4}$	6.0468	0.2442
46	2,2'4,5	5.7870	7.2785	136.050	0.0489	$2.52 \cdot 10^{-4}$	5.9821	0.1951
47	2,2',4,5'	6.2210	7.6295	141.160	0.1152	$2.59 \cdot 10^{-4}$	6.0646	0.1564
48	2,2',4,6	5.6370	6.6210	124.040	1.3559	$4.02 \cdot 10^{-4}$	5.8136	0.1766
49	2,2',4,6'	5.6370	6.7699	126.190	0.0853	$3.62 \cdot 10^{-4}$	5.7589	0.1219
50	2,2',5,5'	6.0910	6.6586	135.000	0.0994	$2.62 \cdot 10^{-4}$	6.1997	0.1087
51	2,2',5,6'	5.6270	6.1810	121.830	0.0035	$9.77 \cdot 10^{-6}$	5.8215	0.1945
52	2,2',6,6'	5.9040	5.3274	109.030	0.1105	$4.85 \cdot 10^{-4}$	5.6047	0.2993
53	2,3,3',4	6.1170	8.3206	146.050	0.0345	$4.93 \cdot 10^{-5}$	5.9921	0.1249
54	2,3,3',4'	6.1170	8.3589	146.460	0.0962	$5.52 \cdot 10^{-5}$	5.9982	0.1188
55	2,3,3',5	6.1770	7.8097	142.860	0.8606	$7.89 \cdot 10^{-5}$	6.1226	0.0544
56	2,3,3',5'	6.1770	7.9437	142.850	0.0479	$5.54 \cdot 10^{-5}$	6.0100	0.1670
57	2,3,3',6	5.9570	6.9758	129.260	0.0476	$9.87 \cdot 10^{-5}$	5.8151	0.1419
58	2,3,4,4'	5.4520	8.2526	145.460	0.0345	$4.82 \cdot 10^{-5}$	5.9947	0.5427
59	2,3,4,5	5.9430	6.7297	133.750	0.0286	$3.72 \cdot 10^{-5}$	6.1181	0.1751
60	2,3,4,6	5.8970	6.2938	123.830	0.0630	$1.40 \cdot 10^{-4}$	5.8617	0.0353
61	2,3,4',5	6.1770	8.0492	147.320	0.0969	$6.32 \cdot 10^{-5}$	6.1662	0.0108
62	2,3,4',6	5.9570	7.2049	133.110	0.0597	$1.63 \cdot 10^{-4}$	5.8873	0.0697
63	2,3,5,6	5.8670	5.6388	122.010	0.1451	$9.11 \cdot 10^{-5}$	6.0644	0.1974
64	2,3',4,4'	5.4520	8.6051	153.380	0.0353	$5.40 \cdot 10^{-5}$	6.1961	0.7441
65	2,3',4,5	6.2070	8.1728	148.280	0.0359	$5.11 \cdot 10^{-5}$	6.1529	0.0541
66	2,3',4,5'	6.2670	8.1003	148.960	0.0334	$5.87 \cdot 10^{-5}$	6.2129	0.0541
67	2,3',4,6	6.0470	7.2353	133.270	0.0631	$1.66 \cdot 10^{-4}$	5.8817	0.1653
68	2,3',4',5	6.2310	7.5977	146.340	0.1246	$5.56 \cdot 10^{-5}$	6.3151	0.0841
69	2,3',4',6	5.9870	6.4091	128.760	0.0684	$1.26 \cdot 10^{-4}$	6.0319	0.0449
70	2,3',5,5'	6.2670	7.2087	143.450	0.0517	$5.87 \cdot 10^{-5}$	6.3459	0.0789
71	2,3',5',6	6.0470	6.0292	126.110	0.0503	$1.18 \cdot 10^{-4}$	6.0737	0.0267
72	2,4,4',5	6.6710	8.2735	152.290	0.0305	$5.04 \cdot 10^{-5}$	6.2873	0.3837
73	2,4,4',6	6.0570	7.4294	137.080	0.1203	$4.40 \cdot 10^{-4}$	5.9621	0.0949
74	2',3,4,5	6.1370	7.3073	139.090	0.0521	$5.68 \cdot 10^{-5}$	6.1119	0.0251
75	3,3',4,4'	6.5230	9.1490	161.430	0.2535	$9.76 \cdot 10^{-5}$	6.3366	0.1864
76	3,3',4,5	6.3570	8.1041	151.240	1.4919	$1.77 \cdot 10^{-4}$	6.4093	0.0523
77	3,3',4,5'	6.4270	8.5725	156.830	0.5807	$1.28 \cdot 10^{-4}$	6.3975	0.0295

78	3,3',5,5'	6.5830	7.9892	152.370	0.3196	$2.56 \cdot 10^{-4}$	6.4229	0.1601
79	3,4,4',5	6.3670	8.5262	157.620	0.1010	$1.19 \cdot 10^{-4}$	6.4188	0.0518
80	2,2',3,3',4	6.1420	8.2976	153.620	0.4084	$3.31 \cdot 10^{-4}$	6.3517	0.2097
81	2,2',3,3',5	6.2670	7.6890	148.730	0.7941	$3.69 \cdot 10^{-4}$	6.4172	0.1502
82	2,2',3,3',6	6.0410	6.9185	136.000	0.0442	$3.88 \cdot 10^{-4}$	6.1260	0.0850
83	2,2',3,4,4'	6.6110	8.4996	159.160	0.2075	$3.32 \cdot 10^{-4}$	6.4975	0.1135
84	2,2',3,4,5	6.2040	7.2394	145.350	0.1531	$2.73 \cdot 10^{-4}$	6.4160	0.2120
85	2,2',3,4,5'	6.3710	7.9248	154.060	0.0658	$3.42 \cdot 10^{-4}$	6.5038	0.1328
86	2,2',3,4,6	7.5160	6.7474	135.070	19.1930	$3.56 \cdot 10^{-4}$	7.4926	0.0234
87	2,2',3,4,6'	6.0770	6.8822	137.480	0.1178	$8.98 \cdot 10^{-4}$	6.1927	0.1157
88	2,2',3,4',5	6.3670	7.9310	154.600	0.1001	$2.65 \cdot 10^{-4}$	6.5303	0.1633
89	2,2',3,4',6	6.1370	7.1280	140.750	0.1762	$3.36 \cdot 10^{-4}$	6.2589	0.1219
90	2,2',3,5,5'	6.3570	7.3803	149.770	0.4452	$3.11 \cdot 10^{-4}$	6.5710	0.2140
91	2,2',3,5,6	6.0470	6.3200	132.670	0.5266	$3.58 \cdot 10^{-4}$	6.2654	0.2184
92	2,2',3,5,6'	6.1370	6.4532	134.510	0.0738	$1.69 \cdot 10^{-4}$	6.2662	0.1292
93	2,2',3,5',6	6.1370	6.6477	136.300	0.4964	$7.70 \cdot 10^{-4}$	6.2704	0.1334
94	2,2',3,6,6	5.7170	5.9325	123.370	0.0920	$3.78 \cdot 10^{-4}$	5.9865	0.2695
95	2,2',3',4,5	6.6710	7.8560	152.750	0.1031	$3.34 \cdot 10^{-4}$	6.4778	0.1932
96	2,2',3',4,6	6.1370	7.0705	139.710	0.1352	$7.49 \cdot 10^{-4}$	6.2187	0.0817
97	2,2',4,4',5	7.2110	8.1412	159.410	0.0577	$2.85 \cdot 10^{-4}$	6.6507	0.5603
98	2,2',4,4',6	6.2370	7.3053	144.740	0.2006	$5.63 \cdot 10^{-4}$	6.3537	0.1167
99	2,2',4,5,5'	7.0710	7.6259	154.880	0.1225	$3.04 \cdot 10^{-4}$	6.6712	0.3998
100	2,2',4,5,6'	6.1670	6.7261	138.380	0.0262	$2.22 \cdot 10^{-5}$	6.3246	0.1576
101	2,2',4,5',6	6.2270	6.8046	140.120	0.0279	$5.27 \cdot 10^{-5}$	6.3674	0.1404
102	2,2',4,6,6	5.8170	6.0829	126.540	0.0922	$3.76 \cdot 10^{-4}$	6.0634	0.2464
103	2,3,3',4,4'	6.6570	9.1546	168.730	0.0365	$4.78 \cdot 10^{-5}$	6.6435	0.0135
104	2,3,3',4,5	6.6470	8.2227	158.650	0.0384	$3.80 \cdot 10^{-5}$	6.5907	0.0563
105	2,3,3',4',5	6.7170	8.5882	164.350	0.0579	$5.64 \cdot 10^{-5}$	6.6895	0.0275
106	2,3,3',4,5'	6.7170	8.6053	163.610	0.0364	$5.31 \cdot 10^{-5}$	6.6482	0.0688
107	2,3,3',4,6	6.4870	7.5286	145.840	0.0862	$1.67 \cdot 10^{-4}$	6.3153	0.1717
108	2,3,3',4',6	6.5320	7.5906	148.130	0.3660	$1.60 \cdot 10^{-4}$	6.4101	0.1219
109	2,3,3',5,5'	6.7670	8.0899	159.630	0.0377	$5.85 \cdot 10^{-5}$	6.6891	0.0779
110	2,3,3',5,6	6.4570	7.1659	142.920	0.1454	$9.27 \cdot 10^{-5}$	6.3458	0.1112
111	2,3,3',5',6	6.5470	7.2338	144.320	0.0712	$1.41 \cdot 10^{-4}$	6.3721	0.1749
112	2,3,4,4',5	6.6570	8.2871	163.400	0.0322	$3.75 \cdot 10^{-5}$	6.7731	0.1161
113	2,3,4,4',6	6.4970	7.7007	150.180	0.1935	$6.47 \cdot 10^{-4}$	6.4241	0.0729
114	2,3,4,5,6	6.3040	6.1425	134.100	0.0371	$1.31 \cdot 10^{-4}$	6.3777	0.0737
115	2,3,4',5,6	6.4670	7.3123	146.830	0.3320	$1.63 \cdot 10^{-4}$	6.4673	0.0003
116	2,3',4,4',5	7.1210	8.7170	168.800	0.0387	$4.74 \cdot 10^{-5}$	6.8309	0.2901
117	2,3',4,4',6	6.5870	7.7307	152.260	0.0954	$5.32 \cdot 10^{-4}$	6.5009	0.0861
118	2,3',4,5,5'	6.7970	8.1850	163.840	0.0386	$5.17 \cdot 10^{-5}$	6.8354	0.0384
119	2,3',4,5',6	6.6470	7.2928	148.060	0.1874	$2.93 \cdot 10^{-4}$	6.5149	0.1321
120	2',3,3',4,5	6.6470	8.3558	159.300	0.0339	$5.32 \cdot 10^{-5}$	6.5626	0.0844
121	2',3,4,4',5	6.7470	8.5776	165.470	0.0506	$1.30 \cdot 10^{-4}$	6.7401	0.0069
122	2',3,4,5,5'	6.7370	7.8399	159.710	0.0350	$5.59 \cdot 10^{-5}$	6.7977	0.0607
123	2',3,4,5,6'	6.5170	6.9321	142.900	0.2070	$7.11 \cdot 10^{-4}$	6.4244	0.0926
124	3,3',4,4'5	6.8970	9.2142	175.600	0.2341	$9.05 \cdot 10^{-5}$	6.9342	0.0372
125	3,3',4,5,5'	6.9570	8.6661	170.560	0.0774	$1.16 \cdot 10^{-4}$	6.9302	0.0268
126	2,2',3,3',4,4'	6.9610	9.3307	179.190	0.7685	$7.25 \cdot 10^{-4}$	7.0572	0.0962
127	2,2',3,3',4,5	7.3210	8.5439	169.150	0.0167	$1.44 \cdot 10^{-3}$	6.8653	0.4557
128	2,2',3,3',4,5'	7.3910	8.6910	172.480	0.0174	$1.04 \cdot 10^{-3}$	6.9659	0.4251

129	2,2',3,3',4,6	6.5870	7.6941	154.760	0.0299	$8.40\cdot10^{-4}$	6.6106	0.0236
130	2,2',3,3',4,6'	6.5870	7.8668	157.060	0.0471	$6.20\cdot10^{-4}$	6.6490	0.0620
131	2,2',3,3',5,5'	6.8670	7.8751	165.430	0.3205	$3.97\cdot10^{-4}$	7.0428	0.1758
132	2,2',3,3',5,6	7.3040	7.1724	150.190	0.2877	$1.26\cdot10^{-3}$	6.6304	0.6736
133	2,2',3,3',5,6'	7.1510	7.1833	151.550	0.0310	$1.92\cdot10^{-4}$	6.7081	0.4429
134	2,2',3,3',6,6'	6.5110	6.4063	138.560	0.0789	$2.98\cdot10^{-4}$	6.4604	0.0506
135	2,2',3,4,4',5'	7.4410	8.6591	174.490	0.3322	$6.17\cdot10^{-4}$	7.1058	0.3352
136	2,2',3,4,4',6	6.6770	7.7162	158.610	0.0244	$6.15\cdot10^{-4}$	6.7795	0.1025
137	2,2',3,4,4',6'	6.6770	7.8664	159.440	17.463	$4.04\cdot10^{-2}$	6.4781	0.1989
138	2,2',3,4,5,5'	7.5920	7.8362	166.540	0.1417	$4.20\cdot10^{-4}$	7.0949	0.4971
139	2,2',3,4,5,6	6.5170	6.7464	146.160	0.0415	$5.16\cdot10^{-4}$	6.6424	0.1254
140	2,2',3,4,5,6'	6.6070	6.9031	149.310	0.5610	$1.94\cdot10^{-4}$	6.7639	0.1569
141	2,2',3,4,5',6	6.6770	7.1854	153.300	0.1587	$1.13\cdot10^{-2}$	6.3768	0.3002
142	2,2',3,4,6,6'	6.2570	6.4146	138.950	0.0774	$2.91\cdot10^{-4}$	6.4743	0.2173
143	2,2',3,4',5,5'	6.8970	8.0823	169.690	0.0963	$6.14\cdot10^{-4}$	7.1201	0.2231
144	2,2',3,4',5,6	6.6470	7.3016	154.700	0.0982	$5.27\cdot10^{-4}$	6.7896	0.1426
145	2,2',3,4',5,6'	6.7370	7.4207	155.500	0.1185	$1.83\cdot10^{-4}$	6.7892	0.0522
146	2,2',3,4',5',6	7.2810	7.1955	153.590	0.0208	$1.48\cdot10^{-3}$	6.7440	0.5370
147	2,2',3,4',6,6'	6.3270	6.5980	141.640	0.0790	$3.00\cdot10^{-4}$	6.5158	0.1888
148	2,2',3,5,5',6	6.6470	6.7917	149.800	0.0450	$1.87\cdot10^{-4}$	6.7967	0.1497
149	2,2',3,5,6,6'	6.2270	6.0509	135.760	0.0815	$2.94\cdot10^{-4}$	6.4866	0.2596
150	2,2',4,4',5,5'	7.7510	8.1709	174.310	0.1986	$4.17\cdot10^{-4}$	7.3015	0.4495
151	2,2',4,4',5,6'	6.7670	7.4824	159.260	0.0188	$9.35\cdot10^{-5}$	6.9258	0.1588
152	2,2',4,4',6,6'	7.1230	6.7120	145.340	0.0787	$2.99\cdot10^{-4}$	6.6313	0.4917
153	2,3,3',4,4',5	7.1870	8.9450	180.720	0.0383	$2.82\cdot10^{-5}$	7.2624	0.0754
154	2,3,3',4,4',5'	7.1870	8.9295	180.060	0.0500	$3.70\cdot10^{-5}$	7.2402	0.0532
155	2,3,3',4,4',6	7.0270	8.1071	166.380	0.3064	$2.87\cdot10^{-4}$	6.9903	0.0367
156	2,3,3',4,5,5'	7.2470	8.4424	175.580	0.0401	$3.17\cdot10^{-5}$	7.2467	0.0003
157	2,3,3',4,5,6	6.9370	7.3571	156.700	0.0739	$1.34\cdot10^{-4}$	6.8677	0.0693
158	2,3,3',4,5',6	7.0870	7.6904	161.910	1.6216	$2.19\cdot10^{-4}$	7.0623	0.0247
159	2,3,3',4',5,5'	7.2470	8.1738	174.530	0.0248	$3.74\cdot10^{-5}$	7.3121	0.0651
160	2,3,3',4',5,6	6.9970	7.5592	161.980	0.2656	$9.21\cdot10^{-5}$	7.0309	0.0339
161	2,3,3',4',5',6	7.0270	7.2468	158.640	0.1114	$1.16\cdot10^{-4}$	7.0031	0.0239
162	2,3,3',5,5',6	7.0570	7.2209	157.930	0.1685	$9.63\cdot10^{-5}$	6.9874	0.0696
163	2,3,4,4',5,6	6.9370	7.4614	161.140	0.5263	$4.59\cdot10^{-4}$	7.0393	0.1023
164	2,3',4,4',5,5'	7.2770	8.4739	180.990	0.0537	$4.30\cdot10^{-5}$	7.4731	0.1961
165	2,3',4,4',5',6	7.1170	7.4940	163.550	0.2913	$5.09\cdot10^{-4}$	7.1138	0.0032
166	3,3',4,4',5,5'	7.4270	9.1962	189.460	0.0633	$8.19\cdot10^{-5}$	7.5426	0.1156
167	2,2',3,3',4,4',5	7.2770	8.9382	188.200	0.1129	$9.03\cdot10^{-5}$	7.5986	0.3216
168	2,2',3,3',4,4',6	6.7040	8.1822	173.820	0.1327	$1.78\cdot10^{-3}$	7.2194	0.5154
169	2,2',3,3',4,5,5'	7.3370	8.4591	183.610	0.0758	$9.65\cdot10^{-4}$	7.5620	0.2250
170	2,2',3,3',4,5,6	7.0270	7.6027	165.040	0.0482	$9.50\cdot10^{-4}$	7.1005	0.0735
171	2,2',3,3',4,5,6'	7.1170	7.5973	167.060	0.0124	$8.87\cdot10^{-5}$	7.2218	0.1048
172	2,2',3,3',4,5',6	7.1770	7.7165	169.220	0.0266	$8.26\cdot10^{-5}$	7.2683	0.0913
173	2,2',3,3',4,6,6'	6.7670	7.0135	155.320	0.0680	$2.45\cdot10^{-4}$	6.9467	0.1797
174	2,2',3,3',4',5,6	7.0870	7.6992	169.170	0.0578	$2.24\cdot10^{-4}$	7.2703	0.1833
175	2,2',3,3',5,5',6	7.1470	7.3049	165.140	0.1383	$2.63\cdot10^{-4}$	7.2622	0.1152
176	2,2',3,3',5,6,6'	6.7370	6.6644	151.780	0.0715	$2.45\cdot10^{-4}$	6.9374	0.2004
177	2,2',3,4,4',5,5'	7.3670	8.6195	188.140	0.0550	$4.81\cdot10^{-3}$	7.5491	0.1821
178	2,2',3,4,4',5,6	7.1170	7.7127	170.270	6.8760	$6.77\cdot10^{-3}$	7.5427	0.4257
179	2,2',3,4,4',5,6'	7.2070	7.8062	171.680	0.6290	$2.78\cdot10^{-4}$	7.3739	0.1669

180	2,2',3,4,4',5',6	7.2070	7.8790	173.190	0.1139	$3.01\cdot10^{-4}$	7.3733	0.1663
181	2,2',3,4,4',6,6'	6.8570	7.1523	158.990	0.0677	$2.43\cdot10^{-4}$	7.0505	0.1935
182	2,2',3,4,5,5',6	7.9330	6.8634	164.000	0.4577	$1.39\cdot10^{-5}$	7.4292	0.5038
183	2,2',3,4,5,6,6'	6.6970	6.2008	148.390	0.0693	$2.38\cdot10^{-4}$	6.9828	0.2858
184	2,2',3,4',5,5',6	7.1770	7.3776	168.560	0.0448	$1.14\cdot10^{-4}$	7.3819	0.2049
185	2,2',3,4',5,6,6'	6.8270	6.7527	155.210	0.0714	$2.43\cdot10^{-4}$	7.0519	0.2249
186	2,3,3',4,4',5,5'	7.7170	9.1221	195.930	0.0488	$3.59\cdot10^{-5}$	7.8604	0.1434
187	2,3,3',4,4',5,6	7.4670	8.2487	178.880	0.5202	$6.32\cdot10^{-4}$	7.4850	0.0180
188	2,3,3',4,4',5',6	7.5570	8.2502	179.590	0.6986	$6.92\cdot10^{-4}$	7.5259	0.0311
189	2,3,3',4,5,5',6	7.5270	7.7471	173.540	0.0626	$2.84\cdot10^{-4}$	7.4413	0.0857
190	2,3,3',4',5,5',6	7.5270	7.6407	174.490	0.4337	$1.22\cdot10^{-4}$	7.5600	0.0330
191	2,2',3,3',4,4',5,5'	8.6830	8.8559	201.230	0.0917	$6.13\cdot10^{-4}$	8.1879	0.4951
192	2,2',3,3',4,4',5,6	7.5670	8.1340	185.280	0.0365	$5.66\cdot10^{-5}$	7.8039	0.2369
193	2,2',3,3',4,4',5',6	7.6570	8.0500	185.640	0.0407	$3.72\cdot10^{-5}$	7.8562	0.1992
194	2,2',3,3',4,4',6,6'	7.3070	7.5022	173.290	0.0595	$1.97\cdot10^{-4}$	7.5363	0.2293
195	2,2',3,3',4,5,5',6	7.6270	7.7337	180.830	0.0627	$1.47\cdot10^{-4}$	7.7742	0.1472
196	2,2',3,3',4,5,6,6'	7.2070	6.9867	166.280	0.0619	$1.96\cdot10^{-4}$	7.4437	0.2367
197	2,2',3,3',4,5',6,6'	7.2770	7.1190	169.460	0.0626	$1.98\cdot10^{-4}$	7.5285	0.2515
198	2,2',3,3',4',5,5',6	7.6270	7.6994	180.700	0.0627	$1.52\cdot10^{-4}$	7.7827	0.1557
199	2,2',3,3',5,5',6,6'	8.4230	6.6004	165.330	0.0656	$1.96\cdot10^{-4}$	7.5645	0.8585
200	2,2',3,4,4',5,5',6	7.6570	7.7846	184.510	0.6727	$3.23\cdot10^{-4}$	7.9513	0.2943
201	2,2',3,4,4',5,6,6'	7.3070	7.0818	170.100	0.0617	$1.93\cdot10^{-4}$	7.5726	0.2656
202	2,3,3',4,4',5,5',6	8.0070	8.1803	191.360	1.1642	$5.30\cdot10^{-4}$	8.1139	0.1069
203	2,2',3,3',4,4',5,5',6	9.1430	7.9885	197.410	0.0270	$6.07\cdot10^{-5}$	8.4003	0.7427
204	2,2',3,3',4,4',5,6,6'	7.7470	7.4690	184.980	0.0550	$1.57\cdot10^{-4}$	8.0680	0.3210
205	2,2',3,3',4,5,5',6,6'	8.1640	7.1318	180.950	0.0577	$1.60\cdot10^{-4}$	8.0319	0.1321
206	2,2',3,3',4,4',5,5',6,6'	9.6030	7.4035	197.030	0.0512	$1.33\cdot10^{-4}$	8.6287	0.9743

Three molecular descriptors take into consideration the geometry of PCBs (IIDDKG**g**, IHDRKE**g**, and aSMMjQ**g**) and one the topology of compounds (aHMmjQ**t**). As atomic property, two descriptors consider the partial change (aHMmj**Q**t, and aSMMj**Q**g), one the group electronegativity (IIDDK**G**g) and one the atomic electronegativity (IHDRK**E**g). Looking at the interaction descriptor (the fifth letter in descriptors name) it can be observed that all descriptors consider the elastic force.

The results of multiple linear regressions associated to the four-varied model (see table 2, and table 3) sustain the estimation and prediction abilities of the best performing SAR model.

In the table 3 are the 95% probability of confidence intervals - lower ($_{95\%}CI_L$) and upper ($_{95\%}CI_U$) boundaries, coefficients, standard error (StdErr) of the coefficient, Student test parameter (t) and Student probability (p_t).

Table 2. Statistics associated with the tetra-varied model

Characteristic	**Notation**	**Values**
Correlation coefficient	r	0.9575
Squared correlation coefficient	r^2	0.9168
Adjusted squared correlation coefficient	r^2_{adj}	0.9151
Standard error of estimated	s_{est}	0.2420
Fisher parameter	F_{est}	554
Probability of wrong model	$p_{est}(\%)$	$< 1\cdot10^{-15}$
Cross-validation leave-one-out (loo) score	$r^2_{cv(loo)}$	0.9093
Fisher parameter for loo analysis	F_{pred}	504
Probability of wrong model for loo analysis	$p_{pred}(\%)$	$< 1\cdot10^{-15}$
Standard error for loo analysis	s_{loo}	0.2526
The difference between r^2 and $r^2_{cv(loo)}$	$r^2 - r^2_{cv(loo)}$	0.0075
Squared correlation coefficients between each descriptor and measured octanol-water partition coefficients or between pairs of descriptors	r^2(IIDDKGg, IHDRKEg)	0.48245
	r^2(IIDDKGg, aHMmjQt)	0.00005
	r^2(IIDDKGg, aSMMjQg)	0.00385
	r^2(IHDRKEg, aHMmjQt)	0.00039
	r^2(IHDRKEg, aSMMjQg)	0.00073
	r^2(aHMmjQt, aSMMjQg)	0.24805
	r^2(IIDDKGg, $\log_{Kow}$)	0.15111
	r^2(IHDRKEg, $\log_{Kow}$)	0.78907
	r^2(aSMMjQg, $\log_{Kow}$)	0.00932
	r^2(aHMmjQt, $\log_{Kow}$)	0.00786

Table 3. Statistics associated with the four-varied model

	$_{95\%}CI_L$	Coefficients	$_{95\%}CI_U$	StdError	t	p_t (%)
Intercept	2.735	3.039	3.343	0.154	19.716	$7.27\cdot10^{-47}$
IIDDKGg	-0.477	-0.421	-0.365	0.028	-14.804	$5.09\cdot10^{-32}$
IHDRKEg	0.042	0.044	0.046	0.001	41.725	$5.97\cdot10^{-99}$
aHMmjQt	0.049	0.070	0.090	0.010	6.639	$2.89\cdot10^{-8}$
aSMMjQg	-47.601	-37.499	-27.397	5.123	-7.319	$5.86\cdot10^{-10}$

The model which consider in estimation four molecular descriptors is significant statistically, having a probability of a wrong model less than $1\cdot10^{-15}$ (%). The estimation ability of the SAR model is sustained by the value of the correlation coefficient ($r = 0.9575$), the confidence boundaries associated with the coefficients (see table 3), and probabilities associated with Student tests (for all coefficients less than 0.001 - see table 3). Almost ninety-two percent ($r^2 = 0.9168$) from variation of octanol-water partition coefficient can be explained by its linear relationship with the variation of the four molecular descriptors used in the model. The probability of wrong model for leave-one-out analysis ($p_{pred}(\%) < 1\cdot10^{-15}$) and its associated Fisher parameter ($F_{pred} = 504$) sustains the estimation ability of the model. The four-varied SAR model is a stable one, stability sustained by the values of difference between correlation coefficient and cross validation leave-one-out correlation score (r^2 - $r^2_{cv(loo)}$ = 0.0075). The power of the four-varied model in octanol-water partition coefficient prediction of PCBs is sustained by the absence of multicolinearity of descriptors used by the model (see the squared correlation coefficients between pairs of descriptors, which always is less than 0.48 - table 2).

The plot of dependency between measured ($\log K_{ow}$) and estimated based on the structure of polychlorinated biphenyls compounds obtained with the tetra-varied model is in figure 1.

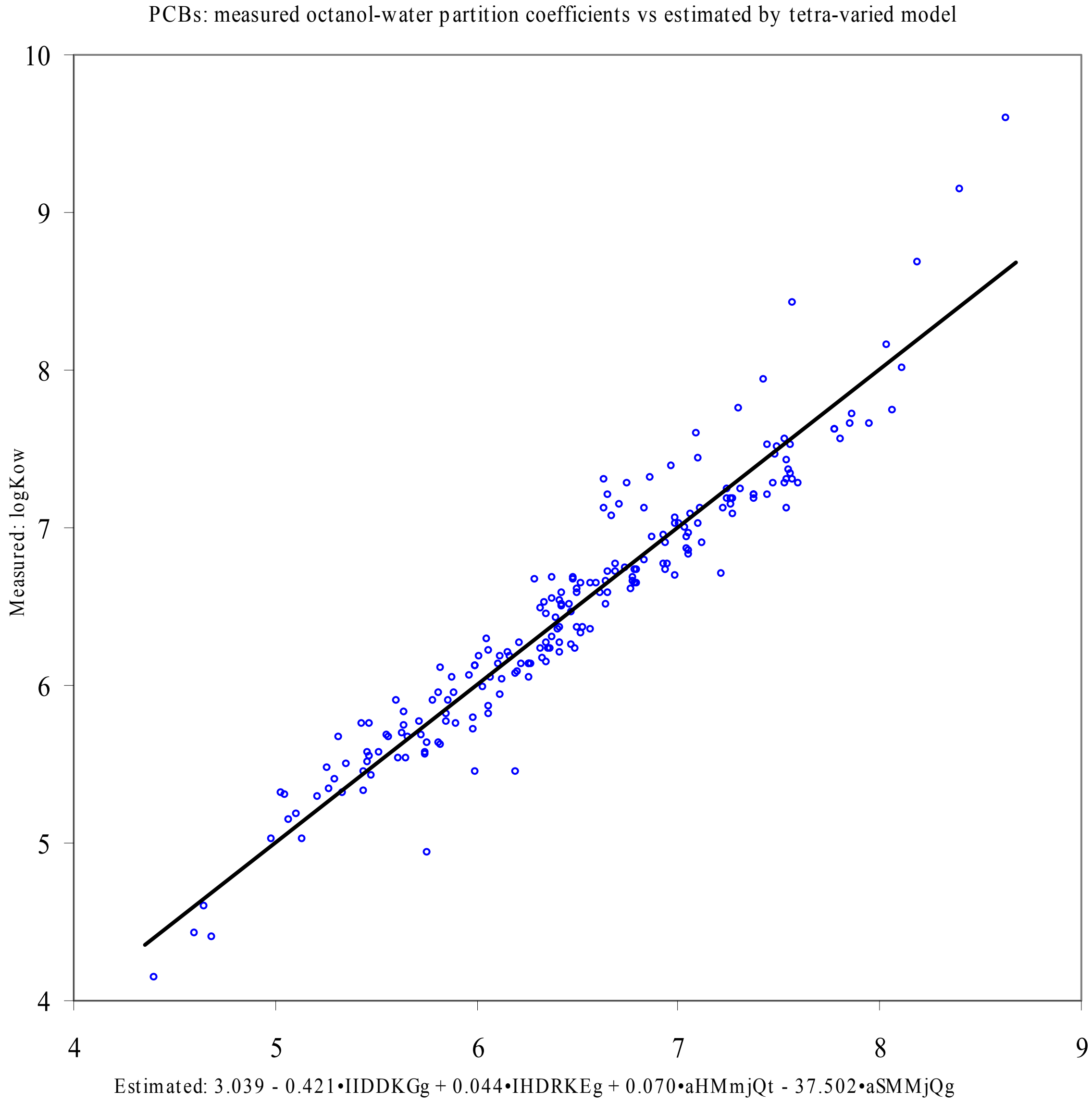

Figure 1. Measured vs. estimated $\log_{Kow}$ by the tetra-varied model

The estimation values of octanol-water partition coefficients by the use of the four-varied model of are less or greater than measured values (see figure 2). Note that, the mean and 95% confidence intervals of the mean and standard error for measured ($m_{Measured}$ = 6.4802, $_{95\%}CI_{Measured}$ = [6.3709, 6.5895], $StdErr_{Measured}$ = 0.0554) and estimated ($m_{Estimated}$ = 6.4806, $_{95\%}CI_{Estimated}$ = [6.3664, 6.5947], $StdErr_{Estimated}$ = 0.0579) octanol-water partition coefficients are almost equal.

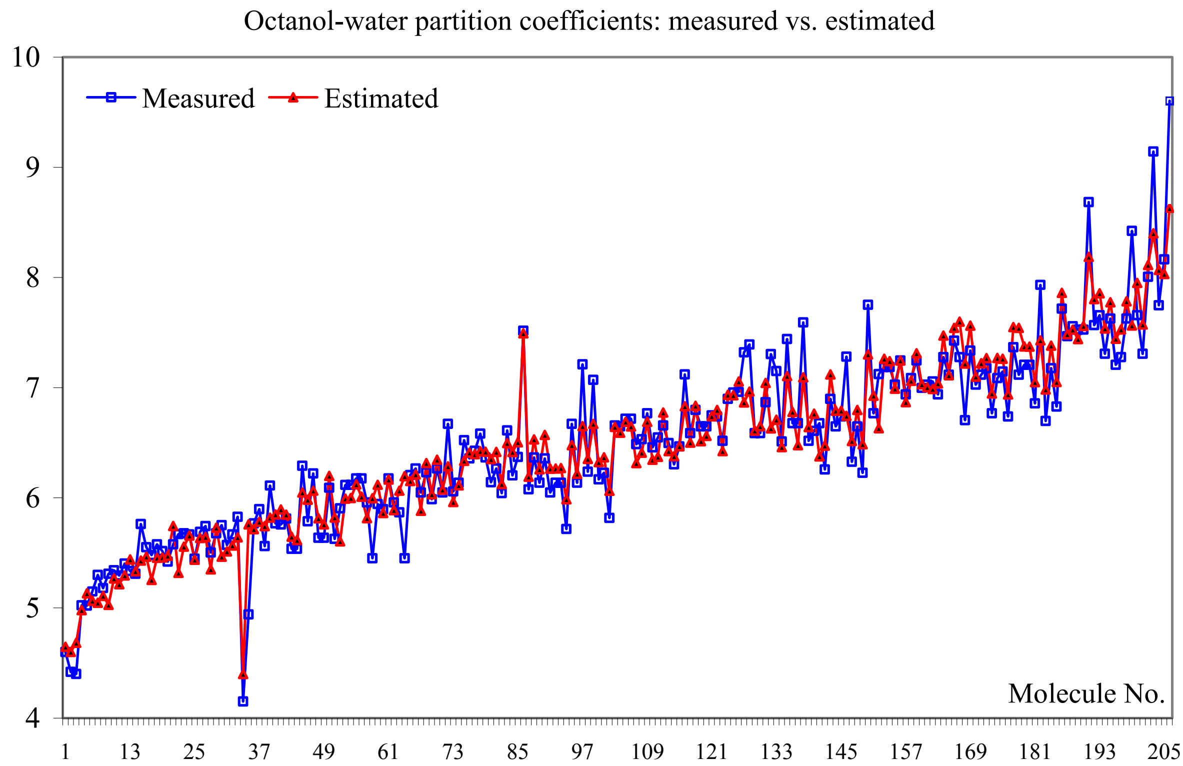

Figure 2. Variation of measured (blue line) and estimated (red line) by the four-varied model of octanol-water partition coefficient for PCBs

In order to seen the estimation abilities of four-varied model, measured and estimated values were sort by the absolute differences between estimated and measured octanol-water partition coefficient of PCBs and split into two subsets (first containing one-hundred PCBs and second containing the other one-hundred and six PBCs). The graphical representations are in figure 3a (one-hundred compounds) and 3b (one-hundred and six compounds), where the PCB number was associated with corresponding estimated and measured values.

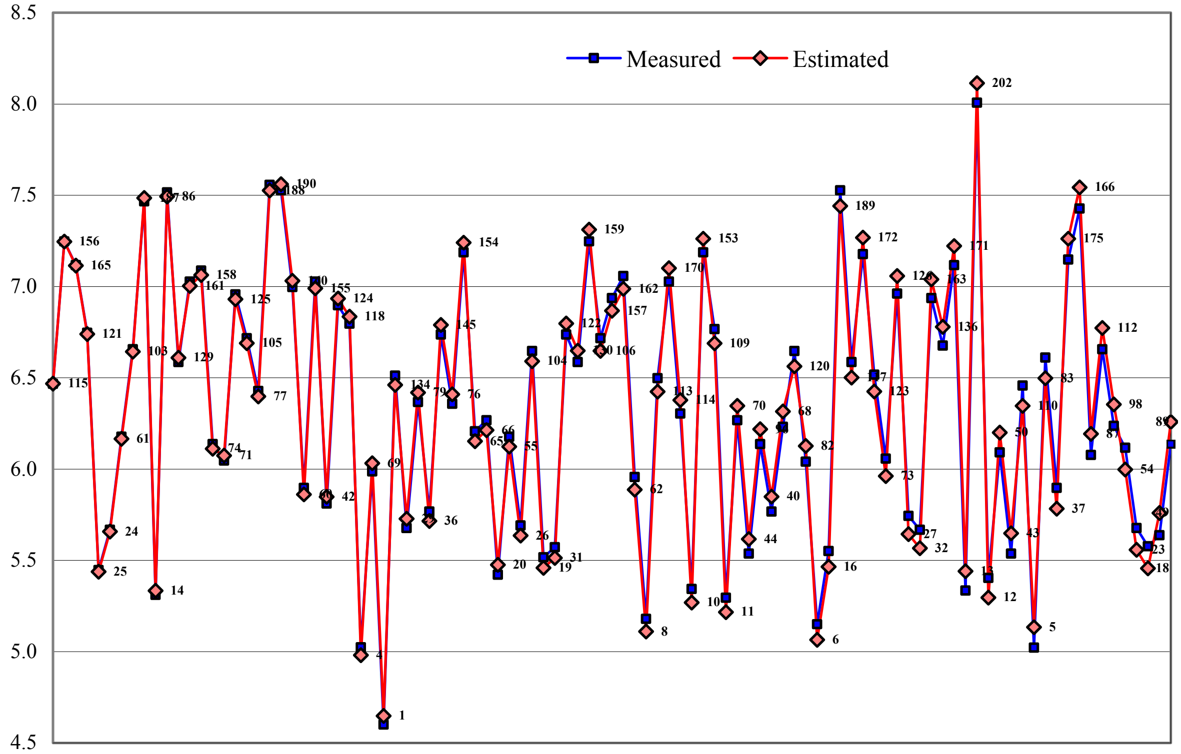

Figure 3a. Measured (blue line) and estimated by the tetra-varied model (red-line) of octanol-water partition coefficients for one-hundred PCBs

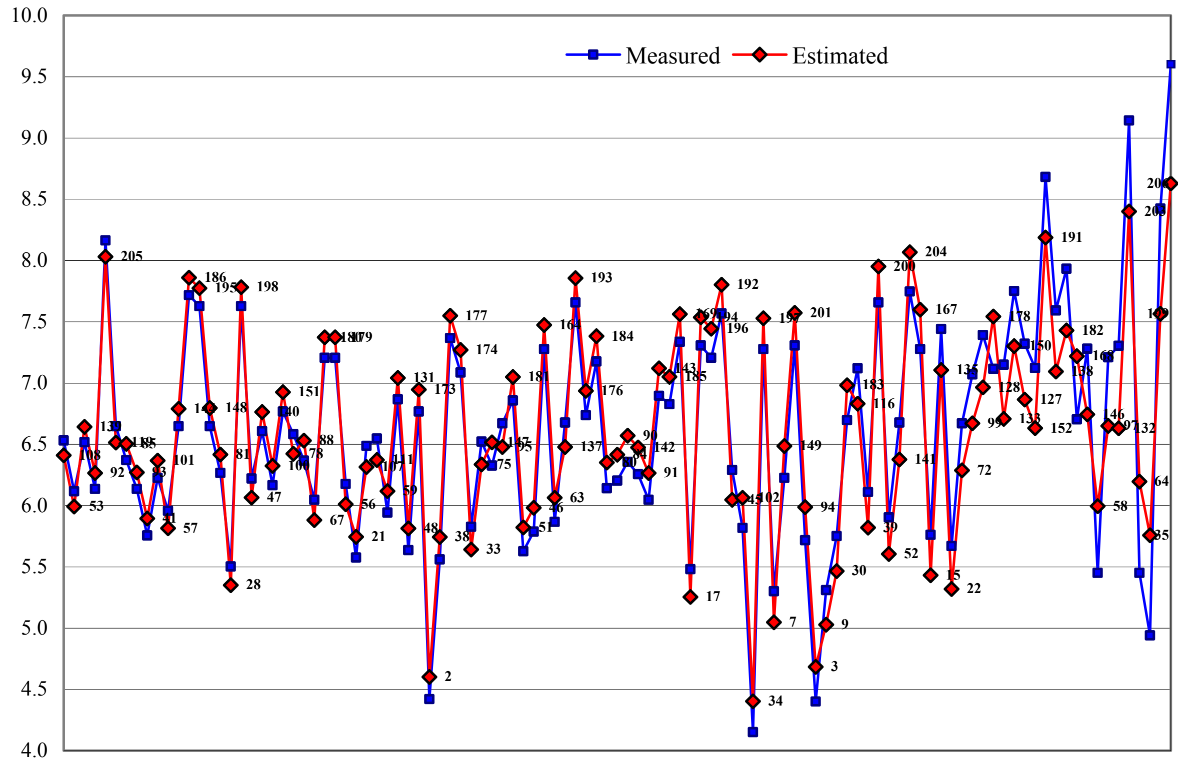

Figure 3b. Measured (blue line) and estimated by the tetra-varied model (red line) of octanol-water partition coefficients for one-hundred and six PCB

Analyzing the residuals of the four-varied model allowed us to assess the suitability of the model. Looking at the differences between measured and estimated octanol-water partition coefficient for PCBs (figure 4) it can be observed that the values vary around zero and most of them between -0.5 and 0.5.

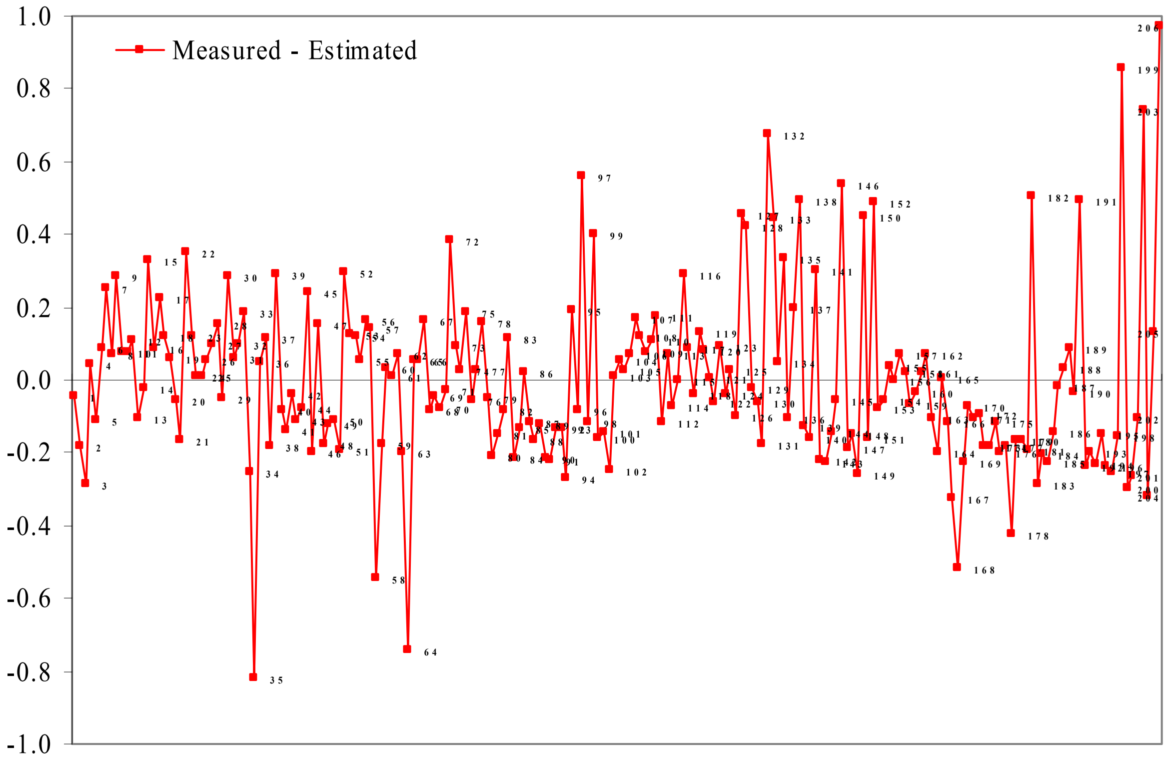

Figure 4. The differences between measured and estimated by the tetra-varied model of octanol-water partition coefficients for PCBs

The prediction abilities of the four-varied SAR model were studied through training and test sets analysis, and the results are in table 4. There were analyzed twelve situations, starting with a training sample size equal with 116 and increasing the number of PCBs included into training sets through randomization with seven until one hundred ninety-three. In table 4, there were included the number of PCBs in training sets (No_{tr}), the coefficients of the model, the squared correlation coefficient for training set (r_{tr}^2), Fisher parameter associated with training set regression (F_{tr}), the number of the PCBs in test sets (No_{ts}), the squared correlation coefficient for test set (r_{ts}^2), Fisher parameter associated with training set regression (F_{ts}), the mean (Mean) and standard deviation (StDev) for squared correlation coefficients and the 95% probability CI [$_{95\%}CI_L$ and $_{95\%}CI_U$] for coeficients.

Table 4. Results of training vs. test sets analysis

No_{tr}	intercept	IIDDKGg	IHDRKEg	aHMmjQt	aSMMjQg	r_{tr}^2	F_{tr}	No_{ts}	r_{ts}^2	F_{ts}
116	3.070	-0.408	0.043	0.064	-34.937	0.9141	295*	90	0.9219	235*
123	3.058	-0.390	0.043	0.064	-43.454	0.9229	353*	83	0.9043	176*
130	2.957	-0.413	0.044	0.067	-33.462	0.9232	376*	76	0.9068	169*
137	3.011	-0.438	0.045	0.064	-32.008	0.9004	298*	69	0.9432	256*
144	3.090	-0.450	0.045	0.062	-45.236	0.9143	371*	62	0.9186	148*
151	3.102	-0.432	0.044	0.062	-42.983	0.9173	405*	55	0.9075	122*
158	3.137	-0.460	0.046	0.073	-37.319	0.9200	440*	48	0.9041	82*
165	3.091	-0.428	0.044	0.070	-37.661	0.9143	427*	41	0.9247	110*
172	3.063	-0.426	0.044	0.069	-36.945	0.9161	456*	34	0.9202	83*
179	3.085	-0.429	0.044	0.069	-37.219	0.9098	439*	27	0.9582	106*
186	2.990	-0.420	0.045	0.070	-37.650	0.9090	452*	20	0.9876	178*
193	3.067	-0.430	0.044	0.074	-37.466	0.9160	513*	13	0.9249	24*

*$p < 0.001$

$_{95\%}CI_L$	3.028	-0.439	0.044	0.065	-40.566	0.9148	Mean	0.9268
$_{95\%}CI_U$	3.092	-0.415	0.045	0.070	-35.490	0.0063	StDev	0.0250

All squared correlation coefficients in training as well as in test sets are greater than 0.9, sustaining the prediction ability of the four-varied model. More, the mean of squared correlation coefficients in test sets is a little bit higher compared with the mean of squared correlation coefficient in training sets, and the dispersions of squared correlation coefficients are very small for both sets. All the regressions in training and test sets are highly significant ($p < 0.001$).

Analyzing the regressions coefficients it can be observed that with no exception the values of coefficients respect the 95% confidence intervals associated to the four-varied model (see table 3 and table 4). More, as it is expected, the 95% CI values (table 4) obtained in training and test sets analyses are contained by the 95% CI values of four-varied model (table 3).

The plot of measured vs. estimated octanol-water partition coefficients in training set (blue line and dots) of sample size equal one-hundred thirty-seven (corresponding with 2/3 from total sample of PCBs) and corresponding test set (red line and dots) of sample size equal with sixty-nine (1/3 from total sample of PCBs) is in figure 5.

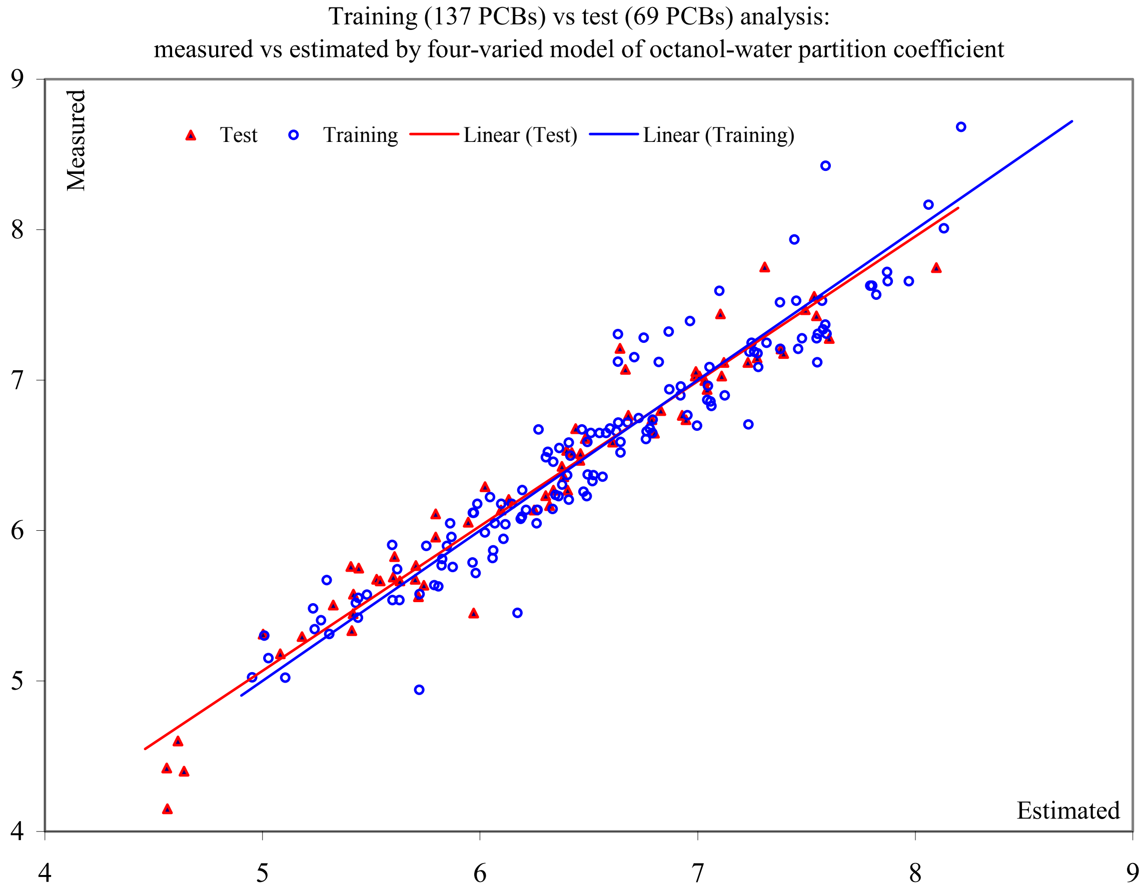

Figure 4. Training (137 PCBs) vs test (69 PCBs) analysis with four-varied model

Starting with the above describe model, and by the use of the original software [9], the octanol-water partition coefficient of new polychlorinated biphenyls can be obtains in a short time, without any experiments, following the next steps: drawing by the use of HyperChem software the three dimensional structure of the new PCB, choosing the model of prediction from the list (in our case PCB_lkow), browsing the *.hin file, and computing the octanol-water partition coefficient based on the four-varied SAR equation.

Conclusions

Modeling the octanol-water partition coefficient of polychlorinated biphenyls by integration of complex structural information provide a stable and performing four-varied model, allowing us to make remarks about relationship between structure of PCBs and associated octanol-water partition coefficients. Thus, the octanol-water partition coefficient of studied PCBs is like to be of geometry and topology nature, depending by the partial change, group and atomic electronegativity as atomic properties, and being in relation with the elastic force.

Acknowledgement

Research was in part supported through project ET36/2005 by UEFISCSU Romania.

References

[1] George C. J., Bennett G. F., Simoneaux D., George W. J., *Polychlorinated biphenyls. A toxicological review*, Journal of Hazardous Materials, 18(2), p. 113-144, 1988.

[2] Borja J., Taleon D. M., Auresenia J., Gallardo S., *Polychlorinated biphenyls and their biodegradation*, Process Biochemistry, 40(6), p. 1999-2013, 2005.

[3] Hertz-Picciotto I., Charles M. J., James R. A. , Keller J. A., Willman E., Teplin S., *In utero polychlorinated biphenyl exposures in relation to fetal and early childhood growth*, Epidemiology, 16(5), p. 648-656, 2005.

[4] Hansen B. G., Paya-Perez A. B., Rahman M., Larsen B. R., *QSARs for K(ow) and K(oc) of PCB congeners: A critical examination of data, assumptions and statistical approaches*, *Chemosphere*, 39(13), p. 2209-2228, 1999.

[5] Zou J.-W., Jiang Y.-J., Hu G.-X., Zeng M., Zhuang S.-L., Yu Q.-S., *QSPR/QSAR studies on the physicochemical properties and biological activities of polychlorinated biphenyls*, Acta Physico - Chimica Sinica, 21(3), p. 267-272, 2005.

[6] Eisler R., Belisle A. A., *Planar PCB Hazards to Fish, Wildlife, and Invertebrates: A Synoptic Review*, Contaminant Hazard Reviews, p. 1-96, 1996.

[7] Jäntschi L., *Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology*, Leonardo Electronic Journal of Practices and Technologies, AcademicDirect, 6, p. 76-98, 2005.

[8] ***, *Leave-one-out Analysis*, © 2005, Virtual Library of Free Software, available at: http://vl.academicdirect.org/molecular_topology/mdf_findings/loo.

[9] ***, *MDF SAR Predictor*, © 2005, Virtual Library of Free Software, available at: http://vl.academicdirect.org/molecular_topology/mdf_findings/sar.