

Prediction of relative response factors of electron-capture detection for some polychlorinated biphenyls using chemometrics

M. Jalali-Heravi^{a,*}, E. Noroozian^b, M. Mousavi^b

^a Department of Chemistry, Sharif University of Technology, P.O. Box 11365-9616, Tehran, Iran

^b Department of Chemistry, Shahid Bahonar University, P.O. Box 76175-133, Kerman, Iran

Received 9 May 2003; received in revised form 17 September 2003; accepted 8 October 2003

Abstract

The relative response factor (RRF) of an electron-capture detection (ECD) system is predicted for a set of 118 polychlorinated biphenyls (PCBs). Due to the wide range of relative retention times of PCB congeners, the RRFs of these compounds were calculated based on two different internal standards. Therefore, the compounds were divided into two molecular subsets. As a first step, multiple linear regression (MLR) was employed to find informative descriptors that can predict the RRFs of these compounds. Two descriptors of molecular ionization potential (MIP) and ionization potential of the molecule (IP) that are related to affinity of the compounds for the electrons show the highest mean effects in subsets 1 and 2, respectively. The descriptors appearing in the MLR models were considered as inputs for developing the back-propagation artificial neural networks (BP-ANN). Two networks with the architectures of 5-5-1 and 7-6-1 were generated for the prediction of RRFs of molecules of subsets 1 and 2, respectively. Comparison of the results indicates the superiority of neural networks over that of the MLR method indicating the nonlinear behaviors of the ECD system. Inspection of the models reveals that the surface of the molecules play different roles in response factors of two subsets due to rotation of one phenyl group with respect to the other for the subset consisting of larger number of chlorine atoms.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Neural networks; Regression analysis; Molecular descriptors; Response factors; Polychlorinated biphenyls

1. Introduction

Polychlorinated biphenyls (PCBs) are comprised of 209 distinct chlorine-substituted biphenyl structures (congeners). Each congener may have between 1 and 10 chlorine atoms, which may be located at various positions on the two linked benzene rings. The physical and chemical properties of each congener vary according to the number and position of the substituted chlorine atoms in the compounds. The dispersion of these PCB congeners by uncontrolled releases into the environment, their long-term stability, their lipophilicity resulting in bio-magnification up food chains, and questions of possible toxicity, together causes concern for their effects on the environment and have engendered a vast body of research over the past three decades [1–3]. The measurement of the large number of PCB congeners in commercial

or environmentally altered PCB mixtures requires the use of high resolution gas chromatography with sensitive and selective detection system such as electron-capture detection (ECD) or selected-ion monitoring mass spectrometry (MS-SIM) [4–8].

The development of sensitive and selective detectors has played a major role in the establishment of chromatography as an unrivaled analytical tool. The retention time can be used for identification of compounds, but it is well accepted that more than one compound can have a similar retention time. However, different detector responses can be used for peak identification of compounds with the same retention time. On the other hand, the response factor (RF) is the fundamental measure of the response of the detector to a given compound and can be considered as a correlation factor. Since numerous compounds are unavailable as standards, the development of a theoretical method for estimating response factor seems to be useful. The first work on the prediction of response factors of substituted benzenes and pyridines, was published by Katritzky and Gordeeva [9]

* Corresponding author. Tel.: +98-21-6005-718;

fax: +98-21-6012-983.

E-mail address: jalali@sharif.edu (M. Jalali-Heravi).

using a multivariate statistical partial least-squares (PLS) method. Also, Katritzky and coworkers applied the multiple linear and nonlinear regression methods to predict the retention time and response factor of different organic compounds [10,11]. Jalali-Heravi and coworkers have used artificial neural networks for predicting flame ionization detection (FID), thermal conductivity detection (TCD) and photoionization detection (PID) response factors for different series of organic molecules [12–14].

The main aim of the present work was the development of a quantitative structure property relationship (QSPR) for predicting relative response factors of PCB compounds obtained using ECD detection system.

2. Methods

Artificial neural networks (ANNs) are mathematical systems that simulate biological neural networks [15–17]. They consist of processing elements (nodes, neurons) organized in layers. Back-propagation neural networks (BNNs) are most often used in analytical applications. The back-propagation network receives a set of inputs, which are multiplied by each node and then a nonlinear transfer function is applied. The goal of training the network is to change the weights between the layers in a direction to minimize the output errors. The changes in the values of the weights can be obtained using Eq. (1):

$$\Delta W_{ij}(n) = \eta \delta_i O_j + \alpha W_{ij}(n-1), \quad (1)$$

where ΔW_{ij} is the change in the weight factor for each network node, δ_i the actual error of node i , and O_j is output of node j . The coefficients η and α are the learning rate and the momentum factor, respectively.

3. Experimental

3.1. Data set

ECD RRFs of different PCB congeners obtained at 300 °C was taken from [18]. The data set includes those PCB congeners that produced non-coeluted and completely resolved chromatographic peaks. Therefore, 118 out of 209 PCB congeners were used in the data set (Table 1). Because of the large scale of relative retention time of PCB congeners (almost 100 min), the relative response factors of these compounds were calculated based on 4-bromobiphenyl (<60 min retention time) and hexa-bromobiphenyl (>60 min retention time) as the internal standards [18]. Hence, the data set was divided into two molecular subsets on the basis of the type of the internal standard, which was used in RRF calculations. The subsets 1 and 2 consist of 52 and 66 compounds, respectively. The training and prediction sets for subsets 1 and 2 consist of 41 and 11, and 51 and 15 molecules, respectively. The training set for each subset was

Table 1
Compounds studied in this work

PCB no.	Structure	PCB no.	Structure
Subset 1			
1	2	44	2, 2', 3, 5'
3	4	45 ^a	2, 2', 3, 6
4	2, 2'	46	2, 2', 3, 6'
6 ^a	2, 3'	48	2, 2', 4, 5
7	2, 4	49	2, 2', 4, 5'
8	2, 4'	51	2, 2', 4, 6'
12 ^a	3, 4	52 ^a	2, 2', 5, 5'
13	3, 4'	53	2, 2', 5, 6'
15	4, 4'	56	2, 3, 3', 4'
16	2, 2', 3	63	2, 3, 4', 5
17	2, 2', 4	66	2, 3', 4, 4'
18 ^a	2, 2', 5	67	2, 3', 4, 5
19	2, 2', 6	70	2, 3', 4', 5
22	2, 3, 4'	74 ^a	2, 4, 4', 5
24	2, 3, 6	84	2, 2', 3, 3', 6
25	2, 3', 4	91	2, 2', 3, 4', 6
26 ^a	2, 3', 5	92	2, 2', 3, 5, 5'
29 ^a	2, 4, 5	95 ^a	2, 2', 3, 5', 6
31	2, 4', 5	96	2, 2', 3, 6, 6'
33	2', 3, 4	99	2, 2', 4, 4', 5
34	2', 3, 5	100	2, 2', 4, 4', 6
35	3, 3', 4	101	2, 2', 4, 5, 5'
37	3, 4, 4'	103	2, 2', 4, 5', 6
40	2, 2', 3, 3'	104 ^a	2, 2', 4, 6, 6'
41	2, 2', 3, 4	113	2, 3, 3', 5', 6
42 ^a	2, 2', 3, 4'		
Subset 2			
77	3, 3', 4, 4'	167 ^a	2, 3', 4, 4', 5, 5'
81	3, 4, 4', 5	169	3, 3', 4, 4', 5, 5'
82	2, 2', 3, 3', 4	170	2, 2', 3, 3', 4, 4', 5
83	2, 2', 3, 3', 5	171	2, 2', 3, 3', 4, 4', 6
85 ^a	2, 2', 3, 4, 4'	172	2, 2', 3, 3', 4, 5, 5'
87	2, 2', 3, 4, 5'	173	2, 2', 3, 3', 4, 5, 6
97	2, 2', 3', 4, 5	174 ^a	2, 2', 3, 3', 4, 5, 6'
105 ^a	2, 3, 3', 4, 4'	175	2, 2', 3, 3', 4, 5', 6
107 ^a	2, 3, 3', 4', 5	176	2, 2', 3, 3', 4, 6, 6'
110	2, 3, 3', 4', 6	177	2, 2', 3, 3', 4', 5, 6
114	2, 3, 4, 4', 5	178	2, 2', 3, 3', 5, 5', 6
118	2, 3', 4, 4', 5	179	2, 2', 3, 3', 5, 6, 6'
119	2, 3', 4, 4', 6	180	2, 2', 3, 4, 4', 5, 5'
122	2', 3, 3', 4, 5	183 ^a	2, 2', 3, 4, 4', 5', 6
126	3, 3', 4, 4', 5	185 ^a	2, 2', 3, 4, 5, 5', 6
128 ^a	2, 2', 3, 3', 4, 4'	187	2, 2', 3, 4, 5, 5', 6
129	2, 2', 3, 3', 4, 5	189	2, 3, 3', 4, 4', 5, 5'
130 ^a	2, 2', 3, 3', 4, 5'	191 ^a	2, 3, 3', 4, 4', 5', 6
131	2, 2', 3, 3', 4, 6	193	2, 3, 3', 4', 5, 5', 6
132	2, 2', 3, 3', 4, 6'	194	2, 2', 3, 3', 4, 4', 5, 5'
134	2, 2', 3, 3', 5, 6	195	2, 2', 3, 3', 4, 4', 5, 6
135	2, 2', 3, 3', 5, 6'	197	2, 2', 3, 3', 4, 4', 6, 6'
136	2, 2', 3, 3', 6, 6'	198	2, 2', 3, 3', 4, 5, 5', 6
137	2, 2', 3, 4, 4', 5	199	2, 2', 3, 3', 4, 5, 5', 6'
138	2, 2', 3, 4, 4', 5'	200	2, 2', 3, 3', 4, 5, 6, 6'
141 ^a	2, 2', 3, 4, 5, 5'	201 ^a	2, 2', 3, 3', 4, 5', 6, 6'
149	2, 2', 3, 4', 5', 6	202 ^a	2, 2', 3, 3', 5, 5', 6, 6'
151	2, 2', 3, 5, 5', 6	203	2, 2', 3, 4, 4', 5, 5', 6
154 ^a	2, 2', 4, 4', 5, 6'	205	2, 3, 3', 4, 4', 5, 5', 6
156	2, 3, 3', 4, 4', 5	206	2, 2', 3, 3', 4, 4', 5, 5', 6
157	2, 3, 3', 4, 4', 5'	207 ^a	2, 2', 3, 3', 4, 4', 5, 6, 6'
165	2, 3, 3', 5, 5', 6	208	2, 2', 3, 3', 4, 5, 5', 6, 6'
166	2, 3, 4, 4', 5, 6	209	2, 2', 3, 3', 4, 4', 5, 5', 6, 6'

^a Prediction set.

used for the generation of models and the prediction set was used for the evaluation of the generated models. The prediction sets were chosen randomly and consist of almost all types of molecules included in the training sets and therefore, are good representatives of the training sets.

3.2. Descriptor generation

A total of 91 separate molecular structure descriptors were calculated for each compound in the data sets. These descriptors can be classified into four major groups: topological, geometric, electronic and physicochemical. Topological descriptors were calculated using two-dimensional representation of the molecules. Geometric and electronic descriptors depend on the three-dimensional coordinates of atoms. Therefore, in order to calculate these types of descriptors one needs to optimize molecular structure for each compound. In the present work, three-dimensional structures of the molecules were optimized using the self-consistence molecular orbital method of AM1 (SCF-MO AM1) implemented in the MOPAC package (version 6) [19]. The topological descriptors consist of constitutional and connectivity indices. The calculations of the constitutional parameters such as number of chlorine atoms in different positions were straightforward. The connectivity indices based on Randic method [20] were calculated using a FORTRAN 77 program written in our laboratory. The boiling point of the compounds was used as a physicochemical descriptor and was taken from the literature. Some of the descriptors generated for each compound encoded similar information about the molecule of interest. Therefore, it was desirable to test each descriptor and eliminate those, which show high correlation ($R > 0.95$) with each other. A total of 21 out of the 91 descriptors showed high correlation and were removed from the consideration. Then, stepwise multiple linear regression (MLR) method was used to build the linear models that re-

late the RRFs to the structural parameters (descriptors). Two selected models for subsets 1 and 2 are presented in Table 2.

3.3. ANN generation

The ANN programs were written in FORTRAN 77 in our laboratory. The networks were generated using the descriptors appearing in the MLR models as inputs. A three-layer network with a sigmoid transfer function was design for each ANN. Before training the networks the input and output values were normalized between 0.1 and 0.9. The number of nodes in the hidden layers, learning rate, and momentum were optimized. The initial weights were selected randomly between -0.3 and 0.3 . As can be seen from Table 2, the MLR model for the subset 1 includes five and that of the subset 2 includes seven descriptors, respectively. Therefore, the number of input for subsets 1 and 2 were five and seven, respectively. The number of nodes in the output layer for both subsets was set to be one. In order to evaluate the performance of the ANNs, the standard error of training (SET) and the standard error of prediction (SEP) were used. All calculations in this work were performed using a Pentium 4, 1.8 GHz PC with 256 M RAM.

4. Results and discussion

Table 2 shows the specifications of two selected MLR models for subsets 1 and 2. The mean effect for each parameter is also included in this table. Inspection of the variables appearing in the MLR models reveals that these parameters encoded different aspects of the molecular structure and properties. In ECD, tendency of a solute for capturing the electrons determines the decreases in the electrical current between the two electrodes that is registered as a response. Two descriptors of molecular ion ionization potential (MIIP) and ionization potential of the molecules (IP) have appeared

Table 2
Specifications of the selected multiple linear regression models


Descriptor	Notation	Coefficient	Mean effect
Subset 1			
Molecular ion ionization potential	MIIP	6.1014 (± 1.0059)	10.6165
Ratio of principle axes of rotation about x and z axes	RPAR (yz)	-0.0799 (± 0.0120)	0.7378
Principle moment of inertia of the molecule about z axis	I_z	$-9.6833E-4$ ($\pm 2.5652E-4$)	-2.5150
Principle moment of inertia of the molecule about x axis	I_x	$-5.1568E-4$ ($\pm 2.1259E-4$)	-0.6563
Surface of the molecule on yz plane	S_{yz}	-0.1110 (± 0.0225)	-4.6296
Constant		-1.1752 (± 1.0350)	
Subset 2			
4X_c Randic connectivity index	RCI	0.6554 (± 0.2425)	0.4194
Ionization potential of the molecule	IP	1.2981 (± 0.3830)	12.5654
Number of Cl atoms in biphenyl ring with lower number of Cl atom	NOCL	-0.1835 (± 0.0887)	-0.4900
Number of configuration  in the molecule	CONF	0.0770 (± 0.0301)	0.2763
Moment of inertia of the molecule	I_{xz}	$2.5151E-4$ ($\pm 9.5544E-4$)	-0.0250
Surface of the molecule on xy plane	S_{xy}	0.0116 (± 0.0092)	0.8270
Surface of the molecule on xz plane	S_{xz}	-0.0013 (± 0.0212)	-0.0774
Constant		-12.7974 (± 3.1156)	

Table 3
Architectures of the ANNs for subsets 1 and 2

	Subset 1	Subset 2
Number of nodes in the input layer	5	7
Number of nodes in the hidden layer	5	6
Number of nodes in the output layer	1	1
Number of iteration in the beginning of over-fitting	160000	29000
Learning rate	0.8	0.5
Momentum	0.9	0.9

in the statistical models of the subsets 1 and 2. The parameter of MIIP is the ionization potential of the negative ion form of the molecules. The differences between this parameter and the ionization potential of corresponding neutral molecule can be considered as a measure of the affinity of the molecules for capturing the electrons. The MIIP parameter shows a high correlation with the mentioned differences. Therefore, one may conclude that this parameter is related to the tendency of the molecules for capturing the electrons. The IP is based on the validity of the Koopmans theorem, $IP = -E_{HOMO}$. These parameters show very high mean effects with comparison to the other descriptors. The surfaces

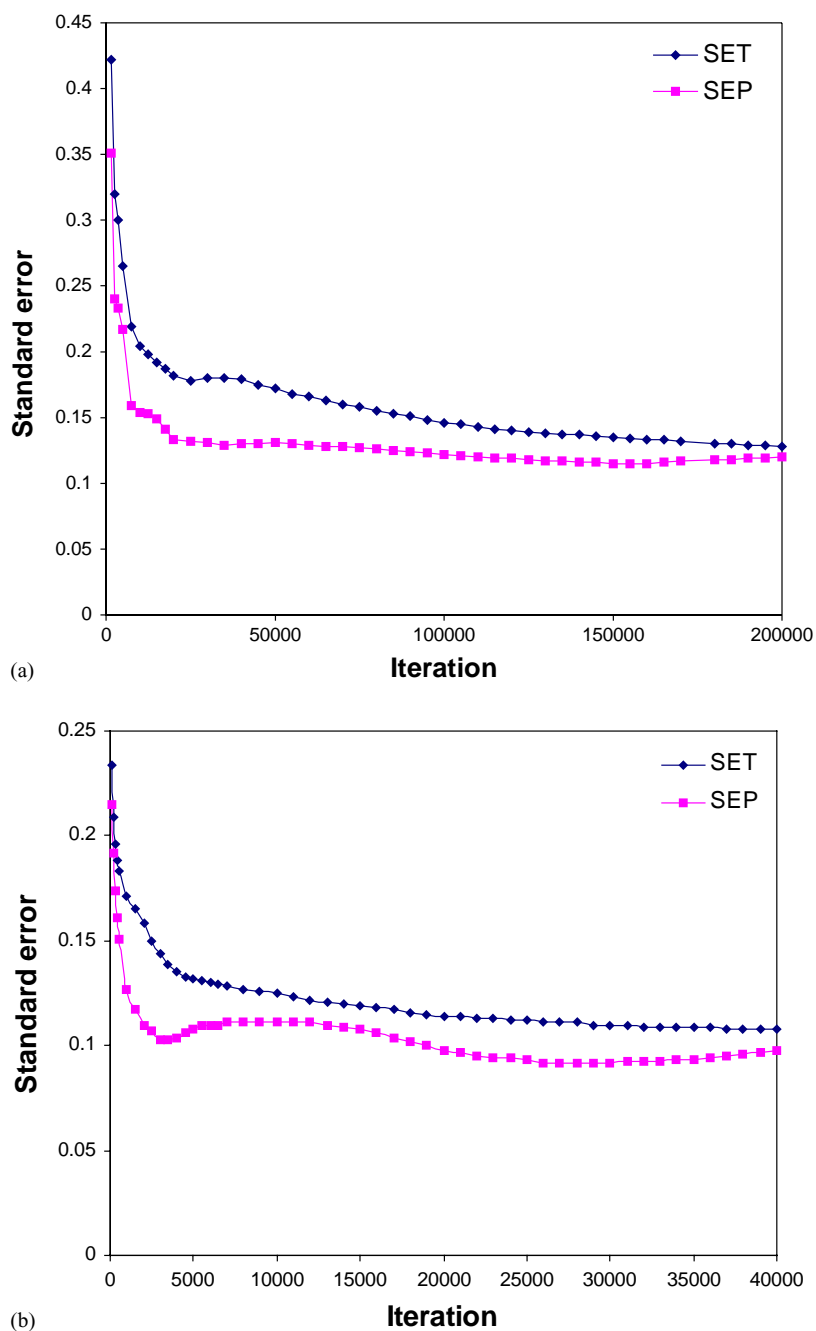


Fig. 1. A typical learning curve for ANN: (a) subset 1, (b) subset 2.

Table 4
Experimental and calculated values of the RRFs for the training and prediction sets

PCB no.	RRF _{exp}	RRF _{NNET}	RRF _{MLR}
Subset 1 ^a			
1	0.025	0.028	0.042
3	0.011	0.000	1.625
4	0.044	0.211	0.116
7	0.919	0.786	0.999
8	0.537	0.705	0.699
13	0.102	0.052	0.165
15	0.102	0.107	0.597
16	0.828	0.492	1.177
17	0.671	0.851	1.181
19	0.358	0.360	0.054
22	1.312	1.280	-0.255
24	1.393	1.387	1.298
25	1.585	1.396	0.333
31	0.677	0.643	1.427
33	1.126	1.405	1.009
34	0.584	0.565	1.030
35	0.454	0.443	0.836
37	0.448	0.476	0.803
40	1.615	1.407	0.667
41	1.560	1.695	0.618
44	0.760	0.860	1.152
46	0.666	0.555	1.502
48	0.539	0.784	1.190
49	0.846	0.971	1.109
51	0.673	0.637	1.353
53	0.545	0.617	0.792
56	1.829	1.760	1.125
63	1.700	1.791	0.203
66	1.128	1.157	1.505
67	1.183	1.204	0.592
70	1.495	1.267	1.353
84	0.870	0.912	0.787
91	0.999	0.823	1.368
92	0.881	0.929	1.168
96	0.582	0.646	1.014
99	0.762	0.800	0.818
100	1.255	1.333	0.613
101	0.951	0.814	0.909
103	0.753	0.864	1.209
113	1.893	1.864	0.891
121	1.606	1.591	1.192
Prediction set 1			
6	0.355	0.256	0.539
12	0.187	0.015	0.088
18	0.459	0.586	0.214
26	0.672	0.645	0.617
29	1.132	1.187	1.092
42	1.648	1.589	1.180
45	0.575	0.825	1.059
52	0.651	0.696	0.531
74	1.527	1.622	1.393
95	0.900	0.773	1.036
104	0.679	0.683	0.798
Subset 2			
77	0.245	0.445	0.350
81	0.369	0.518	0.517
82	0.589	0.640	0.648
83	0.529	0.509	0.472
87	0.776	0.681	0.627

Table 4 (Continued)

PCB no.	RRF _{exp}	RRF _{NNET}	RRF _{MLR}
97	0.393	0.505	0.461
110	0.505	0.597	0.563
114	1.184	1.058	0.892
118	0.440	0.385	0.539
119	0.627	0.445	0.483
122	0.602	0.682	0.723
126	0.374	0.342	0.452
129	0.715	0.616	0.729
131	0.549	0.540	0.622
132	0.541	0.475	0.433
134	0.488	0.522	0.549
135	0.429	0.388	0.342
136	0.364	0.388	0.303
137	1.110	0.772	0.851
138	0.722	0.698	0.727
149	0.384	0.455	0.505
151	0.493	0.441	0.597
156	1.005	1.070	0.873
157	0.876	0.928	0.850
165	0.577	0.769	0.772
166	1.241	1.281	1.264
169	1.015	0.873	0.797
170	0.777	0.791	0.843
171	0.768	0.706	0.783
172	0.756	0.780	0.711
173	1.111	1.101	1.177
175	0.539	0.590	0.543
176	0.481	0.504	0.484
177	0.555	0.656	0.657
178	0.480	0.574	0.527
179	0.413	0.481	0.416
180	0.823	0.937	0.956
187	0.488	0.557	0.692
189	1.304	1.022	0.959
193	1.000	0.991	0.928
194	1.158	1.026	0.942
195	0.907	0.869	0.982
197	0.492	0.525	0.627
198	0.874	0.853	0.803
199	0.776	0.706	0.712
200	0.585	0.528	0.419
203	0.843	1.062	1.029
205	0.845	0.936	0.930
206	1.045	1.094	0.987
208	0.928	0.891	0.737
209	0.853	0.856	1.160
Prediction set 2			
85	0.641	0.724	0.757
105	1.032	0.910	0.812
107	0.511	0.531	0.548
128	0.829	0.770	0.818
130	0.534	0.733	0.612
141	0.871	0.955	0.883
154	0.323	0.375	0.513
167	0.443	0.591	0.524
174	0.522	0.539	0.554
183	0.795	0.715	0.745
185	1.137	1.075	0.985
191	0.860	0.900	0.922
201	0.819	0.756	0.630
202	0.445	0.550	0.435
207	0.729	0.676	0.768

^a Numbers are referred to the molecules given in Table 1.

of the molecules on yz (S_{yz}), xy (S_{xy}) and xz (S_{xz}) planes are also appeared in the models of the subsets 1 and 2, respectively. The values of these surfaces depend upon the number of chlorine atoms of the PCBs. It can be seen from Table 1 that subset 1 mainly consists of the PCB compounds with less than five chlorine atoms, whereas the majority of the molecules included in the subset 2 have more than five chlorine atoms. Table 2 shows that the contribution of S_{yz} to the relative response factor is negligible while S_{yz} and S_{xy} show a considerable contribution to the RRF of the molecules included in the subsets 1 and 2, respectively. Inspection of the variation of these surfaces shows that as the dihedral angle

Table 5
Statistical parameters obtained using the ANN and MLR models for RRFs

Model	SET	SEP	R_{training}	$R_{\text{prediction}}$	F_{training}	$F_{\text{prediction}}$
Subset 1						
MLR	0.260	0.234	0.882	0.845	24	23
ANN	0.129	0.116	0.969	0.971	139	164
Subset 2						
MLR	0.147	0.078	0.860	0.893	17	51
ANN	0.096	0.072	0.915	0.927	253	80

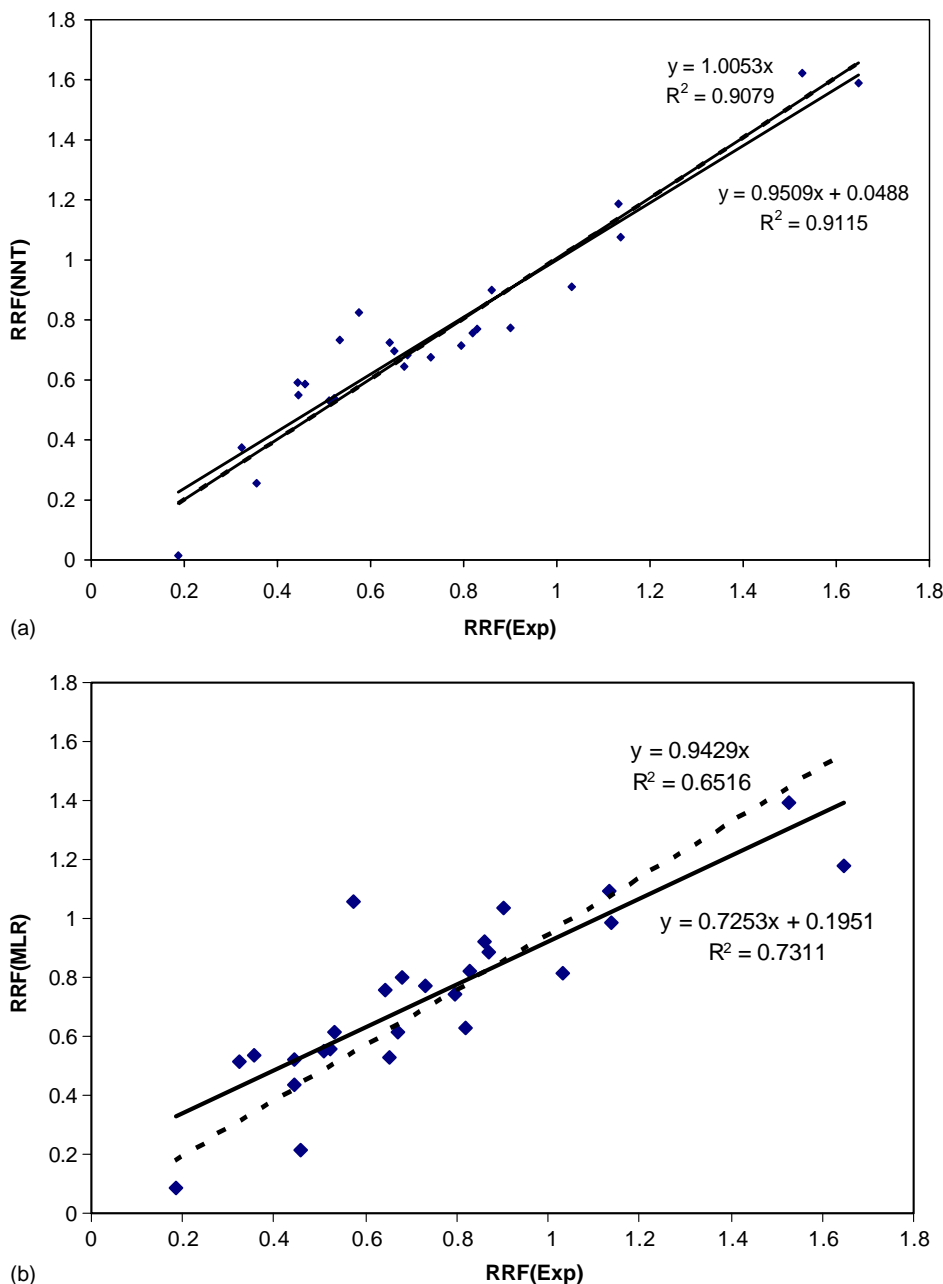


Fig. 2. Plots of the calculated vs. the experimental values of RRFs for the prediction sets. (a) ANN model, (b) MLR model.

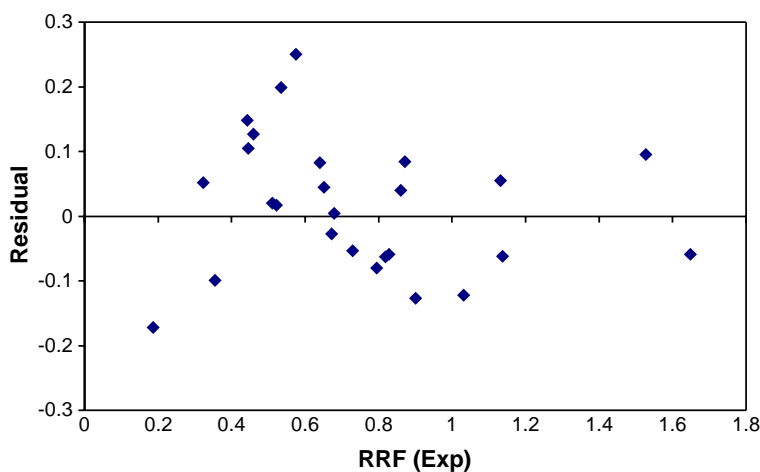


Fig. 3. Plot of residuals vs. the experimental values of RRFs for the prediction sets.

between the two rings of a PCB molecule decreases, S_{xy} increases while S_{yz} and S_{xz} decrease. The appearance of the surface of the molecule parameters in the models indicates that these surfaces play a major role in capturing the electrons in ECD system. The contribution of S_{yz} to the relative response factor is negligible, whereas S_{yz} and S_{xy} show a considerable contribution to the molecules included in the subsets 1 and 2, respectively. As the number of the chlorine atoms increases, the S_{yz} decreases and therefore, the RRF decreases. On the other hand, as the number of chlorine atoms increases, S_{xy} increase and therefore, the RRF increases.

The main goal of the present study was generation of the ANN for modeling of ECD RRFs. Before the training of these networks, the parameters of the number of nodes in the hidden layer, learning rate and momentum were optimized. The procedure for the optimization of these parameters is reported in [12,13,21]. The architectures and specifications of the optimized ANN for ECD system are shown in Table 3. In order to control overfitting of the networks during the training procedure, the SET and SEP values were recorded after each 500 iterations (Fig. 1). Each 500 iterations take a few seconds by using the Pentium 4, 1.8 GHz PC. In the case of subset 1, after 160 000 iterations the values of SEP started to increase and overtraining began and for subset 2 overtraining began after 29 000 iterations. For the evaluation of the prediction ability of the ANNs, the trained ANNs were used to predict the RRFs of the molecules included in the prediction sets. The calculated values of RRFs using the generated ANNs for the training and prediction sets are presented in Table 4. The statistical parameters, such as correlation coefficient (R) between the calculated and experimental values of RRFs and standard errors (SEs) for the training and prediction sets obtained using the ANNs and MLR models are shown in Table 5. Inspection of SET and SEP values for the ANNs and MLR methods reveal the superiority of the neural networks over that of the MLR in predicting of the RRFs. It is noteworthy that the descriptors appearing in

the MLR model were used as inputs for developing of the neural networks. On the other hand, a subset of descriptors was chosen from the reduced descriptor pool by a multiple linear regression analysis. These descriptors were then submitted to a computational neural network to develop a nonlinear model. In fact the linear method of MLR was chosen as a feature selection method for developing the nonlinear model of ANN. However, the improvement in performance of the ANN model means that it adapts to the descriptors better by allowing for a degree of non-linearity. Therefore, one may conclude that RRFs show nonlinear characteristics. As can be seen from Table 5, the values of the SET and SEP are comparable for both subsets and one may conclude that the prediction sets are good representative of the training sets. Fig. 2 shows the plot of the ANN and MLR calculated values of RRFs against the experimental ones. In order to demonstrate the absence of bias, zero intercept unit slope line is also shown as a dashed-line in Fig. 2a for the ANN model. Fig. 2b shows similar plots for the MLR calculated RRFs values. Comparison of zero intercept plots for the MLR and ANN models indicates the superiority of the ANN over that of the MLR model. Fig. 3 shows the plot of residuals against the experimental values of RRFs for the ANN models. The propagation of the residuals in both sides of zero indicates that no systematic error exist in the development of the ANNs.

5. Conclusions

Comparison of the values of the SET and SEP obtained using models of ANN and MLR for predicting of RRFs shows superiority of the neural networks over that of linear regression models. Inspection of the statistical parameters (Table 5) indicates that the use of the linear models in predicting of RRFs of ECD detector is not justified. Since the improvement of the results obtained using nonlinear models of artificial neural networks is considerable, one may conclude

that the nonlinear characteristics of the RRFs are serious. In the present work, ANN models were successfully developed using five and seven descriptors as input for the 52 and 66 molecules of the subsets 1 and 2, respectively. Inspection of the results reveals that the parameters relating to the affinity of the molecules for electrons such as MIIP and IP of the molecules play the major role in capturing the electrons in ECD detection systems. In addition, the surface area of the molecules and therefore, the position of chlorine atoms on the rings is important in the responses of ECD to PCBs.

References

- [1] F. Lepine, S. Milot, N. Vincent, *Bull. Environ. Contam. Toxicol.* 48 (1992) 152.
- [2] D.L. Bedard, J.F. Quensen, in: L.Y. Young, C. Cerniglia (Eds.), *Microbial Transformation and Degradation of Toxic Organic Chemicals*, Wiley, New York, 1995, p. 127.
- [3] D.L. Bedard, R.E. Wagner, M.J. Brennen, M.L. Haberl, J.F. Brown, *Appl. Environ. Microbiol.* 53 (1987) 1094.
- [4] G.M. Frame, J.W. Cochran, S.S. Bowadt, *J. High Resolut. Chromatogr.* 19 (1996) 657.
- [5] B. Larsen, *J. High Resolut. Chromatogr.* 18 (1995) 1.
- [6] G.M. Frame, *Fresenius J. Anal. Chem.* 357 (1997) 701.
- [7] G.M. Frame, *Fresenius J. Anal. Chem.* 357 (1997) 714.
- [8] M.S. Rahman, S. Bowadt, B. Larson, *J. High Resolut. Chromatogr.* 16 (1994) 87.
- [9] A.R. Katritzky, E.V. Gordeeva, *J. Chem. Inf. Comput. Sci.* 33 (1993) 835.
- [10] A.R. Katritzky, E.S. Ignatchenko, R.A. Barcock, V.S. Lobanov, M. Karelson, *Anal. Chem.* 66 (1994) 1799.
- [11] B. Lucic, N. Trinajstic, S. Sild, M. Karelson, A.R. Katritzky, *J. Chem. Inf. Comput. Sci.* 39 (1999) 610.
- [12] M. Jalali-Heravi, M.H. Fatemi, *J. Chromatogr. A* 825 (1998) 161.
- [13] M. Jalali-Heravi, M.H. Fatemi, *J. Chromatogr. A* 897 (2000) 227.
- [14] M. Jalali-Heravi, Z. Garkani-Nejad, *J. Chromatogr. A* 950 (2002) 183.
- [15] D.W. Patterson, *Artificial Neural Networks: Theory and Applications*, Simon and Schuster, New York, 1996.
- [16] J. Zupan, J. Gasteiger, *Anal. Chim. Acta* 248 (1991) 1.
- [17] N.K. Bose, P. Liang, *Neural Network-Fundamentals*, McGraw-Hill, New York, 1996.
- [18] E. Zeng, M. Castillo, A. Khan, Southern Coastal Water Research Project (SCCWRP), Annual Report, 1992–1993; <http://www.sccwrp.org/pubs/annrpt/92-93/ar-03.htm>.
- [19] MOPAC Package, Version 6, US Air Force Academy, Colorado Springs, CO.
- [20] M. Randic, *J. Chem. Phys.* 62 (1975) 309.
- [21] M. Jalali-Heravi, Z. Garakani-Nejad, *J. Chromatogr. A* 927 (2001) 211.