Hindawi Publishing Corporation
Journal of Chemistry
Volume 2013, Article ID 908586, 13 pages
http://dx.doi.org/10.1155/2013/908586

Research Article

Prediction of Gas Chromatography-Mass Spectrometry Retention Times of Pesticide Residues by Chemometrics Methods

Elaheh Konoz, Amir H. M. Sarrafi, Alireza Feizbakhsh, and Zahra Dashtbozorgi

Department of Chemistry, Islamic Azad University, Central Tehran Branch, Tehran 13185-768, Iran

Correspondence should be addressed to Elaheh Konoz; konozelaheh@gmail.com

Received 14 January 2012; Accepted 30 April 2012

Academic Editor: Yenamandra S. Prabhakar

A quantitative structure-retention relationships (QSRRs) method is employed to predict the retention time of 300 pesticide residues in animal tissues separated by gas chromatography-mass spectroscopy (GC-MS). Firstly, a six-parameter QSRR model was developed by means of multiple linear regression. The six molecular descriptors that were considered to account for the effect of molecular structure on the retention time are number of nitrogen, Solvation connectivity index-chi 1, Balaban *Y* index, Moran autocorrelation-lag 2/weighted by atomic Sanderson electronegativity, total absolute charge, and radial distribution function-6.0/unweighted. A 6-7-1 back propagation artificial neural network (ANN) was used to improve the accuracy of the constructed model. The standard error values of ANN model for training, test, and validation sets are 1.559, 1.517, and 1.249, respectively, which are less than those obtained reveals by multiple linear regressions model (2.402, 1.858, and 2.036, resp.). Results obtained the reliability and good predictability of nonlinear QSRR model to predict the retention time of pesticides.

1. Introduction

Pesticides are used on a large scale for agricultural purposes. The adverse effects of pesticides on both human health and the environment are a matter of public concern. Thus, both the actual state and the transition of pesticide residues in various matrices including water, soil, and agricultural products should be extensively monitored. These researches should be undertaken using an efficient analytical system with a laborsaving and cost-effective device, as pesticides as well as applicable fields of research rang over a broad spectrum. Conventional sample preparation methods used to analyze pesticide residues in various matrices require expensive instrumentation, an expert analyst [1–4]. Besides the above mentioned, the experimental determination of chromatographic retention parameters of pestocides is time consuming and expensive. Alternatively, quantitative structure-retention relationship (QSRR) provides a promising method for the estimation of retention time based on descriptors derived solely from the molecular structure to fit experimental data. The advantages of this approach lie in the fact that it requires only the knowledge of chemical structure and is not dependent on any experiment properties.

QSRR studies [5, 6] started from the calculation and selection of descriptors, to finding their relation to retention times and derivation of mathematical models that involve these multivariate data in order to be used for predictive purposes in chromatographic system. Multivariate data consist of the results of observations of many different variables (molecular descriptors) for a number of individuals (molecules). Known methods for this include the multiple regression analysis, experimental design techniques, and nonlinear regression. The drawback, sometimes, of these very popular techniques is their inability to give highly predictive models due to hidden nonlinearity inside the data variables or the prerequisite to specify the mathematical model before the fitting of the data. So there is a need to improve further such kind of models in order to extract the most accurate prediction. To this end, artificial neural networks (ANNs) could be used successfully in QSRR studies providing better results than the conventional regression models.

Today artificial neural networks [7–10] have become an important modeling technique for QSAR and QSPR, and also this technique has been applied in numerous application areas of chemistry and pharmacy [11–16]. The mathematical

adoptability of ANN commends them as powerful tool for pattern classification and building predictive models. A particular advantage of ANNs is their inherent ability to incorporate nonlinear dependencies between the dependent and independent variables without using an explicit mathematical function. There are few reports about the application of QSRR in the chromatographic studies: D'Archivio et al. modeled the combined effect of solute structure and eluent composition on the retention behavior of 26 pesticides in isocratic reversed-phase high-performance liquid chromatography using multilinear regression and artificial neural networks [17]. In another study, they applied a six-parameter nonlinear QSRR model to predict the retention behavior of 26 pesticides including commonly used insecticides, herbicides, and fungicides as well as some metabolites in reversed-phase high-performance liquid chromatography [18]. Also Ghasemi and his coworkers used multiple linear regression and partial least squares regression to QSRR study of the gas chromatography retention time of 38 diverse chlorinated pesticides, herbicides, and organohalides by using molecular descriptors [19]. In the present study, the application of ANN is being described in order to predict accurately the retention time values of 300 pesticides in four groups with different molecular structures [20].

2. Methods

2.1. Dataset. Development of the multiple linear regression and artificial neural networks in the present work relies on a data set taken from reference [20]. This dataset (Table 1) consists of 300 pesticides in animal tissues such as beef, mutton, pork, chicken, and rabbit ranging in retention time from 5.62 to 35.77 min. All of these 300 pesticides divided into four groups, depending upon properties and retention time of each pesticide. Each group is consisting different kind of pesticides such as acaricides, insecticides, and Fungicides. To apply the ANN modeling, the dataset was randomly divided into three groups of training, test, and validation sets consisting of 256, 22, and 22 pesticides, respectively. The training set was used for the model generation. The test set plays a different role in the cases of the MLR and the ANN models. For the ANN model, this set was used for early stopping to optimize learning iteration and avoid overtraining. The validation set was used to assess the accuracy of the ANN predictions. On the other hand, in the case of the MLR model, the test set and the validation set were used to evaluate the model. As can be seen from Table 1, the pesticides in the test and validation sets were chosen in a way that adequately represents the training set in term of retention time.

2.2. Molecular Descriptors. All structures were generated with the HyperChem (Version 7) [21] and optimized with the classic potential MM^+ included. Molecular geometry was optimized with the Austin Model 1 (AM1) method [22], and then the molecular descriptors were calculated by the software Dragon 3.0 [23]. Overall more than 1400 theoretical descriptors were calculated for each molecule by this software. These descriptors can be classified into several groups: 0D: constitutional descriptors; 1D: functional groups, atom-centered fragments, empirical descriptors and molecular properties; 2D: topological descriptors, molecular walk counts, BCUTs descriptors, Galvez topological charge indices, and 2D autocorrelations; 3D: aromaticity indices, Randic molecular profiles from the geometry matrix, geometrical, RDF, 3D-MORSE, WHIMs, and GETAWAYs descriptors. Molecular descriptor meanings and their calculation procedure are explained in Handbook of molecular descriptors by Todeschini. These molecular descriptors of different kinds were used to describe compound chemical diversity.

2.3. Regression Analysis. The main goal of the generation of the MLR model was to choose a set of suitable descriptors that can be used as inputs for construction of the ANN model. Linear models were formed by a stepwise selection of important descriptors and MLR model construction [24]. The best MLR model is one that has high correlation coefficient and F-value, low standard error, and high prediction power. The statistics of the constructed MLR model is presented in Table 2.

2.4. Artificial Neural Network. A detailed description of theory behind artificial neural networks has been adequately described in several publications [25–31]. An ANN program was written in MATLAB 7 in our laboratory. This network was feed-forward fully connected that has three layers with sigmoidal transfer function. Descriptors appearing in the MLR models were used as inputs of network and signal of the output node represent the retention time of interested compound. Thus, this network has six nodes in input layer and one node in output layer. The value of each input was divided into its mean value to bring them into dynamic range of the sigmoid transfer function of the network. The back-propagation algorithm was used for the training of the network [32, 33]. Before training the network, the parameters of the number of nodes in the hidden layer, weights and biases learning rates, and momentum values were optimized [34–36]. Optimized values of these parameters were numbers of nodes in the input layer (=6), numbers of nodes in the hidden layer (=7), numbers of nodes in the output layer (=1), weights learning rates (=0.1), bias learning rate (=0.4), momentum (=0.5), and transfer function was sigmoid. The ANN-calculated values of permeability coefficient for training, test, and prediction sets, are shown in Table 1.

3. Results and Discussion

3.1. Analysis by Multiple Linear Regressions. The best MLR model (Table 2) for the training set includes six descriptors. These descriptors are number of nitrogen, Solvation connectivity index-chi 1, Balaban Y index, Moran autocorrelationlag 2/weighted by atomic Sanderson electronegativity, total absolute charge, and radial distribution function-6.0/unweighted.

Table 1: Experimental, ANN, and MLR predicted values of the retention times (t_R) of training, test, and validation sets.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
Training set					
1	Allidochlor	8.78	10.22	9.83	−1.05
2	Dichlormid	9.74	10.27	9.88	−0.14
3	Etridiazol	10.42	16.11	12.9	−2.48
4	Chlormephos	10.53	11.88	14	−3.47
5	Propham	11.36	14.17	12.13	−0.77
6	Cycloate	13.56	16.32	16.4	−2.84
7	Diphenylamine	14.55	18.13	15.58	−1.03
8	Chlordimeform	14.93	17.44	13.69	1.24
9	Ethalfluralin	15.00	16.17	14.56	0.9
10	Thiometon	16.20	18.1	17.82	−1.62
11	Atrazine-desethyl	16.76	13.96	17.25	−0.49
12	Clomazone	17.00	20.95	18.63	−1.63
13	Fonofos	17.31	20.48	16.24	1.07
14	Simazine	17.85	16.99	18.93	−1.08
15	Propetamphos	17.97	19.72	18.93	−0.96
16	Secbumeton	18.36	20.2	19.21	−0.85
17	Dichlofenthion	18.80	21.02	17.38	1.42
18	Pronamide	18.72	19.61	21.15	−2.43
19	Mexacarbate	18.83	17.05	18.64	0.2
20	Aldrin	19.67	22.92	21.33	−1.66
21	Dinitramine	19.35	18.57	16.45	2.9
22	Ronnel	19.80	22.63	21.2	−1.4
23	Cyprazine	20.18	19.82	18.43	1.75
24	Beta-HCH	20.31	16.11	16.93	3.38
25	Metalaxyl	20.67	20.19	21.92	−1.25
26	Chlorpyrifos (-ethyl)	20.96	25.26	22.59	−1.63
27	Methyl-parathion	20.82	22.85	21.44	−0.62
28	Malathion	21.54	21.2	22.9	−1.36
29	Fenitrothion	21.62	23.17	21.19	0.43
30	Paraoxon-ethyl	21.57	22.19	21.39	0.18
31	Triadimefon	22.22	25.41	25.62	−3.4
32	Parathion	22.32	24	23.28	−0.96
33	Pendimethalin	22.59	20.84	19.77	2.82
34	Linuron	22.44	21.13	19.94	2.5
35	Bromophos-ethyl	23.06	23.18	21.51	1.55
36	trans-chlordane	23.29	23.4	22.61	0.68
37	Phenthoate	23.30	24.91	25.59	−2.29
38	Fenothiocarb	23.79	23.76	24.85	−1.06
39	Prothiophos	24.04	24.99	23.69	0.35
40	Dieldrin	24.43	24.77	24.55	−0.12
41	Procymidone	24.36	22.91	26.47	−2.11
42	Methidathion	24.49	27.18	26.96	−2.47
43	Napropamide	24.84	24.35	23.12	1.72
44	Fenamiphos	25.29	23.4	26.85	−1.56
45	Aramite	25.60	26.63	25.36	0.24
46	Bupirimate	26.00	25.11	24.63	1.37
47	Carboxin	26.25	23.23	25.54	0.71
48	Flutolanil	26.23	22.66	27.39	−1.16

Table 1: Continued.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
49	Ethion	26.69	26.7	26.6	0.09
50	Sulprofos	26.87	25.6	25.14	1.73
51	Etaconazole-2	26.89	22.06	26.1	0.79
52	Myclobutanil	27.19	26.34	26.45	0.74
53	Diclofop-methyl	28.08	28.79	27.35	0.73
54	Propiconazole	28.15	30.15	28.96	−0.81
55	Fensulfothin	27.94	25.84	26.94	1
56	Bifenthrin	28.57	27.56	29.69	−1.12
57	Mirex	28.72	28.48	28.33	0.39
58	Benodanil	29.14	22.77	27.08	2.06
59	Nuarimol	28.90	26.9	26.93	1.97
60	Oxadexyl	29.59	23.49	25.44	4.06
61	Tetramethirn	29.59	28.2	28.88	0.71
62	Phosmet	30.46	29.34	29.94	0.52
63	Oxycarboxin	31.00	25.76	28.63	2.37
64	cis-Permethrin	31.42	31.5	32.3	−0.9
65	trans-Permethrin	31.68	32.74	32.6	−0.92
66	Pyrazophos	31.60	32.87	29.97	1.63
67	Cypermethrin	33.19	33.64	33.73	−0.54
68	Fenvalerate	34.45	35.59	34.07	0.38
69	Deltamethrin	35.77	32.69	33.93	1.84
70	EPTC	8.54	9.64	11.15	−2.61
71	Butylate	9.49	10.01	10.12	−0.63
72	Dichlobenil	9.75	11.27	11.75	−2
73	Pebulate	10.18	11.32	11.54	−1.36
74	Nitrapyrin	10.89	10.95	12.12	−1.23
75	Chloroneb	11.85	15.37	13.01	−1.16
76	Tecnazene	13.54	15.8	16.61	−3.07
77	Heptenophos	13.78	18.84	15.01	−1.23
78	Hexachlorobenzene	14.69	16.1	18.31	−3.62
79	Ethoprophos	14.40	13.64	12.5	1.9
80	cis-Diallate	14.75	14.67	15.7	−0.95
81	Propachlor	14.73	16.12	13.97	0.76
82	Trifluralin	15.23	19.7	14.95	0.28
83	Chlorpropham	15.49	16.41	16.17	−0.68
84	Sulfallate	15.75	12.79	12.78	2.97
85	Alpha-HCH	16.06	15.99	18.04	−1.98
86	Terbufos	16.83	16.38	17.13	−0.3
87	4,4-DDE	23.92	24.19	24.68	−0.76
88	Chlorbufam	17.85	18.15	18.66	−0.81
89	Fluotrimazole	28.39	26.67	28.64	−0.25
90	Terbuthylazine	18.07	18.42	19.7	−1.63
91	Monolinuron	18.15	19.36	20.27	−2.12
92	Cyanophos	18.73	20.34	19.53	−0.8
93	Chlorpyrifos-methyl	19.38	23.86	21.45	−2.07
94	Desmetryn	19.64	17.4	18.43	1.21
95	Alachlor	20.03	19.24	19.27	0.76
96	Terbutryn	20.61	20.03	19.37	1.24
97	Thiobencarb	20.63	20.77	19.73	0.9
98	Dicofol	21.33	23.76	23.45	−2.12
99	Metolachlor	21.34	21.05	21.68	−0.34

Table 1: Continued.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
100	Methoprene	21.71	23.08	22.07	−0.36
101	Bromofos	21.75	23.72	22.31	−0.56
102	Ethofumesate	21.84	24.27	23.8	−1.96
103	Isopropalin	22.10	26.6	24.73	−2.63
104	Propanil	22.68	16.94	19.79	2.89
105	Crufomate	22.93	21.31	19.59	3.34
106	Chlorfenvinphos	23.19	23.53	23.32	−0.13
107	cis-Chlordane	23.55	23.4	21.71	1.84
108	Tolylfluanide	23.45	22.76	23.52	−0.07
109	Butachlor	23.82	22.55	24.27	−0.45
110	Chlozolinate	23.83	25.54	23.06	0.77
111	Crotoxyphos	23.94	24.26	24.8	−0.86
112	Iodofenphos	24.33	23.85	21.87	2.46
113	Chlorbromuron	24.37	21.89	20.71	3.66
114	Profenofos	24.65	22.73	23.34	1.31
115	Buprofezin	24.87	24.53	23.88	0.99
116	2,4′-DDD	24.94	23.16	23.56	1.38
117	Endrin	25.15	24.92	25.1	0.05
118	Hexaconazole	24.92	26.84	26.97	−2.05
119	2,4-DDT	25.56	22.31	23.68	1.88
120	Methoprotryne	25.63	26.87	25.6	0.03
121	Erbon	25.68	25.93	26.83	−1.15
122	Chloropropylate	25.85	25.43	25.32	0.53
123	Nitrofen	26.12	23.08	28.85	−2.73
124	Oxyfluorfen	26.13	25.59	29.37	−3.24
125	Chlorthiophos	26.52	26.26	25.96	0.56
126	Endosulfan I	26.72	23.96	26.39	0.33
127	4,4-DDT	27.22	22.9	25.15	2.07
128	Carbofenothion	27.19	27	26.87	0.32
129	Benalyxyl	27.54	27.7	27.98	−0.44
130	Edifenphos	27.94	25.87	25.96	1.98
131	Triazophos	28.23	27.84	26.84	1.39
132	Chlorbenside sulfone	28.88	26.15	27.85	1.03
133	Endosulfan-sulfate	29.05	28.31	28.93	0.12
134	Bromopropylate	29.30	25.93	27.74	1.56
135	Benzoylprop-ethyl	29.40	28.81	28.84	0.56
136	Fenpropathrin	29.56	28.79	29.79	−0.23
137	Phosalone	31.22	31.11	31.56	−0.34
138	Azinphos-methyl	31.41	31.36	30.37	1.04
139	Fenarimol	31.65	29.56	28.81	2.84
140	Azinphos-ethyl	32.01	32.91	31.79	0.22
141	Prochloraz	33.07	33.03	31.52	1.55
142	Coumaphos	33.22	28.77	30.86	2.36
143	Cyfluthrin	32.94	34.18	33.61	−0.67
144	Dichlorvos	7.80	9.72	7.98	−0.18
145	Biphenyl	9.00	16.62	11.96	−2.96
146	Vernolate	9.82	11.57	10.9	−1.08
147	3,5-Dichloroaniline	11.20	10.42	10.55	0.65
148	Molinate	11.92	14.07	13.92	−2
149	Methacrifos	11.86	15.7	13.49	−1.63
150	2-Phenylphenol	12.47	16.32	12.14	0.33

Table 1: Continued.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
151	cis-1,2,3,6-tetrahydrophthalimide	13.39	12.35	11.96	1.43
152	Fenobucarb	14.60	14.42	13.7	0.9
153	Prometon	16.66	18.42	16.82	−0.16
154	Triallate	17.12	15.55	13.9	3.22
155	Pyrimethanil	17.28	18.45	18.46	−1.18
156	Gamma-HCH	17.48	16.12	18.09	−0.61
157	Disulfoton	17.61	16.97	16.66	0.95
158	Heptachlor	18.49	21.61	19.83	−1.34
159	Isazofos	18.54	22.62	20.07	−1.53
160	Fenpropimorph	19.22	25	21.3	−2.08
161	Transfluthrin	19.04	23.96	22.15	−3.11
162	Tolclofos-methyl	19.69	21.11	18.51	1.18
163	Metobromuron	20.07	19.92	20.38	−0.31
164	HCH, epsilon-	20.78	16.05	17.54	3.24
165	Dipropetryn	20.82	20.32	21.8	−0.98
166	Formothion	21.42	17.67	18.96	2.46
167	Diethofencarb	21.43	20.66	19.42	2.01
168	Dimepiperate	22.28	21.74	21.92	0.36
169	Bioallethrin-1	22.29	22.9	21.57	0.72
170	2,4-DDE	22.64	22.46	24.45	−1.81
171	Fenson	22.54	23.65	21.87	0.67
172	Chlorthion	22.86	24.7	22.65	0.21
173	Prallethrin	23.11	23.96	22.59	0.52
174	Mecarbam	23.46	22	25.24	−1.78
175	Flumetralin	24.10	27.85	24.44	−0.34
176	Triadimenol	24.22	25.2	24.81	−0.59
177	Pretilachlor	24.67	22.65	24.59	0.08
178	Uniconazole	26.15	25.53	26	0.15
179	Flusilazole	26.19	29.21	27.8	−1.61
180	Fluorodifen	26.59	25.19	25.07	1.52
181	Diniconazole	27.03	26.95	27.24	−0.21
182	Piperonyl butoxide	27.46	28.93	26.8	0.66
183	Mepronil	27.91	25.31	27.43	0.48
184	Fenazaquin	28.97	29.38	30.6	−1.63
185	Fenoxycarb	29.57	29.01	30.52	−0.95
186	Sethoxydim	29.63	24.8	27.28	2.35
187	Anilofos	30.68	29.82	28.59	2.09
188	Permethrin	31.57	30.27	32.38	−0.81
189	Pyridaben	31.86	31.57	30.56	1.3
190	Fluoroglycofen-ethyl	32.01	34.85	33.64	−1.63
191	Bitertanol	32.25	32.05	30.82	1.43
192	Etofenprox	32.75	32.57	32.87	−0.12
193	Cycloxydim	33.05	28.55	30.36	2.69
194	Alpha-cypermethrin	33.35	28.9	31.06	2.29
195	Esfenvalerate	34.65	36.73	34	0.65
196	Difenconazole	35.40	36.14	33.03	2.37
197	Flumioxazin	35.50	32.67	32.45	3.05
198	Dimefox	5.62	−0.46	7.13	−1.51
199	Tri-iso-butyl phosphate	11.65	10.47	12.32	−0.67
200	Crimidine	13.13	12.37	12.09	1.04
201	Chlorfenprop-methyl	13.57	17.75	13.6	−0.03

Table 1: Continued.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
202	2,3,5,6-Tetrachloroaniline	14.22	13.5	12.71	1.51
203	Tri-n-butyl phosphate	14.33	16.15	14.42	−0.09
204	2,3,4,5-Tetrachloroanisole	14.66	15.91	14.51	0.15
205	Tebutam	15.30	16.51	17.61	−2.31
206	Dioxabenzofos	16.14	17.74	15.19	0.95
207	Simetone	16.69	17.79	16.84	−0.15
208	Atratone	16.70	18.87	18.68	−1.98
209	Bromocylen	17.43	18.53	17.24	0.19
210	Cycluron	17.95	17.22	19.68	−1.73
211	Musk ambrette	18.62	18.86	17.77	−1.15
212	Musk xylene	18.66	20.74	18.81	−0.15
213	Pentachloroaniline	18.91	15.38	16.2	2.71
214	Aziprotryne	19.11	21.05	19.07	0.04
215	Sebutylazine	19.26	19.07	19.03	0.23
216	Isocarbamid	19.24	16.72	17.15	2.09
217	Musk moskene	19.46	21.79	21.16	−1.7
218	Dimethenamid	19.55	20.54	20.39	−0.84
219	Fenchlorphos oxon	19.72	21.04	18.44	1.28
220	BDMC-2	19.74	15.49	19.3	0.44
221	Paraoxon-methyl	19.83	20.87	20.76	−0.93
222	Monalide	20.02	19.22	21.33	−1.31
223	Isobenzan	20.55	22.18	21.66	−1.11
224	Pyrimitate	20.59	22.79	18.97	1.62
225	Isodrin	21.01	21.88	22.55	−1.54
226	Isomethiozin	21.06	22.14	19.89	1.17
227	Dacthal	21.25	22.38	21.38	−0.13
228	4,4-Dichlorobenzophenone	21.29	21.45	19.45	1.84
229	Nitrothal-isopropyl	21.69	21.85	21.6	0.09
230	Rabenzazole	21.73	21.17	24.36	−2.63
231	Fuberidazole	22.10	20.67	21.26	0.84
232	Isofenphos oxon	22.04	23.32	21.78	0.26
233	Dicapthon	22.44	22.74	22.2	0.24
234	Isocarbophos	22.87	24.01	22.75	0.12
235	Phorate sulfone	23.15	19.74	25.53	−2.38
236	Chlorfenethol	23.29	21.98	22.52	0.77
237	trans-Nonachlor	23.62	25.1	24.51	−0.89
238	Dinobuton	23.88	22.71	22.01	1.87
239	DEF	24.08	21.24	23.61	0.47
240	Flurochloridone	24.31	18.15	21.83	2.48
241	Bromfenvinfos	24.62	25.03	24.94	−0.32
242	Ditalimfos	24.82	25.28	25.31	−0.49
243	4,4-Dibromobenzophenone	25.30	23.06	24.62	0.68
244	Disulfoton sulfone	26.16	20.66	24.24	1.92
245	Cyproconazole	27.23	26.87	26.41	0.82
246	Phthalic acid, benzyl butyl ester	27.56	27.11	26.22	1.34
247	Clodinafop-propargyl	27.74	31.2	29.61	−1.87
248	Fenthion sulfone	28.55	27.46	29.4	−0.85
249	Metamitron	28.63	21.78	27.39	1.24
250	Tebufenpyrad	29.06	30.01	29.27	−0.21
251	Cloquintocet-mexyl	29.32	29.62	29.85	−0.53
252	Lenacil	29.70	21.63	28.28	1.42

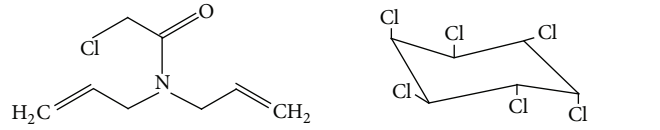
Table 1: Continued.

Number	Pesticide	t_R (EXP)	t_R (MLR)	t_R (ANN)	$(t_R)_{EXP} - (t_R)_{ANN}$
253	Bromuconazole	29.90	28.32	29.66	0.24
254	Fenamiphos sulfone	31.34	29.08	30.17	1.17
255	Fluquinconazole	32.62	32.22	30.22	2.4
256	Fenbuconazole	34.02	33.6	31.53	2.49
Test set					
257	Tetradifon	30.70	29.14	30.26	0.44
258	Fluorochloridone	25.14	19.72	24.4	0.74
259	Cyanofenphos	28.43	27.78	29.74	−1.31
260	EPN	30.06	30.42	31.71	−1.65
261	Benfluralin	15.23	16.21	14.16	1.07
262	Atrizine	17.64	16.45	19.4	−1.76
263	Simetryn	20.18	18.85	19.71	0.47
264	Metribuzin	20.33	17.61	17.78	2.55
265	Bioallethrin-2	22.34	23.58	21.89	0.45
266	Kresoxim-methyl	25.04	27.47	27.78	−2.57
267	Propargite	27.87	26.85	30.06	−2.19
268	Amitraz	30.29	28.83	29.76	0.53
269	Trietazine	17.53	18.78	19.76	−2.23
270	Prosulfocarb	19.51	21.08	21.51	−2
271	Octachlorostyrene	20.60	20.81	23.38	−2.78
272	Methfuroxam	22.45	21.15	24.43	−1.98
273	Flutriafol	25.31	26.82	27.15	−1.84
274	Diclobutrazole	25.95	27.22	27.61	−1.66
275	Triphenyl phosphate	28.65	26.88	30.41	−1.76
276	Desbrom-leptophos	30.15	27.97	28.32	1.83
277	Propisochlor	19.89	16.96	19.98	−0.09
278	Ametryn	20.11	18.98	19.4	0.71
Valid set					
279	Quintozene	16.75	18.1	18.4	−1.65
280	Prometryne	20.13	19.82	19.71	0.42
281	Chlorbenside	22.96	23.73	25.4	−2.44
282	Oxadiazone	25.06	26.67	26.72	−1.66
283	Tetrasul	25.85	25.48	26.27	−0.42
284	Etaconazole-1	26.81	28.41	27.98	−1.17
285	Pyridaphenthion	30.17	29.14	29.64	0.53
286	trans-Diallate	15.29	14.82	15.52	−0.23
287	Propazine	17.67	18.15	19.5	−1.83
288	Pirimiphos-methyl	20.30	23.66	22.81	−2.51
289	Dichlofluanid	21.68	22.03	22.53	−0.85
290	Profluralin	17.36	22.64	17.16	0.2
291	Tetrachlorvinphos	24.36	22.75	24.14	0.22
292	Chlorfenson	25.05	26.07	23.34	1.71
293	2,4-DDD	24.94	26.13	27.65	−1.75
294	Leptophos	30.19	30.5	29.65	0.54
295	Nitralin	30.92	29.65	29.37	1.55
296	Fenamiphos sulfoxide	31.03	26.93	30.38	0.65
297	Dicloran	17.89	26.94	17.14	0.75
298	Perthane	24.81	15.1	26.01	−1.2
299	Cyprodinil	21.94	22.19	22.63	−0.69
300	Mefenacet	31.29	29.75	30.27	1.02

Table 2: Specifications of the selected multiple linear regression model.

Descriptor	Notation	Coefficient	Mean effect
Number of nitrogen atoms	nN	0.980 (±0.123)	0.091
Solvation connectivity index-Chi1	χ1sol	2.191 (±0.125)	1.467
Balaban Y index	Y index	−4.639 (±0.401)	−0.421
Moran autocorrelation-lag 2/weighted by atomic sanderson electronegativity	MATS2e	−7.386 (±1.022)	−0.104
Total absolute charge (electronic charge index—ECI)	Qtot	0.499 (±0.084)	0.211
Radial Distribution-6.0/unweighted	RDF060u	−0.156 (±0.025)	−0.140
Constant		7.045 (±1.346)	

$N = 256$, $F = 240$, $R = 0.925$, SE = 2.456.



Allidochlor (t_R = 8.78, Xlsol = 5.51) HCH-epsilon (t_R = 20.78, Xlsol = 7.71)

Figure 1: Comparison of χ1sol values for HCH-epsilon and allidochlor.

The correlation between these descriptors is shown in Table 3. As shown in this table, there are no significant correlations between these descriptors.

The first descriptor with the larger mean effect is solvation connectivity index—chi 1 (χ1sol) that defined in order to model solvation entropy and describe dispersion interactions in solution. Taking into account the characteristic dimension of the molecules by atomic parameters, they are defined as

$$ {}^{m}\chi_q^s = \left(\frac{1}{2}\right)^{m+1} \cdot \sum_{k+1}^{k} \left(\frac{\left(\prod_{a=1}^{k} L_a\right)_k}{\left(\prod_{a=1}^{n} \delta_a\right)_k^{1/2}} \right), \quad (1) $$

where L_a is the principal quantum number (2 for C, N, O atoms, 3 for Si, S, Cl, …) of the ath atom in the kth subgraph and δ_a the corresponding vertex degree; k is the total number of mth order subgraphs; n is the number of vertices in the subgraph. The normalization factor $1/(2)^{m+1}$ is defined in such a way that the indices ${}^{m}\chi$ and ${}^{m}\chi^s$ for compounds containing only second-row atoms coincide. The first-order solvation connectivity index is

$$ {}^{1}\chi^s = \frac{1}{4} \cdot \left(\frac{\left(L_i \cdot L_j\right)_b}{\left(\delta_i \cdot \delta_j\right)_b^{1/2}} \right), \quad (2) $$

where b runs over all the bonds; L_i and L_j are the principal quantum numbers of the two vertices incident to the considered bond. This index coincides with the Randic connectivity index ${}^{1}\chi$ for the hydrocarbons; $L = 2$ for all the atoms. These molecular descriptors are defined for an H-depleted molecular graph [37].

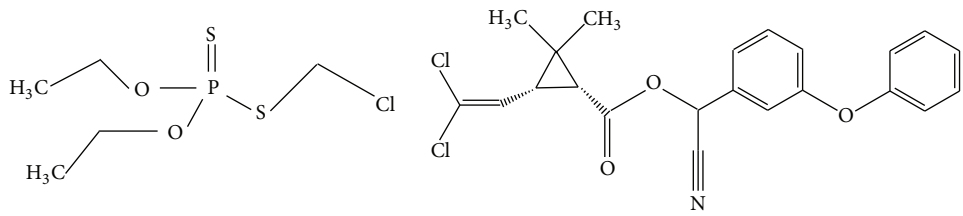
The positive sign for the mean effect of this descriptor reveals that molecules have higher numerical value of χ1sol, therefore, they have longer retention time. For example, HCH-epsilon (compound **164**) with 6 chloride atoms in its structure has bigger χ1sol (7.75) than Allidochlor (compound **1**) with one chloride atom in its structure (5.61) and also has longer retention time than Allidochlor (20.78 and 8.78 min, resp., Figure 1). Compounds that have more atoms in their structure have larger numerical value of χ1sol and so they have longer retention time, for example, compound **194** (Alpha-cypermethrin) has χ1sol value of 13.32 and retention time of 33.35 min but compound **4** (Chlormephos), which has less atoms and shorter structure than Alpha-cypermethrin, has χ1sol and retention time of 7.21 and 10.53 min, respectively (Figure 2). The second important descriptor is Balaban Y index which was calculated by the same formula as the Balaban distance connectivity index J, but by using atomic information indices instead of vertex distance degrees [38]. The Y index is defined based on atomic information indices y_i calculated for vertices of a H-depleted molecular graph as follows:

$$ y_i = \sum_{g-1}^{\eta i} {}^{g}f_i \cdot g \cdot \log_2 g, \quad (3) $$

where g runs over all of the different topological distances from the ith vertex, ${}^{g}f_i$ is the number of distances from the ith vertex equal to g, and η_i is the ith atom eccentricity (i.e., the maximum topological distance from the considered atom) [38]. So, the Balaban Y index is calculated as

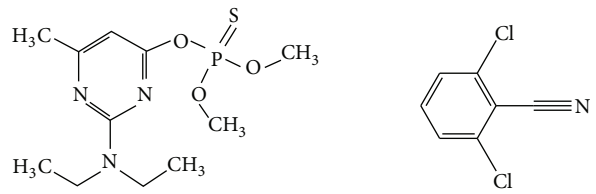
$$ Y\text{index} = \frac{B}{C+1} \cdot \sum_{\text{edge}(i,j)} \left(y_i \cdot y_j\right)^{-1/2}, \quad (4) $$

where B is the number of bonds and C is the cyclomatic number. As the number of cycles, and branching in molecular structures increased (so y_i in (2) increased), the value of Y index decreased. For example, compounds **70** (EPTC), **80** (Cis-Diallate), and **97** (Thiobenarb) which have amine group in their structures with retention time of 8.54, 14.75, 20.63 min have Y index values of 2.27, 2.15, 1.02, respectively. This descriptor has negative sign for its mean effect. Therefore, as the numerical value of this descriptor increased



α-cypermethrin (t_R = 10.53, Xlsol = 7.21) Allidochlor (t_R = 33.35, Xlsol = 13.32)

Figure 2: Comparison of χ1sol values for α-cypermethrin and allidochlor.



Pirimiphos-methyl (t_R = 20.3, Qtot = 10.14) Dichlobenil (t_R = 9.75, Xlsol = 1.02)

Figure 3: Comparison of Qtot values for pirimiphos-methyl and dichlobenil.

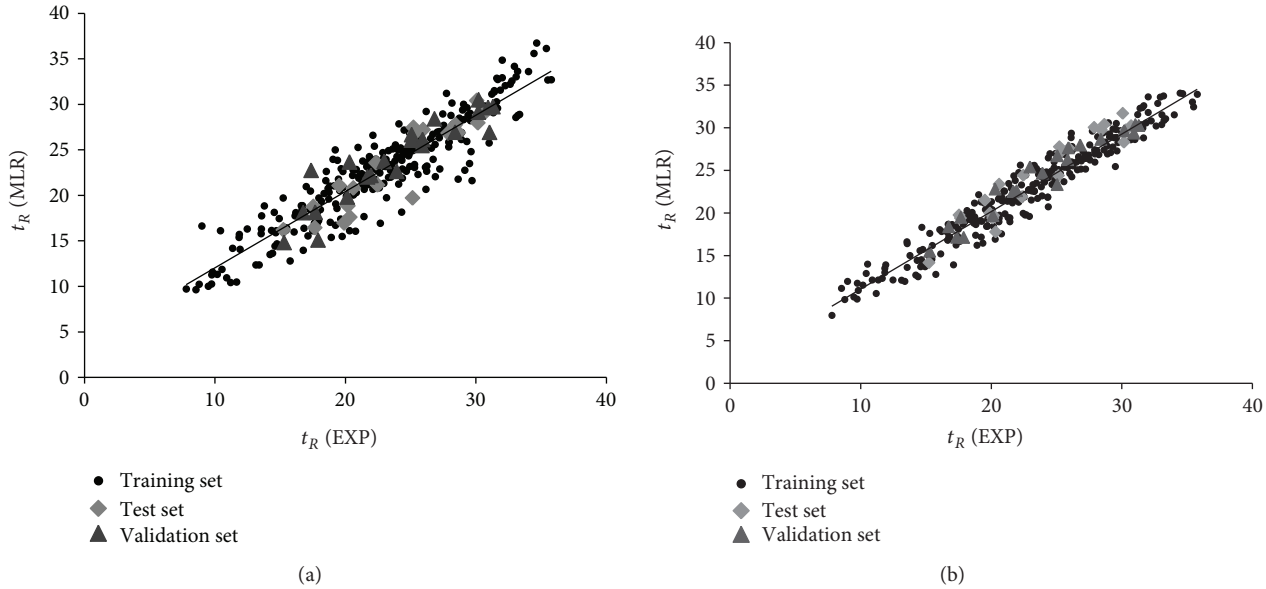


Figure 4: Plot of the calculated t_R by MLR method versus the experimental values for training, test, and validation sets (a) and the plot of retention time values predicted by 6–7–1 feed forward ANN (t_R (ANN)) against the experimental values (t_R (EXP)) (b).

Table 3: Correlation matrix for descriptors applying in this work.

	nN	χ1sol	*Y* index	MATS2e	Qtot	RDF060u
nN	1					
χ1sol	−0.162	1				
*Y*index	−0.120	0.540	1			
MATS2e	0.315	−0.240	0.065	1		
Qtot	−0.082	0.427	0.034	−0.535	1	
RDF060u	0.084	0.634	−0.363	0.162	0.433	1

Table 4: Statistical parameters obtained using the ANN and MLR models.

Model	SE_{tr}	SE_t	SE_v	R_{tr}	R_t	R_v	F_{tr}	F_t	F_v
ANN	1.559	1.517	1.249	0.969	0.951	0.971	3928	188	334
MLR	2.402	1.858	2.036	0.925	0.924	0.922	1508	118	113

tr: training, t: test, v: valid.

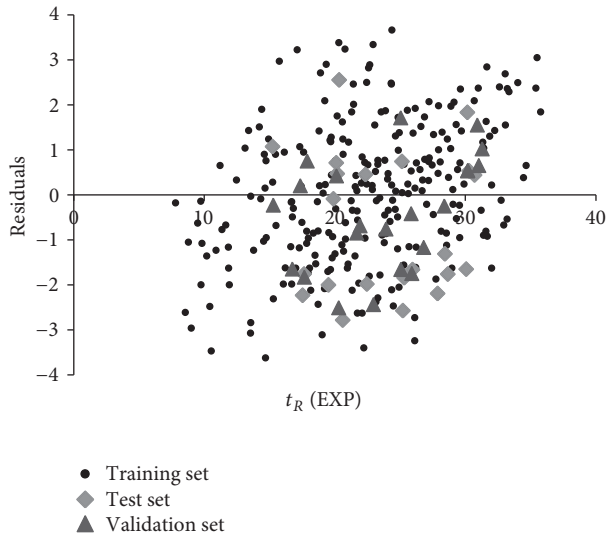


Figure 5: Plot of the residuals versus the experimental values of t_R for training, test and validation sets.

the retention time of compounds decreased. For example in compounds **149** (Methacrifos), **27** (Methyl-parathion), **131** (Triazophos) which are organophosphorous compounds Y index values are 2.08, 1.18, and 0.81, and retention times are 11.86, 20.82, and 28.23 min, respectively.

The total absolute charge is the other descriptor which, also known as the electronic charge index (ECI), is the sum of absolute charge over all atoms in a molecule and is a measure of molecule polarity [37]. Therefore, compounds with polar bonds have more numerical value of Qtot and have longer retention on a polar stationary phase than others. For example, compound **288** (Pirimiphos-methyl) has more retention and polarity (20.30 min and 10.14, resp.) than Dichlobenil (compound **72**) that has Qtot of 1.02 and retention time of 9.75 min (Figure 3), also Compounds **9** (Ethalfuralin), **21** (Dinitramine), and **175** (Flumetralin) that contain *CF3* and two nitro groups in their structure have retention time of 15, 19.35, 24.10, and Qtot value of 6.66, 6.86, and 7.33, respectively. The other molecular descriptors with lower mean effects are number of nitrogen atoms, Moran autocorrelation-lag 2/weighted by atomic sanderson electronegativity and radial distribution-6.0/unweighted. As number of nitrogen atoms increase, in molecule's structure due to increasing the interaction of molecules with polar stationary phase, the retention time of compound increases. For example, in etridiazol, atrazine-desethyl, and fluquinconazole (compounds **3**, **11,** and **255**, resp.) which have 2, 5, and 6 nitrogen atoms in their structure, their retention time is 10.42, 16.76, and 32.62, respectively. A Moran coefficient is a general index of spatial autocorrelation that, if applied to a molecular graph, can be defined as

$$I(d) = \frac{(1/\Delta) \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} \delta_{ij} \cdot (w_i - \overline{w}) \cdot (w_j - \overline{w})}{(1/A) \cdot \sum_{i=1}^{A} (w_i - \overline{w})^2}, \quad (5)$$

where w_i is any atomic property (here Sanderson electronegativity), $\overline{w}$ is its average value on the molecule, A is the atom number, d is a Kronoker delta ($\delta_{ij} = 1$ if $d_{ij} = d$, zero, otherwise). Δ is the sum of the Kronoker deltas, that is, the number of vertex pairs at distance equal to d [39].

So the Moran autocorrelation-lag 2/weighted by atomic sanderson electronegativity can be a factor of electronegativity of moleculs. The last descriptor is radial distribution-6.0/unweighted. The 3D coordinates of the atoms of molecules can be transformed into a structure code that has a fixed number of descriptors irrespective of the size of a molecule. This task is performed by a structure coding technique referred to as radial distribution function code (RDF code) [40]. In general, there are some prerequisites for a structure code: independence from the number of atoms, that is, the size of a molecule, unambiguity regarding the three-dimensional arrangement of the atoms, and invariance against translation and rotation of the entire molecule.

Formally, the radial distribution function of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius r [41]. The equation represents the radial distribution function code as it is used in this investigation:

$$g(r) = f \cdot \sum_{i}^{N-1} \sum_{j>i}^{N} A_i \cdot A_j \cdot e^{-B(r-r_{ij})^2}, \quad (6)$$

where f is a scaling factor and N is the number of atoms. By including characteristic atomic properties A of the atoms i and j, the RDF codes can be used in different tasks to fit the requirements of the information to be represented. The exponential term contains the distance r_{ij} between the atoms i and j and the smoothing parameter B that defines the probability distribution of the individual distances. $g(r)$ was calculated at a number of discrete points with defined intervals.

The atomic properties A_i and A_j used in this equation enable the discrimination of the atoms of a molecule for almost any property that can be attributed to an atom. Such distribution function provides, besides information about interatomic distances in a whole molecule, the opportunity to gain access to other valuable information, for example, bond distance, ring types, planar and nonplanar systems,

and atoms types. This fact is a most valuable consideration for a computer-assisted code elucidation. The radial distribution function in this form meets the entire requirement mentioned above, especially invariance against linear translations. As RDF060u has negative sign for its mean effect, as molecules became larger their RDF factor in 6 Å radius reduces. For example Tebufenpyrad (compound **250**) which has larger structure than Pyrimitate (compound **224**) has lower RDF factor (18.79) than Pyrimitate (29.76); so the retention time of Tebufenpyrad is higer (29.06 min) than Pyrimitate (20.59).

From the above discussion, it can be seen that all descriptors involved in the QSRR model has physical meaning, and these descriptors can account for the structural features that affect the retention time of under studied pesticides.

3.2. Comparison of Neural Network and MLR Models. A graphical comparison of ANN and MLR analysis is given in Figure 4, where the retention time values calculated by means of the respective models are plotted against the experimental values. The statistical parameters obtained by ANN and MLR models for these sets are shown in Table 4. The standard errors of training, test and validation sets for the MLR model are 2.402, 1.858, and 2.036, respectively, which would be compared with the values of 1.559, 1.517, and 1.249, respectively, for the ANN model. Comparison between these values and other statistical parameters in Table 4 reveals the superiority of the ANN model over MLR ones. Figure 5 shows the plot of the residuals against the experimental values of retention time, for the ANN model. Since the residuals are propagated on both sides of the zero line, there is no systematic error in developing of ANN model.

4. Conclusions

Few structure-activity relationships involving pesticides have been published. In this study, we use MLR and ANN to predict the retention time of 300 pesticides that were different in molecular structure. The results of this study demonstrate that QSRRs method using ANN techniques can generate a suitable model for prediction of gas chromatographic retention of pesticides.

Also the results obtained in this work indicate that the regression and ANN models exhibit reasonable prediction capabilities. Descriptors which appeared in the obtained QSRR models reveal that electronic interactions as well as steric parameters can be affected on the gas chromatographic retention time of pesticides.

References

[1] M. Skamoto and T. Tsutsumi, "Applicability of headspace solid-phase microextraction to the determination of multi-class pesticides in waters," *Journal of Chromatography A*, vol. 1028, no. 1, pp. 63–74, 2004.

[2] F. Hernández, O. J. Pozo, J. V. Sancho, L. Bijlsma, M. Barreda, and E. Pitarch, "Multiresidue liquid chromatography tandem mass spectrometry determination of 52 non gas chromatography-amenable pesticides and metabolites in different food commodities," *Journal of Chromatography A*, vol. 1109, no. 2, pp. 242–252, 2006.

[3] C. Gonçalves and M. F. Alpendurada, "Solid-phase microextraction—gas chromatography—(tandem) mass spectrometry as a tool for pesticide residue analysis in water samples at high sensitivity and selectivity with confirmation capabilities," *Journal of Chromatography A*, vol. 1026, no. 1-2, pp. 239–250, 2004.

[4] L. Alder, S. Lüderitz, K. Lindtner, and H. J. Stan, "The ECHO technique—the more effective way of data evaluation in liquid chromatography-tandem mass spectrometry analysis," *Journal of Chromatography A*, vol. 1058, no. 1-2, pp. 67–79, 2004.

[5] A. R. Katritzky, V. S. Lobanov, and M. Karelson, "QSPR: the correlation and quantitative prediction of chemical and physical properties from structure," *Chemical Society Reviews*, vol. 24, no. 4, pp. 279–287, 1995.

[6] R. Kaliszan, *Structure and Retention in Chromatographic Approach*, Harwood Academic, Amsterdam, The Netherlands, 1997.

[7] C. Bishop, *Neural Networks for Pattern Recognition*, University Press, Oxford, UK, 1995.

[8] L. Fausett, *Fundamentals of Neural Networks*, Preentice Hall, New York, NY, UK, 1994.

[9] G. C. Looney, *Pattern Recognition Using Neural Networks*, Oxford University Press, New York, NY, USA, 1997.

[10] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing. Experiments in the Microstructure of Cognition*, MIT Press, Cambridge, Mass, USA, 1986.

[11] S. Goll and P. Jurs, "Prediction of vapor pressures of hydrocarbons and halohydrocarbons from molecular structure with a computational neural network model," *Journal of Chemical Information and Computer Sciences*, vol. 39, no. 6, pp. 1081–1089, 1999.

[12] J. Tetteh, T. Suzuki, E. Metcalfe, and S. Howells, "Quantitative structure-property relationships for the estimation of boiling point and flash point using a radial basis function neural network," *Journal of Chemical Information Computer Science*, vol. 39, no. 3, pp. 491–507, 1999.

[13] Z. J. Gasteiger, *Neural Networks for Chemists: an Introduction*, Wiley-VCH, Weinheim, Germany, 1993.

[14] J. A. Burns and G. Whitesides, "Feed-forward neural networks in chemistry: Mathematical systems for classification and pattern recognition," *Chemical Reviews*, vol. 93, no. 8, pp. 2583–2601, 1993.

[15] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997.

[16] S. Agatonovic-Kustrin, L. H. Ling, S. Y. Tham, and R. G. Alany, "Molecular descriptors that influence the amount of drugs transfer into human breast milk," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 29, no. 1-2, pp. 103–119, 2002.

[17] A. A. D'Archivio, F. Ruggieri, P. Mazzeo, and E. Tettamanti, "Modelling of retention of pesticides in reversed-phase high-performance liquid chromatography: quantitative structure-retention relationships based on solute quantum-chemical descriptors and experimental (solvatochromic and spin-probe) mobile phase descriptors," *Analytica Chimica Acta*, vol. 593, no. 2, pp. 140–151, 2007.

[18] M. Aschi, A. A. D'Archivio, M. A. Maggi, P. Mazzeo, and F. Ruggieri, "Quantitative structure-retention relationships of

pesticides in reversed-phase high-performance liquid chromatography," *Analytica Chimica Acta*, vol. 582, no. 2, pp. 235–242, 2007.

[19] J. Ghasemi, S. Asadpour, and A. Abdolmaleki, "Prediction of gas chromatography/electron capture detector retention times of chlorinated pesticides, herbicides, and organohalides by multivariate chemometrics methods," *Analytica Chimica Acta*, vol. 588, no. 2, pp. 200–206, 2007.

[20] G. Pang, Y. Cao, J. Zhang et al., "Validation study on 660 pesticide residues in animal tissues by gel permeation chromatography cleanup/gas chromatography—mass spectrometry and liquid chromatography—tandem mass spectrometry," *Journal of Chromatography A*, vol. 1125, no. 1, pp. 1–30, 2006.

[21] HyperChem and Autodesk, "Release 3 for windows," 1993.

[22] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. Stewart, "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model," *Journal of the American Chemical Society*, vol. 107, no. 13, pp. 3902–3909, 1985.

[23] R. Todeschini, V. Consonni, M. Pavan, and V. Pisani, 13-20124, Milano, Italy, Dragon software version 3.0.

[24] N. Draper and H. Smith, *Applied Regression Analysis*, Wiley Interscience, New York, NY, USA, 2nd edition, 1981.

[25] S. Haykin, *Neural Network*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1994.

[26] J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, Germany, 1999.

[27] N. K. Bose and P. Liang, *Neural Network, Fundamentals*, McGraw-Hill, New York, NY, USA, 1996.

[28] L. S. Anker and P. C. Jurs, "Prediction of carbon-13 nuclear magnetic resonance chemical shifts by artificial neural networks," *Analytical Chemistry*, vol. 64, no. 10, pp. 1157–1164, 1992.

[29] M. T. Beal, H. B. Hagan, and M. Demuth, *Neural Network Design*, PWS, Boston, Mass, USA, 1996.

[30] J. Zupan and J. Gasteiger, *Neural Networks for Chemists: An Introduction*, Wiley-VCH, Weinheim, Germany, 1993.

[31] P. K. Hopke and X. Song, "Classification of single particles by neural networks based on the computer-controlled scanning electron microscopy data," *Analytica Chimica Acta*, vol. 348, no. 1–3, pp. 375–388, 1997.

[32] S. Haykin, *Neural Networks. AComprehensive Foundation:*, vol. 1, Pearson Education, Saddle River, NJ, USA, 2nd edition, 1999.

[33] T. Masters, *Practical Neural Network Recipes in C++*, Academic Press, 1993.

[34] M. Jalali-Heravi and M. H. Fatemi, "Prediction of flame ionization detector response factors using an artificial neural network," *Journal of Chromatography A*, vol. 825, no. 2, pp. 161–169, 1998.

[35] M. Jalali-Heravi and M. H. Fatemi, "Prediction of thermal conductivity detection response factors using an artificial neural network," *Journal of Chromatography A*, vol. 897, no. 1-2, pp. 227–235, 2000.

[36] M. H. Fatemi, M. Jalali-Heravi, and E. Konuze, "Predictions of bioconcentration factors using genetic algorithm and artificial neural network," *Analytica Chimica Acta*, vol. 486, pp. 101–108, 2003.

[37] R. Todeschini and V. Consonni, *Hand book of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, 2000.

[38] A. T. Balaban and T. S. Balaban, "New vertex invariants and topological indices of chemical graphs based on information on distances," *Journal of Mathematical Chemistry*, vol. 8, no. 1, pp. 383–397, 1991.

[39] P. A. P Moran, "Notes on continuous stochastic phenomena," *Biometricka*, vol. 37, pp. 17–23, 1950.

[40] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, and V. Steinhauer, "Chemical information in 3D space," *Journal of Chemical Information and Computer Sciences*, vol. 36, no. 5, pp. 1030–1037, 1996.

[41] M. C. Hemmer, V. Steinhauer, and J. Gasteiger, "Deriving the 3D structure of organic molecules from their infrared spectra," *Vibrational Spectroscopy*, vol. 19, no. 1, pp. 151–164, 1999.