ELSEVIER

Journal of Chromatography A, 897 (2000) 227–235

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Prediction of thermal conductivity detection response factors using an artificial neural network

M. Jalali-Heravi*, M.H. Fatemi

Sharif University of Technology, Department of Chemistry, P.O. Box 11365-9516, Tehran, Iran

Received 20 January 2000; received in revised form 18 July 2000; accepted 19 July 2000

Abstract

The main aim of the present work was the development of a quantitative structure–activity relationship method using an artificial neural network (ANN) for predicting the thermal conductivity detector response factor. As a first step a multiple linear regression (MLR) model was developed and the descriptors appearing in this model were considered as inputs for the ANN. The descriptors of molecular mass, number of vibrational modes of the molecule, molecular surface area and Balaban index appeared in the MLR model. In agreement with the molecular diameter approach, molecular mass and molecular surface area play a major role in estimating the thermal conductivity detector response factor (TCD-RF). A 4-7-1 neural network was generated for the prediction of the TCD-RFs of a collection of 110 organic compounds including hydrocarbons, benzene derivatives, esters, alcohols, aldehydes, ketones and heterocyclics. The mean absolute error between the ANN calculated and the experimental values of the response factors was 0.02 for the prediction set. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Neural networks; Thermal conductivity detection; Detection, GC; Heat transfer; Quantitative structure–activity relationships; Regression models; Response factors; Chemometrics

1. Introduction

The peak area in gas chromatography (GC) using thermal conductivity detection (TCD) is a function of the molecular properties of the solute and the carrier gas. Numerous investigations have stated that the signal strength arising from the presence of a solute in the TCD system depends on the nature of the solute [1–3]. On the other hand, accurate quantitative analysis using GC requires knowledge of the response factor (RF) for each compound. This factor is essentially a correction factor that measures the response of a given compound to the detecting device.

Since numerous compounds are unavailable as standards, the development of a theoretical method for estimating response factors appears to be useful. Among the chemometric methods, QSRR (quantitative structure–retention relationship) has been the most popular method used for the prediction of chromatographic factors [4,5]. Katritzky et al. applied the multiple linear regression (MLR) method to predict the GC retention times and the flame ionization detection (FID) response factors of a series of organic compounds [6].

Recently, artificial neural networks (ANNs) have

*Corresponding author. Tel.: +98-21-600-5718; fax: +98-21-601-2983.

E-mail address: jalali@ch.sharif.ac.ir (M. Jalali-Heravi).

0021-9673/00/$ – see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0021-9673(00)00793-7

been applied to a wide variety of chemical problems such as multispectral interpretation [7,8], prediction of carbon-13 NMR chemical shifts [9], quantitative structure–activity relationship (QSAR) studies [10,11] and response surface modeling in HPLC optimization [12]. Also, the FID response factors of a diverse set of organic molecules were predicted using an artificial neural network developed in our laboratory [13]. The main aim of the present work was the development of a QSAR method using an ANN for the modeling of the TCD response factor. The generated ANN was evaluated and applied for the prediction of the RFs of a wide series of organic compounds. As far as we are aware this is the first QSAR study using an ANN for the prediction of the TCD response factor.

2. Methods

A detailed description of the theory behind a neural network has been adequately described elsewhere [14–16]. The relevant principle of supervised learning in an ANN is that it takes numerical inputs (the training data) and transfers them into desired outputs. The input and output nodes may be connected to the 'external world' and to other nodes within the network. The way in which each node transforms its input depends on the so-called 'connection weights' or 'connection strength' and bias of the node, which are modifiable. The output values of each node depend on both the weight strengths and biase values. In addition, the outputs depend on the weighted sum of all its inputs which are normally transformed by a nonlinear weighting function. For the present purposes, the great power of ANNs stems from the fact that it is possible to train them. Training is done by continually presenting the networks with known inputs and outputs and modifying the connection weights and biases between the individual nodes. This process is confirmed until the output nodes of the network match the desired outputs to a stated degree of accuracy. However, training can be performed by using the back-propagation algorithm. In order to train the network using the back-propagation algorithm, the differences between the ANN output and its desired value are calculated after each iteration. The changes in the values of the weights can be obtained using the equation:

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

where Δw_{ij} is the change in the weight factor for each network node, δ_i is the actual error of node i, and O_j is the output of node j. The coefficients η and α are the learning rate and the momentum factor, respectively. These coefficients control the velocity and the efficiency of the learning process. These parameters would be optimized before training the network. Equation like Eq. (1) can be used for the bias settings.

The ANN can use qualitative as well as quantitative inputs, and it does not require an explicit relationship between the inputs and the outputs. Although in statistics the analysis is limited to a certain number of possible interactions, more terms can be examined for interactions by the ANN. Also, by allowing more data to be analyzed at the same time, more complex and subtle interactions can be studied using this technique.

In this paper, the generation of a three layer neural network consisting of four inputs, seven nodes in the hidden layer and one node in the output layer (a 4-7-1 neural network) is reported and is used for the first time for the prediction of the thermal conductivity detector response factor of a variety of organic compounds.

3. Experimental

3.1. Data set

A collection of 110 organic compounds was chosen as the data set taken from Ref. [17]. These molecules consist of hydrocarbons, benzene derivatives, esters, alcohols, aldehydes, ketones and heterocyclic compounds. The data set was randomly divided into two groups: a training set (Table 1) and a prediction set (Table 2). The training and prediction sets consist of 90 and 20 compounds, respectively. The training set was used for model generation and the prediction set was used for the evaluation of the model. For the training and prediction sets given

Table 1
Experimental and ANN and MLR calculated values of the RFs for the training set, together with the calculated values of the descriptors[a]

No.	Compound	M_r	NVIB	MSA	BAL	RF_{MLR}	RF_{ANN}	RF_{exp}
1	2-Octanone	128	36	199.116	2.750	1.44	1.49	1.47
2	2-Propanol	60	30	103.752	2.324	0.83	0.85	0.85
3	3-Pentanone	88	48	147.978	2.754	1.13	1.14	1.09
4	Propene	42	21	82.152	1.632	0.66	0.70	0.65
5	Piperidine	85	45	128.916	2.000	1.08	1.09	1.02
6	Isopropylbenzene	120	57	171.684	2.231	1.41	1.47	1.42
7	2,4,4-Trimethyl-1-pentene	112	66	187.128	3.445	1.40	1.41	1.58
8	2,2-Dimethylpropane	72	45	134.208	3.024	1.00	0.98	0.99
9	Pentylacetate	130	63	190.746	2.750	1.52	1.50	1.46
10	Ethylbenzene	106	48	146.034	2.154	1.26	1.32	1.29
11	2,2,3-Trimethylbutane	100	63	174.222	3.541	1.29	1.34	1.29
12	*n*-Propylcyclohexane	126	75	200.952	2.128	1.56	1.61	1.58
13	Cyclopentane	70	39	113.688	2.083	0.94	0.94	0.97
14	1-Pentene	70	39	125.352	2.191	0.96	1.00	0.99
15	3-Methyl-2-butanone	100	51	163.584	3.144	1.24	1.20	1.18
16	Cyclopentanone	84	36	116.388	2.187	1.03	1.05	1.06
17	1-Butane	56	30	103.368	1.975	0.81	0.85	0.81
18	Cyclopentene	68	33	104.562	2.083	0.90	0.89	0.81
19	Methanol	32	12	56.718	1.000	0.53	0.54	0.55
20	2,4-Dimethylpentane	100	63	178.380	2.953	1.31	1.34	1.29
21	5-Decanol	158	93	259.920	3.017	1.91	1.86	1.84
22	1,3,5-Trimethylbenzene	120	57	169.308	2.358	1.41	1.45	1.49
23	2-Methylpropane	58	36	113.202	2.324	0.85	0.88	0.82
24	Tetrahydrofuran	72	33	103.536	2.083	0.92	0.91	0.83
25	2-Hexanone	100	51	158.292	2.678	1.24	1.23	1.30
26	Pyrrole	67	26	90.792	2.083	0.84	0.83	0.86
27	Naphthalene	128	48	153.594	1.714	1.41	1.43	1.39
28	2-Hexanol	102	57	170.712	2.678	1.29	1.28	1.30
29	Ethyl-*n*-butyl ether	102	57	168.714	2.448	1.29	1.29	1.30
30	3-Hexanone	100	51	158.346	2.832	1.23	1.22	1.23
31	Pyrolidine	71	36	107.91	2.083	0.93	0.92	0.91
32	*m*-Xylene	106	48	150.786	2.642	1.26	1.26	1.31
33	1-Heptanol	116	66	187.452	2.530	1.44	1.44	1.28
34	2,2-Dimethylbutane	86	54	154.836	3.038	1.14	1.15	1.16
35	3,3-Dimethyl-2-butanone	114	51	156.510	3.541	1.32	1.21	1.18
36	1,1-Dimethylcyclopentane	98	57	156.402	2.399	1.25	1.25	1.24
37	2-Methyl-2-pentene	70	48	144.846	2.407	1.02	1.06	0.96
38	Octane	114	72	198.522	2.530	1.46	1.52	1.6
39	1,2,4-Trimethylcyclopentane	112	66	180.54	2.436	1.40	1.43	1.43
40	1-Methylcyclohexane	96	51	147.222	2.123	1.20	1.22	1.15
41	3-Methyl-1-pentanol	86	57	166.824	2.832	1.18	1.21	1.07
42	2-Methylpentane	86	54	158.616	2.407	1.16	1.19	1.20
43	Ethane	30	18	69.031	0.510	0.56	0.56	0.51
44	Di-*n*-butyl ether	130	75	209.538	2.595	1.59	1.63	1.60
45	*n*-Butanol	74	39	123.084	2.191	0.98	1.01	0.95
46	3,5-Dimethylpentane	100	63	178.326	2.832	1.31	1.34	1.33
47	1,3-Dimethylcyclopentane	98	57	158.292	2.304	1.25	1.26	1.25
48	1,2,4-Trimethylbenzene	120	57	170.280	2.290	1.41	1.46	1.50
49	Propane	44	27	90.306	1.632	0.70	0.72	0.65

Table 1. Continued

No.	Compound	M_r	NVIB	MSA	BAL	RF_{MLR}	RF_{ANN}	RF_{exp}
50	Benzene	78	30	107.964	2.000	0.96	1.00	1.00
51	2,2,4-Trimethylpentane	114	72	198.846	3.445	1.45	1.55	1.47
52	*n*-Butylamine	73	42	129.078	2.191	0.99	1.03	1.14
53	1,1-Dimethylcyclohexane	112	66	176.004	2.312	1.40	1.44	1.41
54	Ethanol	46	21	79.398	1.632	0.68	0.69	0.72
55	2-Butanol	74	39	127.566	1.583	0.98	1.00	0.96
56	2-Pentanol	88	48	146.898	2.407	1.14	1.16	1.10
57	1-Hexanol	102	57	164.88	2.447	1.28	1.29	1.18
58	2-Butanone	72	33	113.418	2.583	0.93	0.94	0.98
59	3-Methylpentane	86	54	156.672	2.754	1.16	1.17	1.19
60	Cyclohexane	84	48	135.396	2.000	1.10	1.11	1.14
61	Ethylcyclopentane	98	57	157.374	2.140	1.25	1.26	1.26
62	2-Propanone	58	24	92.790	2.324	0.80	0.81	0.86
63	Hexylamine	101	60	172.926	2.448	1.30	1.30	1.04
64	1,4-Dimethylcyclohexane	112	66	178.974	2.192	1.40	1.44	1.46
65	Ethylacetate	88	36	126.324	2.407	1.06	1.11	1.11
66	2-Methyly-1-propanol	74	39	125.622	2.583	0.98	0.99	0.97
67	Butane	58	36	110.988	1.975	0.85	0.87	0.85
68	Ethanal	44	15	71.298	1.632	0.63	0.67	0.65
69	Di-isopropyl ether	102	57	169.902	2.953	1.29	1.27	1.30
70	Toluene	92	39	129.294	1.123	1.12	1.19	1.16
71	Ethylcyclohexane	112	66	177.948	2.154	1.40	1.44	1.45
72	2-Methylhexane	100	63	181.188	2.678	1.31	1.34	1.36
73	Aniline	93	36	121.734	2.123	1.09	1.16	1.14
74	Pyrroline	69	30	100.728	2.083	0.88	0.89	0.83
75	2-Methyl-1-butene	70	39	123.246	2.583	0.95	0.96	0.99
76	Biphenyl	154	60	192.474	2.699	1.67	1.65	1.69
77	2-Nonanone	142	78	224.442	2.778	1.70	1.69	1.61
78	*trans*-2-Butane	56	30	101.538	1.975	0.80	0.83	0.82
79	3-Hexanol	102	57	170.496	2.837	1.29	1.27	1.25
80	2-Methylbutane	72	45	136.530	2.583	1.00	1.02	1.02
81	1,2,3-Trimethylbenzene	120	57	224.442	2.381	1.49	1.47	1.49
82	Nonane	128	81	218.448	2.595	1.61	1.71	1.77
83	Cyclopentadiene	66	27	96.624	2.083	0.85	0.86	0.68
84	Diethyl ether	74	39	123.624	2.191	0.98	1.01	1.10
85	Pentane	72	45	133.020	2.191	1.00	1.04	1.05
86	*n*-Butylacetate	116	54	170.334	2.678	1.37	1.36	1.35
87	Di-*n*-pentyl ether	158	93	254.952	2.690	1.90	1.85	1.83
88	*cis*-2-Butene	56	30	103.590	1.975	0.81	0.85	0.87
89	2,2-Dimethylpentane	100	63	184.266	3.633	1.31	1.32	1.33
90	Methane	16	9	46.566	0.000	0.41	0.37	0.36

[a] Definition of the descriptors is given in Table 3.

in Tables 1 and 2, benzene acts as a reference with a defined RF of 1.00.

3.2. Descriptor generation

Several descriptors were calculated by considering experiment driven principles. In a thermal conductivity detector cell the solute vapor interferes with the heat transportation process at the sensing filament in proportion to its total cross-sectional area. Experiments show that increases in molecular mass for substances belonging to a homologues series increase the RF and that the RF of a branched compound is lower than that of a normal isomer [18]. In addition, the RF of a molecule decreases as its molecular symmetry increases [19]. A total of 40 separate molecular descriptors were calculated for each compound in the data set. These descriptors consist of

Table 2
Experimental and ANN and MLR calculated values of the RFs for the prediction set, together with the calculated values of the descriptors[a]

No.	Compound	M_r	NVIB	MSA	BAL	RF_{MLR}	RF_{ANN}	RF_{exp}
1	Heptane	100	63	175.302	2.448	1.30	1.33	1.43
2	2,3-Dimethylbutane	86	54	154.782	2.993	1.15	1.16	1.16
3	2,3-Dimethylpentane	100	63	174.870	3.187	1.30	1.34	1.35
4	3-Ethylpentane	100	63	179.514	2.992	1.31	1.33	1.31
5	2-Methyl-1-butene	70	39	127.404	2.583	0.96	1.00	0.99
6	1,3-Butadiene	54	24	95.382	1.975	0.76	0.82	0.80
7	*n*-Propylbenzene	120	57	173.466	2.078	1.42	1.49	1.45
8	*p*-Ethyltoluene	120	57	170.388	2.199	1.41	1.48	1.50
9	Methylcyclopentane	84	48	137.394	2.236	1.09	1.12	1.15
10	Methylcyclohexane	98	57	157.104	2.123	1.25	1.27	1.20
11	Hexane	86	54	151.380	2.339	1.15	1.18	1.23
12	Pyridine	79	27	103.266	2.000	0.94	1.01	1.00
13	3-Pentanone	86	42	135.396	2.754	1.08	1.10	1.10
14	2-Heptanone	114	60	178.596	2.792	1.39	1.36	1.33
15	Cyclohexanone	98	45	137.502	2.123	1.18	1.24	1.25
16	1-Propanol	60	30	100.620	1.975	0.83	0.85	0.85
17	2-Methyl-2-propanol	74	39	123.948	3.024	0.97	0.96	0.96
18	2-Methyl-2-butanol	88	48	125.784	2.629	1.10	1.07	1.06
19	Isopropylacetate	102	45	149.544	2.953	1.22	1.21	1.21
20	Di-*n*-propyl ether	102	47	166.716	2.448	1.28	1.29	1.31

[a] Definition of the descriptors is given in Table 3.

geometric, electronic and topological parameters. Geometric descriptors were calculated using optimized Cartesian coordinates and van der Waals radii of each atom in the molecule [20,21]. Electronic descriptors were calculated using the MOPAC program (version 6) [22]. Topological descriptors were estimated from two-dimensional representations of the molecules. Some of the 40 descriptors generated for each compound were highly correlated and encoded similar information. It was therefore desirable to test each descriptor and eliminate those with a high correlation coefficient ($R = 0.90$). Using this criterion, 12 of the original 40 descriptors were eliminated. For the selection of important descriptors the linear regression technique was used based on the construction of a linear mathematical model relating the observed RF to numerically encoded structural parameters. This equation was formed by a stepwise deletion of terms procedure (backward method) [23]. The parameters appearing in the best equation showed that four descriptors are the most important for the prediction of the thermal conductivity detector response factor. Table 3 presents the selected model together with the definition of the descriptors appearing in the model and their mean effects. These descriptors are used as inputs for the generation of the ANN.

3.3. Artificial neural network generation

The ANN program was written in FORTRAN 77 in our laboratory. All of the calculations presented in

Table 3
Specification of the multiple linear regression model

Descriptor	Notation	Coefficient	Mean effect
Molecular mass	M_r	$+0.00648(\pm0.00075)$	0.586
No. of vibrational modes	NVIB	$+0.00336(\pm0.00160)$	0.162
Molecular surface area	MSA	$+0.00153(\pm0.00087)$	0.494
Balaban index	BAL	$-0.01130(\pm0.01790)$	−0.004
Constant		$+0.2074(\pm0.03608)$	

this work were carried out on a Hewlett-Packard 133 MHz Pentium computer, Model HP Vectra VL. The number of inputs in the ANN was four, which was equal to the number of descriptors appearing in the MLR model, and the number of nodes in the output layer was set to one. The number of nodes in the hidden layer was optimized. The initial weights were selected randomly from a uniform distribution that ranged between -0.3 and $+0.3$. The initial bias values were set to one. These values were optimized during the training of the network.

Before training, the network was optimized for the number of nodes in the hidden layer, learning rate and momentum. In order to evaluate the performance of the ANN, the standard error of calibration (SEC) and the standard error of prediction (SEP) were used [24]. The network was then trained using the training set by the back-propagation strategy for optimization of the weights and bias values. The procedure for optimization of each parameter is given elsewhere [13]. It should be noted that it is common to plot SEC versus the number of iterations for optimization of the ANN parameters. However, we have used a new procedure in this work that is discussed in the next section.

4. Results and discussion

Tables 1 and 2 show that the data set consists of a large and diverse set of molecules with the TCD response factors ranging between 0.36 and 1.84 for methane and 5-decanol, respectively. Table 3 demonstrates the specification of the selected MLR model. The variables in this model encode different aspects of the molecular structure and properties and, as shown in Table 3, different types of parameters such as fragment, electronic, geometric and topological descriptors affect the TCD response factor. The molecular mass, which is a fragment descriptor, shows a positive mean effect of 0.586, which is the largest effect among the effects of the other descriptors. This is in agreement with experimental data indicating that relative response is a linear function of molecular mass within a homologous series [18]. According to the molecular diameter approach suggested by Littlewood and extended by Barry et al., the relative molar response (RMR) data of hydrocarbons and oxygen-containing compounds depend upon the molecular diameter and molecular mass. This is in agreement with our MLR model (Table 3) which shows that both molecular mass and molecular surface area play a major role in predicting the TCD response factors. This is due to the fact that the geometric descriptor of the molecular surface area (MSA) shows a high correlation with the molecular cross section. Table 3 shows that MSA with a mean effect of 0.494 demonstrates a large contribution to the TCD response factor. The Balaban index (BAL) with negative mean effects encodes the size, degree of branching and compactness of the molecule. Among isomers, the most branched compound with the lowest cross section has the higher Balaban index values and lowest response, therefore the sign of the coefficient of this index in the MLR model and also its mean effect are negative. The negative sign of this parameter confirms the conclusion obtained from the experimental data indicating that the RMR of a branched compound is lower than that of the response of the normal (*n*-)isomer [19]. As can be seen from Table 3, the mean effect of BAL is very small compared with the other parameters appearing in the MLR model. However, deletion of this parameter reduces the prediction ability of the ANN generated using the descriptors appearing in the MLR model. In general,

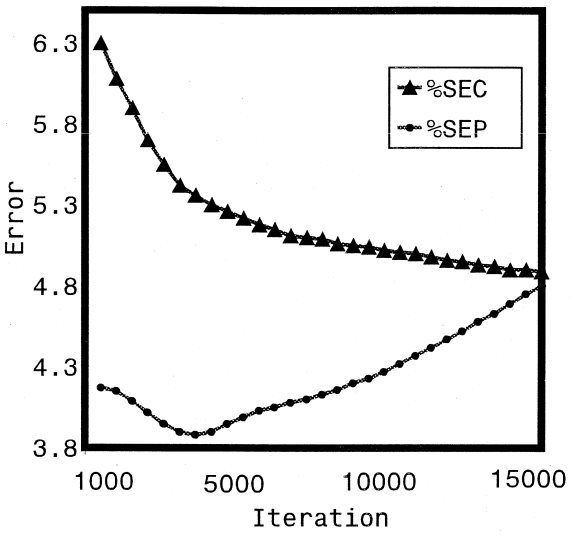

Fig. 1. A typical learning curve.

both descriptors MSA and BAL show some correlations with the cross-sectional area of the molecule and their appearances in the MLR model reveals that the TCD response factor of organic compounds depends on the size and the shape of the molecule. The appearance of the number of vibrational (NVIB) modes with a positive mean effect in the model explains the overall increase in RF with increasing vibrational motion. This indicates that the heated sensor of a thermal conductivity detector transfers more caloric energy to the molecules as the number of vibrational modes increases.

The next step was the generation of an ANN using the descriptors appearing in the MLR model as inputs. Fig. 1 shows a plot of SEC and SEP versus the number of iterations, which represents the 'learning curve' and was used to estimate the extent of training. It can be seen that while the learning curve for the training set continues to decrease during the progression of iteration, the prediction set's learning curve initially decreases and then starts to increase after about 4400 iterations. This situation is called overtraining and causes the ANN to be overfitted to the training set and then loses its predictive power. Therefore, during optimization of the ANN parameters, it is desirable that iterations be stopped when overtraining begins.

In order to determine the optimum number of nodes in the hidden layer, several training sessions were conducted with different numbers of hidden nodes. The values of SEC and SEP were calculated after each 100 iterations and calculation was stopped when overtraining began; the SEP and SEC values were then recorded. The recorded values of SEP and

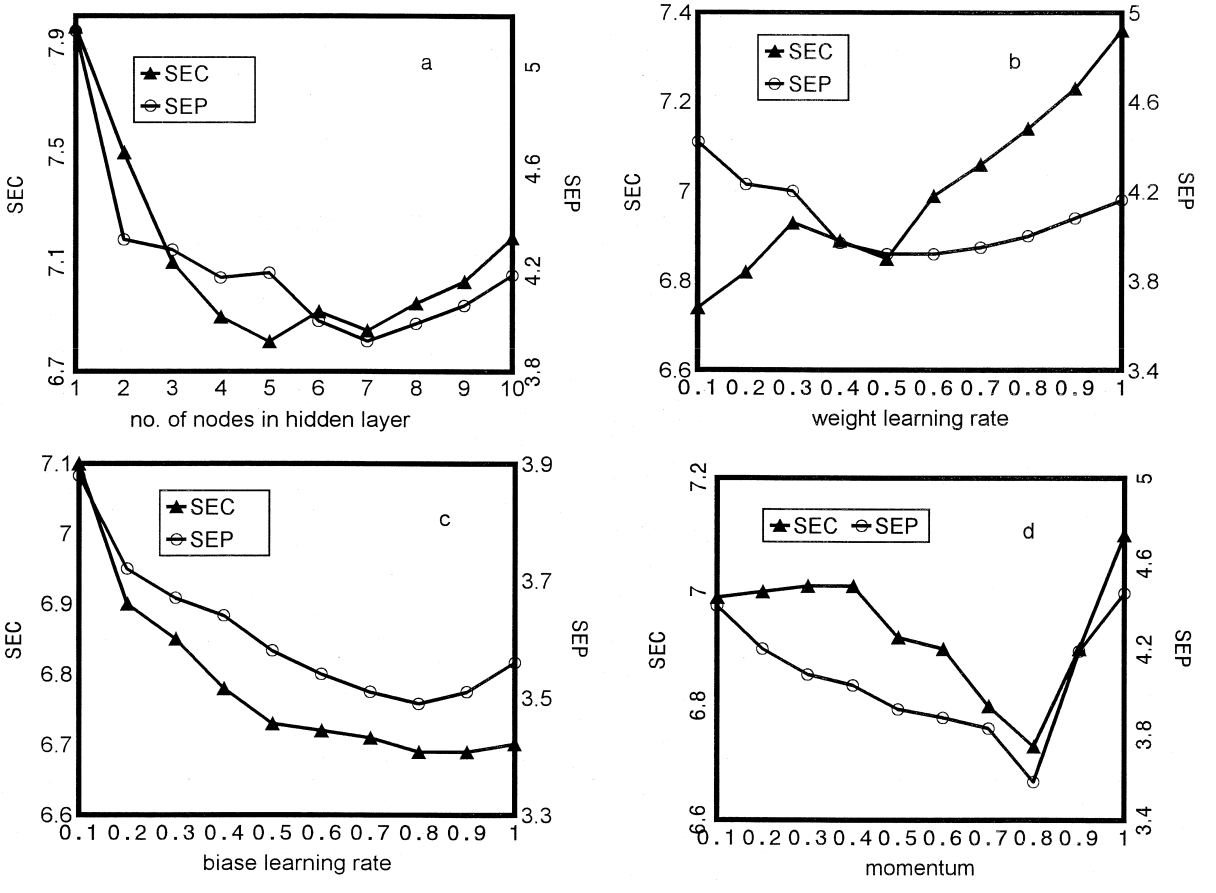

Fig. 2. Values of SEC and SEP when overfitting began against (a) the number of nodes in the hidden layer, (b) the weight learning rate, (c) the momentum, (d) the bias learning rate.

SEC were plotted against the number of nodes in the hidden layer, and the number of hidden nodes with minimum values of SEC and SEP was chosen as the optimum (Fig. 2a). It can be seen that seven nodes in the hidden layer were sufficient for a good performance of the network. It is common to use only the plot of SEC against the number of iterations for optimization of the ANN. However, as a new strategy we have used both SEC and SEP values for obtaining the optimum values for the number of hidden layers, learning rates and momentum [25]. It is noteworthy that weight and bias values were optimized using only the training set and the prediction set plays the role of choosing criteria for stopping training of the network. Learning rates of weights and biases and the value of the momentum were also optimized in a similar way and the results obtained are shown in Fig. 2b–d, respectively. As can be seen the optimum values of the weights and biases learning rates and momentum were 0.5, 0.8 and 0.8, respectively. The generated ANN was then trained using the training set for the optimization of the weights and biases. However, training was stopped when overtraining began. For evaluation of the prediction power of the network, the trained ANN was used to predict the TCD response factors of the molecules included in the prediction set (Table 2).

Tables 1 and 2 present the experimental and calculated values of the RFs using the generated ANN and MLR models for the training and the prediction sets, respectively. The calculated values of the descriptors appearing in the MLR and ANN models are also included in these tables. Table 4 compares the results obtained using the MLR and ANN models. The SEP values of the ANN and MLR models are 3.4 and 4.6%, respectively. Comparison between these values and other statistical parameters included in Table 4 shows the superiority of the ANN over that of the MLR model for the prediction of the TCD response factor. It is worth noting that, in the absence of BAL as an input, the values of SEC and SEP for the ANN model were 6.4 and 4.5%, respectively. Comparison of these values with the values of 5.7 and 3.4% obtained in the presence of the Balaban index justifies the inclusion of BAL in the MLR model. It can be seen from Table 2 that the ANN model was able to predict accurately most of the RFs, especially for molecules 2, 13, 17 and 19. The largest difference between the calculated and experimental values of the RFs is 0.1, which is due to heptane. As both the MLR and ANN models predict a too low value for the RF of this molecule, one may conclude that the experiment overestimates the TCD-RF of this molecule. The mean absolute error between the calculated and the experimental values of the RFs is 0.02 for the prediction set.

Fig. 3 shows a plot of the calculated versus the experimental values of RFs for the prediction set. The correlation coefficient of 0.984 for this plot confirms the ability of the ANN model to predict TCD-RFs. The residuals of the ANN calculated values of the RFs are plotted against the experimental values in Fig. 4. The propagation of the residuals

Table 4
Comparison between results obtained from the ANN and MLR models[a]

Model	SEC (%)	SEP (%)	R_t	R_p	F_t	F_p
ANN	5.7	3.4	0.980	0.984	2168	563
MLR	6.7	4.6	0.974	0.971	1650	300

[a] t, training set; p, prediction set; *R*, correlation coefficient; *F*, statistical *F* value.

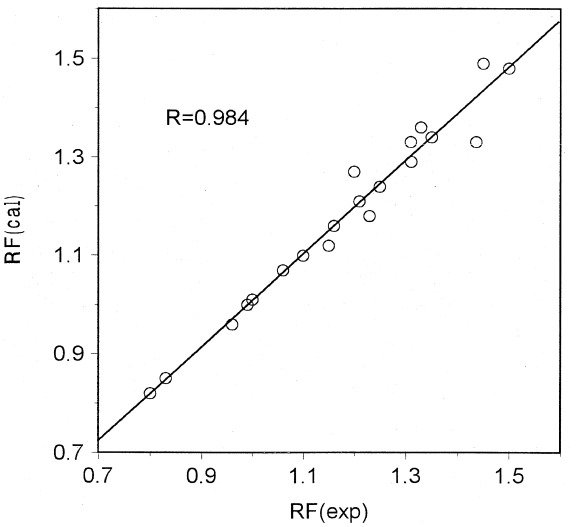

Fig. 3. Plot of calculated response factors against the experimental values.

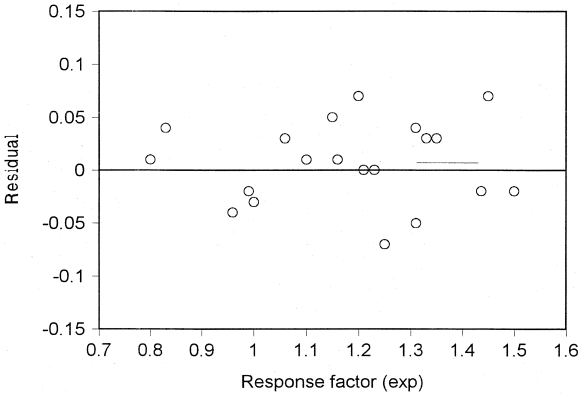

Fig. 4. Plot of residuals versus experimental response factors.

on both sides of zero indicates that no systematic error exists in the development of the neural network.

In general, the good agreement between the experimental results and the predicted values using the ANN model confirms its validity and shows that it can be used successfully for a better understanding of the mechanism of heat transfer in a thermal conductivity detector.

References

[1] D.A. Skoog, F.J. Holler, T.A. Nieman, Principles of Instrumental Analysis, 5th ed., Saunders College, New York, 1998, Chapter 26.
[2] J.T. Scanlon, D.E. Willis, J. Chromatogr. Sci. 23 (1985) 333.
[3] R.L. Grob, Modern Practice of Gas Chromatography, 3rd ed., Wiley, New York, 1995, Chapter 5.
[4] D.T. Stanton, P.C. Jurs, Anal. Chem. 61 (1989) 1328.
[5] D.T. Stanton, P.C. Jurs, Anal. Chem. 62 (1990) 2323.
[6] A.R. Katritzky, E.S. Ignatchenko, R.A. Barcock, V.S. Lobanov, M. Karelson, Anal. Chem. 66 (1994) 1799.
[7] M.E. Munk, M.S. Madison, E.W. Rubb, J. Chem. Inf. Comput. Sci. 36 (1996) 231.
[8] C. Klawum, C.L. Wilkins, J. Chem. Inf. Comput. Sci. 36 (1996) 249.
[9] S.L. Anker, P.C. Jurs, Anal. Chem. 64 (1992) 1157.
[10] W.-L. Xing, X.-W. He, Anal. Chim. Acta 349 (1997) 283.
[11] K.L. Peterson, Anal. Chem. 64 (1992) 379.
[12] S. Agatonovic Kustrin, M. Zecevic, Lj. Zivanovic, I.G. Tucker, Anal. Chim. Acta 364 (1998) 265.
[13] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 825 (1998) 161.
[14] M.T. Hagan, H.B. Demuth, M. Beal, Neural Network Design, PWS, Boston, 1996.
[15] S. Haykin, Neural Network, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[16] N.K. Bose, P. Liang, Neural Network, Fundamentals, McGraw-Hill, New York, 1996.
[17] O.R. Anino, R. Villalobos, Process Gas Chromatography, Instrument Society of America, NC, 1992, Chapter 9.
[18] E.F. Barry, R.S. Fischer, D.M. Rosie, Anal. Chem. 44 (1972) 1559.
[19] E.G. Hoffmann, Anal. Chem. 34 (1962) 1216.
[20] E.K. Whalden-Pederson, P.C. Jurs, Anal. Chem. 53 (1981) 484.
[21] T.R. Stouch, P.C. Jurs, J. Chem. Inf. Comput. Sci. 26 (1986) 26.
[22] J.J.P. Stewart, MOPAC, A semiempirical molecular orbital program, QCPE, 455 (1983), Version 6 (1990).
[23] N. Draper, H. Smith (Eds.), Applied Regression Analysis, 2nd ed., Wiley-Interscience, New York, 1981, p. 307.
[24] T.B. Blank, S.T. Brown, Anal. Chem. 65 (1993) 3084.
[25] M. Jalali-Heravi, M.H. Fatemi, Anal. Chim. Acta 415 (2000) 95.