

**PREDICTION OF PROGRAMMED-TEMPERATURE  
RETENTION INDICES OF POLYCYCLIC AROMATIC  
HYDROCARBONS IN THE LEE INDEX SCALE BY  
ARTIFICIAL NEURAL NETWORK**

B. Škrbić<sup>a,\*</sup> and A. Onjia<sup>b</sup>

<sup>a</sup> Faculty of Technology, University of Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad,  
Serbia

<sup>b</sup> Vinča Institute of Nuclear Sciences, POB 522, 11001 Belgrade, Serbia

(Received September 13, 2005)

**Abstract**

A feed-forward artificial neural network (ANN) model was used to predict the programmed-temperature retention indices RIs in the Lee index scale of polycyclic aromatic hydrocarbons (PAHs). The data used in this paper include 96 RIs in Lee index scale of 48 parent and alkylated PAHs obtained on SE-52 and DB-5 slightly polar stationary phases with three different temperature programmes. Four parameters: boiling point, molecular weight, connectivity index and F-number were used as input parameters. The data containing 96 RIs

---

\* Corresponding author. Tel.: +381/21/450288; fax: +381/21/450413; e-mail address: [biljana@tehnol.ns.ac.yu](mailto:biljana@tehnol.ns.ac.yu) (B. Škrbić)

were randomly divided into three sets: a training set (including 32 RIs obtained on SE-52 stationary phase), a validation set (including 16 RIs for the same SE-52 phase) and testing sets (including 48 RIs obtained on DB-5 stationary phase and with two different temperature programmes, 30 at one and 18 RIs at the other programmes, respectively). The structures of networks and the number of learning epochs were optimized. The best network structure is 4-6-1. The optimum number of learning epoch is 1000. The results obtained in this study showed that the average percentage deviation between the predicted ANN values and the experimental values of Lee retention indices for the validation and two testing sets were 1.42% on the SE-52 and 1.32 and 1.43% on the DB-5 stationary phases, respectively. The result illustrated that the prediction performance of ANN in the field of investigating the programmed-temperature retention behavior of polycyclic aromatic hydrocarbons is very satisfactory.

## **Introduction**

Polycyclic aromatic hydrocarbons (PAHs) represent a very important group of chemical carcinogens, which identification makes possible determination of the environmental pollution level [1, 2], and as well as evaluation and control of the overall public health hazard. In some cases programmed-temperature gas-chromatography (GC) may be the sole means of identification based on direct comparison of retention times with standards or precise knowledge of retention indices (RI values). Among presentations of gas-chromatographic data for PAHs, the programmed-temperature retention index [3] in Lee index scale [4, 5] based on PAH internal standards (naphthalene, phenanthrene, chrysene, picene) is preferred because of its much better reproducibility.

PAHs retention indices are calculated by a linear relationship suggested by van Den Dool and Kratz [3] for linear-temperature programmed GC:

$$RI=100 \cdot \frac{T_{R(\text{substance})} - T_{R(Cz)}}{T_{R(Cz+1)} - T_{R(Cz)}} + 100 \cdot z$$

where  $T_{R(\text{substance})}$  is the retention time or elution temperature of the substance for which the retention index is to be determined,  $T_{R(Cz)}$  and  $T_{R(Cz+1)}$  are retention times or elution temperatures for the PAH internal standards of interest, and  $z$  is the number of rings in the PAH standard that elutes prior to the substance of interest.

Retention is a phenomenon that is primary dependent on the interactions between solute and the stationary phase molecules. The kind of interaction depends on the structure and properties of the stationary phase and the solute molecules. The retention index of a solute is an important parameter in the study of the quantitative structure-retention relationships (QSRRs). Typically, either empirical physico-chemical parameters or non-empirical structural descriptor parameters have been used in order to obtain quantitative or semi-quantitative relationships which allow the prediction of the retention behavior of an individual solute of a given class. The QSRRs are often expressed in the form of a linear equation whose independent variables are the empirical physico-chemical or non-empirical structural descriptor parameters and whose depend variable is a retention index. The equation is obtained by performing a multivariate linear regression on data for compounds of a given class on a given stationary phase. A number of theoretical and empirical models which predict the retention data of solutes have been published [6-8].

Artificial neural networks (ANN) have been found to be powerful tools in chemistry and have been applied to wide variety of chemical problems. A comprehensive review of ANN and its application in chemistry is given by Zupan and Gasteiger [9].

Prediction of gas chromatographic retention index data by neural networks has been carried for different class of compounds: alkanes, alkenes, alcohols, esters, ketones and ethers by Yan et al. [10], alkylbenzenes by Zhang et al. [11], noncyclic and monocyclic terpenes by Jalali-Heravi and Fatemi [12], disulfides by Gao et al. [13].

In this report, a back propagation ANN with delta-bar-delta learning algorithm model was employed to generate QSRRs between the molecular-based structural parameters and programmed-temperature Lee retention indices. The usefulness of these ANNs in the modeling and predicting of programmed-temperature retention indices from boiling point  $T_b$ , molecular weight  $M$ , connectivity index  $\chi$  and F-number were studied for polycyclic aromatic hydrocarbons on two different stationary phases: SE-52 and DB-5 and three different temperature programmes. The result showed that the prediction performance of ANN is very satisfactory.

## **Method**

### **Brief description of neural networks**

A neural network model is composed of a large number of simple processing elements or neuron nodes organized into a sequence of layers.

The first layer is the input layer with one node for each variable or feature of the data. The last layer is the output layer consisting of one node for each variable to be investigated. In between are series of one or more hidden layer(s) consisting of number of nodes, which are responsible for learning. Nodes in any layer are fully or randomly connected to nodes of a succeeding layer. Each connection is represented by a number called a weight.

Multilayer feed forward networks are most often used to analyze nonlinear multivariable data. In these networks, signals are propagated from the input layer through the hidden layer(s) to the output layer. A node thus receives signals via connections from other nodes or the outside world in the case of the input layer. The net input for a node  $j$  is given by:

$$\text{net}_j = \sum_i w_{ji} \cdot o_i$$

where  $i$  represents nodes in the previous layer,  $w_{ji}$  is the weight associated with the connection from node  $i$  to node  $j$ , and  $o_i$  is the output of node  $i$ . The output of a node is determined by the transfer function and the net input of the node. The following sigmoidal transfer function in the hidden layer is used

$$f(\text{net}_j) = \frac{1}{1 + e^{-(\text{net}_j + \Theta_j)}}$$

where  $\Theta_j$  is a bias term of threshold value of node  $j$  responsible for accommodation nonzero offsets in the data.

The adequate functioning of a neural network strongly depends on the way the signals are propagated through the network. The weights play an important role in this propagation and a proper setting of these weight factors is essential. Generally, such a setting is not known beforehand and the weights are initially given small, random values. The process of adapting the weights to an optimum set of values is called training the neural network. This training is usually done by means of supervised learning. A representative training set with examples is presented iteratively to the neural network and the difference between the desired solution and the one obtained is used to adapt the weights in small steps, according to a learning algorithm. There is a number of learning algorithms used to train a neural network. A frequently used one is called the back propagation learning rule [11].

One of the main problems with back propagation ANN is that long training sessions are often required in order to find an acceptable weight solution because of the well known difficulties inherent in gradient descent optimization. There exist a number of modifications to this algorithm which are designed to overcome this problem. The delta-bar-delta learning algorithm is one of the modifications to this algorithm, and it is used here [14].

### **The neural network models and data**

The data set of programmed-temperature Lee retention indices of polycyclic aromatic hydrocarbons was taken from Lee et al. [4], Lundstedt et al. [15] and Takada et al. [16]. The retention indices of all polycyclic aromatic hydrocarbons included in the data set were obtained on two slightly polar stationary phases SE-52 and DB-5 with different temperature programmes. The Lee retention data measured by Lee [4] were obtained on SE-52 with temperature programmed from 50 to 250°C, at 2°C min<sup>-1</sup>; while the ones determined on DB-5 stationary phase were with two different temperature programmes: from 80°C for 2 min increased at 8°C min<sup>-1</sup> to 300°C held for 10 min [15]; and with the temperature range from 70°C held 2 min, with 30°C min<sup>-1</sup> to 150°C, then with the run of 5°C min<sup>-1</sup> to 200°C and 4°C min<sup>-1</sup> to 310°C, held for 5 min [16], respectively. The boiling points and molecular weights, the physico-chemical parameters of these compounds, were taken from [17, 18]. The connectivity index, that is a topological parameter, was calculated according to ref. [19]. F-number, a molecular size descriptor suggested by Schabron et al. [20], was calculated as follows:

$$F = (\text{number of double bounds}) + (\text{number of primary and secondary carbons}) \\ - 0.5 (\text{number of non-aromatic rings})$$

The F-numbers included here are partly values found in the literature [21-24] and partly calculated values.

The compounds and their parameters: boiling point Tb, molecular weight M, connectivity index  $\chi$ , F-number and retention index RI, were given in Table 1.

**Table 1**

The compounds and data used in this paper

No.	Compound	M	Tb, °C	$\chi$	F	RI
1	2-Methylnaphthalene	142.2	241	3.815	6	218.14 [4]
2	2-Methylnaphthalene	142.2	241	3.815	6	220.10 [15]
3	1-Methylnaphthalene	142.2	245	3.821	6	221.04 [4]
4	1-Methylnaphthalene	142.2	245	3.821	6	223.10 [15]
5	2-Ethyl naphthalene	156.23	258	4.376	7	236.08 [4]
6	1-Ethyl naphthalene	156.23	259	4.486	7	236.56 [4]
7	2,6-Dimethylnaphthalene	156.23	262	4.226	7	237.58 [4]
8	2,6-Dimethylnaphthalene	156.23	262	4.226	7	239.70 [15]
9	2,7-Dimethylnaphthalene	156.23	262	4.226	7	237.71 [4]
10	1,3-Dimethylnaphthalene	156.23	265	4.232	7	240.25 [4]
11	1,7-Dimethylnaphthalene	156.23	263	4.232	7	240.66 [4]
12	1,6-Dimethylnaphthalene	156.23	266	4.232	7	240.72 [4]
13	2,3-Dimethylnaphthalene	156.23	268	4.232	7	243.55 [4]
14	1,4-Dimethylnaphthalene	156.23	268	4.238	7	243.57 [4]
15	1,5-Dimethyl naphthalene	156.23	269	4.238	7	244.98 [4]
16	Acenaphthylene	152.21	270	4.149	6.5	244.63 [4]
17	Acenaphthylene	152.21	270	4.149	6.5	247.40 [15]
18	1,2-Dimethylnaphthalene	156.23	271	4.238	7	246.49 [4]
19	Acenaphthene	154.21	279	4.445	5.5	251.29 [4]
20	Acenaphthene	154.21	279	4.445	5.5	253.30 [15]
21	Fluorene	166.23	294	4.612	6.5	268.17 [4]
22	Fluorene	166.23	294	4.612	6.5	269.60 [15]
23	2-Methylfluorene	180.25	318	5.022	7.5	288.21 [4]
24	2-Methylfluorene	180.25	318	5.022	7.5	287.70 [15]
25	1-Methylfluorene	180.25	318	5.028	7.5	289.03 [4]
26	1-Methylfluorene	180.25	318	5.028	7.5	288.70 [15]
27	Anthracene	178.24	340	4.809	7	301.69 [4]
28	Anthracene	178.24	340	4.809	7	301.40 [15]
29	Anthracene	178.24	340	4.809	7	301.20 [16]
30	3-Methylphenanthrene	192.26	352	5.226	8	319.46 [4]
31	3-Methylphenanthrene	192.26	352	5.226	8	318.60 [15]
32	3-Methylphenanthrene	192.26	352	5.226	8	315.60 [16]
33	2-Methylphenanthrene	192.26	355	5.226	8	320.17 [4]
34	2-Methylphenanthrene	192.26	355	5.226	8	319.50 [15]
35	2-Methylphenanthrene	192.26	355	5.226	8	316.40 [16]
36	2-Methylanthracene	192.26	359	5.22	7	321.57 [4]
37	2-Methylanthracene	192.26	359	5.22	7	320.90 [15]
38	4H-Ciklopenta(def)phenanthrene	190.24	359	5.356	7.5	322.08 [4]
39	4H-Ciklopenta(def)phenanthrene	190.24	359	5.356	7.5	322.30 [15]
40	9-Methylphenanthrene	192.26	355	5.232	8	323.06 [4]
41	9-Methylphenanthrene	192.26	355	5.232	8	319.20 [16]
42	1-Methylanthracene	192.26	363	5.226	8	323.33 [4]
43	1-Methylphenanthrene	192.26	359	5.232	8	323.90 [4]

44	1-Methylphenanthrene	192.26	359	5.232	8	323.60 [15]
45	1-Methylphenanthrene	192.26	359	5.232	8	320.00 [16]
46	3,6-Dimethylphenanthrene	206.29	363	5.637	9	337.83 [4]
47	Fluoranthene	202.26	383	5.565	8.5	344.01 [4]
48	Fluoranthene	202.26	383	5.565	8.5	344.90 [15]
49	Fluoranthene	202.26	383	5.565	8.5	340.10 [16]
50	Pyrene	202.26	393	5.559	8	351.22 [4]
51	Pyrene	202.26	393	5.559	8	352.80 [15]
52	Pyrene	202.26	393	5.559	8	348.10 [16]
53	Benzo(a)fluorene	216.29	407	6.022	8.5	366.74 [4]
54	Benzo(a)fluorene	216.29	407	6.022	8.5	366.50 [15]
55	Benzo(b)fluorene	216.29	402	6.017	8.5	369.39 [4]
56	Benzo(b)fluorene	216.29	402	6.017	8.5	368.90 [15]
57	4-Methylpyrene	216.29	410	5.976	9	369.54 [4]
58	2-Methylpyrene	216.29	410	5.97	9	370.15 [4]
59	1-Methylpyrene	216.29	410	5.976	9	373.55 [4]
60	1-Methylpyrene	216.29	410	5.976	9	374.20 [15]
61	1-Methylpyrene	216.29	410	5.976	9	370.00 [16]
62	Benzo(ghi)fluoranthene	226.28	432	6.309	9.5	389.60 [4]
63	Benzo(ghi)fluoranthene	226.28	432	6.309	9.5	390.90 [15]
64	Benzo(ghi)fluoranthene	226.28	432	6.309	9.5	389.60 [16]
65	Benzo(a)anthracene	228.3	435	6.22	9	398.50 [4]
66	Benzo(a)anthracene	228.3	435	6.22	9	398.60 [15]
67	Benzo(a)anthracene	228.3	435	6.22	9	398.40 [16]
68	Benzo(j)fluoranthene	252.32	480	6.976	9.5	440.92 [4]
69	Benzo(b)fluoranthene	252.32	481	6.976	9.5	441.74 [4]
70	Benzo(b)fluoranthene	252.32	481	6.976	9.5	442.10 [15]
71	Benzo(b)fluoranthene	252.32	481	6.976	9.5	442.70 [16]
72	Benzo(k)fluoranthene	252.32	481	6.97	9.5	442.56 [4]
73	Benzo(k)fluoranthene	252.32	481	6.97	9.5	443.60 [15]
74	Benzo(k)fluoranthene	252.32	481	6.97	9.5	442.80 [16]
75	Benzo(e)pyrene	252.32	493	6.975	9	450.73 [4]
76	Benzo(e)pyrene	252.32	493	6.975	9	451.80 [15]
77	Benzo(e)pyrene	252.32	493	6.975	9	452.70 [16]
78	Benzo(a)pyrene	252.32	496	6.97	10	453.44 [4]
79	Benzo(a)pyrene	252.32	496	6.97	10	453.40 [15]
80	Benzo(a)pyrene	252.32	496	6.97	10	454.30 [16]
81	Perylene	252.32	497	6.975	10	456.22 [4]
82	Perylene	252.32	497	6.975	10	456.30 [15]
83	Perylene	252.32	497	6.975	10	457.50 [16]
84	Pentacene	278.35	529	7.619	11	486.81 [4]
85	Dibenzo(ac)anthracene	278.36	535	7.637	11	495.01 [4]
86	Dibenzo(ac)anthracene	278.36	535	7.637	11	495.10 [16]
87	Dibenzo(ah)anthracene	278.36	535	7.631	11	495.45 [4]
88	Dibenzo(ah)anthracene	278.36	535	7.631	11	494.50 [15]
89	Dibenzo(ah)anthracene	278.36	535	7.631	11	499.00 [16]
90	Benzo(b)chrysene	278.36	541	7.631	11	497.66 [4]
91	Benzo(b)chrysene	278.36	541	7.631	11	498.50 [15]
92	Benzo(ghi)perylene	276.34	542	7.72	11	501.32 [4]
93	Benzo(ghi)perylene	276.34	542	7.72	11	502.90 [15]
94	Benzo(ghi)perylene	276.34	542	7.72	11	501.30 [16]
95	Anthanthrene	276.34	547	7.714	11	503.89 [4]
96	Anthanthrene	276.34	547	7.714	11	508.40 [15]

data obtained: on SE-52 [4], on DB-5 [15,16] stationary phases and three different temperature-programmes, respectively

The parameters were chosen on the base of Kitty et al. study [25] who applied PCA – principal component analysis to investigate the influence of 13 physico-chemical properties on retention time of 60 PAHs obtained on DB-5 stationary phase. Namely, it was shown that above used parameters as input data mainly effecting PC1 that describes 68% of all data. The first four parameters were used as input parameters and the retention index was used as output. The data containing 96 RIs were randomly divided into three sets: a training set (32 RIs obtained on SE-52 stationary phase with temperature programme: from 50 to 250°C, at 2°C min<sup>-1</sup>); a validation set (16 RIs obtained on the same SE-52 with the same temperature programme) and testing sets (including 48 RIs obtained on DB-5 stationary phase and two different temperature programmes, 30 at one and 18 RIs at the other temperature programme [15, 16], respectively). They were shown in Table 2.

**Table 2**

Training, validation and testing sets \*

Sets	Compound numbers
Training set	1, 5, 7, 10, 11, 12, 14, 15, 19, 21, 23, 27, 30, 36, 38, 40, 43, 46, 47, 53, 55, 59, 62, 65, 68, 72, 75, 81, 84, 85, 90, 95
Validation set	3, 6, 9, 13, 16, 18, 25, 33, 42, 50, 57, 58, 69, 78, 87, 92
Testing set I	2, 4, 8, 17, 20, 22, 24, 26, 28, 31, 34, 37, 39, 44, 48, 51, 54, 56, 60, 63, 66, 70, 74, 76, 79, 82, 88, 91, 93, 96
Testing set II	29, 32, 35, 41, 45, 49, 52, 61, 62, 67, 71, 73, 77, 80, 83, 86, 89, 94

\*See Table 1 for name of compound from No.

The training set is used to train a neural net. The error of this data set is minimized during training. The validation set is used to determine the performance of a neural network on patterns that are trained during learning. A testing set for finally checking the over all performance of a neural net. All the data were scaled to (0,1) before training using following equation:

$$a_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$$

where  $x_i$  is the any one of the data in the training set.  $x_{\max}$  and  $x_{\min}$  are the maximum and minimum values of the data in the training set.

The ANN systems were simulated using a QwikNet ANN simulator (Craig Jensen, Redmond, USA).

## Results and discussion

### Network Optimization

The neural network had four input nodes, boiling point, molecular weight, connectivity index and F-number, one hidden layer of  $x$  nodes, and a single output node. Such an ANN may be designated as 4-x-1 net to indicate the number of nodes in input, hidden and output layers, respectively.

The data were combined into three sets: training, validation and testing (see Table 2). The neural network was trained directly on the training set, and its performance was monitored using the validation set. Through this process, we can select the best networks. The best ANNs determined by the validation set were used for testing 48 parent and alkylated polycyclic aromatic hydrocarbons from the test sets.

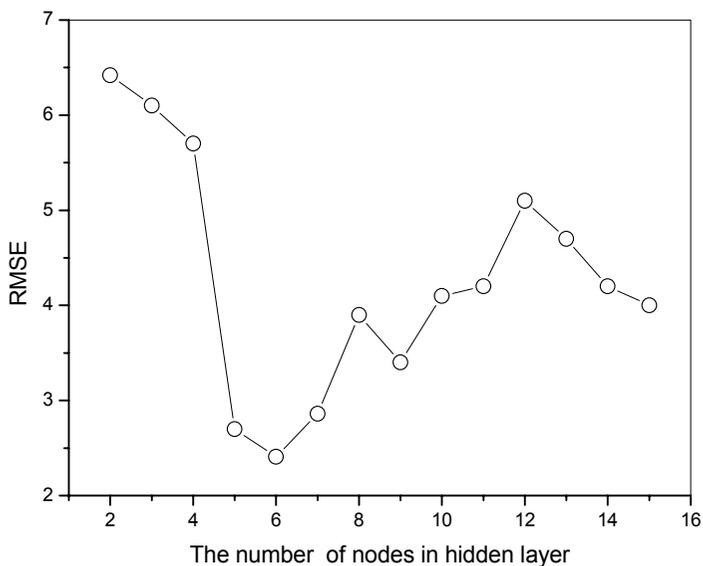
Optimization of a neural network is difficult and time-consuming, and one has to use a trial and error method to find the best neural network architecture. In order to determine the optimal number of hidden layer nodes, ANNs with different numbers of hidden nodes were trained. The number of hidden nodes was varied from 2 to 15 and the root mean square error (RMSE) function was calculated:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (o_i - d_i)^2}{n}}$$

where  $d_i$  is a desired output (experimental values),  $o_i$  is the actual output (ANN predicted values),  $n$  is the number of compounds in the analyzed set. According to Fig. 1, ANN with 6 hidden nodes had the lowest RMSE error, and that number of nodes was chosen for further optimization.

**Fig. 1.** The hidden node numbers vs. average RMS error on validation set

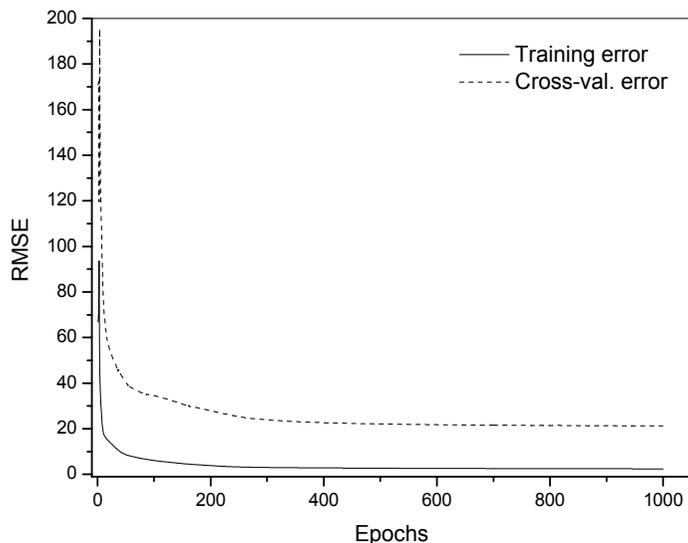
**ANN**



RMSE(2)=6.42, RMSE(6)=2.41, RMSE(12)=5.10, RMSE(15)=4.00

To select the best learning times, we made the curves of RMSE values of training set and validation set versus learning epochs when the number of hidden layer nodes was six. They are shown in Fig. 2.

**Fig. 2.** The average RMS error vs. the number of training epochs for training and validation data sets



RMSE\_train(0)=66.82, RMSE\_train(1000)=2.41,  
RMSE\_cross-val(0)=170.83, RMSE\_cross-val (1000)=21.27

From the validation curve, we can see that the optimum learning epochs is about 1000.

### **Prediction of programmed-temperature retention index data using neural networks**

Through the above process, we learn that the best number of hidden layer nodes is 6 and the optimum number of learning epochs is about 1000.

For the evaluation of the predictive power of the network, a trained ANN was used to predict the temperature programmed Lee retention indices of the molecules included in the validation and testing sets. Table 3 represents the experimental and ANN predicted values of retention indices on SE-52 stationary phase, while Tables 4 and 5 represent the comparison of the

experimental and ANN predicted values on DB-5 stationary phase for two testing sets, respectively. Accuracy as a degree of agreement between the experimental and the predicted retention indices in the Lee scale for all studied PAHs included in Tables 3-5 was expressed by relative error (RE) and was calculated as:

$$RE = \frac{RI_{pred} - RI_{exp}}{RI_{exp}} \cdot 100$$

**Table 3**

Experimental and predicted programmed-temperature retention indices in the Lee scale of the validation set [4]

No.	Experimental data	Predicted data	Relative Error (%)
3	221.04	226.01	2.25
6	236.56	241.49	2.08
9	237.71	238.52	0.34
13	243.55	241.31	-0.92
16	244.63	239.96	-1.91
18	246.49	242.84	-1.48
25	289.03	292.58	1.22
33	320.17	321.76	0.50
42	323.33	326.53	0.99
50	351.22	351.09	-0.04
57	369.54	371.86	0.63
58	370.15	371.92	0.48
69	441.74	443.14	0.32
78	453.44	451.89	-0.34
87	495.45	474.47	-4.23
92	501.32	475.60	-5.13

**Table 4**

Experimental and predicted programmed-temperature retention indices in the Lee scale of the testing set I [15]

No.	Experimental data	Predicted data	Relative Error (%)
2	220.1	225.19	2.31
4	223.1	226.01	1.30
8	239.7	238.52	-0.49
17	247.4	239.96	-3.01
20	253.3	250.88	-0.95
22	269.6	266.34	-1.21
24	287.7	292.53	1.68
26	288.7	292.58	1.34
28	301.4	301.92	0.17
31	318.6	320.02	0.44
34	319.5	321.76	0.70
37	320.9	322.05	0.36
39	322.3	321.43	-0.26
44	323.6	324.12	0.16
48	344.9	345.26	0.10
51	352.8	351.09	-0.48
54	366.5	368.49	0.54
56	368.9	364.85	-1.09
60	374.2	371.86	-0.62
63	390.9	393.47	0.66
66	398.6	398.60	0.00
70	442.1	443.14	0.23
74	442.8	443.21	0.09
76	451.8	451.31	-0.11
79	453.4	451.89	-0.33
82	456.3	452.41	-0.85
88	494.5	474.47	-4.05
91	498.5	476.24	-4.46
93	502.9	475.60	-5.43
96	508.4	477.02	-6.17

**Table 5**

Experimental and predicted programmed-temperature retention indices in the Lee scale of the testing set II [16]

No.	Experimental data	Predicted data	Relative Error (%)
29	301.2	301.92	0.24
32	315.6	320.02	1.40
35	316.4	321.76	1.69
41	319.2	321.76	0.80
45	320.0	324.12	1.29
49	340.1	345.26	1.52
52	348.1	351.09	0.86
61	370.0	371.86	0.50
62	389.6	393.47	0.99
67	398.4	398.60	0.05
71	442.7	443.14	0.09
73	443.6	443.21	-0.09
77	452.7	451.32	-0.30
80	454.3	451.89	-0.53
83	457.5	452.42	-1.11
86	495.1	474.45	-4.17
89	499.0	474.47	-4.91
94	501.3	475.60	-5.12

The maximum relative error for predicted retention indices in the Lee scale on SE-52 stationary phase is -5.13% for benzo(ghi)perylene and the minimum value is -0.04% for pyrene; while on DB-5 stationary phases with two different temperature programmes the maximum percentage error are -6.17% for anthanthrene in test set I [15] and -5.12% for benzo(ghi)perylene in the test set II [16], respectively, and the minimum values are 0.10% for fluoranthene (test I) and 0.05% for benzo(a)anthracene (test II), respectively. The average percentage deviation between the programmed-temperature predicted and the experimental values of Lee retention indices are 1.43% on SE-52; and 1.32% and 1.42% on DB-5 stationary phases, respectively. However, it is worth noting that these values are in agreement with the results obtained by experiment.

**Fig. 3.** Predicted vs. experimental programmed-temperature retention indices in the Lee scale

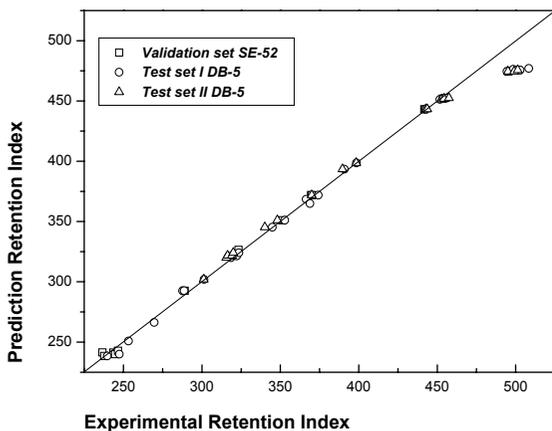


Fig. 3 shows the plot of the ANN predicted versus the experimental values for the programmed-temperature Lee retention indices of the validation and the test sets and illustrate that the predicted data of the validation and testing sets are in good agreement with those of the experimental data.

## Conclusion

The results of this study demonstrate that QSRRs method using ANN techniques can generate a suitable model for prediction of programmed-temperature Lee retention indices on SE-52 and DB-5 stationary phases, simultaneously. The boiling point, molecular weight, connectivity index and F-number can be considered as comprehensive descriptors for

predicting the programmed-temperature Lee retention indices of polycyclic aromatic hydrocarbons on slightly polar stationary phases SE-52 and DB-5.

## Acknowledgements

This work was partly supported by the Serbian Ministry for Science and Environmental Protection.

## References

1. Škrbić B, Miljević N (2002) *J. Environ. Sci. Health A37*:1029-1039
2. Škrbić B, Cvejanov J, Đurišić-Mladenović N (2005) *J. Environ. Sci. Health A40*:29-42
3. van Den Dool H, Kratz P (1963) *J. Chromatogr.* 11:463-471
4. Lee M, Vassilaros D, White C, Novotny M (1979) *Anal. Chem.* 51:768-773
5. Babushok V I, Linstrom P J (2004) *Chromatographia* 60:725-728
6. Heberger K (1989) *Anal. Chim. Acta* 223:161-174
7. Heberger K (1990) *Chromatographia* 29:375-384
8. Škrbić B, Đurišić-Mladenović N, Cvejanov J (2004) *Chemometr. Intell. Lab. Syst.* 72:167-171
9. Zupan J, Gasteiger J (1991) *Anal. Chim. Acta* 248:1-30
10. Yan A X, Zhang R S, Liu M C, Hu Z D, Hooper M A (1998) *Comput. Chem.* 22:405-412
11. Zhang R, Yan A, Liu M, Liu H, Hu Z (1999) *Chemometr. Intell. Lab. Syst.* 45:113-120
12. Jalali-Heravi M, Fatemi M H (2001) *J. Chromatogr.* 915:177-183
13. Gao Y, Wang Y, Yao X, Zhang X, Liu M, Hu Z, Fan B (2003) *Talanta* 59:229-237

14. Vasiljević T, Onjia A, Čokeša Đ, Laušević M (2004) *Talanta* 64:785-790
15. Lundstedt S, Haglund P, Oberg L (2003) *Environ. Toxicol. Chem.* 22:1413-1420
16. Takada H, Onda T, Ogura N (1990) *Environ. Sci. Technol.* 24:1179-1186
17. Bjorseth A (ed) (1983) *Handbook of Polycyclic Aromatic Hydrocarbons*. Marcel Dekker, New York
18. <http://chrom.tutms.tut.ac.jp/JINNO/DATABASE/00database.html>.
19. Randić M (1975) *J. Am. Chem. Soc.* 97:6609-6615
20. Schabron J F, Hurtubise R J, Silver H F (1977) *Anal. Chem.* 49:2253-2260
21. Jinno Lab (1999) Available from  
<http://chrom.tutms.tut.ac.jp/JINNO/DATABASE00alphabet.html>.
22. <http://wlapwww.gov.bc.ca>
23. Lee M, Novotny M, Bartle K (1981) *Analytical Chemistry of Polycyclic Aromatic Hydrocarbons*. Academic Press, New York
24. <http://www.epa.gov/oilspill/pdfs/neffpresent.pdf>
25. Kittl A, Harju M, Tysklind M, van Bavel B (2003) *Chemosphere* 50:627-637