

Optimal parametrization of structure for prediction of properties of alkanes[†]

M Kuanar, R K Mishra & B K Mishra*

Center of Studies in Surface Science and Technology, Department of Chemistry, Sambalpur University,
Jyoti Vihar 768 019, India

Received 15 April 1996; revised 13 June 1996

A generalised statistical model has been developed using algorithmically derived independent variables based on chemical structure for quantitative prediction of various physical properties of some hydrocarbons. Out of the fourteen principal components generated from principal component analysis (PCA) of the 14 topological parameters, by the use of only four principal components, the physical properties boiling point, molar volume, molecular refraction, heat of vaporisation, critical temperature and surface tension have been predicted reasonably well.

Application of statistical tools in predicting properties of organic molecules including drug molecules has led to efficient synthesis of new drugs with specific activity. Various techniques like principal component analysis (PCA), cluster analysis, discriminant analysis and multiple regression analysis (MRA) are in much use. Among these tools PCA and MRA have versatile applications in correlation analysis in chemistry.

Alkaline homologs are mostly used for quantitative structure property relationship (QSPR) studies because many properties of the alkanes vary in a regular manner with molecular mass and extent of branching. There is a small variation in intermolecular interactions and hence small perturbation is assumed during prediction of physical properties through models. Needham *et al.*¹ have carried out molecular modeling of the physical properties of alkanes using various topological and ad hoc descriptors. These descriptors can easily be calculated for known and unknown structures. Though the descriptors do not differentiate stereoisomers, the thermodynamic properties differ a little between diastereomers and not at all among enantiomers². Various models have already been proposed for the prediction of physical properties of alkanes. Basak *et al.*³⁻⁶ have made several attempts for the prediction of boiling point of alkanes by using various topological indices. In the present report we have used PCA for generating a new set of parameters to obtain a general set of descriptors for molecular modeling.

Method

The physical properties selected for the present investigation are boiling points (BP), molar volumes at 20°C (MV), molar refractions at 20°C (MR), heats of vaporisation at 25°C (HV), critical temperatures (TC) and surface tensions at 20°C (ST). All the values were taken from ref. 1.

Calculation of parameters

The Wiener number (W)⁷, proposed by Wiener is the first topological index and represents the sum of all possible lengths in a hydrogen-suppressed molecular graph. From the distance matrix (N × N matrix, where N is the number of vertices in the hydrogen depleted graph) W is obtained using Eq. 1.

$$W = \sum_{ij} d_{ij} / 2 \quad \dots (1)$$

where $d_{ij} = d_{ji}$ are the entries of the distance matrix representing the number of bonds between vertices i and j by the shortest path.

Shannon⁸, Wiener⁹, Ashby¹⁰, and Kolmogorov¹¹ were the pioneer workers in information theory. In Shannon's statistical information theory, information is measured as the reduced uncertainty of the system. Information indices have been used successfully in a variety of applications in chemistry^{12,13}. For this purpose, a total molecular graph (including hydrogen) is constructed for the molecule and each vertex (atom) in the graph, partition co-ordinates are assigned according to the bonding of the connected atoms in immediate neighbourhood. The co-ordinates bear the information

[†]Dedicated to H. Wiener in the Golden Jubilee year of the first topological parameter, W.

regarding the types of bonds between the concerned atom and the adjacent atom, and also the valency of the adjacent atom. The co-ordinates are then classified according to their partition co-ordinates. On the basis of Shannon's information theory, Kier's information index, I , is defined as¹⁴

$$I = -N \sum P_i \log P_i \quad \dots (2)$$

$$= N \log N - \sum n_i \log n_i$$

where $P_i = n_i/N$ is the probability of class i , n_i is the number of members in class i , and N is the total number of atoms in the compound.

Wiener also introduced the parameter P_3 , the number of paths with three C-C bond length which was originally as a "polarizability" factor^{7,15} but later was found to be associated with steric crowding¹⁶. The parameter 'f' introduced by Platt¹⁶ has been calculated by taking the number of adjacent bonds to each bond and summing these values for all bonds in the carbon skeleton.

Randic's connectivity index χ , has wide application in correlation analysis in chemistry¹⁷⁻¹⁹. To evaluate the parameter each carbon atom 'i' in the carbon skeleton of a hydrocarbon is assigned a valence, δ_i , equal to the number of carbon bonds to that atom. The zeroth-order connectivity index ${}^0\chi$ is the sum of all atoms i .

$${}^0\chi = \sum_{\text{atoms}} 1/\delta_i^{1/2} \quad \dots (3)$$

The first order connectivity index ${}^1\chi$ is the sum of all bonds $i-j$ (Eq. 4).

$${}^1\chi = \sum_{\text{bonds}} 1/(\delta_i \delta_j)^{1/2} \quad \dots (4)$$

The second order connectivity index ${}^2\chi$ is the sum of all bonds $i-j-k$ (Eq. 5).

$${}^2\chi = \sum_{\text{bonds}} 1/(\delta_i \delta_j \delta_k)^{1/2} \quad \dots (5)$$

The third order connectivity index ${}^3\chi$ is the sum of all bonds $i-j-k-l$ (Eq. 6).

$${}^3\chi = \sum_{\text{bonds}} 1/(\delta_i \delta_j \delta_k \delta_l)^{1/2} \quad \dots (6)$$

The total structure index, χ_t , can be calculated with all the carbon valencies of the compound (Eq. 7).

$$\chi_t = 1/(\delta_1, \delta_2, \dots, \delta_m)^{1/2} \quad \dots (7)$$

Some recent studies on specific ad hoc descriptors have been shown to give a good account of several properties²⁰⁻²². Needham *et al.*¹ have reported very good correlation of the aforesaid physical properties using specific ad hoc descriptors.

Table 1—Simple correlation matrix of various topological and ad hoc descriptors used to derive principal components

P_0	P_1	P_3	f	N_c	T_m	T_3	W	${}^0\chi$	${}^1\chi$	${}^2\chi$	${}^3\chi$	χ_t	I	
P_0	1.0													
P_1	.99	1.0												
P_3	.84	.84	1.0											
f	.85	.86	.84	1.0										
N_c	.99	1.0	.84	.86	1.0									
T_m	.52	.54	.65	.87	.54	1.0								
T_3	.15	.15	.47	.45	.15	.62	1.0							
W	.94	.94	.72	.72	.94	.34	-.03	1.0						
${}^0\chi$.98	.99	.86	.92	.99	.64	.23	.91	1.0					
${}^1\chi$.98	.98	.80	.74	.98	.35	.02	.96	.94	1.0				
${}^2\chi$.78	.80	.66	.95	.80	.85	.35	.68	.86	.66	1.0			
${}^3\chi$.79	.78	.97	.76	.78	.57	.53	.66	.80	.76	.56	1.0		
χ_t	-.92	-.93	-.71	-.70	-.93	-.34	-.04	-.83	-.90	-.94	-.65	-.69	1.0	
I	.83	.83	.65	.56	.83	.21	-.02	.86	.78	.88	.48	.63	-.80	1.0

The simplest representative of ad hoc descriptors of hydrocarbons or their analogs is N_c , the number of carbon atoms. T_m is a crude measure of branching referring to the number of terminal methyl groups. Randic first suggested the use of T_3 , the number of terminal methyl groups separated by 3 bonds, as a steric parameter in analysis of chromatographic retention data²³. We have introduced P_0 and P_1 as path indices for C-H bond connection and single C-C bond connections respectively which may also be treated as ad hoc descriptors.

The PCA and MRA have been performed by using SAS software package.

Results and Discussion

The correlation among the topological and ad hoc descriptors is presented in Table 1. It is ap-

parent that many of these parameters are themselves intercorrelated. This interrelationship demands reduction of descriptors and to this effect, principal component analysis was carried out involving all the topological indices calculated for the 74 hydrocarbons. The eigenvalues and cumulative variances for 14 principal components are given in Table 2. Analysis of Table 2 reveals that the first principal component (PC_1) is associated with 75.7% of total variance of all the topological parameters and the second principal component (PC_2) is associated with 14.3% of total variance. The cumulative variance of PC_1 and PC_2 is 90.0%. Upto the 5th principal component the cumulative variance is 98.8%. These PCs have orthogonal relationship. To know the characteristics of the PCs all the 14 topological indices have been correlated with the PCs and the resulting statistical par-

Table 2—Eigenvalues, percent of variance and cumulative percent of variance derived from PCA

PCs	Eigenvalues	Variance (%)	Cummulative variance (%)
1	10.59	75.7	75.7
2	2.00	14.3	90.0
3	0.79	5.6	95.6
4	0.237	1.7	97.3
5	0.205	1.5	98.8
6	0.118	0.8	99.6
7	0.024	0.2	99.8
8	0.0157	0.1	99.9
9	0.0136	0.1	100.0
10	0.0022	0.0	100.0
11	0.000	0.0	100.0
12	0.000	0.0	100.0
13	0.000	0.0	100.0
14	0.000	0.0	100.0

Table 3—Correlation coefficients of topological and ad hoc descriptors with principal components (PCs) for alkanes

	PC1	PC2	PC3	PC4	PC5
P_0	0.98	-0.14	0.02	-0.02	-0.03
P_1	0.99	-0.14	-0.04	-0.01	-0.03
P_3	0.89	0.22	-0.30	0.19	0.11
f	0.91	0.34	-0.20	-0.04	0.02
N_c	0.99	-0.14	-0.04	-0.01	-0.03
T_m	0.63	0.69	-0.28	-0.04	0.10
T_3	0.28	0.84	0.36	0.26	-0.11
W	0.90	-0.34	-0.05	0.05	0.14
χ^0	0.99	-0.01	-0.09	-0.02	-0.01
χ^1	0.94	-0.31	0.07	-0.03	-0.03
χ^2	0.83	-0.30	-0.04	0.08	-0.02
χ^3	0.84	0.22	0.45	-0.15	0.02
χ_i	0.89	0.28	-0.03	-0.01	0.33
I	0.81	-0.41	0.17	0.20	0.18

Table 4—Statistical parameters (as defined for Eq. 8) derived from multiple regression analysis of various physical properties of alkanes

Eq. no.	Properties	P^0	a	b	c	d	R^2	F
9	BP	106.4	13.8	-6.2	3.4	-14.4	0.99	8665
10	MV	158.5	7.1	-3.4	-3.4	-2.7	0.99	2050
11	MR	38.2	2.2	-0.81	-0.53	-0.95	0.99	7317
12	HV	37.4	2.2	-1.5	0.68	-1.80	0.99	1923
13	TC	280.9	17.1	-4.8	6.2	-18.5	0.99	1568
14	ST	20.3	0.92	-0.28	0.79	-1.31	0.96	340.5

Table 5—Observed and predicted boiling points (BP), molar volumes (MV), molar retractions (MR) for alkanes

Sl. No.	Alkanes ^a	BP (°C)			MV		MR	
		Obsv.	Prdc ^b	Prdc ^c	Obsv.	Prdc. ^d	Obsv.	Prdc. ^e
1.	2	-88.63	-90.93	-103.0	-	64.0	--	8.95
2.	3	-42.07	-40.51	-51.4	-	87.3	-	15.87
3.	4	-0.5	3.04	-2.2	-	100.9	-	20.58
4.	2M3	-11.7	-13.35	-6.7	-	102.1	-	20.25
5.	5	36.1	37.41	40.2	115.2	117.9	25.27	25.65
6.	2M4	27.8	26.84	29.6	116.4	114.7	25.29	24.88
7.	22MM3	7.5	6.60	24.0	122.1	118.5	25.72	24.83
8.	6	68.7	68.61	75.8	130.7	132.8	29.91	30.21
9.	2M5	60.3	59.94	62.0	131.9	134.1	29.95	30.38
10.	3M5	63.3	65.03	61.9	129.7	130.6	29.80	30.01
11.	22MM4	49.7	52.92	57.7	132.7	130.2	29.94	29.63
12.	23MM4	58.0	58.05	64.0	130.2	128.7	29.81	29.56
13.	7	98.4	96.65	104.9	146.5	147.3	34.55	34.64
14.	2M6	90.0	89.29	94.4	147.7	148.9	34.59	34.92
15.	3M6	91.8	92.32	94.6	145.8	147.1	34.46	34.80
16.	3E5	93.5	92.71	88.0	143.5	145.8	34.28	34.46
17.	22MM5	79.2	79.92	86.0	148.7	150.4	34.62	35.06
18.	23MM5	89.8	89.16	90.5	144.1	145.0	34.32	34.57
19.	24MM5	80.5	80.86	85.0	148.9	149.0	34.62	34.68
20.	33MM5	86.1	89.32	86.6	144.5	144.1	34.33	34.44
21.	223MMM4	80.9	82.34	86.2	145.2	143.4	34.37	34.28
22.	8	125.7	122.5	128.6	162.6	161.5	39.19	38.93
23.	2M7	117.6	116.0	118.6	163.7	163.6	39.23	39.34
24.	3M7	118.9	119.1	121.0	161.8	161.6	39.10	39.20
25.	4M7	117.7	114.1	118.1	162.1	162.4	39.12	38.96
26.	3E6	118.5	118.0	119.0	160.1	161.1	38.94	39.04
27.	22MM6	106.8	108.2	115.2	164.3	165.3	39.25	39.59
28.	23MM6	115.6	114.6	116.4	160.4	160.9	38.98	39.13
29.	24MM6	109.4	111.7	115.6	163.1	162.7	39.13	39.27
30.	25MM6	109.1	109.2	112.2	164.7	163.4	39.26	39.12
31.	33MM6	112.0	113.1	115.5	160.7	160.1	39.01	38.91
32.	34MM6	117.7	118.4	114.4	158.8	157.9	38.85	38.74
33.	23ME5	115.6	113.1	111.6	158.8	161.4	38.84	39.10
34.	33ME5	118.3	118.2	110.8	157.0	156.5	38.72	38.50
35.	223MMM5	109.8	109.5	107.8	159.5	158.3	38.92	38.71
36.	224MMM5	99.2	99.2	104.2	165.1	164.6	39.26	39.20
37.	233MMM5	114.8	114.5	108.5	157.3	157.0	38.76	38.85
38.	234MMM5	113.5	111.2	117.5	158.9	158.2	38.87	38.68

(Contd)

Table 5—Observed and predicted boiling points (BP), molar volumes (MV), molar refractions (MR) for alkanes—*Contd.*

Sl. No.	Alkanes ^a	BP (°C)			MV		MR	
		Obsv.	Prdc ^b	Prdc ^c	Obsv.	Prdc ^d	Obsv.	Prdc ^e
39.	2233MMMM4	106.5	107.9	106.3	—	154.9	—	38.35
40.	9	150.8	146.9	149.1	178.7	176.1	43.84	43.29
41.	2M8	143.3	141.0	139.8	179.8	178.2	43.88	43.73
42.	3M8	144.2	143.2	142.1	177.9	176.4	43.73	43.55
43.	4M8	142.5	142.1	141.6	178.1	177.3	43.77	43.66
44.	3E7	143.0	143.7	142.0	176.4	175.6	43.64	43.38
45.	4E7	141.2	143.4	141.5	175.7	175.3	43.49	43.29
46.	22MM7	132.7	134.1	135.7	180.5	180.1	43.91	44.04
47.	23MM7	140.5	137.1	137.8	176.7	176.0	43.63	43.42
48.	24MM7	133.5	135.4	136.0	179.1	178.4	43.74	43.77
49.	25MM7	136.0	137.7	137.1	179.4	177.3	43.85	43.70
50.	26MM7	135.2	134.8	132.0	180.9	178.2	43.93	43.56
51.	33MM7	137.3	139.9	139.6	176.9	176.3	43.69	43.73
52.	34MM7	140.6	141.5	138.3	175.3	174.7	43.55	43.49
53.	35MM7	136.0	140.6	137.3	177.4	174.4	43.64	43.25
54.	44MM7	135.2	136.1	135.8	176.9	177.8	43.60	43.69
55.	23ME6	138.0	138.1	138.4	175.4	176.6	43.66	43.61
56.	24ME6	133.8	137.1	137.6	177.4	176.7	43.65	43.50
57.	33ME6	140.6	142.2	138.5	173.1	174.2	43.27	43.47
58.	34ME6	140.4	141.3	136.3	172.8	174.1	43.37	43.33
59.	2233MM6	133.6	135.1	132.6	175.9	175.8	43.62	43.68
60.	224MMM6	126.5	129.9	131.4	179.2	179.3	43.76	43.96
61.	225MMM6	124.1	127.7	130.9	181.3	181.0	43.94	44.13
62.	2331MMM6	137.7	137.6	132.7	173.8	174.1	43.43	43.53
63.	234MMM6	139.0	139.0	139.6	173.5	173.0	43.39	43.34
64.	235MMM6	131.3	133.4	138.9	177.7	176.6	43.65	43.63
65.	244MMM6	130.6	132.4	131.7	177.2	177.5	43.66	43.81
66.	334MMM6	140.5	141.0	132.0	172.1	172.2	43.34	43.41
67.	33EE5	146.2	142.5	130.0	170.2	172.7	43.11	43.10
68.	223MME5	133.8	132.0	125.6	174.5	177.4	43.46	43.76
69.	233MME5	142.0	140.8	128.4	170.1	171.9	42.95	43.32
70.	234MEM5	136.7	132.5	133.0	173.8	176.2	43.40	43.46
71.	2233(M)5	140.3	137.4	130.6	169.5	170.7	43.21	43.31
72.	2234(M)5	133.0	131.8	133.1	173.6	174.0	43.44	43.48
73.	244(M)5	122.3	118.3	123.3	178.3	182.0	43.87	44.01
74.	2334(M)5	141.6	138.5	132.5	169.9	168.8	43.20	42.87

^a 2 = ethane, 3 = propane, 4 = butane, 5 = pentane, 6 = hexane, 7 = heptane, 8 = octane, 9 = nonane, M = methyl, E = ethyl; ^b Using Eq. 9; ^c Reference 3 (Basak and Niemi) ^d Using Eq. 10; ^e Using Eq. 11.

Table 6—Observed and predicted heats of vaporisation (HV), critical temperatures (TC) and surface tensions (ST) for alkanes

Sl. No.	Alkanes ^a	HV		TC		ST	
		Obsv.	Prdc. ^f	Obsv.	Prdc. ^g	Obsv.	Prdc. ^h
1.	2	—	6.54	32.27	38.36	—	6.78
2.	3	—	14.02	96.80	98.90	—	10.13
3.	4	—	20.78	155.01	154.2	—	13.80
4.	2M3	—	17.97	134.98	132.1	—	11.79
5.	5	26.42	26.51	196.62	194.5	16.0	16.03
6.	2M4	24.59	24.25	187.80	184.7	15.0	15.29
7.	22MM3	21.78	20.46	160.60	156.1	—	12.47
8.	6	31.55	31.59	234.70	232.2	18.42	18.16
9.	2M5	29.86	30.05	224.90	221.6	17.38	17.11
10.	3M5	30.27	30.51	231.20	232.0	18.12	18.16
11.	22MM4	27.69	27.90	216.20	218.2	16.30	16.86
12.	23MM4	29.12	28.38	227.10	226.8	17.37	17.63
13.	7	36.55	36.38	267.01	265.6	20.26	19.98
14.	2M6	34.80	35.02	257.90	256.7	19.29	19.04
15.	3M6	35.08	35.52	262.40	262.6	19.79	19.66
16.	3E5	35.22	35.34	267.60	263.7	20.44	19.81
17.	22MM5	32.43	32.86	247.70	245.1	18.02	17.86
18.	23MM5	34.24	34.25	264.60	263.1	19.96	19.68
19.	24MM5	32.88	32.91	247.10	247.3	18.15	18.08
20.	33MM5	33.02	33.76	263.00	265.1	19.59	19.81
21.	223MMM4	32.04	31.84	258.30	259.2	18.76	19.17
22.	8	41.48	40.85	296.20	296.2	21.76	21.59
23.	2M7	39.68	39.69	288.00	288.2	20.60	20.71
24.	3M7	39.83	40.19	292.00	294.3	21.17	21.35
25.	4M7	39.67	39.24	290.00	285.9	21.00	20.57
26.	3E6	39.40	40.14	292.00	293.7	21.51	21.34
27.	22MM6	37.29	37.74	279.00	279.5	19.60	19.68
28.	23MM6	38.79	38.84	293.00	291.5	20.99	21.01
29.	24MM6	37.76	38.51	282.00	285.7	20.05	20.43
30.	25MM6	37.86	37.72	279.00	281.3	19.73	19.95
31.	33MM6	37.93	38.57	290.84	290.1	20.63	20.94
32.	34MM6	39.02	39.08	298.00	299.3	21.64	21.82
33.	23ME5	38.52	38.81	295.00	288.3	21.52	20.75
34.	33ME5	37.99	38.93	305.00	300.4	21.99	21.98
35.	223MMM5	36.91	37.16	294.00	289.7	20.67	20.85
36.	224MMM5	35.13	35.92	271.15	269.0	18.77	18.75
37.	233MMM5	37.22	37.59	303.00	279.2	21.56	21.65
38.	234MMM5	37.61	37.22	295.00	292.0	21.14	21.01
39.	2233MMMM4	—	34.75	270.80	296.1	—	21.18

(Contd)

Table 6—Observed and predicted heats of vaporisation (HV), critical temperatures (TC) and surface tensions (ST) for alkanes—*Contd.*

Sl. No.	Alkanes ^a	HV		TC		ST	
		Obsv.	Prdc. ^f	Obsv.	Prdc. ^g	Obsv.	Prdc. ^h
40.	9	46.44	45.28	322.00	324.5	22.92	23.01
41.	2M8	44.65	44.20	315.00	317.3	21.88	22.18
42.	3M8	44.75	44.45	318.00	322.1	22.34	22.71
43.	4M8	44.75	44.52	318.30	319.5	22.34	22.45
44.	3E7	44.81	44.62	318.00	323.0	22.81	22.84
45.	4E7	44.81	44.46	318.30	322.7	22.81	22.81
46.	22MM7	42.28	42.36	302.00	309.9	20.80	21.25
47.	23MM7	43.79	42.91	315.00	316.5	22.34	22.11
48.	24MM7	42.87	42.93	306.00	311.8	21.30	21.60
49.	25MM7	42.87	43.14	307.80	316.2	21.30	22.02
50.	26MM7	42.82	42.31	306.00	311.2	20.83	21.48
51.	33MM7	42.66	43.19	314.00	321.3	22.01	22.45
52.	34MM7	43.84	43.78	322.70	323.9	22.80	22.84
53.	35MM7	42.98	43.21	312.30	322.6	21.77	22.70
54.	44MM7	42.66	42.61	317.80	314.1	22.01	21.76
55.	23ME6	43.84	43.45	322.70	317.0	22.80	22.18
56.	24ME6	42.98	43.05	330.3	315.4	21.77	22.00
57.	33ME6	43.04	43.56	327.2	326.1	23.22	23.00
58.	34ME6	43.95	43.76	312.3	323.5	23.27	22.86
59.	223MMM6	41.91	41.73	318.1	318.1	21.86	22.05
60.	224MMM6	40.57	41.16	301.1	307.3	20.51	20.96
61.	225MMM6	40.17	40.82	296.6	302.7	20.04	20.45
62.	233MMM6	42.23	42.06	326.1	323.3	22.41	22.59
63.	234MMM6	42.93	42.58	324.2	325.0	22.80	22.87
64.	235MMM6	41.42	41.80	309.4	313.7	21.27	21.70
65.	244MMM6	40.84	41.48	309.1	312.6	21.17	21.51
66.	334MMM6	42.28	42.60	333.6	329.7	23.27	23.26
67.	33EE5	43.46	43.48	342.8	327.0	23.75	23.17
68.	223MME5	42.02	41.61	322.6	311.5	22.38	21.46
69.	233MME5	42.55	42.59	338.6	329.4	23.87	23.26
70.	234MEM5	42.93	41.81	324.2	312.0	22.80	21.62
71.	2233(M)5	41.00	40.70	334.5	330.6	23.38	23.15
72.	2234(M)5	41.00	40.41	319.6	318.1	21.98	21.99
73.	244(M)5	38.10	38.20	301.6	292.1	20.37	19.26
74.	2334(M)5	41.75	40.80	334.5	332.5	23.31	23.44

^a 2 = ethane, 3 = propane, 4 = butane, 5 = pentane, 6 = hexane, 7 = heptane, 8 = octane, 9 = nonane, M = methyl, E = ethyl; ^f Using Eq. 12; ^g Using Eq. 13; ^h Using Eq. 14.

ameters are given in Table 3. By analyzing the data of Table 3 it is clear that PC₁ and PC₂ are loaded significantly with all the topological parameters. Except T_m and T₃ all the other variables have good correlation (R²=0.83-0.99) with PC₁. The parameters T_m and T₃ correlate better with PC₂ than PC₁ indicating PC₂ to be a steric component.

All the five PCs have been correlated with six physical properties namely boiling point, molar volume, molar refraction, heat of vaporisation, critical temperature and surface tension. The results of 't' test reveals PC₄ to be a poor candidate to be incorporated in the regression analysis. Hence a four-dimensional model (Eq. 8) has been proposed to predict the physical properties (P) and the relevant statistical parameters have been calculated.

$$P = P^0 + aPC_1 + bPC_2 + cPC_3 + dPC_5 \quad \dots 8)$$

where a, b, c and d are the sensitivities of PC₁, PC₂, PC₃ and PC₅ respectively towards the physical properties.

The values of P⁰, a, b, c, d and R² are given in Table 4. All the physical properties are reasonably well reproduced by the present parameters (PCs). The predicted values of all the physical properties obtained using appropriate equations are presented in Tables 5 and 6. For comparison the predicted data of boiling point by Basak and Niemi³ are also given in Table 5.

The analysis of the Tables 5 and 6 clearly shows that the parameters generated through principal component analysis are suitable for the prediction of various physical properties. One should stress that topological indices are valuable because they can be easily obtained for known and unknown molecules.

Acknowledgement

The authors thank the Department of Science and Technology, New Delhi for funding the present work.

References

- 1 Needham D E, Wei I C & Seybold P G, *J Am chem Soc*, 110 (1986) 4186.
- 2 Balaban A T, Basak S C, Colburn T & Grunwald D G, *J chem Inf comput Sci*, 34 (1994) 1118.
- 3 Basak S C & Niemi G J, *J math Chem*, 7 (1991) 243.
- 4 Basak S C & Grunwald D G, *Math Modelling Sci Computing*, 2 (1993) 735.
- 5 Basak S C & Grunwald D G, *SAR & QSAR in Environmental Research*, 3 (1995) 265.
- 6 Basak S C & Grunwald D G, *J chem Inf comput Sci*, 35 (1995) 366.
- 7 Wiener H, *J Am chem Soc*, 69 (1947) 17, 2636.
- 8 Shannon C E, *Bell Syst Tech J*, 27 (1948) 379.
- 9 Wiener N, *Cybernetics* (Wiley, New York) 1948.
- 10 Ashby W, *An introduction to cybernetics* (Wiley, New York) 1956.
- 11 Kolmogorov A N, *Probl Peridachi Inf*, 5 (1969) 3.
- 12 Bonchev D, *Information theoretic indices for characterisation of chemical structure* (Wiley Interscience, New York) 1983.
- 13 Basak S C, Harris D K & Magnuson V R, *J pharm Sci*, 73 (1984) 429.
- 14 Kier L B, *J pharm Sci*, 69 (1980) 807.
- 15 Wiener H, *J chem Phys*, 15 (1947) 766.
- 16 Platt J R, *J phys Chem*, 56 (1952) 328.
- 17 Randic M, *J Am chem Soc*, 97 (1975) 6609.
- 18 Kier L B & Hall L H, *Molecular connectivity in chemistry and drug research* (Academic, New York) 1976.
- 19 Kier L B & Hall L H, *Molecular connectivity in chemistry and drug research* (Wiley, New York) 1986.
- 20 Seybold P G, May M A & Bagal U A, *J chem Educ*, 64 (1987) 575.
- 21 Seybold P G, May M A & Gargas M L, *Acta pharm Jugosl*, 36 (1986) 253.
- 22 Hinkckley D A, Seybold P G & Borris D P, *Spectrochim Acta Part A*, 42A (1986) 747.
- 23 Randic M, *J Chromatogr*, 161 (1978) 1.