

List Operations on Chemical Graphs. 3. Development of Vertex and Edge Models for Fitting Retention Index Data

Ch. Duvenbeck and P. Zinn*

Lehrstuhl für Analytische Chemie, Ruhr-Universität Bochum, Postfach 102148, D-4630 Bochum 1, Germany

Received June 12, 1992

A general method for fitting the gas chromatographic retention index data is presented. The method bases on vertex and edge models developed from graph theory. The resulting multilinear regression models use atomic-level topological indexes as the atomic-level molecular connectivity or the electrotopological index. By fitting a standard data set of 217 acyclic and cyclic alkanes, alkenes, alcohols, esters, ketones, and ethers the mean absolute errors of the edge models were observed in a range from 7 to 9 retention index units. The models can be used to predict retention index data of compounds that incorporate the same bond types as those of the standard data set.

INTRODUCTION

The estimation of chromatographic retention index data is a widely investigated topic in Quantitative Structure-Property Relationship (QSPR) studies. Gas chromatographic retention indexes measured on nonpolar stationary phases are an especially well-suited object of QSPR studies. In contrast to other molecular properties, interactions between eluents can widely be neglected. Other reasons are the high precision of index measurement and the large amount of data reported in literature.¹⁻⁴

Several aspects of the development of valid estimation models and the significance of model parameters are discussed by Kaliszan.⁵ Among model parameters, topological indexes are of special interest because they can be derived directly from the molecular structure without any experimental effort.

There are two types of topological indexes that can be used as model parameters. Conventionally topological indexes characterize a molecule as a whole. These indexes are most popular in QSPR studies and other applications of graph theory in chemistry.⁶ The index developed by Wiener⁷ bases on the distance matrix of a molecule. Hosoya⁸ counts the combinations of disconnected bonds, and Randić⁹ describes indexes for molecular connectivity. Indexes characterizing molecular branching are given by Balaban.¹⁰ Kier and Hall¹¹ show concepts of topological indexes for molecular shape.

In developing estimation models for QSPR studies, topological indexes are frequently used. Jurs reported about the prediction of boiling points¹² and surface tensions¹³ using a wide spectrum of topological indexes among other descriptors as model parameters. Trinajstić¹⁴ recently compared topological indexes for the estimation of boiling points. Several authors discuss the estimation of retention index data using topological model parameters.¹⁵⁻²⁰ In most cases, the developed estimation models work well for a limited class of substances. Therefore, a more general approach seemed to be a supplementary contribution to retention index data estimation.

The second type of topological indexes describes structure information at the atomic level of a molecule. One of the most interesting indexes of this type is the electrotopological state (E-state), developed by Kier and Hall.²¹ The E-state can be calculated for each atom in a molecule, including heteroatoms. The electronic interactions of atoms in a molecular field are quantified for each atom by this index. It

is to be expected that E-state-based estimation models can be applied to different substance classes.

In order to make a more general approach for the retention index estimation, incremental models are also of interest. A typical example of an incremental model was introduced by Takács.²²⁻²⁴ Takács' approach is an additive model that summarizes increments corresponding to atoms and bonds of a molecule. Gautzsch et al.²⁵ showed that the application of this method is limited due to the proliferation of bond types. Another more practical incremental model is described by Herndon²⁶ for the estimation of enthalpic properties of alkanes. According to Herndon, we have developed regression models that consider all atoms of a molecule, the so-called vertex models, and those considering all bonds, the so-called edge models. The used regression parameters are based on atomic-level topological indexes. To derive atomic-level-based descriptors and to build the descriptor matrix, we used the list operating system, described in ref 27. The calculation of topological indexes is a typical application of list operations on chemical graphs.²⁸ As an interface to the list operating system, the list structured line notation referring to ref 27 is used. Table I summarizes the compounds of the used calibration data set and their corresponding line notations. The output of the list operating system is a matrix containing topological indexes belonging to atom or bond types of a molecule. The following multilinear regression analysis was carried out with commercial statistic software products.

In the first step, we present the developed models for the retention data set of 71 acyclic and cyclic alkanes measured by Rijks¹ on squalane at 50 °C. In the second step, we extend the models in order to investigate more complex compounds such as alkenes, alcohols, ethers, ketones, and esters. The origin of the retention indexes of these compounds measured on OV1 at 60 °C is the Sadler catalogue.² All compounds and the corresponding retention indexes are listed in Table I. In this paper we will discuss the fitting of model-calculated retention indexes versus data from the literature. In a following paper we will demonstrate the predictive power of the models for the retention data estimation of terpenes and related compounds.²⁹

VERTEX MODELS FOR FITTING RETENTION INDEX DATA

Vertex models are multilinear regression models for the estimation of physicochemical molecular properties. As regression parameters, atomic-level topological indexes are

Table I. Compounds of Used Calibration Data Set and Corresponding Retention Indexes from Literature *I*[lit.] and Those Resulting from Model Estimation *I*[est.]^a

no.	compound	<i>I</i> [lit.]	<i>I</i> [est.]				list structured line notation
			<i>M</i> model	ξ model	<i>S</i> model	<i>R</i> model	
1	2,2-dimethylpropane	412.3	408.9	421.3	423.8	409.7	(C C (C) (C) C)
2	2-methylbutane	475.3	465.2	475.1	473.9	469.0	(C C (C) C C)
3	2,2-dimethylbutane	536.8	524.5	531.2	533.7	529.3	(C C (C) (C) C C)
4	2,3-dimethylbutane	567.3	551.6	556.7	562.0	554.2	(C C (C) C (C) C)
5	2-methylpentane	569.7	565.2	568.3	572.0	566.9	(C C (C) C C C)
6	3-methylpentane	584.2	567.7	574.8	573.7	573.8	(C C C (C) C C)
7	2,2-dimethylpentane	625.6	624.5	625.3	631.1	626.7	(C C (C) (C) C C C)
8	2,4-dimethylpentane	629.8	630.3	630.4	634.8	627.1	(C C (C) C C (C) C)
9	2,2,3-trimethylbutane	639.7	624.6	629.3	631.8	627.1	(C C (C) (C) C (C) C)
10	3,3-dimethylpentane	658.9	640.0	645.0	644.3	645.0	(C C C (C) (C) C C)
11	2-methylhexane	666.6	665.2	668.3	670.7	666.4	(C C (C) C C C C)
12	2,3-dimethylpentane	671.7	654.1	657.9	662.2	658.2	(C C (C) C (C) C C)
13	3-ethylpentane	686.0	670.2	681.9	673.4	674.7	(C C C (C) C C)
14	2,2,4-trimethylpentane	689.9	689.6	686.0	691.4	686.7	(C C (C) (C) C C (C) C)
15	2,2-dimethylhexane	719.4	724.5	725.3	729.2	726.2	(C C (C) (C) C C C C)
16	2,5-dimethylhexane	728.4	730.3	736.5	733.5	727.3	(C C (C) C C C (C) C)
17	2,4-dimethylhexane	731.9	732.8	730.1	734.5	731.9	(C C (C) C C (C) C C)
18	2,2,3-trimethylpentane	737.1	727.2	731.1	732.7	731.1	(C C (C) (C) C (C) C C)
19	3,3-dimethylhexane	743.5	740.0	739.1	742.2	742.5	(C C C (C) (C) C C C)
20	2,3,4-trimethylpentane	752.4	740.6	741.0	747.9	742.4	(C C (C) C (C) C (C) C)
21	2,3,3-trimethylpentane	759.4	740.1	742.9	743.6	742.7	(C C (C) C (C) (C) C C)
22	2-methylheptane	764.9	765.2	768.3	769.6	765.8	(C C (C) C C C C C)
23	4-methylheptane	767.2	767.7	761.1	770.5	769.6	(C C C C (C) C C C)
24	3,4-dimethylhexane	770.6	756.7	759.2	762.7	762.1	(C C C (C) C (C) C C)
25	3-methylheptane	772.3	767.7	767.9	770.9	771.1	(C C C (C) C C C C)
26	2,2,4,4-tetramethylpentane	772.7	748.9	742.1	744.7	746.1	(C C (C) (C) C C (C) (C) C)
27	3-methyl-3-ethylpentane	774.0	755.5	762.7	755.6	757.0	(C C C (C) (C) C C C)
28	2,2,5-trimethylhexane	776.3	789.6	793.5	790.4	787.1	(C C (C) (C) C C C (C) C)
29	2,2,4-trimethylhexane	789.1	792.1	785.7	790.9	791.4	(C C (C) (C) C C (C) C C)
30	2,4,4-trimethylhexane	807.7	805.1	799.8	802.2	802.4	(C C (C) C C (C) (C) C C)
31	2,3,5-trimethylhexane	812.0	819.3	813.3	821.2	816.3	(C C (C) C (C) C C (C) C)
32	2,2-dimethylheptane	815.4	824.5	825.3	827.6	825.6	(C C (C) (C) C C C C C)
33	2,2,3,4-tetramethylpentane	819.6	813.6	814.0	814.5	815.5	(C C (C) (C) C (C) C (C) C)
34	2,2,3-trimethylhexane	821.6	827.2	824.3	830.3	829.1	(C C (C) (C) C (C) C C C)
35	2,2-dimethyl-3-ethylpentane	822.2	829.7	840.5	833.6	831.4	(C C (C) (C) C (C) C C C)
36	3,3-dimethylheptane	835.8	840.0	839.1	840.5	841.9	(C C C (C) C C C C C)
37	2,4-dimethyl-3-ethylpentane	836.5	843.1	851.3	848.6	841.8	(C C (C) C (C) C (C) C C)
38	2,3,4-trimethylhexane	849.1	843.1	842.3	848.6	846.4	(C C (C) C (C) C (C) C C)
39	2,3,3,4-tetramethylpentane	858.0	840.3	839.8	838.5	840.7	(C C (C) C (C) (C) C (C) C)
40	3-methyloctane	870.2	867.7	867.9	870.0	870.6	(C C C (C) C C C C C)
41	3,3-diethylpentane	877.2	871.0	884.4	867.6	865.2	(C C C (C C) (C C) C C)
42	<i>n</i> -propane	300	300.0	300.0	308.9	304.1	(C C C)
43	<i>n</i> -butane	400	400.0	400.0	408.2	406.5	(C C C C)
44	<i>n</i> -pentane	500	500.0	500.0	507.3	506.0	(C C C C C)
45	<i>n</i> -hexane	600	600.0	600.0	606.5	605.5	(C C C C C C)
46	<i>n</i> -heptane	700	700.0	700.0	705.9	704.9	(C C C C C C C)
47	<i>n</i> -octane	800	800.0	800.0	805.4	804.4	(C C C C C C C C)
48	<i>n</i> -nonane	900	900.0	900.0	905.0	903.8	(C C C C C C C C C)
49	<i>n</i> -decane	1000	1000.0	1000.0	1004.6	1003.3	(C C C C C C C C C C)
50	ethylcyclopropane	510.2	521.6	528.9	520.8	523.0	(C C C 1 C C 1)
51	cyclopentane	565.7	551.4	560.6	553.3	556.2	(C 1 C C C C 1)
52	ethylcyclobutane	621.1	621.6	628.9	620.0	622.4	(C C C 1 C C C 1)
53	methylcyclopentane	627.9	619.1	621.7	619.6	620.9	(C C 1 C C C C 1)
54	cyclohexane	662.7	651.4	660.6	653.1	655.7	(C 1 C C C C C 1)
55	1,1,2-trimethylcyclopentane	763.2	794.0	792.5	801.6	794.0	(C C 1 (C) C (C) C C C 1)
56	1,1,3-trimethylcyclopentane	723.6	759.0	747.3	756.6	755.0	(C C 1 (C) C C (C) C C 1)
57	methylcyclohexane	725.8	719.1	721.7	718.6	720.4	(C C 1 C C C C C 1)
58	ethylcyclopentane	733.8	721.6	728.9	719.2	721.9	(C C C 1 C C C C 1)
59	1,1-dimethylcyclohexane	787.0	791.3	793.7	791.4	790.7	(C C 1 (C) C C C C C 1)
60	isopropylcyclopentane	812.1	808.1	813.6	810.6	805.5	(C C (C) C 1 C C C C 1)
61	<i>n</i> -propylcyclopentane	830.3	821.6	822.0	818.4	819.8	(C 1 (C C C) C C C C 1)
62	ethylcyclohexane	834.3	821.6	828.9	818.7	821.4	(C C C 1 C C C C C 1)
63	1,1,3-trimethylcyclohexane	840.4	859.0	847.3	853.5	854.4	(C C 1 (C) C C (C) C C C 1)
64	1- <i>cis</i> -3-dimethylcyclopentane	682.7	686.7	676.7	684.6	684.9	(C C 1 C C (C) C C 1)
65	1- <i>trans</i> -3-dimethylcyclopentane	686.8	686.7	676.7	684.6	684.9	(C C 1 C C (C) C C 1)
66	1- <i>cis</i> -2-dimethylcyclopentane	720.9	708.1	706.1	713.0	709.8	(C C 1 C (C) C C C 1)
67	1- <i>trans</i> -2-dimethylcyclopentane	689.2	708.1	706.1	713.0	709.8	(C C 1 C (C) C C C 1)
68	1- <i>trans</i> -2-dimethylcyclohexane	801.8	808.1	806.1	811.2	809.3	(C C 1 C (C) C C C C 1)
69	1- <i>cis</i> -2-dimethylcyclohexane	829.3	808.1	806.1	811.2	809.3	(C C 1 C (C) C C C C 1)
70	1- <i>trans</i> -3-dimethylcyclohexane	805.6	786.7	776.7	782.8	784.4	(C C 1 C C (C) C C C 1)
71	1- <i>cis</i> -3-dimethylcyclohexane	785.0	786.7	776.7	782.8	784.4	(C C 1 C C (C) C C C 1)
72	2-methyl-2-butanol	626.2	613.0	621.8	626.3	608.6	(C C (C) (O) C C)
73	1-butanol	646.5	653.6	654.6	657.7	653.9	(O C C C C)
74	3-methyl-2-butanol	666.0	656.3	665.3	656.9	647.8	(C C (C) C (C) O)

Table I (Continued)

no.	compound	I[lit.]	I[est.]				list structured line notation
			M model	$^1\xi$ model	S model	R model	
75	2-pentanol	682.7	669.8	680.8	682.9	659.1	(C C (O) C C C)
76	2-methyl-2-pentanol	717.6	713.0	715.8	721.2	706.1	(C C (C) (O) C C C)
77	3-methyl-1-butanol	719.3	718.7	722.8	720.4	714.8	(O C C C (C) C)
78	4-methyl-2-pentanol	744.1	735.0	743.0	743.1	719.3	(C C (O) C C (C) C)
79	1-pentanol	750.4	753.6	755.6	755.1	753.4	(O C C C C C)
80	2-methyl-3-pentanol	758.0	758.8	757.7	760.6	766.8	(C C C (O) C (C) C)
81	2,4-dimethyl-2-pentanol	775.9	778.1	776.6	779.0	766.0	(C C (C) (O) C C (C) C)
82	3,3-dimethyl-1-butanol	778.8	778.0	779.9	773.0	774.6	(C C (C) (C) C C O)
83	3-hexanol	780.4	772.4	771.6	786.4	778.0	(C C C (O) C C C)
84	2-methyl-2-hexanol	817.3	813.0	815.8	817.2	805.5	(C C (C) (O) C C C C)
85	2-methyl-1-pentanol	818.4	821.2	807.9	821.0	827.9	(O C C (C) C C C)
86	2,4-dimethyl-3-pentanol	821.2	845.3	836.8	828.6	852.5	(C C (C) C (O) C (C) C)
87	4-methyl-1-pentanol	821.2	818.7	822.8	817.7	814.3	(O C C C C (C) C)
88	2,3-dimethyl-3-pentanol	823.7	828.7	825.4	801.3	833.9	(C C C (C) (O) C (C) C)
89	2-ethyl-1-butanol	825.9	823.8	821.9	824.7	830.9	(O C C (C) C C)
90	3-methyl-1-pentanol	828.8	821.2	822.5	820.1	819.6	(O C C C (C) C C)
91	5-methyl-3-hexanol	838.2	837.5	833.8	846.3	839.2	(C C C (O) C C (C) C)
92	3-ethyl-3-pentanol	843.1	844.0	842.0	850.8	867.8	(C C C (O) (C) C C C)
93	1-hexanol	853.0	853.6	854.6	853.0	852.8	(O C C C C C C)
94	4-heptanol	875.4	872.4	864.7	882.2	876.9	(C C C C (O) C C C)
95	2,2,4-trimethyl-3-pentanol	881.5	918.3	907.4	863.8	922.2	(C C (C) C (O) C (C) (C) C)
96	3,5-dimethyl-3-hexanol	883.1	893.6	884.8	891.1	897.5	(C C C C C (O) (C) C C)
97	2-methyl-2-heptanol	916.4	913.0	915.8	914.0	905.0	(C C C C C C (C) (C) O)
98	6-methyl-2-heptanol	951.1	935.0	949.1	938.0	918.9	(C C (C) C C C C (O) C)
99	4-ethyl-3-hexanol	953.3	963.9	967.7	961.7	971.0	(C C C (C) C (O) C C C)
100	4-octanol	975.5	972.4	964.7	978.9	976.3	(C C C C C (O) C C C)
101	3-octanol	982.0	972.4	971.6	980.5	977.9	(C C C C C C (O) C C)
102	3,6-dimethyl-3-heptanol	986.6	993.6	992.3	989.0	997.9	(C C (C) C C C (C) (O) C C)
103	3-methyl-2-butanone	640.9	630.9	631.0	637.8	628.2	(C C (C) C (=O) C)
104	2-pentanone	666.3	666.8	667.8	672.6	661.1	(C C (=O) C C C)
105	3-pentanone	676.4	659.4	662.1	671.3	662.9	(C C C (=O) C C)
106	3,3-dimethyl-2-butanone	693.1	685.6	685.5	678.8	681.1	(C C (C) (C) C (=O) C)
107	4-methyl-2-pentanone	721.2	732.0	731.8	731.5	722.7	(C C (=O) C C (C) C)
108	2-methyl-3-pentanone	733.0	723.5	724.2	729.0	729.5	(C C C (=O) C (C) C)
109	4-methyl-3-pentanone	733.0	723.5	724.2	729.0	729.5	(C C C (=O) C (C) C)
110	3-methyl-2-pentanone	734.8	733.4	736.1	736.4	734.8	(C C C (C) C (=O) C)
111	3-hexanone	764.8	759.4	759.3	764.7	764.0	(C C C (=O) C C C)
112	2-hexanone	767.9	766.8	767.8	767.5	760.5	(C C (=O) C C C C)
113	2,4-dimethyl-3-pentanone	779.0	787.6	786.8	784.5	795.4	(C C (C) C (=O) C (C) C)
114	5-methyl-3-hexanone	816.7	824.6	823.3	823.0	825.6	(C C C (=O) C C (C) C)
115	2-methyl-3-hexanone	820.0	823.5	821.5	821.8	830.6	(C C (C) C (=O) C C C)
116	5-methyl-2-hexanone	836.5	832.0	836.0	827.5	821.4	(C C (=O) C C C (C) C)
117	4-heptanone	853.4	859.4	856.6	858.3	865.1	(C C C (=O) C C C C)
118	3-heptanone	865.8	859.4	859.3	859.7	863.4	(C C C (=O) C C C C)
119	2-heptanone	868.7	866.8	867.8	863.6	860.0	(C C (=O) C C C C C)
120	2,2,4,4-tetramethyl-3-pentanone	900.0	897.0	897.9	857.4	897.0	(C C (C) C (=O) C (C) (C) C)
121	2,6-dimethyl-4-heptanone	954.7	989.7	984.5	973.1	988.3	(C C (C) C C (=O) C C (C) C)
122	2,2-dimethyl-3-heptanone	964.7	978.2	976.8	958.3	981.6	(C C C C C (=O) C (C) (C) C)
123	3-octanone	966.0	959.4	959.3	955.9	962.9	(C C C (=O) C C C C C)
124	2-octanone	968.8	966.8	967.8	960.6	959.4	(C C (=O) C C C C C C)
125	acetic acid, ethyl ester	600.0	591.4	596.6	614.6	579.3	(C C (=O) O C C)
126	propionic acid, methyl ester	615.2	612.5	612.2	622.4	615.1	(C C C (=O) O C)
127	isobutyric acid, methyl ester	671.0	676.6	676.4	675.9	684.5	(C C (C) C (=O) O C)
128	propionic acid, ethyl ester	694.2	684.0	692.8	701.8	685.4	(C C C (=O) O C C)
129	acetic acid, propyl ester	696.3	691.4	687.0	703.5	683.1	(C C (=O) O C C C)
130	butyric acid, methyl ester	705.6	712.5	709.4	711.4	716.2	(C C C (=O) O C)
131	isobutyric acid, ethyl ester	744.6	748.2	757.1	754.3	754.7	(C C (C) C (=O) O C C)
132	acetic acid, sec-butyl ester	743.8	726.3	732.1	739.7	726.4	(C C (=O) O C (C) C C)
133	acetic acid, isobutyl ester	757.7	756.6	751.0	760.0	747.6	(C C (=O) O C C (C) C)
134	isopentanoic acid, methyl ester	761.3	777.6	773.4	767.4	777.8	(C C (C) C C (=O) O C)
135	butyric acid, ethyl ester	784.0	784.0	790.1	790.9	786.5	(C C C C (=O) O C C)
136	propionic acid, propyl ester	792.6	784.0	783.3	790.8	789.2	(C C C (=O) O C C C)
137	acetic acid, butyl ester	796.2	791.4	787.0	795.6	782.6	(C C (=O) O C C C C)
138	butyric acid, isopropyl ester	827.6	816.3	823.4	819.4	822.6	(C C C C (=O) O C (C) C)
139	isopentanoic acid, ethyl ester	838.4	849.2	854.0	846.3	848.1	(C C (C) C C (=O) O C C)
140	propionic acid, isobutyl ester	852.8	849.2	847.3	846.8	853.7	(C C C (=O) O C C (C) C)
141	butyric acid, propyl ester	881.5	884.0	880.5	880.0	890.3	(C C C C (=O) O C C C)
142	acetic acid, 1,3-dimethylbutyl ester	885.1	891.4	887.4	886.8	884.6	(C C (=O) O C (C) C C (C) C)
143	propionic acid, butyl ester	891.4	884.0	883.3	882.8	888.6	(C C C (=O) O C C C C)
144	acetic acid, pentyl ester	896.4	891.4	887.0	889.6	882.0	(C C (=O) O C C C C C)
145	isobutyric acid, isobutyl ester	900.0	913.3	911.5	897.9	923.0	(C C (C) C (=O) O C C (C) C)
146	hexanoic acid, methyl ester	907.0	912.5	909.4	897.4	915.1	(C C C C C C (=O) O C)
147	butyric acid, isobutyl ester	940.3	949.2	944.5	935.7	954.8	(C C C C (=O) O C C (C) C)
148	acetic acid, 2-ethylbutyl ester	957.0	961.7	957.9	951.8	953.4	(C C (=O) O C C (C) C C)
149	butyric acid, butyl ester	979.4	984.0	980.5	972.1	989.7	(C C C C (=O) O C C C C)

Table I (Continued)

no.	compound	<i>I</i> [lit.]	<i>I</i> [est.]				list structured line notation
			<i>M</i> model	¹ ξ model	<i>S</i> model	<i>R</i> model	
150	hexanoic acid, ethyl ester	982.9	984.0	990.1	977.1	985.4	(C C C C C C (=O) O C C)
151	propionic acid, pentyl ester	990.5	984.0	983.3	976.9	988.1	(C C C (=O) O C C C C C)
152	acetic acid, hexyl ester	996.5	991.4	987.0	985.1	981.5	(C C (=O) O C C C C C C)
153	propyl ether	681.6	691.4	684.8	689.2	697.6	(C C C O C C C)
154	butylethyl ether	688.5	691.4	694.4	690.0	693.3	(C C C C O C C)
155	butyl ether	876.3	891.4	884.8	883.5	896.5	(C C C C O C C C C)
156	1-octene	781.2	786.0	783.4	782.3	785.3	(C = C C C C C C C)
157	2-methyl-3-ethyl-2-pentene	778.4	794.8	802.2	794.5	778.6	(C C (C) = C (C C) C C)
158	2,3,4-trimethyl-2-pentene	765.9	766.3	764.9	766.6	759.4	(C C (C) = C (C) C (C) C)
159	2,5-dimethyl-2-hexene	749.9	760.6	770.1	756.4	757.3	(C C (C) = C C C (C) C)
160	2,3-dimethyl-1-hexene	739.3	750.5	744.8	748.7	750.8	(C = C (C) C (C) C C C)
161	2-methyl-3-ethyl-1-pentene	735.0	753.0	764.3	751.3	755.6	(C = C (C) C (C) C C C)
162	2,2-dimethyl- <i>cis</i> -3-hexene	716.8	707.3	702.6	708.5	707.5	(C C (C) (C) C = C C C)
163	2,2-dimethyl- <i>trans</i> -3-hexene	692.8	707.3	702.6	708.5	707.5	(C C (C) (C) C = C C C)
164	2,4,4-trimethyl-2-pentene	715.4	717.0	723.8	711.9	713.7	(C C (C) = C C (C) (C) C)
165	2,4,4-trimethyl-1-pentene	704.3	708.3	700.9	701.6	709.8	(C = C (C) C C (C) (C) C)
166	2,3-dimethyl-2-pentene	703.4	702.2	701.6	701.9	705.8	(C C (C) = C (C) C C)
167	3-ethyl-2-pentene	697.2	696.3	695.2	699.4	690.2	(C C = C (C C) C C)
168	2-methyl-2-hexene	691.2	695.4	700.5	692.8	696.5	(C C (C) = C C C C)
169	<i>cis</i> -3-heptene	690.4	685.8	683.7	686.3	686.1	(C C C = C C C C)
170	<i>trans</i> -3-heptene	687.5	685.8	683.7	686.3	686.1	(C C C = C C C C)
171	3-methyl- <i>trans</i> -3-hexene	691.2	687.9	695.2	687.7	686.5	(C C C (C) = C C C)
172	3-methyl- <i>cis</i> -3-hexene	684.6	687.9	695.2	687.7	686.5	(C C C (C) = C C C)
173	1-heptene	681.8	686.0	683.4	683.7	685.8	(C = C C C C C C)
174	2-ethyl-1-pentene	681.8	676.5	678.7	676.1	674.3	(C = C (C C) C C C)
175	2-methyl-1-hexene	678.1	683.9	680.3	682.2	687.6	(C = C (C) C C C C)
176	3,4-dimethyl- <i>trans</i> -2-pentene	678.3	667.8	660.9	671.5	664.7	(C C = C (C) C (C) C)
177	3,4-dimethyl- <i>cis</i> -2-pentene	670.6	667.8	660.9	671.5	664.7	(C C = C (C) C (C) C)
178	3-methyl-2-ethyl-1-butene	659.1	640.6	646.7	643.0	630.8	(C = C (C C) C (C) C)
179	4-methyl-1-hexene	657.9	653.6	652.7	651.5	652.5	(C = C C C (C) C C)
180	4-methyl- <i>trans</i> -2-hexene	656.7	660.5	662.9	657.8	659.5	(C C = C C (C) C C)
181	4-methyl- <i>cis</i> -2-hexene	654.9	660.5	662.9	657.8	659.5	(C C = C C (C) C C)
182	2,3-dimethyl-1-pentene	650.4	650.5	651.7	651.8	652.8	(C = C (C) C (C) C C)
183	5-methyl-1-hexene	650.0	651.1	651.7	649.5	647.3	(C = C C C C (C) C)
184	3-ethyl-1-pentene	646.9	647.5	661.4	644.4	647.1	(C = C C (C C) C C)
185	3-methyl-1-hexene	644.7	645.0	638.5	643.4	646.0	(C = C C (C) C C C)
186	2,4-dimethyl-2-pentene	640.6	651.9	655.2	649.6	649.3	(C C (C) = C C (C) C)
187	2,4-dimethyl-1-pentene	637.7	649.0	644.3	646.9	649.8	(C = C (C) C C (C) C)
188	3,4-dimethyl-1-pentene	636.9	631.5	632.4	631.8	631.3	(C = C C (C) C (C) C)
189	4,4-dimethyl- <i>cis</i> -2-pentene	635.5	623.0	622.0	623.3	622.2	(C C = C C (C) (C) C)
190	4,4-dimethyl- <i>trans</i> -2-pentene	614.7	623.0	622.0	623.3	622.2	(C C = C C (C) (C) C)
191	3,3-dimethyl-1-pentene	626.1	623.0	627.5	620.2	624.7	(C = C C (C) (C) C C)
192	2,3-dimethyl-2-butene	625.1	609.6	606.7	608.5	629.3	(C C (C) = C (C) C)
193	4,4-dimethyl-1-pentene	604.6	610.4	610.8	610.1	607.2	(C = C C C (C) (C) C)
194	<i>cis</i> -2-hexene	603.6	601.5	603.0	599.8	600.7	(C C = C C C C)
195	<i>trans</i> -2-hexene	596.9	601.5	603.0	599.8	600.7	(C C = C C C C)
196	3-methyl- <i>trans</i> -2-pentene	612.7	603.7	599.4	605.4	607.5	(C C = C (C) C C)
197	3-methyl- <i>cis</i> -2-pentene	602.8	603.7	599.4	605.4	607.5	(C C = C (C) C C)
198	2-methyl-2-pentene	597.8	595.4	597.6	595.9	598.9	(C C (C) = C C C)
199	<i>cis</i> -3-hexene	592.6	585.8	580.7	588.9	588.5	(C C C = C C C)
200	<i>trans</i> -3-hexene	592.1	585.8	580.7	588.9	588.5	(C C C = C C C)
201	2-ethyl-1-butene	592.0	576.5	581.5	579.1	573.2	(C = C (C C) C C)
202	1-hexene	582.3	586.0	583.4	585.5	586.4	(C = C C C C C)
203	2-methyl-1-pentene	580.1	583.9	580.3	584.4	588.1	(C = C (C) C C C)
204	2,3-dimethyl-1-butene	558.8	548.0	546.5	552.3	546.3	(C = C (C) C (C) C)
205	4-methyl- <i>trans</i> -2-pentene	561.9	558.0	554.4	558.8	556.7	(C C = C C (C) C)
206	4-methyl- <i>cis</i> -2-pentene	556.2	558.0	554.4	558.8	556.7	(C C = C C (C) C)
207	3-methyl-1-pentene	551.4	545.0	545.3	547.1	548.0	(C = C C (C) C C)
208	4-methyl-1-pentene	549.4	551.1	552.0	552.7	547.7	(C = C C C (C) C)
209	2-methyl-2-butene	514.3	511.1	509.2	510.6	521.2	(C C (C) = C C)
210	3,3-dimethyl-1-butene	506.8	507.5	507.8	512.5	510.7	(C = C C (C) (C) C)
211	<i>trans</i> -2-pentene	500.0	501.5	500.1	502.4	503.2	(C C = C C C)
212	<i>cis</i> -2-pentene	504.9	501.5	500.1	502.4	503.2	(C C = C C C)
213	2-methyl-1-butene	488.0	483.9	483.0	487.3	487.0	(C = C (C) C C)
214	1-pentene	481.8	486.0	483.4	488.2	486.9	(C = C C C C)
215	3-methyl-1-butene	450.3	442.5	436.8	449.8	445.1	(C = C C (C) C)
216	<i>trans</i> -2-butene	406.6	417.2	419.5	415.4	417.8	(C C = C C)
217	<i>cis</i> -2-butene	416.9	417.2	419.5	415.4	417.8	(C C = C C)

^a Additionally, the used line notations²⁷ are shown.

calculated and summarized for identical atom types (vertices). The resulting model parameters are based on four different atomic-level indexes:

the number of identical vertices *N*
the atomic-level χ indexes $^0\xi$ and $^1\xi$ ⁹
the electrotopological state *S*

$$I_k = A_{0k} + \sum_{j=1}^4 A_{jk} \alpha_j^k$$

$$\alpha_j^1 = N_j$$

$$\alpha_j^2 = \sum_i {}^0\xi_{ij}$$

$$\alpha_j^3 = \sum_i {}^1\xi_{ij}$$

$$\alpha_j^4 = \sum_i S_{ij}$$

Figure 1. Vertex equations I_k modeling retention indexes of alkane compounds. A_{jk} 's are the regression coefficients; α_j^k 's are the regression parameters based on the atomic level topological indexes ${}^0\xi$, ${}^1\xi$, and S . The index j is ranging from 1 to 4 corresponding to the atom types CH_3 - to $-\text{C}\llcorner$.

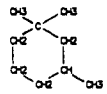
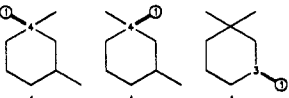
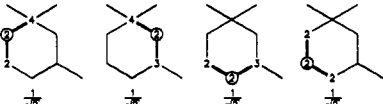


1,1,3-Trimethylcyclohexane				
				
CH3		$N = 3$		
$\sum {}^0\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		3.000	
$\sum {}^1\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		1.577	
$\sum S$	2.362		7.160	
CH2		$N = 4$		
$\sum {}^0\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		2.828	
$\sum {}^1\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		3.524	
$\sum S$	1.442		5.795	
CH		$N = 1$		
$\sum {}^0\xi$	$\frac{7}{12}$		0.577	
$\sum {}^1\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		1.394	
$\sum S$	0.890		0.980	
C		$N = 1$		
$\sum {}^0\xi$	$\frac{7}{12}$		0.500	
$\sum {}^1\xi$	$\frac{7}{12} + \frac{7}{12} + \frac{7}{12} + \frac{7}{12}$		1.707	
$\sum S$	0.890		0.650	

Figure 2. Example of the calculation of the regression parameters of the vertex models in Figure 1 for 1,1,3-trimethylcyclohexane. In the first step, the atomic-level indexes ${}^0\xi$, ${}^1\xi$, and S are calculated for each atom. In the second step, the indexes corresponding to the same atom types are summed up to form regression parameters.

Figure 1 shows the vertex equations, modeling the retention indexes of alkane compounds. The coefficients A are obtained by regression analysis. As an example for the calculation of the regression parameters, α_j^k corresponding to the molecular structure 1,1,3-trimethylcyclohexane in Figure 2 is chosen. In this example, vertices of degree 1 are appearing three times, of degree 2 four times, and once for degrees 3 and 4. The atomic-level indexes are calculated for each atom and are summed up according to the vertex degree.

The results of fitting acyclic and cyclic alkanes with the four different vertex models are shown in Figure 3. The fitted

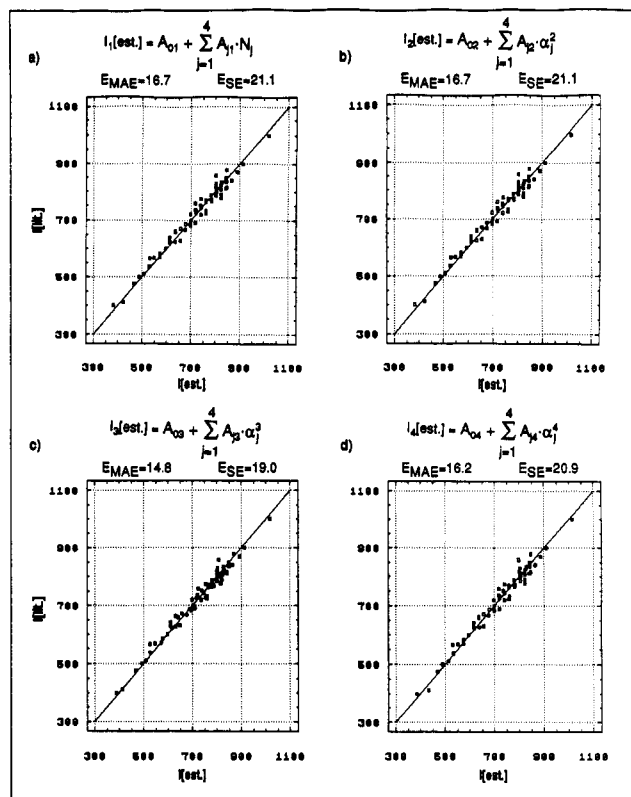


Figure 3. Comparison of fitted retention indexes $I[\text{est.}]$ of 71 acyclic and cyclic alkanes using vertex models with data from literature $I[\text{lit.}]$. E_{MAE} = mean absolute error; E_{SE} = standard error.

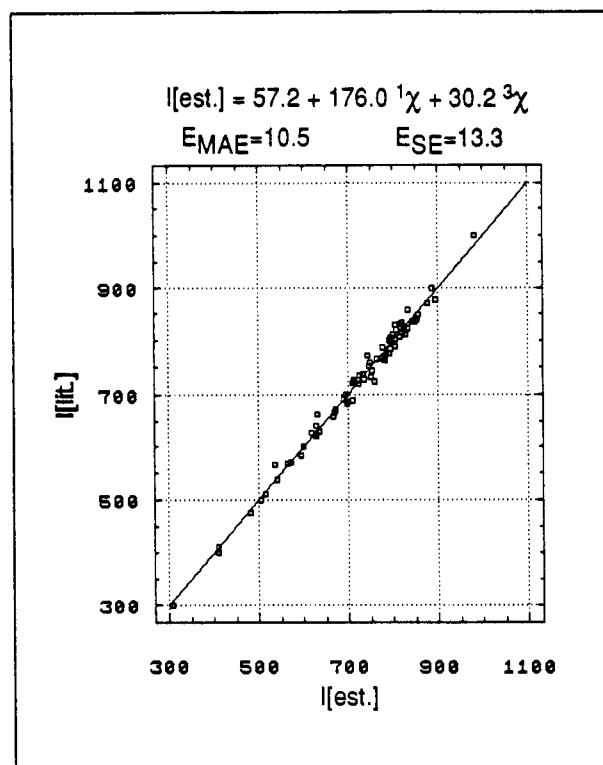


Figure 4. Comparison of fitted retention indexes $I[\text{est.}]$ of 71 acyclic and cyclic alkanes using the reference model with data from literature $I[\text{lit.}]$. E_{MAE} = mean absolute error; E_{SE} = standard error.

retention indexes $I[\text{est.}]$ are plotted vs literature data $I[\text{lit.}]$. The results are very similar with respect to the mean absolute error E_{MAE} and the standard error of estimation E_{SE} . The resulting coefficients are always highly significant with the exception of the A_0 coefficients that must be neglected. With

Table II. Bond Types, Corresponding Regression Coefficients, and Regression Parameters Used in Edge Equations Modeling Retention Indexes of Alkanes

bond	j	coeff	parameter	note
CH ₃ -CH ₃				ethane only
CH ₃ -CH ₂ -	1	B _{1k}	β ₁ ^k	
CH ₃ -CH<	2	B _{2k}	β ₂ ^k	
CH ₃ -C≡	3	B _{3k}	β ₃ ^k	
-CH ₂ -CH ₂ -	4	B _{4k}	β ₄ ^k	
-CH ₂ -CH<	5	B _{5k}	β ₅ ^k	
-CH ₂ -C≡	6	B _{6k}	β ₆ ^k	
>CH-CH<	7	B _{7k}	β ₇ ^k	
>CH-C≡	8	B _{8k}	β ₈ ^k	
>C-C<				not present in data set

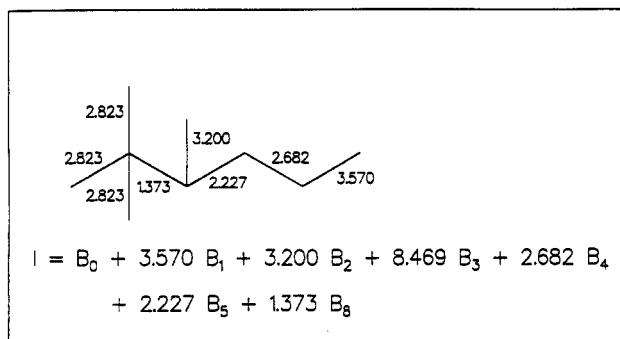
$$I_k = B_{0k} + \sum_{j=1}^8 B_{jk} \beta_j^k$$

$$\beta_1^k = M_i \quad \text{M-model}$$

$$\beta_2^k = \sum_i ({}^1\xi^i + {}^1\xi^i)_{ij} \quad {}^1\xi\text{-model}$$

$$\beta_3^k = \sum_i (S^i + S^i)_{ij} \quad \text{S-model}$$

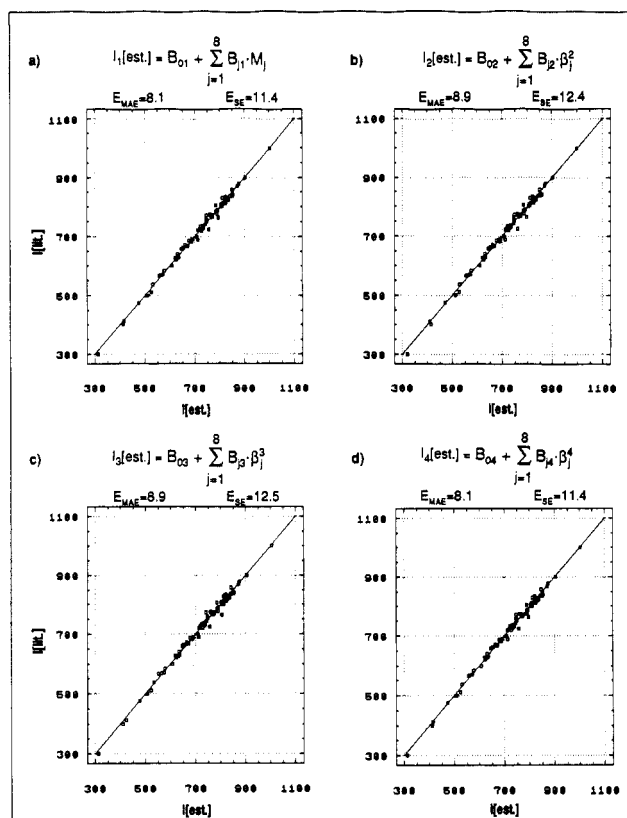
$$\beta_4^k = \sum_i (R^i + R^i)_{ij} \quad \text{R-model}$$

Figure 5. Edge models I_k for the calculation of retention indexes of alkane compounds. B_{jk} 's are the regression coefficients; β_j^k 's are the regression parameters based on the atomic-level indexes ${}^1\xi$, S , and R . The superscripts l and r mean 'left' and 'right' atoms of a bond. Bond types j correspond to Table II.**Figure 6.** Edge equation of 2,2,3-trimethylhexane for the S model. The bond indexes are assigned to the edges of the compound's skeleton. They are built by summarizing the electrotopological indexes of both atoms of the bond. The regression parameters in the edge equation are the sums of the bond indexes corresponding to the same bond types. The bond types $\text{-CH}_2\text{-C}\equiv$ ($j = 6$) and >CH-CH< ($j = 7$) are of no relevance for this compound.

respect to the reference model in Figure 4, no improvement in fitting the index data can be observed. The reference model is based on the molecular connectivity indexes ${}^1\chi$ and ${}^3\chi$ and fits the same data set as used for the vertex models. The chosen reference model gave the best results within a group of several models using topological indexes on molecular level.

EDGE MODELS FOR FITTING RETENTION INDEX DATA

To improve the quality of fitting, the vertex models were extended in order to form edge models. The building of edge models is analogous to vertex models considering molecular bonds (edges) instead of atoms. For each bond, a bond index is calculated and summed up to a regression parameter. Table

**Figure 7.** Comparison of fitted retention indexes $I[\text{est.}]$ of 71 acyclic and cyclic alkanes using edge models including a regression constant B_{0k} with data from literature $I[\text{lit.}]$. E_{MAE} = mean absolute error; E_{SE} = standard error.

II gives an overview of the used bond types, corresponding parameters, and coefficients. With respect to the used data set, the number of regression coefficients is twice the number in the case of the vertex models. The equations for edge models fitting alkane data and the formulation of the associated regression parameters are summarized in Figure 5. The edge models differ in the types of the regression parameters β_j^k from the vertex models. Because of the same results which were obtained by using ${}^0\xi$ and N indexes in the case of the vertex models, the ${}^0\xi$ model was discarded and substituted by a model considering the R index. R is a reduced form of the electrotopological index S , considering only bonded contributions as described by Hall et al.²¹

An example of an edge model equation based on the S index for 2,2,3-trimethylhexane is given in Figure 6. The topological indexes on bond level are assigned to the edges of the compound. The regression parameters are derived by summing up bond indexes for the same bond types. The bond types $\text{-CH}_2\text{-C}\equiv$ ($j = 6$) and >CH-CH< ($j = 7$) are of no relevance for this compound.

FITTING RETENTION INDEXES OF ACYCLIC AND CYCLIC ALKANES

The described edge models were used to fit retention indexes of the standard data set of 71 acyclic and cyclic alkanes. The results are shown in Figure 7, where the fitted values are plotted versus the literature data. The mean absolute errors E_{MAE} and the standard errors E_{SE} of the four edge models are smaller in comparison with the reference and vertex models. Great similarity is to be mentioned between the edge models. A statistical analysis of the resulting regression coefficients shows that they are all significantly different from '0', even the B_0 coefficient.

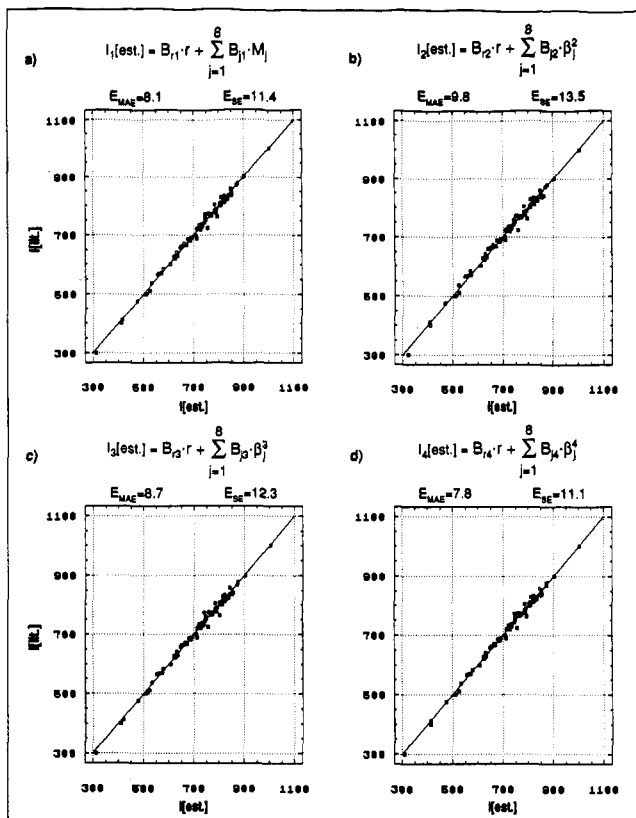


Figure 8. Comparison of fitted retention indexes $I[\text{est.}]$ of 71 acyclic and cyclic alkanes using edge models including a ring constant B_r with data from literature $I[\text{lit.}]$. E_{MAE} = mean absolute error; E_{SE} = standard error.

If we take a closer look at the used compounds, only the 49 acyclic alkanes of the data set were fitted. In this case the regression constant B_0 is no longer of significance. It is obvious that B_0 for the fitting of the total data set is mainly affected by cyclic compounds. Therefore the edge models were altered by introducing a so-called ring parameter r (1 for cyclic, 0 for acyclic compounds) and a ring constant B_r instead of the regression constant B_0 . The results of these models are shown in Figure 8.

The errors of fitting the B_r and B_0 models are not of significant difference. The advantage of the B_r models can be clarified by looking at the theoretical values of the regression coefficients for the $\text{CH}_3\text{-CH}_2\text{-}$ and $\text{-CH}_2\text{-CH}_2\text{-}$ bond types. In the case of the M model, these values should be 150 and 100 with respect to the Kovats' definition of retention indexes. They were calculated respectively to be 156.2 and 98.4 using the B_r model and to be 124.9 and 98.4 using the B_0 model. Because of the good agreement between the theoretical values and calculated coefficients using the B_r model, they were combined in a structure-dependent parameter Δ . Similar considerations were made for the 1ξ model while the Δ parameter was not taken into account for the S and R models.

The resulting models show high significance of all the calculated regression coefficients. Table III gives an overview of the model coefficients and a statistical evaluation. The t values of the coefficients vary with the frequency of the bond types in the data set given in Table IV. This is independent of the regarded model. No model can be a preferred fitting retention index of alkanes. Therefore, all the four models were expanded in order to consider double bonds and compounds containing oxygen as heteroatoms.

Table III. Coefficients and Corresponding Statistical Parameters from Regression Analysis of Alkane Data Set Using Edge Models of Figure 8

model parameter	coeff type	coeff	SE	t value	sig level
M	B_{r1}	62.57	8.16	7.67	0.0000
	$\text{CH}_3\text{-CH}_2\text{-}$	156.18	2.53	61.80	0.0000
	$\text{CH}_3\text{-CH}<$	123.65	1.81	68.39	0.0000
	$\text{CH}_3\text{-C}<$	103.51	1.36	75.95	0.0000
	$\text{-CH}_2\text{-CH}_2\text{-}$	98.37	1.25	78.44	0.0000
	$\text{-CH}_2\text{-CH}<$	69.46	2.51	27.71	0.0000
	$\text{-CH}_2\text{-C}<$	62.46	3.01	20.72	0.0000
	$>\text{CH-CH}<$	64.97	4.66	13.94	0.0000
	$>\text{CH-C}<$	72.41	5.45	13.30	0.0000
	B_{r2}	54.23	9.35	5.80	0.0000
β^2	$\text{CH}_3\text{-CH}_2\text{-}$	76.16	1.63	46.73	0.0000
	$\text{CH}_3\text{-CH}<$	53.73	1.06	50.85	0.0000
	$\text{CH}_3\text{-C}<$	40.79	0.68	59.90	0.0000
	$\text{-CH}_2\text{-CH}_2\text{-}$	49.35	0.74	67.10	0.0000
	$\text{-CH}_2\text{-CH}<$	36.70	1.29	28.41	0.0000
	$\text{-CH}_2\text{-C}<$	33.94	1.35	25.13	0.0000
	$>\text{CH-CH}<$	37.71	2.05	18.38	0.0000
	$>\text{CH-C}<$	40.34	2.09	19.35	0.0000
	B_{r3}	54.81	9.58	5.720	0.0000
	$\text{CH}_3\text{-CH}_2\text{-}$	46.03	0.77	59.62	0.0000
β^3	$\text{CH}_3\text{-CH}<$	41.56	0.60	69.40	0.0000
	$\text{CH}_3\text{-C}<$	39.12	0.51	76.05	0.0000
	$\text{-CH}_2\text{-CH}_2\text{-}$	33.10	0.48	68.75	0.0000
	$\text{-CH}_2\text{-CH}<$	25.37	1.16	21.92	0.0000
	$\text{-CH}_2\text{-C}<$	25.04	1.68	14.91	0.0000
	$>\text{CH-CH}<$	27.73	2.78	9.98	0.0000
	$>\text{CH-C}<$	40.98	4.29	9.55	0.0000
	B_{r4}	64.48	7.94	8.12	0.0000
	$\text{CH}_3\text{-CH}_2\text{-}$	45.84	0.69	66.88	0.0000
	$\text{CH}_3\text{-CH}<$	40.72	0.53	76.91	0.0000
β^4	$\text{CH}_3\text{-C}<$	38.23	0.46	83.53	0.0000
	$\text{-CH}_2\text{-CH}_2\text{-}$	32.75	0.41	80.72	0.0000
	$\text{-CH}_2\text{-CH}<$	24.22	0.91	26.71	0.0000
	$\text{-CH}_2\text{-C}<$	22.16	1.24	17.82	0.0000
	$>\text{CH-CH}<$	22.38	1.99	11.24	0.0000
	$>\text{CH-C}<$	25.71	2.82	9.13	0.0000

FITTING RETENTIONS INDEXES OF OXYGEN AND DOUBLE-BOND CONTAINING COMPOUNDS

The aim of the fitting retention index data of oxygen and double-bond containing compounds was to develop a more general model, which is able to estimate retention indexes of foreign compounds not belonging to a training data set. Therefore, the standard data set of the alkane retention indexes was extended by monofunctional compounds including 62 alkenes, 31 alcohols, 28 esters, 22 ketones, and 3 ethers from the Sadtler catalogue.² The bond types of the resulting calibration data set and the corresponding frequencies are shown in Table V. Considering these bond types, the edge models have to be extended in order to calculate 29 regression coefficients. Because of the strongly differing frequencies of the bond types, it is to be expected that the significance levels of the coefficients and the calculated t values of the M model, given in Table VI, are also different. In contrast to these expectations, nearly all coefficients are of high significance. Exceptions are the bond types $=\text{CH-CH}<$ and $=\text{CH-C}<$, which are relevant for special alkenes and aldehydes. The $>\text{C=C}<$ bond was found as a linear combination with other bond types in the used calibration data set. The comparison between fitted retention indexes of the calibration data set using edge models and those from literature is illustrated in Figure 9. The errors E_{MAE} and E_{SE} for all edge models are nearly the same as in the cases of fitting the alkane data set only. The edge models have been found to be stable during the expansion from fitting alkanes to the calibration data set.

In order to derive robust model coefficients, only monofunctional compounds were used. In cases of intramolecular

Table IV. Frequencies of Bond Types Incorporated in Fitted Alkane Data Set

CH ₃ -CH ₂ -	CH ₃ -CH<	CH ₃ -C<	-CH ₂ -CH ₂ -	-CH ₂ -CH<	-CH ₂ -C<	>CH-CH<	>C-C<
66	93	67	113	88	32	16	9

Table V. Frequencies of Bond Types Incorporated in Fitted Calibration Data Set

bond types (alkanes)	frequencies	bond types (alcohols)	frequencies	bond types (alkenes)	frequencies	bond types (esters)	frequencies
CH ₃ -CH ₂ -	208	OH-CH ₂ -	9	>C-CH ₃	37	>C-CH ₃	19
CH ₃ -CH<	209	OH-CH<	13	>C-CH ₂ -	17	>C-CH ₂ -	42
CH ₃ -C<	124	OH-C<	9	>C-CH<	8	>C-CH<	10
-CH ₂ -CH ₂ -	233			=CH-CH ₃	22	>C-C<	4
-CH ₂ -CH<	154			=CH-CH ₂ -	29	O=C<	53
-CH ₂ -C<	49			=CH-CH<	10	-O-C<	31
>CH-CH<	23			=CH-C<	7	-O-CH ₃	9
>CH-C<	11			CH ₂ =CH-	15	-O-CH ₂ -	30
				-CH=CH-	20	-O-CH<	4
				>C=CH ₂	12		
				>C=CH-	11		
				>C=C<	4		

Table VI. Coefficients and Corresponding Statistical Parameters from Regression Analysis of Calibration Data Set Using *M* Model^a

coeff type	coeff	SE	t value	sig level
<i>B</i> ₁	51.38	3.20	16.06	0.0000
CH ₃ -CH<	120.88	0.87	138.50	0.0000
CH ₃ -C<	102.24	0.98	104.50	0.0000
-CH ₂ -CH<	73.40	1.00	73.08	0.0000
-CH ₂ -C<	67.74	1.59	42.50	0.0000
>CH-CH<	68.11	2.71	25.09	0.0000
>CH-C<	76.17	3.73	20.43	0.0000
OH-CH ₂ -	303.55	3.75	80.94	0.0000
OH-CH<	225.56	3.46	65.27	0.0000
OH-C<	190.76	4.13	46.24	0.0000
O=C<	169.42	4.60	36.82	0.0000
-O-C<	23.91	4.67	5.12	0.0000
-O-CH ₃	174.15	3.87	45.03	0.0000
-O-CH ₂ -	95.72	2.60	36.76	0.0000
-O-CH<	36.27	6.23	5.82	0.0000
>C-CH ₃	152.39	1.51	101.11	0.0000
>C-CH ₂ -	94.99	2.38	39.87	0.0000
>C-CH<	67.35	3.71	18.14	0.0000
>C-C<	57.09	5.98	9.55	0.0000
=CH-CH ₃	102.13	8.84	11.55	0.0000
=CH-CH ₂ -	36.43	8.31	4.38	0.0000
=CH-CH<	1.20	9.85	0.12	0.9029
=CH-C<	1.26	10.47	0.12	0.9039
CH ₂ =CH-	199.53	9.21	21.66	0.0000
-CH=CH-	212.91	18.01	11.82	0.0000
>CH=CH ₂	86.49	4.67	18.55	0.0000
>C=CH-	104.20	9.59	10.87	0.0000
>C=C<				

^a The coefficients *B*_{CH₃-CH₂-} and *B*_{CH₂-CH₂-} are set to '150' and '100'.

interactions, the fitting results are not satisfactory. As an example, the calibration data set was extended by three 1,2-diols and three 1,3-diols where intramolecular hydrogen bonds are to be expected. The fitted retention indexes of the diols were found to be about a value of 20 too small. Another group of compounds which yield the largest fitting errors, are the lactones. In Figure 10 the calibration data set is extended by three lactones. The lactones are lying dramatically outside all calibration compounds. On the other hand, outliers may be helpful to point out unconsidered molecular properties like the intramolecular hydrogen bonds in the case of the mentioned diols.

Comparing the four edge models, a preference cannot be given to one of them. The fitting errors have been found within a small range. This is even true in the case of the simple incremental *M* model. The use of atomic-level indexes as basis for model parameters has no advantage with respect to fitting errors. A less important advantage of these models

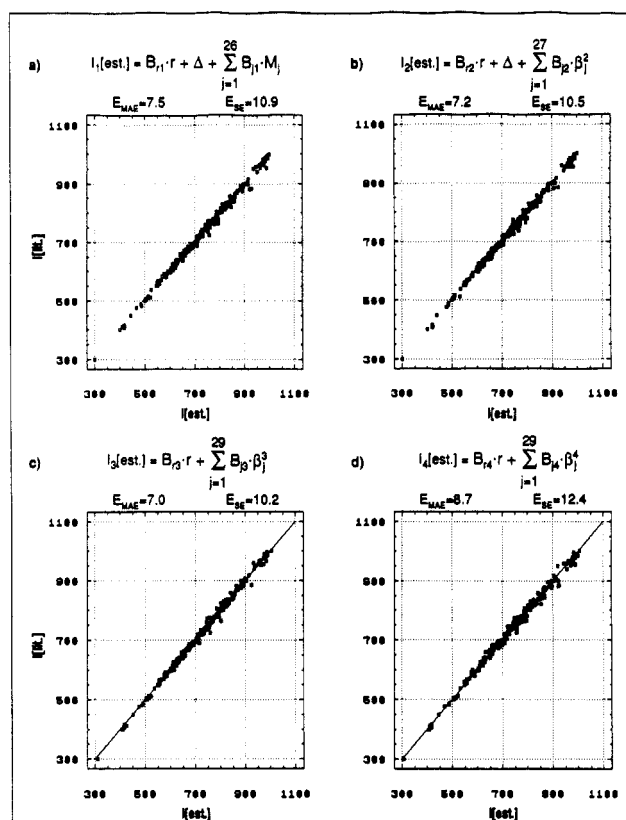


Figure 9. Comparison of fitted retention indexes *I*[est.] of the calibration data set shown in Table I including 217 acyclic and cyclic alkanes, alkenes, alcohols, esters, ketones, and ethers using edge models with data from literature *I*[lit.]. *E*_{MAE} = mean absolute error; *E*_{SE} = standard error.

is that linear combinations can widely be avoided, while the *M* model offers an incremental method for the prediction of retention indexes. In mixing edge and vertex models, we found no significant improvement for the fitting in the errors *E*_{MAE} and *E*_{SE}. This is in contradiction to Herndon,²⁶ who counts vertices, edges, and additional molecular paths including two neighbored bonds. In fitting physical properties, it is important to control the frequencies of atom, bond, and path types. Both of the additionally used path types and one of the bond types in Herndon's data set appear only once. This leads of course to a drastic improvement of the fitting errors. When building a calibration data set for retention indexes, we discarded all bond types with low frequencies and the related compounds to obtain robust coefficients.

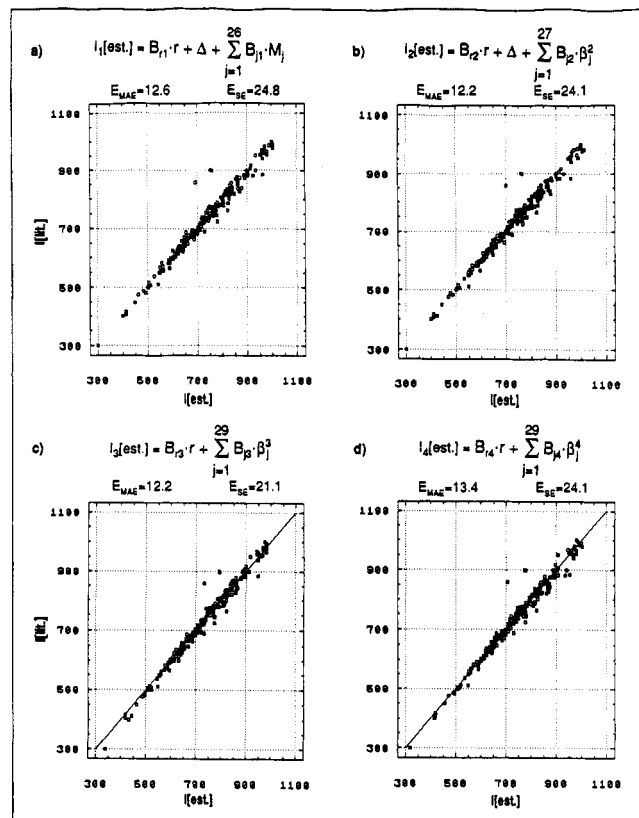


Figure 10. Comparison of fitted retention indexes $I[\text{est.}]$ of the calibration data set expanded by three lactones (outliers) using edge models with data from literature $I[\text{lit.}]$. E_{MAE} = mean absolute error; E_{SE} = standard error.

CONCLUSIONS

The presented vertex and edge models are of interest for fitting physical properties of chemical compounds. As an example, gas chromatographic retention index data of mono-substituted compounds gave the best fitting results when using edge models. In other cases vertex models, mixed models, or those using the additional counting of molecular paths may be advantageous. In fitting retention data with the four different edge models based on atomic-level topological indexes, no preference is to be given to one of them.

In order to test the predictive power of the edge models, it is necessary to investigate test data that were not incorporated in the development of the models. As a necessary supposition, the bond types of the test data set must be included in the calibration data set. In this case, the prediction of test data is based on proposed similarities to the calibration data set. The prediction of the retention indexes of compounds not incorporated in the calibration data set is the main topic of a following paper.²⁹ The results of the prediction in connection with the results of the fitting retention index data will allow us to evaluate the quality of the discussed models.

REFERENCES AND NOTES

- (1) Rijks, J. A.; Cramers, C. A. High precision capillary gas chromatography of hydrocarbons. *Chromatographia* **1974**, *77*, 99–106.

- (2) *The Sadtler standard gas chromatography retention index library*; Sadtler Research Laboratories: Philadelphia, PA, 1985.
- (3) Davies, N. W. Gas chromatographic retention indices of monoterpenes and sesquiterpenes on methyl silicone and Carbowax 20M phases. *J. Chromatogr.* **1990**, *503*, 1–24.
- (4) McReynolds, W. O. *Gas chromatographic retention data*; Preston Technical Abstracts: Evanston, IL, 1966.
- (5) Kalisz, R. *Quantitative structure–chromatographic retention relationships*; John Wiley & Sons: New York, 1987.
- (6) Balaban, A. T. Applications of graph theory in chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 334–343.
- (7) Wiener, H. J. Correlation of heats of isomerization and differences in heat of vaporization of isomers, among the paraffin hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636–2638.
- (8) Hosoya, H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (9) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (10) Balaban, A. T. Chemical graphs. XXXIV. Five new topological indices for the branching of tree-like graphs [1]. *Theor. Chim. Acta* **1979**, *53*, 355–375.
- (11) Hall, L. H.; Kier, L. B. The molecular connectivity χ indexes and κ shape indexes in structure–property modeling. In *Reviews Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; p 367.
- (12) Stanton, D. T.; Jurs, P. C.; Hicks, M. G. Computer-assisted prediction of normal boiling points of furans, tetrahydrofurans, and thiophenes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301–310.
- (13) Stanton, D. T.; Jurs, P. C. Computer-assisted study of the relationship between molecular structure and surface tension of organic compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 109–115.
- (14) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative study of molecular descriptors derived from the distance matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28–37.
- (15) Michotte, Y.; Massart, D. L. Molecular connectivity and retention indexes. *J. Pharm. Sci.* **1977**, *66*, 1630.
- (16) Millership, J. S.; Woolfson, A. D. The relation between molecular connectivity and gas chromatographic retention data. *J. Pharm. Pharmacol.* **1978**, *30*, 483.
- (17) Randić, M. The structural origin of chromatographic retention data. *J. Chromatogr.* **1978**, *161*, 1.
- (18) Bonchev, D.; Mekenjan, O.; Protić, G.; Trinajstić, N. Application of topological indices to gas chromatographic data: Calculation of the retention indices of isomeric alkylbenzenes. *J. Chromatogr.* **1979**, *176*, 149.
- (19) Kier, L. B.; Hall, L. H. Molecular connectivity analyses of structure influencing chromatographic retention indexes. *J. Pharm. Sci.* **1979**, *68*, 120.
- (20) Sabljčić, A. Calculation of retention indices by molecular topology. Chlorinated benzenes. *J. Chromatogr.* **1985**, *319*, 1.
- (21) Hall, L. H.; Mohnen, B.; Kier, L. B. The electrotopological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76–82.
- (22) Takács, J.; Szila, C.; Tarjan, G. Contribution to the theory of the retention index system. III. Retention index and molecular structure. Calculation of retention indices of paraffin hydrocarbons on the basis of their molecular structure. *J. Chromatogr.* **1971**, *56*, 1–12.
- (23) Lombosi, T. S.; Lombosi, E. R.; Bernát, I.; Bernát, Zs. Sz.; Takács, E. C.; Takács, J. M. General contribution to the theory of retention index systems in gas–liquid chromatography. I. Pre-calculation of the retention indices of alkanes on the basis of their molecular structures and thermodynamics in gas–liquid chromatography. *J. Chromatogr.* **1976**, *119*, 307–319.
- (24) Takács, E. C.; Vörös, J.; Takács, J. M. Determination of bond- and/or interaction increment codes and pre-calculation of retention indices for alkanes in gas–liquid chromatography by computer. *J. Chromatogr.* **1978**, *150*, 297–307.
- (25) Gautzsch, R.; Zinn, P.; Haffer, Chr. M. Experiences with symbolic programming in the automated estimation of retention index data. *Chem. Intell. Lab. Sys.* **1991**, *12*, 81–90.
- (26) Garbalena, M.; Herndon, W. C. Optimum graph-theoretical models for enthalpic properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 37–42.
- (27) Gautzsch, R.; Zinn, P. List operations on chemical graphs. 1. Basic list structures and operations. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 541–550.
- (28) Gautzsch, R.; Zinn, P. List operations on chemical graphs. 2. Combining basic list operations. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 551–555.
- (29) Duvenbeck, Ch.; Zinn, P. List operations on chemical graphs. 4. Using edge models for prediction of retention index data. *J. Chem. Inf. Comput. Sci.* **1993**, following paper in this issue.