Liquid Density and Critical Properties of Hydrocarbons Estimated from Molecular Structure

William A. Wakeham,*,† Georgi St. Cholakov,‡ and Roumiana P. Stateva§

University of Southampton, Highfield, Southampton SO17 1BJ, U.K., Department of Organic Synthesis and Fuels, University of Chemical Technology and Metallurgy, Sofia 1756, Bulgaria, and Institute of Chemical Engineering, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria

The paper provides new correlations for the estimation of the relative liquid density and critical parameters of hydrocarbons. The correlations employ molecular descriptors from computer simulation of molecular mechanics, which proved valuable in our previous work. High precision is achieved, without the use of unduly complex descriptors and rules, as a result of improved methodology. The new methodology includes a limit upon the functionality of the property database and its predesign by expanding the range and improving the distribution of the repeating structural features in the selected compounds.A compilation of estimated data for the liquid density and critical properties of some hydrocarbons of industrial importance for which no values have been available before is provided.

1. Introduction

The estimation of the properties of pure components is an important prerequisite for the design of processes and equipment, environmental impact assessment, HAZOP studies, and other major chemical engineering activities related to phase equilibrium. Large commercial databases of such properties are compiled and their population has to include compounds not yet synthesized and/or compounds for which an experimental determination of the properties is unreliable or impossible.[1] Consequently, new methods for the interpolation and extrapolation of existing data are needed.

Modern computer simulation packages for design often require the normal boiling point and a standard relative liquid density (d_4^{20}) of the pure components of a mixture that occur in a process. These parameters are fundamental characteristics of a fluid by themselves, but they are also often the input from which all the other physical properties required for a particular design can be estimated. Our earlier work[2] was devoted to the estimation of the normal boiling points of pure hydrocarbons. In this paper, we present new, quantitative structure−property relationships (QSPRs) for the evaluation of d_4^{20} and the critical properties of hydrocarbons.

Group contribution methods are the most common tool employed for the estimation of properties from molecular structure. They typically begin with a decomposition of the molecular structure into particular groups and the counting of atoms in those groups (atom counts). Characteristic incremental values are assigned to the groups by regression of known data for the particular property. The whole molecule is then “restored” by summation of the contributions of all groups, and the property is determined as a function of the summations.[3] The early group contribution methods (i.e. Lydersen) have become very popular because of their inherent advantages−explicit and comprehensive description of the molecules, straightforward restoration of the whole structure of the molecule, and simple calculation of the desired property. An important additional advantage is their direct application for generation of molecules having predesigned properties.[3,4] Conversely, the earlier group contribution models using atom counts do not distinguish among isomers, and they tend to produce unreliable results for compounds which are not similar to those used in their derivation.[5] In attempts to achieve greater accuracy, the number of groups has been increased, more sophisticated groups have been introduced, and more complicated rules for the restoration of the molecule from the groups have been tried. The increase of the number of groups means that statistically the predictions are even more closely linked to the available information. The sophistication of the groups themselves has led to representations of the molecules that are rather complex.[5]

A significant development has been the employment of graph topological indices in physical property correlations. Topological indices are nonempirical. They use a comprehensible representation of the molecular structure by an adjacency and/or a distance matrix but introduce complicated rules for extracting the information from the matrixes and the restoration of the whole molecule.[6] The fact that predictive schemes containing only topological descriptors have relatively modest accuracy has led to the increase in the number of descriptors to more than 150 and still further sophistication of the rules.

Considerable improvements of predictive accuracy have been reported recently with the introduction of group interactions[5] and second- and third-order group contributions[7,8] and−in a separate development−the introduction of combining topological, electronic, geometrical, and other descriptors.[3,9−12] In both cases, the cost was even greater sophistication and complexity.

The introduction of second and third levels of description of the molecular structure and combinations of descriptors marks the emerging of a new method, which we are going to call henceforth the “significant common-feature method”. Typically,[2,9−12] this method aims at the selection on the basis of experimental data of a small combination of

---

* Corresponding author. Telephone: +44 (0)23 8059 2801. Fax: +44 (0)23 8059 3159. E-mail: vice-chancellor@soton.ac.uk.

† University of Southampton.

‡ University of Chemical Technology and Metallurgy.

§ Bulgarian Academy of Sciences.

10.1021/je010308l CCC: $22.00

descriptors, which comprise the most significant common features of the molecular structures, relevant to the particular property being targeted. The traditional advantages of the early group contribution methods are lessened by the two developments described, but the improved prediction that results is not always commensurate with the additional effort.

The present work is an attempt to find a compromise between the above extremes, i.e., to develop correlations, which ensure reasonably high precision of the estimated properties, but do not involve unduly complex descriptors and rules. Because this is a first step in this direction, and for reasons that are going to be discussed later, the compounds targeted in the present work are limited to the most industrially important classes of hydrocarbons. To what extent such a compromise can be found for other more complex molecules, especially when polarity and solvation effects are involved, remains an open question.

We continue the approach of our previous work[2] by examining the use of molecular descriptors from computer simulation of molecular mechanics. We also address the practical problem arising from the lack of an adequate amount of experimental data. An opportunity to ameliorate the problem limiting the functional groups in the targeted compounds, and by expanding their range and improving the variation of their structural elements with extrapolated (pseudo-experimental) data and a controlled small amount of nonexperimental data is suggested.

We employ our methodology to compile estimated data for d_4^{20} and critical properties of some hydrocarbons of industrial importance, for which no such data are available experimentally.

2. Methodology

The correlations described in this paper employ two databases—the *property database* and the *structure database*. These databases are interrelated since both contain the same selection of molecules. The quantitative relationship between the databases is established in the form of a *model*, the constants of which are determined by minimizing the standard deviation of the predicted data from the available experimental data in the property database.

2.1. Property Database. QSPRs establish relations between a set of molecules and their properties. If a correlation is recommended as applicable to all known organic compounds, the relevant set of the latter will consist of millions of molecules grouped in clusters of structurally related compounds. Recent work[13,14] has shown that the derivation of QSPRs should obey the general rules of statistical experimental designs. Consequently, the model should be developed from a property database, which adequately represents the set of compounds that are being targeted for prediction.

The property database should contain more than a critical amount of experimental data and representatives of all relevant structural features of the set in carefully selected proportions. Much of the earlier work in this field has employed databases that include compounds selected arbitrarily, merely on the basis of what exists. However, following our previous successful experience,[2] we have tried to approach the requirement advocated above. Databases of representative molecules, selected to reflect the main structural molecular features of the targeted groups of chemical compounds, we shall further call “predesigned databases” to distinguish them from databases of arbitrary compounds.

The first step in our methodology is defining the design of the structural relations in the database. For the purposes of the present study, it should consist of hydrocarbons, representing four main industrially important groups (*n*- and isoalkanes, 1-alkenes, cycloalkanes, and arenes). It should contain also hybrid compounds, related to the main groups by structural elements, which appear partially or in full in the four groups (i.e., *n*- and isoalkylarenes and cycloarenes, aliphatic and cyclic dienes, polyarenes, etc).

In ref 2, we have shown that the compounds in the database used for the normal boiling points are structurally related and fulfill the requirement defined above. This was demonstrated by projection of the molecules of the database on the plane of the first two principal components of the molecular features selected for estimating normal boiling points. That is why, in the present study, we use the database from our previous investigation on the normal boiling points.[2] Compound no. 233 (*m*-ethyltoluene), which was wrongly listed also under no. 196 in the earlier database, has been deleted, and the compounds after it have been renumbered accordingly here. Compound no. 218, which was wrongly printed in ref 2 as α-methylbenzene, is given its correct name—α-methylstyrene.

The names and number of the compounds in the present database are shown in Table 1. The available data for their properties are listed in Table 2. The density is expressed as relative liquid density, d_4^{20}—the ratio of the density of the hydrocarbon at 20 °C and the density of water at 4 °C. The data for compounds, which are solid or gaseous at that temperature, are extrapolated densities of the undercooled liquid, values for the triple point or the saturated state.

All data for the 260 hydrocarbons were obtained from the literature. The values in Table 2 are divided into three groups—experimental, extrapolated, and nonexperimental. The experimental values for d_4^{20}were compiled from refs 15−17. Values for the critical temperature (T_c) and the critical pressure (P_c), taken from ref 18 were considered experimental, as reported by the authors. The rest of the experimental values for the critical properties are those recommended in refs 19−23, complemented with recently obtained experimental values.[17,24−26] For compounds for which recommended and/or more recent values than those in ref 18 were available, the former were taken. As noted in Table 2, only three critical temperatures and three critical pressures from the available experimental values for the compounds in the database were not used. The reasons for the exclusions were the relatively high experimental errors reported for these compounds or the use of unquoted values of the normal boiling point in the extrapolation to the critical values for compounds unstable at the critical state. Tables 1 and 2 are used to illustrate the fundamental methodological problems in the development of QSPRs below.

The success of any model depends on the amount of data used in determining the contribution of the independent variables (descriptors), as compared to the number of the selected variables. QSPRs depend also on the extent and uniformity of the distribution of the variation of the descriptors. Ideally, a QSPR should be developed only from reliable experimental data for the properties of structurally related molecules. The available experimental data for our database as seen from Table 2 are relatively low in number. It should be underlined also, that experimental data for the critical properties are available altogether only for less than 200 hydrocarbons.[19] The total number of organic compounds for which there is some experimental data is between 700 and 800.[8]

Table 1. Names of Hydrocarbons Included in the Database[a,b,c]

no.	name	no.	name	no.	name	no.	name
1	ethane	66	3-methyl-3-ethylpentane	131[b]	3-methyl-1-butene	196[a,b]	*m*-ethyltoluene
2[b]	propane	67	2,2,3-trimethylhexane	132	2,3-dimethylbutadiene	197	pentylbenzene
3[b]	*n*-butane	68	2,2,4-trimethylhexane	133	3,3-dimethyl-1-butene	198	hexylbenzene
4[b]	*n*-pentane	69	2,2,5-trimethylhexane	134	2-methyl-2-pentene	199	heptylbenzene
5[b]	*n*-hexane	70[b]	3,3-diethylpentane	135	3-methyl-1-pentene	200	octylbenzene
6[a,b]	*n*-heptane	71[a,b]	2,2,3,3-tetramethylpentane	136[b]	1,5-hexadiene	201	nonylbenzene
7[b]	*n*-octane	72	2,2,3,4-tetramethylpentane	137	limonene	202	decylbenzene
8[b]	*n*-nonane	73	2,2,4,4-tetramethylpentane	138	α-pinene	203	undecylbenzene
9[b]	*n*-decane	74	2,3,3,4-tetramethylpentane	139	squalene	204	dodecylbenzene
10	*n*-undecane	75	2-methyloctane	140	lycopene	205	tridecylbenzene
11	*n*-dodecane	76	2-methylnonane	141	*β*-carotene	206	tetradecylbenzene
12	*n*-tridecane	77	3,3,5-trimethylheptane	142	cyclopropane	207	pentadecylbenzene
13[a,b]	*n*-tetradecane	78	2,2,3,3-tetramethylhexane	143	cyclobutane	208	hexadecylbenzene
14	*n*-pentadecane	79	2,5-dimethyldecane	144[a,b]	cyclopentane	209	heptadecylbenzene
15	*n*-hexadecane	80	2,5,-dimethyldodecane	145[b]	cyclohexane	210	octadecylbenzene
16	*n*-heptadecane	81	2,6,10-trimethyldodecane	146	cycloheptane	211	nonadecylbenzene
17[a,b]	*n*-octadecane	82	2,6,10-trimethyltetradecane	147	cyclooctane	212	eicosylbenzene
18[b]	*n*-nonadecane	83	pristane	148[a,b]	methylcyclohexane	213	heneicosylbenzene
19	*n*-eicosane	84	phytane	149[b]	ethylcyclohexane	214	docosylbenzene
20	*n*-heneicosane	85	squalane	150[b]	propylcyclohexane	215	tricosylbenzene
21	*n*-docosane	86	lycopane	151	butylcyclohexane	216	tetracosylbenzene
22	*n*-tricosane	87[b]	propylene	152[b]	methylcyclopentane	217	styrene
23	*n*-tetracosane	88[a,b]	1-butene	153[b]	ethylcyclopentane	218	α-methylstyrene
24	*n*-pentacosane	89[b]	1-pentene	154	propylcyclopentane	219[b]	cumene
25	*n*-hexacosane	90[b]	1-hexene	155	butylcyclopentane	220[b]	*o*-ethyltoluene
26	*n*-heptacosane	91[a,b]	1-heptene	156	pentylcyclopentane	221	*p*-ethyltoluene
27	*n*-octacosane	92[a,b]	1-octene	157	hexylcyclopentane	222[b]	mesitylene
28	*n*-nonacosane	93	1-nonene	158	heptylcyclopentane	223[b]	1,2,3-trimethylbenzene
29	*n*-triacontane	94[a,b]	1-decene	159	octylcyclopentane	224[b]	1,2,4-trimethylbenzene
30	*n*-dotriacontane	95	1-undecene	160	nonylcyclopentane	225[b]	1,2,3,4-tetrahydronaphthalene
31	*n*-pentatriacontane	96	1-dodecene	161	decylcyclopentane	226	*tert*-butylbenzene
32	*n*-hexatriacontane	97	1-tridecene	162	undecylcyclopentane	227[b]	*p*-cymene
33	*n*-tetracontane	98	1-tetradecene	163	dodecylcyclopentane	228	*m*-diethylbenzene
34	*n*-tetratetracontane	99	1-pentadecene	164	tridecylcyclopentane	229	*o*-diethylbenzene
35	*n*-hexacontane	100	1-hexadecene	165	tetradecylcyclopentane	230[b]	isobutylbenzene
36[b]	*i*sobutane	101	1-heptadecene	166	pentadecylcyclopentane	231	*m*-diisopropylbenzene
37[b]	2-methylbutane	102	1-octadecene	167	hexadecylcyclopentane	232	diphenylmethane
38[b]	2,2-dimethylpropane	103	1-nonadecene	168	heptadecylcyclopentane	233	*sec*-butylbenzene
39[b]	2-methylpentane	104	1-eicosene	169	octadecylcyclopentane	234[a,b]	*p*-diethylbenzene
40[b]	3-methylpentane	105	1-heneicosene	170	nonadecylcyclopentane	235	*p*-diisopropylbenzene
41[b]	2,2-dimethylbutane	106	1-docosene	171	eicosylcyclopentane	236	diphenyl
42[b]	2,3-dimethylbutane	107	1-tricosene	172	heneicosylcyclopentane	237	1,1-diphenylethane
43[b]	2-methylhexane	108	1-tetracosene	173	docosylcyclopentane	238	1,2-diphenylethane
44[b]	3-methylhexane	109	1-pentacosene	174	tricosylcyclopentane	239	naphthalene
45	2,2-dimethylpentane	110	1-hexacosene	175	tetracosylcyclopentane	240	anthracene
46[b]	2,3-dimethylpentane	111	1-heptacosene	176	pentacosylcyclopentane	241	phenanthrene
47	2,4-dimethylpentane	112	1-octacosene	177	*cis*-1,2-dimethylcyclohexane	242	*m*-terphenyl
48	3,3-dimethylpentane	113	1-nonacosene	178[b]	*trans*-1,2-dimethylcyclohexane	243	*p*-terphenyl
49	3-ethylpentane	114	1-triacontene	179[b]	*cis*-1,3-dimethylcyclohexane	244	1,2-benzo-[*a*]-pyrene
50[a,b]	2,2,3-trimethylbutane	115[a,b]	1,3-butadiene	180[a,b]	*trans*-1,3-dimethylcyclohexane	245	pyrene
51	2-methylheptane	116[b]	*cis*-2-butene	181	*cis*-1,4-dimethylcyclohexane	246	chrysene
52	3-methylheptane	117[b]	*trans*-2-butene	182[b]	*trans*-1,4-dimethylcyclohexane	247	*o*-terphenyl
53	4-methylheptane	118[b]	isobutene	183	cyclopentene	248	triphenylmethane
54	2,2-dimethylhexane	119	isoprene	184[a,b]	cyclohexene	249	acenaphthylene
55[b]	2,3-dimethylhexane	120	2,3-dimethyl-1-butene	185	1,3-cyclohexadiene	250	acenaphthene
56	2,4-dimethylhexane	121	2,3-dimethyl-2-butene	186	5-methyl-1,3-cyclopentadiene	251	1,1,2,2-tetraphenylethane
57	2,5-dimethylhexane	122	2-ethyl-1-butene	187	1,3-cyclopentadiene	252	4-methyloctane
58	3,3-dimethylhexane	123	*cis*-2-hexene	188[b]	benzene	253	2,2,3,3-tetramethylbutane
59	3,4-dimethylhexane	124	*trans*-2-hexene	189[b]	toluene	254	2-ethylhexene
60	3-ethylhexane	125	2-methyl-1-pentene	190[b]	ethylbenzene	255	adamantane
61[b]	2,2,3-trimethylpentane	126	4-methyl-1-pentene	191[b]	propylbenzene	256	1,5-cyclooctadiene
62[a,b]	2,2,4-trimethylpentane	127	2,4,4-trimethyl-1-pentene	192[b]	butylbenzene	257	2,5-methyl-1,5-hexadiene
63[b]	2,3,3-trimethylpentane	128	2,4,4-trimethyl-2-pentene	193[a,b]	*o*-xylene	258	c-1-propenylbenzene
64	2,3,4-trimethylpentane	129[a,b]	2-methyl-1-butene	194[b]	*m*-xylene	259	1-phenylnaphthalene
65[b]	2-methyl-3-ethylpentane	130[b]	2-methyl-2-butene	195[b]	*p*-xylene	260	indane

[a] Members of the prediction set (pset) of this work. [b] Members of the comparison set, the same as in ref 18. [c] All compounds, which are not shown above as members of the pset, comprise the learning set (lset) of this work.

From the point of view of mathematical statistics, an equal number of the independent variables (structural descriptors) and the property data used in estimating their contribution lead to an exact solution. The derived model, however, is a chance correlation, which should not be used for any predictive purposes. A high ratio of the number of property data to the number of descriptors decreases the probability of developing chance correlations. Thus, the number of independent variables, which can be used in critical property QSPRs is rather limited, for example, to significantly less than 200 variables for hydrocarbons.

The available experimental data are also limited to compounds of relatively low and medium molecular masses. The data for alkylcycloalkanes and alkylbenzenes, for instance, do not span further than several carbon atoms in the side chain; no densities for polycyclic hydrocarbons are available; and the critical properties of only several of the latter have been determined. A chemical engineer may need the critical properties of high molecular mass alkanes in order to simulate the dewaxing of high boiling petroleum fractions, of polyarenes to simulate their carcinogenic activity, or of *β*-carotene to simulate its supercritical

Table 2. Data for the Properties of the Compounds in the Data Base[d]

		T_c	P_c	V_c			T_c	P_c	V_c			T_c	P_c	V_c			T_c	P_c	V_c
no.	d_4^{20}	K	MPa	$cm^3\ mol^{-1}$	no.	d_4^{20}	K	MPa	$cm^3\ mol^{-1}$	no.	d_4^{20}	K	MPa	$cm^3\ mol^{-1}$	no.	d_4^{20}	K	MPa	$cm^3\ mol^{-1}$
1	**0.3390**[c]	**305.3**	**4.872**	**145.5**	66	**0.7274**	**576.5**	**2.81**	455	131	**0.6272**[c]	**452.7**	**3.53**	**304.9**	196	**0.8645**	**637.1**	**2.837**	490
2	**0.5004**[c]	**369.8**	**4.248**	**200**	67	**0.7295**	588	2.493	498	132	**0.7267**	527.2	3.536	333	197	**0.8585**	**678.9**[c]	**2.634**[c]	**552**[c]
3	**0.5786**[c]	**425.1**	**3.796**	**255**	68	**0.7155**	573.7	2.371	507	133	**0.6529**	490.2	3.151		198	**0.8583**[c]	**695.8**[c]	**2.394**[c]	**614**[c]
4	**0.6262**	**469.7**	**3.37**	**311**	69	**0.7072**	**569.8**	2.249	519	134	**0.6863**	518	3.283	351	199	**0.8575**[c]	**711.3**[c]	**2.189**[c]	**677**[c]
5	**0.6593**	**507.6**	**3.025**	**368**	70	**0.7535**	**610.1**	**2.675**	473	135	**0.6675**				200	**0.8569**[c]	**725.6**[c]	**2.009**[c]	**741**[c]
6	**0.6838**	**540.2**	**2.74**	**428**	71	**0.7566**	**607.5**	**2.74**	478	136	**0.6923**	**508**	3.4454	328	201	**0.8564**[c]	**738.8**[c]	**1.854**[c]	**805**[c]
7	**0.7027**	**568.7**	**2.49**	**492**	72	**0.7389**	**592.6**	**2.6**	490	137	**0.8422**	650	2.75		202	**0.8561**[c]	**750.9**[c]	**1.716**[c]	**870**[c]
8	**0.7177**	**594.6**	**2.29**	**555**	73	**0.7194**	**574.6**	**2.49**	504	138	**0.8582**	632	2.76		203	**0.8558**[c]	**762.2**[c]	**1.595**[c]	**935**[c]
9	**0.7301**	**617.7**	**2.11**	**624**	74	**0.7547**	**607.5**	**2.72**	481	139	**0.8562**				204	**0.8556**[c]	**772.7**[c]	**1.485**[c]	**1001**[c]
10	**0.7402**	**639**	**1.98**	**689**	75	**0.7134**	**582.8**	**2.31**	541	140					205	**0.8554**[c]	**782.4**[c]	**1.388**[c]	**1066**[c]
11	**0.7488**	**658**	**1.82**	**754**	76	0.7261	610.4	2.097	596	141					206	**0.8552**[c]	**791.4**[c]	**1.301**[c]	**1132**[c]
12	**0.7561**	**675**	**1.68**	**823**	77	0.7432	**609.5**	**2.32**	537	142		**398**	**5.495**	**162**	207	**0.8551**[c]	**799.9**[c]	**1.221**[c]	**1198**[c]
13	**0.7626**	**693**	**1.57**	**894**	78	0.7644	**623**	**2.51**	518	143	0.6873	460	5.056	210	208	**0.8551**[c]	**807.7**[c]	**1.149**[c]	**1263**[c]
14	**0.7684**	**708**	**1.48**	**966**	79					144	**0.7454**	**511.7**	**4.51**	**259**	209	**0.8550**[c]	**815.1**[c]	**1.083**[c]	**1328**[c]
15	**0.7735**	**723**	**1.40**	**1034**	80					145	**0.7785**	**553.8**	**4.08**	**308**	210	**0.8550**[c]	**822**[c]	**1.023**[c]	**1393**[c]
16	**0.7780**	**736**	**1.34**	**1103**	81	0.7714				146	**0.8098**	**604.2**	**3.82**	**353**	211	**0.8549**[c]	**828.4**[c]	**0.968**[c]	**1457**[c]
17	**0.7828**[c]	**747**	**1.29**	**1189**	82	0.7806				147	**0.8349**	**647.2**	**3.56**	**410**	212	**0.8549**[c]	**834.4**[c]	**0.917**[c]	**1521**[c]
18	**0.7835**[c]	**755**	**1.16**	**1249**[c]	83	0.7845				148	**0.7694**	**572.1**	**3.48**	**369**	213	**0.8548**[c]	**840.1**[c]	**0.870**[c]	**1585**[c]
19	**0.7843**[c]	**768**	**1.07**	**1319**[c]	84	0.7935				149	**0.7879**	**609.2**	**2.97**	**450**	214	**0.8548**[c]	**845.4**[c]	**0.827**[c]	**1645**[c]
20	**0.7916**[c]	**778**	**1.03**	**1390**[c]	85	**0.8087**	*b*	*b*		150	**0.7935**	**639.1**	**2.65**	**477**[c]	215	**0.8547**[c]	**850.4**[c]	**0.787**[c]	**1711**[c]
21	**0.7943**[c]	**786**	**0.98**	**1461**[c]	86					151	**0.7996**	**667.2**[c]	**2.385**[c]	**534**[c]	216	**0.8547**[c]	**855.2**[c]	**0.750**[c]	**1773**[c]
22	**0.7968**[c]	**790**	**0.92**	**1530**[c]	87	**0.5133**[c]	**364.9**	**4.60**	**184**	152	**0.7486**	**532.7**	**3.79**	**318**	217	**0.9057**	**635.2**	**3.992**	
23	**0.7990**[c]	**800**	**0.87**	**1600**[c]	88	**0.5945**[c]	**419.5**	**4.02**	**240**	153	**0.7665**	**569.5**	**3.40**	**375**	218	**0.9090**	654[b]	**3.405**	**410**
24	**0.8011**[c]	**808**[c]	**0.846**[c]	**1669**[c]	89	**0.6405**	**464.8**	**3.56**	**298**	154	**0.7763**	**594.7**[c]	**2.962**[c]	**426**[c]	219	**0.8617**	**631.2**	**3.209**	428
25	**0.8031**[c]	**816**	**0.795**	**1737**[c]	90	**0.6732**	**504**	**3.21**	**355**	155	**0.7845**[c]	**618.9**[c]	**2.65**[c]	**484**[c]	220	**0.8806**	**651.1**	**3.04**	460
26	**0.8049**[c]	**821**[c]	**0.769**[c]	**1805**[c]	91	**0.6970**	**537.3**	**2.92**	**409**	156	**0.7915**[c]	**640.8**[c]	**2.385**[c]	**543**[c]	221	**0.8611**	**640.2**	**3.23**	470
27	**0.8066**[c]	**824**	**0.744**	**1873**[c]	92	**0.7149**	**567**	**2.68**	**468**	157	**0.7976**[c]	**660.7**[c]	**2.162**[c]	**603**[c]	222	**0.8651**	**637.3**	**3.127**	433
28	**0.8082**[c]	**832**[c]	**0.703**[c]	**1940**[c]	93	**0.7292**	**594**	**2.429**[c]	**526**[c]	158	**0.8016**[c]	**678.9**[c]	**1.971**[c]	**665**[c]	223	**0.8944**	**664.5**	**3.454**	430
29	**0.8096**[c]	**843**	**0.636**	**2006**[c]	94	**0.7408**	**617**	**2.22**	**584**	159	**0.8056**[c]	**695.5**[c]	**1.806**[c]	**727**[c]	224	**0.8758**	**649.1**	**3.232**	430
30	**0.8123**[c]	**847**[c]	**0.588**[c]	**2136**[c]	95	**0.7503**	**638.7**[c]	**2.062**[c]	**652**[c]	160	**0.8086**[c]	**710.7**[c]	**1.663**[c]	**789**[c]	225	**0.9702**	**720**	**3.648**	408
31	**0.8157**[c]	**859**[c]	**0.517**[c]	**2327**[c]	96	**0.7584**	**658**	**1.93**	**716**[c]	161	**0.8116**[c]	**724.8**[c]	**1.537**[c]	**852**[c]	226	**0.8665**	660	2.969	
32	**0.8168**[c]	**872**	**0.472**	**2389**[c]	97	**0.7653**	**673**	**1.73**	**781**[c]	162	**0.8147**[c]	**737.7**[c]	**1.426**[c]	**916**[c]	227	**0.8573**	**652**	**2.8**	
33	**0.8204**[c]	**874**[c]	**0.445**[c]	**2632**[c]	98	**0.7713**	**691.7**	**1.59**	**846**[c]	163	**0.8167**[c]	**749.7**[c]	**1.326**[c]	**979**[c]	228	**0.8639**			
34	**0.8219**[c]	**885**[c]	**0.394**[c]	**2866**[c]	99	**0.7764**	**705**	**1.54**	**912**[c]	164	**0.8187**[c]	**760.8**[c]	**1.238**[c]	**1043**[c]	229	**0.8800**			
35	**0.8300**[c]	**909**[c]	**0.284**[c]	**3730**[c]	100	**0.7811**	**718**	**1.39**	**978**[c]	165	**0.8207**[c]	**771.1**[c]	**1.158**[c]	**1107**[c]	230	**0.8532**	**650**	**3.05**	480
36	**0.5569**[c]	**407.8**	**3.64**	**259**	101	**0.7852**	**734**	**1.34**	**1045**[c]	166	**0.8217**[c]	**780.6**[c]	**1.087**[c]	**1170**[c]	231	**0.8559**			
37	**0.6193**	**460.4**	**3.38**	**306**	102	**0.7888**	**748**	**1.30**	**1111**[c]	167	**0.8237**[c]	**789.6**[c]	**1.021**[c]	**1234**[c]	232	**1.0060**[c]	**760**	*b*	**563**
38	**0.6135**[c]	**433.8**	**3.196**	**307**	103	**0.7920**	**755**	**1.19**	**1177**[c]	168	**0.8247**[c]	**797.9**[c]	**0.963**[c]	**1297**[c]	233	**0.8621**	664	2.949	
39	**0.6532**	**497.7**	**3.04**	**368**	104	**0.7950**	**772**	**1.14**	**1244**[c]	169	**0.8257**[c]	**805.6**[c]	**0.908**[c]	**1360**[c]	234	**0.8620**	**657.9**	**2.807**	480
40	**0.6645**	**504.6**	**3.12**	**368**	105	**0.7977**	**776.5**[c]	**1.083**[c]	**1310**[c]	170	**0.8267**[c]	**812.9**[c]	**0.858**[c]	**1423**[c]	235	**0.8568**			
41	**0.6485**	**489**	**3.1**	**358**	106	**0.8001**	**785.3**[c]	**1.026**[c]	**1375**[c]	171	**0.8287**[c]	**819.7**[c]	**0.813**[c]	**1485**[c]	236		**773**	**3.38**	**497**
42	**0.6616**	**500**	**3.15**	**361**	107	**0.8023**	**793.5**[c]	**0.975**[c]	**1441**[c]	172	**0.8287**[c]	**826.1**[c]	**0.771**[c]	**1547**[c]	237	**0.9997**			
43	**0.6787**	**530.4**	**2.74**	**421**	108	**0.8045**	**801.1**[c]	**0.926**[c]	**1506**[c]	173	**0.8298**[c]	**832**[c]	**0.732**[c]	**1608**[c]	238				
44	**0.6872**	**535.2**	**2.81**	**404**	109	**0.8063**	**808.3**[c]	**0.882**[c]	**1570**[c]	174	**0.8308**[c]	**837.6**[c]	**0.696**[c]	**1669**[c]	239	**1.0253**[c]	**748.4**	**4.05**	**407**
45	**0.6738**	**520.5**	**2.77**	**416**	110	**0.8082**	**815**[c]	**0.840**[c]	**1634**[c]	175	**0.8318**[c]	**842.9**[c]	**0.663**[c]	**1729**[c]	240		883		
46	**0.6950**	**537.3**	**2.91**	**393**	111	**0.8097**	**821.3**[c]	**0.801**[c]	**1698**[c]	176	**0.8328**[c]	**847.8**[c]	**0.631**[c]	**1789**[c]	241		**869**		
47	**0.6726**	**519.8**	**2.74**	**418**	112	**0.8114**	**827.2**[c]	**0.765**[c]	**1761**[c]	177	**0.7962**	**606.1**	**2.938**		242		**883**	*b*	**724**
48	**0.6933**	**536.4**	**2.95**	**414**	113	**0.8127**	**832.7**[c]	**0.732**[c]	**1823**[c]	178	**0.7760**	**596.1**	**2.938**		243		**908**	**2.99**	**729**
49	**0.6983**	**540.6**	**2.89**	**416**	114	**0.8141**	**837.9**[c]	**0.700**[c]	**1886**[c]	179	**0.7660**	**591.1**	**2.9380**		244				
50	**0.6990**	**531.1**	**2.95**	**398**	115	**0.6211**	**425**	**4.32**	**221**	180	**0.7847**	**598**	**2.938**	408.5	245				
51	**0.6979**	**559.7**	**2.5**	**488**	116		**435.5**	**4.21**	**233.8**	181	**0.7828**	598	2.938		246				
52	**0.7058**	**563.6**	**2.55**	**464**	117	**0.6051**[c]	**428.6**	**4.10**	**237.8**	182	**0.7626**	**587.7**	**2.938**		247		*b*	2.99	731
53	**0.7046**	**561.7**	**2.54**	**476**	118	**0.5953**[c]	**417.9**	**4.00**	**238.8**	183	**0.7720**	**506.5**	**4.80**	**245**	248				
54	**0.6953**	**549.8**	**2.53**	**478**	119	**0.6805**	484	3.85	276	184	**0.8102**	**560.4**	**4.35**	292	249				
55	**0.7121**	**563.5**	**2.63**	**468**	120	**0.6779**	501	3.242	343	185	**0.8406**				250				
56	**0.7004**	**553.5**	**2.56**	**472**	121	**0.7080**	524	3.364	351	186					251				
57	**0.6936**	**550**	**2.49**	**482**	122	**0.6894**				187	**0.8021**				252	**0.7202**			
58	**0.7100**	**562**	**2.65**	**443**	123	**0.6869**	518	3.283	351	188	**0.8789**	**562.1**	**4.895**	**256**	253	0.7242	567.8	2.908	461
59	**0.7192**	**568.8**	**2.69**	**466**	124	**0.6784**	516	3.273	351	189	**0.8669**	**591.8**	**4.108**	**316**	254	**0.7253**			
60	**0.7135**	**565.5**	**2.61**	**455**	125	**0.6799**				190	**0.8669**	**617.2**	**3.609**	**374**	255				
61	**0.7160**	**563.5**	**2.73**	**436**	126	**0.6642**	**495**	**3.29**	**353.6**	191	**0.8620**	**638.4**	**3.20**	**440**	256	**0.8811**			
62	**0.6919**	**543.8**	**2.57**	**468**	127	**0.7150**	560.2	2.645	457	192	**0.8601**	**660.5**	**2.89**	**497**	257	**0.7512**			
63	**0.7262**	**573.5**	**2.82**	**455**	128	**0.7218**	563.2	2.655	457	193	**0.8801**	**630.3**	**3.732**	**370**	258	**0.9090**			
64	**0.7190**	**566.4**	**2.73**	**460**	129	**0.6504**	**465**	**3.40**	**294**	194	**0.8642**	**617**	**3.541**	**375**	259	**1.0960**[c]			
65	**0.7193**	**567.1**	**2.70**	**442**	130	**0.6580**	**470**	**3.42**	**318**	195	**0.8610**	**616.2**	**3.511**	**378**	260	**0.9639**	**684.9**	**3.952**	

[a] The experimental data are given in bold without any superscripts. [b] Experimental values are published, but have not been used in the database (see text). [c] Data determined by extrapolation (pseudo-experimental data), given with bold, but with a superscript. [d] Summary: d_4^{20}, total of 237 data, out of which 155 are experimental, 9 are nonexperimental (3.8%), and 72 are extrapolated with various correlations to the triple point, at saturation pressure or undercooled liquid state; T_c, total of 222 data, out of which 137 are experimental, 22 (9.9%) are nonexperimental, and 63 are extrapolated with ABCs; P_c, total of 219 data, out of which 132 are experimental, 21 (9.6%) are nonexperimental, and 66 are extrapolated with ABCs. V_c, total of 208 data, out of which 87 are experimental, 36 are nonexperimental (17.8%), and 85 are calculated with ABCs. Nonexperimental data: taken from refs 3 and 29. Experimental and extrapolated data given in boldface type.

extraction. A correlation based only on available experimental data is not likely to produce the desired values with a satisfactory precision for such compounds. The reason is that data for compounds with the needed high molecular masses and variation of the structural elements cannot be included in the development of the experiment-only based correlation.

The available experimental data are also limited with respect to the variation of their structural features (groups, bonds, etc.). Irrespective of the design of the description of the molecular structures, the predictive ability of a structure−property correlation depends on the distribution of the repeated structural features in the molecules selected for the database. The importance of this assumption can be illustrated by the fact that a correlation derived by whatever method with a database, consisting only of arenes, cannot be expected to predict adequately the properties of alkanes. Accordingly, the proportions between arenes, alkanes, alkenes, etc., determine the goodness of the prediction for each of these groups.

The above limitations have led to the idea of generating pseudo-experimental data as well as the use of a small amount of nonexperimental data. This gives an opportunity to increase the total amount of data, to improve the distribution of the repeated structural features in the database and extrapolate predictions beyond the boundaries of the existing experimental data with lower uncertainty.

Pseudo-experimental data were generated by extrapolation of the known experimental data for the critical properties of the members of the major homologous series (*n*-alkanes, 1-alkenes, alkylcyclopentanes, and alkylarenes) with asymptotic behavior correlations (ABCs) of the highest precision.[27,28] For the density of *n*-alkanes, which are gaseous or solid at 20 °C, we have employed the values listed in ref 15 (Table 2).

The consideration of data extrapolated within homologous series as pseudo-experimental is based on several observations:

•Extrapolation (i.e., of vapor pressure−temperature curves) is widely used in the "experimental" determination of the properties of unstable compounds.[15,25]

•Extrapolation is often used to check the consistency of new experimental data with the previously known data for the lower members of homologous series.[24,25]

•Extrapolation of data for a limited set of related compounds, such as a homologous series, typically achieves a precision which is better than the precision of any more general correlation (including ours) and/or experimental determinations for unstable substances.

•For many high molecular mass substances, which are unstable in the critical region, extrapolation is the only tool for estimation of their properties.

The process of extrapolation within a homologous series does not fill all gaps in our database. There remain areas in the molecular space, which can only be filled with estimated data. We postulate that to fill them with estimated data is preferable to leaving them void, so long as the number of such data is not excessive. We have compiled the necessary nonexperimental data from recognized sources.[3,29]

The nonexperimental values allowed in our database (Table 2) are less than 10% of all T_c and P_c data and 3.8% of the density data. The nonexperimental data for V_c are a higher proportion (17.8%), because of the very small amount of experimental data for that property. In all cases, the selection of the particular molecules is based strictly on structural considerations. As will be shown hereunder, the use of nonexperimental data does not influence significantly the prediction of the rest of the data.

The critical properties and densities in the database are thus for the most part recommended experimental data, or data extrapolated from experimental values with the highest available precision. Their molecular features represent the widest practical range. The number of carbon atoms spans from 2 to 60 for the *n*-alkanes, from 3 to 40 for the series finishing with β-carotene, and up to 30 for the rest of the homologous series. A considerable number of industrially important hydrocarbons, which do not have published values for at least one of the critical parameters, are also included, so that the final objective of the present work can be satisfied.

2.2. Structure Database. Two types of descriptors, carbon atom/bond descriptors and molecular energy descriptors, have been employed as independent variables in the present investigation.

Carbon atom descriptors of various levels of sophistication have been used. The least sophisticated atom counts are Joback's first level descriptors,[3] used in our previous work.[2] These are complemented by three relevant additional descriptors for arenes suggested for the estimation of normal boiling points by Stein and Brown.[30] The descriptors proposed by Ambrose[3] and Somayajulu[31] have also been tested. The most sophisticated descriptors of carbon atoms and their bonds used are graph topological indices, derived from the adjacency and distance matrixes of the hydrocarbon structures. As in our earlier work,[2] only a limited number of the most popular indices, calculated as suggested in refs 6 and 32 were tried in the present work. The molecular energy descriptors obtained through molecular mechanics simulation[33] were the same as in our work on boiling points.[2]

The total number of descriptors, including the topological ones, as well as molecular mass, amounted to 75. Electronic descriptors were deliberately avoided in order to be consistent with our intention to use a relatively simple and comprehensible representation of molecular features.

2.3. Modeling. A "stepwise" multiple regression procedure[34] was employed to select the most influential structural features and determine their optimal number for the representation by a particular model. The procedure is applied to equations in which the dependent variable is represented by models of the form

$$Y_j = b_0 + \sum_{i=1}^{N} (b_i X_{ij}) \qquad (1)$$

where Y_j is the appropriate property of the compound j, b_0 is the intercept term, and b_i is the coefficient for a descriptor X_{ij}. A linear contribution of the structural descriptors was adopted for all variables, except for one, nonlinear, exponential independent variable. A limit of pairwise correlations was set as ±0.85 as was the case in our earlier work.[2] The typical algorithm used for obtaining the best models with an increasing number of independent variables has been described in detail elsewhere.[35]

Accordingly, the compounds in the database were divided into a learning set (lset), a prediction set (pset), and a comparison set (Table 1). The compounds in the learning set were used for the development of the models. The 18 compounds in the prediction set (shown also in Table 6) were not used in the derivation of any of the models.

The aim of our representation of the data in Table 2 was to achieve a quality of prediction of these values, close to that of the best, but much more complicated representations for each property achieved by other authors. To make such an assessment, we compare the goodness of fit of our

Table 3. Model for d_4^{20} [a]

independent variables	b_i	standard error (±)	Fisher's criterion for removal from model
$C_{tot}^{-0.90}$	−0.681800	0.013622	5147.2
DBC	0.010018	0.000615	265.2
$N_{CH_2}^{c}$	0.016728	0.000536	974.4
DBC/*Chi0*	0.040096	0.004159	93.0
$N_{CH}^{a}/Chi0$	−0.072527	0.005945	148.8
$N_C^{a}/Chi0$	−0.173974	0.011958	211.7
C_{tot}/V_{vdw}	6.685916	0.479571	194.4
J	0.052551	0.003026	301.7
b_0	0.522669	0.013622	

[a] Statistical data: $N = 219$; SD = 0.0064; coefficient of multiple correlation = 0.997; calculated Fisher's ratio = 5476.0; maximum AE = 0.0195; minimum AE = −0.0231; ARE = 0.82%; maximum RE = 2.65%; minimum RE = − 3.36%; Z = 0.0019. Predictions with the simplified model (eq 5): $N = 219$; SD = 0.0066; maximum AE = 0.0196; minimum AE = −0.0252; ARE = 0.84%; maximum RE = 2.88%; minimum RE = − 2.71%.

Table 4. Model for the Critical Temperature, T_c

independent variables[a]	(b_i/K)	standard error (±)	Fisher's criterion for removal from model
$C_{tot}^{1.50}$	1.009509	0.036099	782.0
T_{bmix}	1.267090	0.013738	8506.4
Chi0	−25.908528	1.149981	507.6
Chi1	18.995625	1.548637	150.5
$N_{CH_2}^{a}$	−5.448293	0.192488	801.2
N_{CH}^{a}	−3.853217	0.633129	37.0
J/C_{tot}	−133.402832	11.609532	132.0
J	39.297725	1.352483	844.0
b_0	121.730214	6.550642	

[a] Statistical data: N = 204; SD = ±4.12 K; coefficient of multiple correlation = 0.999; calculated Fisher's ratio; 25859.9; maximum AE = 13.96 K; minimum AE = − 17.95 K; ARE = 0.61%; maximum RE = 2.12%; minimum RE = − 2.40%; Z = 1.14 K.

Table 5. Model for the Critical Pressure, P_c [a]

independent variables	(b_i/MPa)	standard error (±)	Fisher's criterion for removal from model
$C_{tot}^{-0.38}$	13.563667	0.688180	388.5
d_4^{20}(mix)	4.895707	0.194406	634.2
d_4^{20}(mix)/*Chi0*	6.487280	0.419904	238.7
$N_{DC}^{a}/Chi0$	−0.370022	0.070495	27.5
IWDM	0.070433	0.032886	4.6
E_{tot}	0.004025	0.000542	52.9
E_{tor}	−0.004813	0.000895	28.9
J	0.209842	0.013509	241.3
b_0	−8.573061	0.502136	

[a] Statistical data: $N = 201$; SD = ±0.0567 MPa, coefficient of multiple correlation = 0.999; calculated Fisher's ratio = 10002.7; maximum AE = 0.302 MPa; minimum AE = −0.206 MPa; ARE = 2.33%; maximum RE = 10.57%; minimum RE = − 6.74%; Z = 0.0334 MPa. Simplified model for P_c (eq 6): N = 201; SD = ±0.0588 MPa, coefficient of multiple correlation = 0.999; calculated Fisher's ratio = 9279.8; maximum AE = 0.2452 MPa; minimum AE = −0.190 MPa; ARE = 3.37%; maximum RE = 10.15%; minimum RE = −34.37%.

models with the representation of the same comparison set of data for each property generated by other authors.

The original comparison set comprises 75 hydrocarbons studied in ref 18 that also occur in our database with experimental data for T_c and P_c. The pset and the comparison set are chosen to be typical of other published work and contain hydrocarbons covering the center of the range of molecular masses, so that the comparison is meaningful.

Nonexperimental values had to be excluded from the original comparison set, for the V_c models. However, the resulting reduced set was very small and had to be complemented with 12 additional compounds (nos. 14 to 16, 146, 147, 218, 232, 236, 239, 242, 243, 247 from Table 1 and Table 2), for which experimental information was available.

Table 6. Model for the Critical Volume, V_c [a]

independent variables	(b_i/cm^3 mol^{-1})	standard error (±)	Fisher's criterion for removal from model
$V_{vdw}^{1.20}$	0.607051	0.002996	10000.0
N_{CH}^{a}/J	−43.533274	5.750584	57.3
W	−0.017811	0.000403	1953.2
N_C^{a}	−42.576755	5.061075	70.8
Chiz3/J	−669.265392	33.830456	391.4
N_{CH_3}	9.258427	1.744670	28.2
N_{CH}^{aa}/J	15.787892	0.791447	397.9
N_C^{aaa}/J	−8.778870	3.002696	8.5
b_0	10.620489	3.455389	

[a] Statistical data: N = 190; SD = ±13.8 cm^3 mol^{-1}; coefficient of multiple correlation = 0.999; calculated Fisher's ratio = 44232.5 maximum AE = 32.8 cm^3 mol^{-1}; minimum AE = −32.6 cm^3 mol^{-1}; ARE = 2.95%; maximum RE = 11.0%; minimum RE = −10.7%; Z = 5.6 cm^3 mol^{-1}. Predictions with the simplified model (eq 7): N = 190; SD = ±15.80 cm^3 mol^{-1}; maximum AE = 50.0 cm^3 mol^{-1}; minimum AE = −37.0 cm^3 mol^{-1}; ARE = 3.63%; maximum RE = 13.51%; minimum RE = −12.56%.

In the case for the density model there is an additional difficulty. The best available model from elsewhere[7] is for densities at 25 °C, but we have chosen to predict densities at 20 °C, owing to the greater volume of experimental data at that temperature.

For this reason, our comparison must be slightly less direct, so that we have corrected the densities we have predicted at 20 °C to densities at 25 °C using the standard formulas widely used by the petroleum processing industry.[36] As for the rest of the properties, only experimental data were used for the comparison of the density models.

3. Results and Discussion

The coefficients of eq 1 established for each of the various properties, are presented in Tables 3−6: for d_4^{20} in Table 3; for the critical temperature, T_c, in Table 4; for the critical pressure, P_c, in Table 5, and for the critical volume, V_c, in Table 6. Table 7 shows the predicted and experimental values for all properties of the compounds in the prediction set.

We have tried to select the best models for the comparison with the correlations developed in this work. For T_c, and P_c the results obtained with the better models of the authors, derived with neural networks were chosen.[18] The comparison for the density and V_c was performed with the latest group contribution methods.[7,8] Figure 1 compares the absolute errors of the models for d_4^{25} with those from a recent work[7] for the compounds in the comparison set. Figures 2−4 provide the same comparison for T_c, P_c, and V_c.

The representations for the four properties, derived in this work permit us to generate values for the compounds in the database for which no experimental values exist and for which no estimates have hitherto been made. Table 8 lists the newly estimated values for a set of industrially important hydrocarbons.

Finally, an example which illustrates the practical application of the correlations proposed in this work, for the estimation of d_4^{20} and the critical properties of 1,2,3,4-tetrahydronaphthalene, is presented in the appendix.

3.1. Predictive Quality of the Models. The predictive quality of the models is evaluated through their ability to predict the known values, used in their derivation (the lset) as well as known values not used in their derivation (the

Table 7. Published Data and Predicted Properties of the Compounds in the pset[a]

		d_4^{20}		T_c/K		P_c/MPa		V_c/cm^3 mol^{-1}	
no.	name	publ	pred	publ.	pred	publ	pred	publ	pred
5	*n*-hexane	0.6593	0.6568	507.6	504.2	3.025	3.044	368	384.7
17	*n*-octadecane	0.7828	0.7809	747.0	742.7	1.29	1.229	1189	1167.3
50	2,2,3-trimethylbutane	0.6990	0.6882	531.1	529.9	2.95	2.959	398	392.6
62	2,2,4-trimethylpentane	0.6920	0.7031	543.8	545.9	2.57	2.518	468	454.2
71	2,2,3,3-tetramethylpentane	0.7566	0.7385	607.5	605.2	2.74	2.754	478	479.4
88	1-butene	0.5945	0.5928	419.5	422.0	4.02	4.126	240	234.6
91	1-heptene	0.6970	0.6981	537.3	541.0	2.92	2.9	409	398.4
92	1-octene	0.7149	0.7157	567.0	570.9	2.68	2.628	468	458.7
94	1-decene	0.7408	0.7415	617.0	621.3	2.22	2.205	584	581.0
115	1,3-butadiene	0.6211	0.6256	425.0	434.0	4.32	4.54	221	190.9
129	2-methyl-1-butene	0.6504	0.6566	465.0	469.8	3.40	3.581	294	294.0
144	cyclopentane	0.7454	0.7290	511.7	514.5	4.51	4.379	259	256.2
148	methylcyclohexane	0.7694	0.7812	572.1	574.3	3.48	3.307	369	365.8
180	*trans*-1,3-dimethylcyclohexane	0.7847	0.7855	598.0	598.1	2.938	2.996	408.5	427.8
184	cyclohexene	0.8102	0.7979	560.4	558.0	4.35	4.196	292	277.8
193	*m*-xylene	0.8801	0.8804	630.3	626.2	3.732	3.698	370	378.6
196	*m*-ethyltoluene	0.8645	0.8719	637.1	640.6	2.837	3.194	490	450.3
234	*p*-diethylbenzene	0.8620	0.8691	657.9	661.6	2.807	2.906	480	505.6

[a] Statistical data for the prediction of the pset follow. Model d_4^{20}: SD = 0.0087; maximum AE = 0.0118; minimum AE = −0.0181; ARE = 1.17%; maximum RE = 1.60%; minimum RE = −2.39%. Simplified model (eq 5): SD = 0.0082; maximum AE = 0.0120; minimum AE = −0.0180; ARE = 1.09%; maximum RE = 1.40%; minimum RE = −2.38%. Model T_c: SD = 3.64 K; maximum AE = 9.0 K; minimum AE = − 4.3 K; ARE = 0.71%; maximum RE = 2.12%; minimum RE = −0.67%. Model P_c: SD = 0.135 MPa; maximum AE = 0.357 MPa; minimum AE = −0.172 MPa; ARE = 4.25%; maximum RE = 12.57%; minimum RE = −4.95%. Simplified model (eq 6): Standard deviation: 0.121 MPa; maximum AE = 0.330 MPa.; minimum AE = −0.153 MPa; ARE = 3.85%; maximum RE = 11.64%; minimum RE = −4.40%. Model V_c: SD = 16.4 cm^3 mol^{-1}; maximum AE = 25.6 cm^3 mol^{-1}; minimum AE = −39.4 cm^3 mol^{-1}; ARE = 4.53%; maximum RE = 5.33%; minimum RE = − 13.62%. Simplified model (eq 7): SD = 17.6 cm^3 mol^{-1}; maximum AE = 27.5 cm^3 mol^{-1}; minimum AE = −36.7 cm^3 mol^{-1}; ARE = 4.97%; maximum RE = 5.73%; minimum RE = −12.26%.

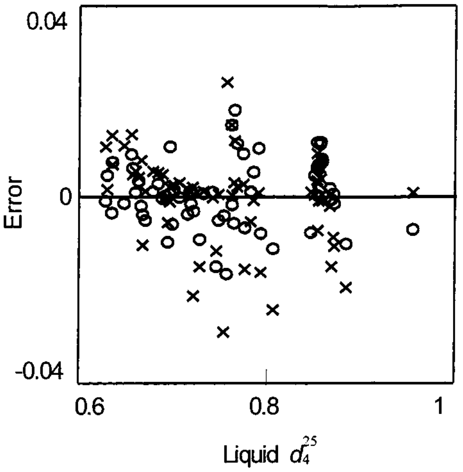

Figure 1. Absolute errors in the estimation of d_4^{25} of the compounds in the comparison set: ×, ref 7; ○, this work. Statistical data for d_4^{25} models. Reference 7: N = 64; standard deviation, 0.0103; maximum AE = 0.0259; minimum AE = −0.0314; ARE = 1.38%; maximum RE = 3.44%; minimum RE = −4.19%. This work: N = 64; standard deviation, 0.0077; maximum AE = 0.0195; minimum AE = −0.0181; ARE = 1.01%; maximum RE = 2.56%; minimum RE = −2.40%. Simplified model (eq 5): N = 64; standard deviation, 0.0079; maximum AE = 0.0196; minimum AE = −0.0180; ARE = 1.03%; maximum RE = 2.58%; minimum RE = −2.39%.

pset). The prediction of the latter is particularly indicative because it characterizes the suitability of the models for interpolation to generate unknown values and the quality of the predictions within the whole range of compounds. In view of the above, the relations between the structures of the compounds in the lset and the pset are important. As seen from Table 1 and Table 7, the pset includes compounds with structural elements, which occur frequently in the lset, such as *n*-alkanes, 1-alkenes, cyclopentane, etc. But it also contains compounds such as 1, 3-butadiene and cyclohexene, which are exemplified in the lset by only a few similar structures. A model of good predictive quality should predict the properties of the latter compounds with an error comparable to its prediction of the former.

A variety of statistical criteria[34] have been applied in the present work; but only the most familiar statistical data are summarized and interpreted here in the interests of brevity.

The statistical data for the models include

standard deviation,

$$\mathrm{SD} = \sqrt{\sum (Y_{\mathrm{est}} - Y_{\mathrm{exp}})^2/(N-1)} \quad (2)$$

average relative error,

$$\mathrm{ARE}\ (\%) = \frac{1}{N}\sum |(Y_{\mathrm{est}} - Y_{\mathrm{exp}})/Y_{\mathrm{exp}}| \times 100 \quad (3)$$

where N is the number of data points, Y_{est} is the estimated value of the property Y, and Y_{exp} is the experimental value of the property Y. Maximum and minimum errors are also reported. The values for Fisher's criterion,[36] for the stepwise selection of independent variables in the model, and Fisher's ratio, defining together with the coefficient of multiple correlation the adequacy of the statistical description, are also given.

The mean-square residual Z, defined as

$$Z = \sqrt{\frac{\sum (Y_{i,\mathrm{all}} - Y_{i,\mathrm{lset}})^2}{N}} \quad (4)$$

where $Y_{i,\mathrm{all}}$ is the property value of the compound estimated with a regression derived from all compounds in the lset and pset, $Y_{i,\mathrm{lset}}$ is the property value of the compound estimated only from the lset, and N is the number of data points in the pset, was used as a measure for the stability of the prediction, as suggested in refs 7 and 8.

Model for the Liquid Density, d_4^{20}. Table 3 shows that the model for d_4^{20} is of high statistical quality. The high precision achieved is important for our subsequent work

Table 8. d_4^{20} and Critical Properties of Industrially Important Hydrocarbons for Which New Estimated Values Are Recommended

			T_c	P_c	V_c
no.	name	d_4^{20}	K	MPa	$cm^3 mol^{-1}$
79	2,5-dimethyldecane	0.7502	642.2	1.878	728
80	2,5,-dimethyldodecane	0.7646	672.5	1.633	854
81	2,6,10-trimethyldodecane	0.7714^a	693.7	1.573	900
82	2,6,10-trimethyltetradecane	0.7806^a	721.1	1.404	1035
83	pristane	0.7845^a	740.0	1.348	1154
84	phytane	0.7935^a	768.8	1.289	1209
85	squalane	0.8087^a	849.0	0.806	1885
86	lycopane	0.8277	864.2	0.677	2533
122	2-ethyl-1-butene	0.6859^a	512.6	3.285	350
125	2-methyl-1-pentene	0.6809^a	506.5	3.18	352
133	3,3-dimethyl-1-butene	0.6537^a	490.2^a	3.151^a	316
135	3-methyl-1-pentene	0.6678^a	501.1	3.214	323
137	limonene	0.8422^a	650^a	2.75^a	483
138	α-pinene	0.8582^a	632^a	2.76^a	443
139	squalene	0.8562^a	886.1	0.936	1777
140	lycopene	0.8958	962.6	0.897	2190
141	*β*-carotene	0.9720	1041.9	1.021	1890
177	*cis*-1,2-dimethylcyclohexane	0.7966^a	606.1^a	2.938^a	428
178	*trans*-1,2-dimethylcyclohexane	0.7764^a	596.2^a	2.938^a	432
179	*cis*-1,3-dimethylcyclohexane	0.7664^a	591.1^a	2.938^a	428
180	*trans*-1,3-dimethylcyclohexane	0.7855^a	598^a	2.938^a	428
181	*cis*-1,4-dimethylcyclohexane	0.7835^a	598^a	2.978	432
182	*trans*-1,4-dimethylcyclohexane	0.7634^a	587.7^a	2.938^a	440
185	1,3-cyclohexadiene	0.8406^a	555.6	4.473	248
186	5-methyl-1,3-cyclopentadiene	0.8008	542.4	4.267	261
187	1,3-cyclopentadiene	0.8021^a	506.3	5.099	198
217	styrene	0.9082^a	635.2^a	3.87^a	338
226	*tert*-butylbenzene	0.8680^a	660^a	2.969^a	468
227	*p*-cymene	0.8579^a	652^a	2.8^a	486
228	*m*-diethylbenzene	0.8650^a	659.5	2.855	506
229	*o*-diethylbenzene	0.8810^a	664.2	3.038	499
231	*m*-diisopropylbenzene	0.8559^a	676.3	2.392	581
232	diphenylmethane	1.0078^a	760^a	3.088	563^a
233	*sec*-butylbenzene	0.8629^a	664^a	2.949^a	477
235	*p*-diisopropylbenzene	0.8568^a	683.6	2.421	581
236	diphenyl	1.0378	773^a	3.38^a	497^a
237	1,1-diphenylethane	0.9997^a	778.0	2.893	604
238	1,2-diphenylethane	1.0019	781.3	2.855	633
240	anthracene	1.0886	883^a	3.42	492
241	phenanthrene	1.0920	869^a	3.489	492
242	*m*-terphenyl	1.1224	883^a	2.983	724^a
243	*p*-terphenyl	1.1136	908^a	2.99^a	729^a
244	1,2-benzo[*a*]pyrene	1.1865	1044.4	3.404	634
245	pyrene	1.1372	947.5	3.377	517
246	chrysene	1.1535	999.3	3.057	593
247	*o*-terphenyl	1.1224	908.8	2.99^a	731^a
248	triphenylmethane	1.1067	887.4	2.86	806
249	acenaphthylene	1.0680	802.2	3.925	421
250	acenaphthene	1.0400	804	3.625	436
251	1,1,2,2-tetraphenylethane	1.1738	961.9	2.653	1102
252	4-methyloctane	0.7202^a	589.0	2.355	553
254	2-ethylhexene	0.7253^a	573.4	2.647	463
255	adamantane	0.8656	693.9	3.064	425
256	1,5-cyclooctadiene	0.8811^a	641.2	3.711	343
257	2,5-methyl-1,5-hexadiene	0.7512^a	573.3	2.774	455
258	*cis*-1-propenylbenzene	0.9090^a	658.2	3.569	405
259	1-phenylnaphthalene	1.0960^a	882.1	3.444	595
260	indane	0.9639^a	684.9^a	3.953^a	371

[a] Published values.

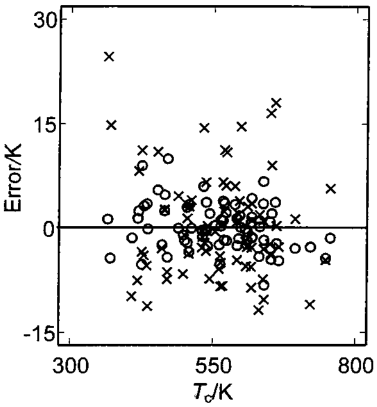

Figure 2. Absolute errors for the prediction of the critical temperatures of the compounds in the comparison set: ×, ref 18; ○, this work. Statistical data for T_c models. Reference 18: N = 75; standard deviation, 7.56 K; maximum AE = 24.6 K; minimum AE = −11.8 K; ARE = 1.54%; maximum RE = 6.94%; minimum RE = −2.57%. This work: N =75; standard deviation, 3.36 K; maximum AE = 10.0 K; minimum AE = −8.2 K; ARE = 0.65%; maximum RE = 2.13%; minimum RE = −1.28%.

because the error in the estimation of densities is carried through to the model for P_c. The residual, Z, is several times lower than the standard deviation of the model, confirming the reliability of the prediction. The standard deviation in the prediction of d_4^{20} for the compounds in the pset is somewhat higher (Table 7) than that for the lset. All deviations, however, are normally distributed within the error bounds of the lset. The comparison of the predictions for the two groups of compounds in the pset, defined above, shows (Table 7) that the error for cyclohexene is higher than that, which is typical for the 1-alkenes. Even so, the errors for cyclohexene and 1,3 butadiene are smaller than the largest error for compounds with structural elements, which occur frequently in the lset (i.e., cyclopentane). Using the logic presented earlier, this provides support for our selected representation.

As seen from Figure 1 the present model for the liquid density compares well with the recent two-level group contribution method,[7] although it is considerably simpler.

Model for the Critical Temperature, T_c. The statistical quality of this model is also high (Table 4). The error distribution for the prediction values for the learning set is statistically normal. The small value of the residual, Z, confirms the reliability of the prediction. On average, the errors for the pset (Table 7) are smaller than for the lset. The critical temperatures of 1,3-butadiene and cyclohexene are comparable with that of the compounds which have a better structural representation in the lset. Again, this result supports the chosen representation.

The comparison of the model with the model of Turner et al.,[18] for the compounds presented in Figure 2, is favorable.

Model for the Critical Pressure, P_c. The critical pressure is predicted with higher errors (especially for the pset) than for the remainder of the properties targeted in this study. The residual, Z, is only half of the error of the regression, derived with the lset. As can be seen from Table 7, the largest errors in the pset come from one particular compound—*m*-ethyltoluene. The critical pressures of 1,3-butadiene and cyclohexene are predicted with relatively high errors, but they are comparable to those of, for example, cyclopentane and methylcyclopentane. It should be noted that the critical pressure is usually more difficult to predict.[3] When compared with the results of the method of Turner et al.,[18] the predictions of the model developed in this work are of similar quality (Figure 3).

Model for the Critical Volume, V_c. The critical volumes of the compounds in the lset are predicted with a low level of uncertainty, as seen from the statistical parameters in Table 6. The maximum and minimum relative errors exceed 10%, although as seen from the average relative error these values are exceptional. The errors of the pset (Table 7) are within the same boundaries, though the average relative error is somewhat higher. The residual, Z, is less than half of the SD of the model. The comparison with the most recent group contribution method[8] shows absence of systematic deviations and a good predictive ability for the model developed in this work (Figure 4).

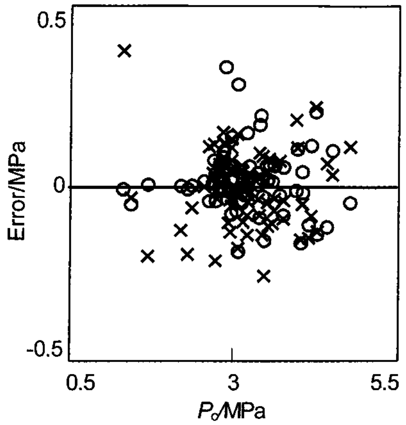

Figure 3. Absolute errors for the prediction of the critical pressures of the compounds in the comparison set: ×, ref 18; ○, this work. Statistical data for P_c models. Reference 18: $N = 75$; standard deviation, 0.120 MPa; maximum AE = 0.406 MPa; minimum AE = − 0.280 MPa; ARE = 5.58%; maximum RE = 35.02%; minimum RE = −13.94%. This work: $N = 75$; standard deviation, 0.096 MPa; maximum AE = 0.357 MPa; minimum AE = − 0.206 atm; ARE = 2.99%; maximum RE = 12.57%; minimum RE = −6.74%. Simplified model (eq 6): $N = 75$; standard deviation, 0.090 MPa; maximum AE = 0.330 MPa; minimum AE = − 0.190 atm; ARE = 2.79%; maximum RE = 11.63%; minimum RE = −6.23%.

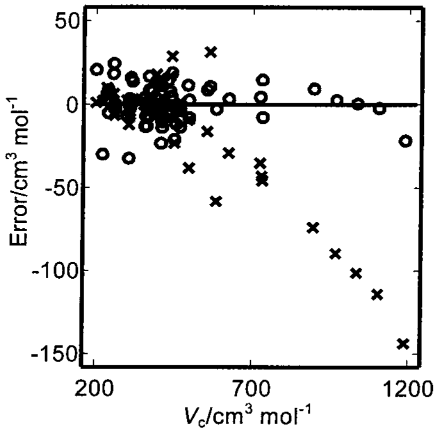

Figure 4. Absolute errors in the estimation of the critical volumes of compounds in the comparison set: ×, ref 8; ○, this work. Statistical data for V_c models. Reference 8: $N = 65$; standard deviation, 31.8 cm^3 mol^{-1}; maximum AE = 31.1 cm^3 mol^{-1}; minimum AE = −143.9 cm^3 mol^{-1}; ARE = 3.90%; maximum RE = 6.47%; minimum RE = −12.10%. This work: $N = 65$; standard deviation, 12.1 cm^3 mol^{-1}; maximum AE = 24.4 cm^3 mol^{-1}; minimum AE = −32.5 cm^3 mol^{-1}; ARE = 3.85%; maximum RE = 10.35%; minimum RE = −13.63%. Simplified model (eq 7): $N = 65$; standard deviation, 14.1 cm^3 mol^{-1}; maximum AE = 33.1 cm^3 mol^{-1}; minimum AE = −35.3 cm^3 mol^{-1}; ARE = 4.46%; maximum RE = 10.77%; minimum RE = −12.27%.

3.2. Extrapolation Ability of the Models. Statistical models of the type developed in the present work should, in principle, be used for interpolation within the boundaries defined by the data involved in their derivation. However, models involving chemical compounds, which are closely related through repeating molecular structures, usually allow also for extrapolations for a short range beyond the boundaries of the input data.[13,14] Such extrapolations must be done carefully for the purpose of comparison when it is impossible to obtain the necessary data from other sources. Extrapolated data for the critical properties might be needed, for instance, for the description of experimental data on phase equilibria.[38] In that case, they may be verified indirectly by the proper prediction of experimental phase equilibrium data.

Table 8 lists estimated values of the properties of 58 hydrocarbons, for some of which the particular data have not been published before.

3.3. Simplified Models. One of the great advantages of the earlier group contribution correlations (e.g., Ambrose, Somayajulu, Joback) is that they do not require the use of specialized software, so that properties can be estimated by simple calculations.

The model for the critical temperature, developed in this work, is represented by eq 1 with the parameters of Table 4. Since it needs only atom counts and topological indices, the critical temperatures are easily calculated by hand. Our other three models use descriptors that have to be estimated with the program for molecular mechanics simulation.[33] The calculations with this program[33] are performed automatically once the structural formula of the compound is introduced, and its total molecular energy is minimized. However, to increase the applicability of our work, we have developed models that make use of descriptors that are easily calculated by hand. They are presented below.

Simplified Model for Density.

$$d_4^{20} = 0.187725 - 0.693402\, C_{\mathrm{tot}}^{-0.84} + 0.008862\,(\mathrm{DBC}) + 0.014754 N_{\mathrm{CH}_2}^{\ \mathrm{c}} + 0.052909\left(\frac{\mathrm{DBC}}{Chi0}\right) - {-0.080005}\left(\frac{N_{\mathrm{CH}}^{\ \mathrm{a}}}{Chi0}\right) - 0.192955\left(\frac{N_{\mathrm{C}}^{\ \mathrm{a}}}{Chi0}\right) + 7.329044\left(\frac{C_{\mathrm{tot}}}{M}\right) + 0.073820J + + 0.028720(\mathrm{MI}) \quad (5)$$

Simplified Model for P$_c$.

$$\frac{P_{\mathrm{c}}}{\mathrm{MPa}} = -139.425429 + 198.294544 C_{\mathrm{tot}}^{-0.34} + 53.174751 d(\mathrm{mix}) + 60.072056\left(\frac{d(\mathrm{mix})}{Chi0}\right) - {-4.712113}\left(\frac{N_{\mathrm{DC}}^{\ \mathrm{c}}}{Chi0}\right) + 2.171253J - 4.475980\left(\frac{N_{\mathrm{DC}}^{\ \mathrm{a}}}{Chi0}\right) + 4.062432(\mathrm{IWDM}) - 0.302460 N_{\mathrm{CH}}^{\ \mathrm{c}} \quad (6)$$

Simplified Model for V$_c$.

$$\frac{V_{\mathrm{c}}}{\mathrm{cm}^3\ \mathrm{mol}^{-1}} = 59.541346 + 2.186624 M^{1.14} - 27.948690 N_{\mathrm{DCH}_2}^{\ \mathrm{a}} - 12.106196 N_{\mathrm{CG}} - 20.964877 N_{\mathrm{Ph}}^{\ \mathrm{S}} - 0.014900W - 68.433164(\mathrm{MI}) - 6.299214(\mathrm{IEDM}) - 601.415699\left(\frac{Chiz3}{J}\right) \quad (7)$$

The errors involved in the prediction of each property with the simplified models are presented within the respective tables. The simplified models allow for the estimation of the respective properties with a precision comparable to that of the models, requiring specialized software for calculation of the descriptors. The calculations of the values of the descriptors, and of the topological indices in particular,[32] used in the simplified models can be done with a hand calculator. The fact that considerable precision can be achieved even with the simplified models confirms the usefulness of the methodological contributions suggested in the present work.

It should be emphasized that our achievements with simple methods for property prediction for a limited set of compounds does not diminish the utility of the earlier methods[7,8,18] that have a much wider scope and use only molecular structure parameters.

The abilities of the present methods to describe polar compounds, especially when hydrogen bonds are present, still have to be evaluated.

Conclusions

The aim of the correlations, developed in the present study, is to provide a tool for chemical engineers to use in the design and simulation of industrial processes and installations, especially when phase equilibria are involved. The models to be employed should, therefore be ideally characterized by simple rules for representing molecular features and for restoring the whole molecule, while predicting the properties with sufficient precision for practical purposes. Tables 1−8, Figures 1−4, and the example shown in the appendix illustrate the achievement of the aims of the present work:

•First, the use of predesigned databases with limited functionality of the selected compounds, expanded range of the property data and improved structural variance has allowed the development of relatively simple and comprehensive models.

•Second, the present correlations for hydrocarbons compare favorably with both simple and sophisticated atom count methods. They do not require the representation of the molecule by means of elaborate group contributions. On the other hand, despite their simplicity, their precision is comparable to that of methods, which need a knowledge of the electronic structure of the molecules and quantum-chemical methodology.

•Finally, the relatively low number of descriptors, the increased amount of data, and the improved variance of the structures used in the development of the present correlations provide sound statistical foundations for their predictive ability especially for high molecular mass hydrocarbons.

The high precision achieved by the present correlations opens up possibilities for future work in a direction of high relevance to chemical engineering studies. In particular, it becomes possible to investigate the influence of uncertainties in calculated thermophysical parameters on the final solution of designs and simulations of existing applications, in which the effects of small changes in the properties are important.[39]

Appendix: An Example for Estimation of the Properties of 1,2,3,4-Tetrahydronaphthalene (Compound No. 225) from Its Molecular Structure with the New Correlations

The properties of any compound are calculated with the formula

$$Y_j = b_0 + \sum_{i=1}^{N} (b_i X_{ij}) \qquad (1)$$

where Y_j is the calculated property of the compound j, b_0 is the intercept term, and b_i is the coefficient for descriptor X_{ij}. The values of the intercept terms and the coefficients are taken from the tables for the respective models (Table 1 for d_4^{20}, Table 2 for T_c, Table 3 for P_c, and Table 4 for V_c). The values of the descriptors X_{ij} for the particular compound are taken from the descriptors' database.

Table 9. Molecular Energy Descriptors of 1,2,3,4-Tetrahydronaphthalene Participating in the Correlations

row	name	E_{tot} kJ mol^{-1}	E_{tor} kJ mol^{-1}	V_{VdW} Å^3
225	1,2,3,4-tetrahydronaphthalene	52.435	23.57	234

Table 10. Molecular Mass and Atom Counts of 1,2,3,4-tetrahydronaphthalene

descriptor	value	descriptor	value
M	132.206	$N_{CH_2}{}^{a}$	0
C_{tot}	10	$N_{DCH_2}{}^{a}$	0
N_{CH_3}	0	$N_{DC}{}^{a}$	0
DBC	6	$N_{DC}{}^{c}$	2
$N_{CH_2}{}^{c}$	4	N_{CG}	0
$N_{CH}{}^{a}$	0	$N_{Ph}{}^{S}$	0
$N_{CH}{}^{c}$	0	$N_{CH}{}^{aa}$	4
$N_{C}{}^{a}$	0	$N_{C}{}^{aaa}$	0

Table 11. Topological Indices of 1,2,3,4-Tetrahydronaphthalene Participating in the Correlations

descriptor	value	descriptor	value
C_{HI0}	5.619	W	158.02
C_{HI1}	3.449	IWDM	5.335
C_{HIZ3}	0.1172	IEDM	4.189
J	1.33	MI	2

The necessary excerpts from the descriptors' database for 1,2,3,4-tetrahydronaphthalene are given below.

The relevant molecular energy descriptors of 1,2,3,4-tetrahydronaphthalene, calculated with simulated molecular mechanics software are given in Table 9.

The atom count excerpts from the descriptors' database are given in Table 10.

The values for the topological indices of 1,2,3,4-tetrahydronaphthalene calculated with the formulas described in sufficient detail and examples in refs 6 and 32 are given in Table 11.

The normal boiling point of 1,2,3,4-tetrahydronaphthalene, needed for the calculation of the critical temperature, is $T_{b\,mix} = 480.77$ K [2]. Substitution of all needed values for 1,2,3,4-tetrahydronaphthalene in the respective correlations developed in this work (Tables 3−6) gives

$$d_4^{20} = 0.9623; \quad T_c = 717.29\ \text{K}; \quad P_c = 3.703\ \text{MPa}; \quad V_c = 419.3\ \text{cm}^3\ \text{mol}^{-1}$$

The calculation of the same properties with the simplified models (eqs 5−7) gives

$$d_4^{20} = 0.9662; \quad P_c = 3.654\ \text{MPa}; \quad V_c = 414.4\ \text{cm}^3\ \text{mol}^{-1}$$

The experimental values for the properties of 1,2,3,4-tetrahydronaphthalene (Table 2) are

$$d_4^{20} = 0.9702; \quad T_c = 720.0\ \text{K}; \quad P_c = 3.648\ \text{MPa}; \quad V_c = 408\ \text{cm}^3\ \text{mol}^{-1}$$

Nomenclature

AE = absolute error
b_0 = regression intercept of eq 1
b_i = regression coefficients of eq 1

Chi0 = connectivity topological index of order zero (Randic): single atoms [6, 32]

Chi1 = connectivity topological index of order one (Randic): duplets of atoms[6,32]

Chi3 = connectivity topological index of order three (Randic): quadruples of atoms[6,32]

Chiz3 = path connectivity index of order three (Randic): quadruples of atoms, calculated with atom numbers in the diagonal of the adjacency matrix[6,32]

C_{tot} = total number of carbon atoms in the molecule

DBC = total number of carbon atoms in double bonds in rings

$d_4{}^{20}$ = relative liquid density at 20 °C

$d_4{}^{20}$(mix) = relative liquid density with published data and missing data, estimated with the correlation developed in the present work (Table 1)

E_{tot} = total energy of the molecule, calculated with molecular mechanics simulation[32]

E_{tor} = torsional energy of the molecule, calculated with molecular mechanics simulation[32]

IEDM = mean information content on the distribution of distances in the molecule[32]

IWDM = mean information on the realized distances in the molecule[32]

J = average distance sum index (Balaban's index)[32]

lset = learning set of compounds used for derivation of a model

M = molecular mass

MI = cyclomatic number[32]

N = number of the data points

$N_{CH_2}{}^{a}$ = number of carbon atoms in aliphatic CH_2 groups

$N_{CH_2}{}^{c}$ = number of carbon atoms in $CH_2{=}CH_2$ double bonds in rings

N_{CH_3} = number of carbon atoms in CH_3 groups

$N_{CH}{}^{a}$ = number of carbon atoms in aliphatic CH groups

$N_{CH}{}^{c}$ = number of carbon atoms in CH groups in rings.

$N_{C}{}^{aach}$ = number of carbon atoms in CH atoms, engaged in two aromatic bonds[30]

$N_{C}{}^{a}$ = number of carbon atoms in aliphatic C groups

$N_{C}{}^{aaa}$ = number of carbon atoms, engaged in three aromatic bonds[30]

$N_{DC}{}^{a}$ = number of carbon atoms in C= aliphatic double bonds

$N_{DC}{}^{c}$ number of carbon atoms in C= double bonds in rings

$N_{DCH}{}^{a}$ = number of carbon atoms in CH= in aliphatic double bonds

N_{CG} = number of C atoms in quadruples in gauche C−C−C−C structural features[31]

$N_{DCH_2}{}^{a}$... = number of CH_2= atoms in aliphatic double bonds

$N_{Ph}{}^{S}$... = number of phenyl substituents[31]

P_c = critical pressure

pset = prediction set of compounds, not used for derivation of models

T_c = critical temperature

T_{bmix} normal boiling point, published data, except for compounds nos. 140, 141, and 244, estimated with the model developed by the authors[2]

V_c = critical volume

V_{vdw} = van der Waals volume

W = Wiener's topological index[6,32]

Literature Cited

(1) Daubert, T. E. Strengths and Weaknesses of Predictive Methods for Estimating Thermophysical Properties. *J. Chem. Eng. Data* **1996**, *41*, 942−946.

(2) Cholakov, G. St.; Wakeham, W. A.; Stateva, R. P. Estimation of Normal Boiling Points of Hydrocarbons from Descriptors of Molecular Structure. *Fluid Phase Equilib.* **1999**, *163*, 21−42.

(3) Poling, B. E.; Prauznitz, J. M.; O'Connel, J. P. *Properties of Gases and Liquids*, 5th ed.; McGraw-Hill: New York, 2001.

(4) Harper, P. M.; Gani, R.; Kolar, P.; Ishikawa, T. Computer-Aided Molecular Design with Combined Molecular Modeling and Group Contribution. *Fluid Phase Equilib.* **1999**, *158−160*, 337−347.

(5) Marrero-Morejon, J.; Pardillo-Fontdevila, E. Estimation of Pure Compound Properties Using Group-Interaction Contributions. *AIChE J.* **1999**, *45*, 615−621.

(6) Horvath, A. L. *Molecular Design*; Elsevier: Amsterdam, 1992.

(7) Constantinou L.; Gani, R.; O'Connell, J. P. Estimation of the Acentric Factor and the Liquid Molar Volume at 298-K Using a New Group-Contribution Methodology. *Fluid Phase Equilib.* **1995**, *103*, 11−22.

(8) Marrero, J.; Gani, R. Group-contribution Based Estimation of Pure Component Properties. *Fluid Phase Equilib.* **2001**, *183−184*, 183−208.

(9) Iwai, Y.; Yamanaga, Sh.; Arai, Y. Calculation of Normal Boiling Points for Alkane Isomers by a Second-Order Group Contribution method. *Fluid Phase Equilib.* **1999**, *163*, 1−8.

(10) Grigoras, S. A Structural Approach to Calculate Physical Properties of Pure Organic Substances−The Critical Temperature, Critical Volume and Related Properties. *J. Comput. Chem*. **1990**, *11*, 593−510.

(11) Egolf, L. M.; Wessel, M. D.; Jurs, P. C. Prediction of Boiling Points and Critical Temperatures of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947−956.

(12) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure−Property Relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28−41.

(13) Geladi, P.; Tosato, M.-L. In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Acad. Publ.: Dordrecht, The Netherlands, 1990; pp 170−179.

(14) Tosato, M.-L.; Geladi, P.; In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Acad. Publ.: Dordrecht, The Netherlands, 1990; pp 317−341.

(15) TRC (Thermodynamic Research Center). *TRC Thermodynamic Tables-Hydrocarbons*; The Texas A&M University: College Station, TX, 1997 revision.

(16) *Properties of Organic Compounds*; Potehin, A. A. Ed.; Chimia Publ. Ho.: Leningrad, Russia, 1984 (in Russian).

(17) Steele, W. V.; Chirico, R D.; Cowell, A. B.; Knipmeyer, S. E.; Nguyen, A. Thermodynamic Properties and Ideal-Gas Enthalpies of Formation for 2-aminoisobutyric acid (2-methylalanine), acetic acid, (Z)-5-ethylidene-2-norbornene, mesityl oxide (4-methyl-3-penten-2-one), 4-methylpent-1-ene, 2, 2′-bis(phenylthio)propane, and glycidyl phenyl ether (1,2-epoxy-3-phenoxypropane). *J. Chem. Eng. Data* **1997**, *42*, 1053−1066.

(18) Turner, B. E.; Costello, C. L.; Jurs, P. Prediction of Critical Temperatures and Pressures of Industrially Important Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 639−645.

(19) Ambrose D.; Young, C. L. Vapor-Liquid Critical Properties of Elements and Compounds. 1. An Introductory Survey. *J. Chem. Eng. Data* **1995**, *40*, 345−357.

(20) Ambrose D.; Tsonopoulos, C. Vapor-Liquid Critical Properties of Elements and Compounds. 2. Normal Alkanes. *J. Chem Eng. Data* **1995**, *40*, 531−546.

(21) Tsonopoulos, C.; Ambrose, D. Vapor-Liquid Critical Properties of Elements and Compounds. 3. Aromatic Hydrocarbons. *J. Chem Eng. Data* **1995**, *40*, 547−558.

(22) Daubert, T. E. Vapor-Liquid Critical Properties of Elements and Compounds. 5. Branched Alkanes and Cycloalkanes. *J. Chem Eng. Data* **1996**, *41*, 365−372.

(23) Tsonopoulos, C.; Ambrose, D. Vapor-Liquid Critical Properties of Elements and Compounds. 6. Unsaturated Aliphatic Hydrocarbons. *J. Chem. Eng. Data* **1996**, *41*, 645−656.

(24) Nikitin, E. D.; Pavlov, P. A.; Popov, A. P. Vapour-Liquid Critical Temperatures and Pressures of Normal Alkanes with from 19 to 36 Carbon Atoms, Naphthalene and *m*-terphenyl Determined by the Pulse-Heating Technique. *Fluid Phase Equilib.* **1997**, *141*, 155 − 164.

(25) Nikitin, E. D.; Popov, A. P. Critical Temperatures and Pressures of Linear Alk-1-enes with 13 to 20 Carbon Atoms Using the Pulse-Heating Technique. *Fluid Phase Equilib.* **1999**, *166*, 237 − 243.

(26) VonNiederhausern, D. M.; Wilson, G. M.; Giles, N. F. Critical Point and Vapor Pressure Measurements at High Temperatures by Means of a New Apparatus with Ultralow Residence Times. *J. Chem. Eng. Data* **2000**, *45*, 157−160.

(27) Marano, J. J.; Holder, G. D. General Equation for Correlating the Thermophysical Properties of *n*-paraffins, *n*-olefins, and Other

Homologous Series. 1. Formalism for Developing Asymptotic Behavior Correlations. *Ind. Eng. Chem. Res.* **1997**, *36*, 1887−1894.

(28) Marano, J. J.; Holder, G. D. General Equation for Correlating the Thermophysical Properties of *n*-paraffins, *n*-olefins, and Other Homologous Series. 2. Asymptotic Behavior Correlations for PVT Properties. *Ind. Eng. Chem. Res.* **1997**, *36*, 1895−1907.

(29) *API Technical data Book—Petroleum Refining*, 4th ed.; American Petroleum Institute: Washington, DC, 1983.

(30) Stein, S. E.; Brown, R. L. Estimation of Normal Boiling Points from Group Contributions. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581 − 587.

(31) Somayajulu, G. R. Estimation Procedures for Critical Constants. *J. Chem. Eng. Data* **1989**, *34*, 106 − 120.

(32) Labanowski, J. K.; Motoc, I.; Damkoehler, R. A. The Physical Meaning of Tological Indexes. *Comput. Chem.* **1991**, *15*, 47−53.

(33) PCMODEL, 5th ed. Serena Software, Bloomington, IN, 1992.

(34) STATGRAPHICS for DOS, 7th ed. STSC, Inc., and Manugistics, Inc., Rockville, MD.

(35) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, *100*, 10400−10407.

(36) Aleksandrov, I. A. *Peregonka i Rektifikatzia w Neftepererabotke;* Chimia: Moscow, 1981 (In Russian).

(37) Fisher, R. A. *The Design of Experiments*, 6th ed.; Oliver and Boyd: London, 1951.

(38) Cygnarowicz, M. L.; Maxwell, R. J.; Seider, W. D. Equilibrium Solubilities of Beta-Carotene in Supercritical Carbon Dioxide. *Fluid Phase Equilib.* **1990**, *59*, 57−71.

(39) Wakeham, W. A.; Cholakov, G. St., Stateva, R. P. Consequences of property errors on the design of distillation columns, *Fluid Phase Equilib.* **2001**, *185*, 1−12.

Received for review November 20, 2001. Accepted February 20, 2002. The authors acknowledge with gratitude the financial support of the Royal Society.

JE010308L