Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks†

Lowell H. Hall* and C. T. Story‡

Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts 02170

Received April 30, 1996⊗

Two sets of heterogeneous organic compounds were analyzed with artificial neural networks using atom type electrotopological state indices. The first set contains the boiling point for 298 compounds; 30 were placed in a testing set. The neural network model used atom type E-state indices for the 19 atom types present in the data set; the actual network used for prediction had a 19:5:1 architecture. This model produced a mean absolute error (MAE) of 3.93 K for the overall set, 3.86 for the training set, and 4.57 for the test set. The average relative percent error for 10 runs is 0.94% for the whole data set and 1.12% for the test set. The second set contains critical temperatures for 165 compounds; 18 were placed in the testing set. The neural network possessed a 19:4:1 architecture and produced an MAE of 4.52 K for the whole set, 4.39 K for the training set, and 5.59 K for the test set. The average relative percent error for 5 runs is 0.77% for the whole data set and 0.95% for the test set.

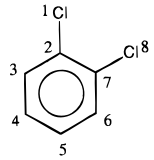
INTRODUCTION

Boiling point and critical temperature are significant properties in revealing the intermolecular aspects of molecules. Further, they are useful for testing development of QSAR models. For both these reasons, we have chosen to examine the relationship of boiling point and critical temperature to a new set of molecular structure descriptors, the atom type electrotopological state indices.

The atom level topological index, called the electrotopological state index (E-state), was introduced in 1990. Other topological indices represent or encode the whole molecule, usually as a sum over subgraphs of the molecular graph. In contrast, the E-state index is computed as a graph invariant for each atom in the molecular graph. The index combines the electronic state of the bonded atom within the molecule with its topological nature in the context of the whole molecular skeleton. The E-state indices have been used for a variety of QSAR studies.[1]

Recently, an extension of the E-state indices, called the atom type E-state, has been introduced.[2] Each atom in the molecule is categorized according to a valence state classification scheme. All the atoms of the same type are taken together, and their E-state values are summed to make the atom type E-state index for that atom type. This atom type index lends itself for use in a group additive type scheme in which an index appears for each atom type in the molecule (together with its E-state contribution). In some cases, a small number of atom type indices may be used for a particular investigation, particularly in a biological study in which only a few atom types are required for a quality QSAR equation. For biological QSARs reported to date, a type of skeletal superposition has been used so that the E-state values for corresponding atoms may be entered as variables in regression analysis.[1,3−9] The current development of atom type E-state values provides the basis for application to a wider range of problems to which the E-state formalism is applicable without the need for superposition.

Table 1. The Electrotopological State Indices Calculated for *o*-Dichlorobenzene Along with the Atom-Type Electrotopological Indices



atom	atom id	E-state values
-Cl	1	5.577
-C<	2	0.606
-CH-	3	1.754
-CH-	4	1.841
-CH-	5	1.841
-CH-	6	1.754
-C<	7	0.606
-Cl	8	5.577

atom type	atom type E-state value
SsCl	11.154
SaaCH	7.190
SaasC	1.212

The E-state indices have been used for ethers, aldehydes, and ketones to correlate ^{17}O NMR frequencies:[1,3−5] binding studies including binding of a series of indolealkylamines to 5-HT_2 receptors[1,7] and binding of barbiturates to β-cyclodextrin;[3] receptor binding QSAR including affinity of β-carbolines[6] and dopamine D-2 receptor binding of salicylamides;[7] and inhibition studies including inhibition of flu virus by benzimidazoles[4] and inhibition of MAO by hydrazides.[6] The MAO inhibition study was extensively developed to include a careful analysis of the inhibitor molecules using semiempirical MO computations. The model based on the E-state indices was found to be superior to that based on MO computed charges;[8] the time requirements for the MO study was at least 1000 times more than that required for the E-state analysis. The atom type E-state indices were recently introduced with a QSAR regression

* Person to whom correspondence should be addressed.

† Keywords: QSAR/QSPR, electrotopological state (E-state), atom type, molconn, artificial neural network, boiling point, critical temperature.

‡ Current address: Department of Chemistry, Kansas State University, Manhattan, KS 66506.

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1996.

S0095-2338(96)00375-7 CCC: $12.00

Table 2

a. Summary for the Modeling of Boiling Point for 298 Mixed Organic Compounds Using 19 Atom Type Electrotopological Indices with a 19:5:1 Artificial Neural Network

quantity	training set	testing set	total set
no. of compounds	268	30	298
mean absolute error, MAE	3.86	4.57	3.93
root-mean-square error, RMS	5.30	5.87	5.36
correlation coefficient, *r* (calculated vs predicted)	0.9975	0.9967	0.9975
slope of line (calculated vs predicted)	1.00	0.98	1.00

b. Statistical Summary for the Modeling of Critical Temperature for 165 Mixed Organic Compounds Using 19 Atom Type Electrotopological Indices with a 19:4:1 Artificial Neural Network

quantity	training set	testing set	total set
no. of compounds	147	18	165
mean absolute error, MAE	4.39	5.59	4.52
root-mean-square error, RMS	6.61	6.71	6.62
correlation coefficient, *r* (calculated vs predicted)	0.9965	0.9949	0.9964
slope of line (calculated vs predicted)	1.00	1.00	1.00

on a data set of 245 alkanes and alcohols.[2] In a very recent application of the atom type E-state index, a new measure of molecular similarity was developed and applied to database searching.[10]

Jurs et al.[11] have collected a set of data from the DIPPR database for the purpose of testing a particular approach to QSPR modeling. For this paper we have chosen to use this same data set for modeling with the atom type E-state indices both as a test of their QSAR capability and as a means of providing a model for prediction of boiling point and critical temperature. It is expected that the relation between these temperature measures and some of the molecular descriptors is nonlinear. For this reason, we elected to use artificial neural networks for modeling. This modeling has been described in a book[12] and used in the development of QSAR.[13−15]

METHOD

Structure Descriptors. The structure descriptors used in this study to represent the molecular structure of the alkanes, alcohols, and chloroalkanes are the atom type electrotopological state indices. The set of indices was calculated with a recently released new version of the Molconn software, Molconn-Z.[16] An example of the atom type E-state indices is given in Table 1 for *o*-dichlorobenzene. The symbols for the 19 atom types represented in this data set are as follows: SsCH3, SssCH2, SsssCH, SssssC, SaaCH, SaasC, SdCH2, SdsCH, SdssC, StsC, SsNH2, SssNH, SsssN, SaaN, SsOH, SssO, SdO, SsCl, and SsBr.[2] In this symbol, "S" stands for the sum of the E-state values for a given atom type in a molecule. The set of bonds to a skeletal atom is given by a string of lower case letters: s (single), d (double), t (triple), and a (aromatic). The element is given by its symbol together with the number of hydrogen atoms. For example, SdCH2 represents the E-state values for a terminal CH_2 group on a double bond, while the SssCH2 symbol represents the methylene group with two single bonds; SaasC stands for an aromatic carbon to which a substituent is bonded. Structure input for the 298 compounds was from SMILES code.

Data Sets. The data sets for this study were taken from Jurs et al.,[11] because the data had been carefully selected for quality experimental information. For set 1 the list of 298 compounds and their boiling points (K) are given in Table 3. The last 30 compounds are marked, indicating that they comprised the testing set. For set 2 the list of 165 compounds and their critical temperatures (K) are given in Table 4. The last 18 compounds comprised the testing set.

Statistical Method. The boiling points (or critical temperatures for set 2) together with the 19 atom type E-state indices were entered into the spreadsheet of the Neuralyst program[17] (a modification of the EXCEL spreadsheet software). After entering data into the Neuralyst spreadsheet, the neural network architecture was selected, and the learning parameters set. A training set was designed the same as that used by Jurs.[11] The network weights were then assigned as random numbers. Training was initiated and followed by examining the *RMS* error (*RMS* stands for the root mean error, that is, the square root of the average residual) for the total set and for both the training and test set as well. Training was terminated when there was no further improvement in the test set *RMS* error. In addition, other statistics were recorded, including the mean average error (*MAE*) for the total data set as well as both the training and test sets. We also computed the relative error for the whole data set and the correlation coefficient between the observed and predicted values.

Several network architectures were examined in order to find a useful model. We examined only three layer networks with 19 input nodes (for the 19 structure variables), one output node (for the boiling point or critical temperature), and one hidden layer with *n* nodes. For these 19-*n*-1 networks, we found that the *RMS* error improved significantly up to values for *n* around 4 or 5. We decided to develop further a 19:5:1 network for boiling point and 19:4:1 for critical temperature. In these studies the learning rate was set to 0.80 and the momentum to 0.60. The fraction of the data set selected for training was approximately 90%. For each training set, we examined several sets of randomly selected weights. Typical runs took 20 000−50 000 cycles.

RESULTS

Boiling Point. For the studies performed in this investigation for boiling point, the overall relative error (for the whole data set of 298 compounds) ranged from 0.94% to 1.17%. The run with the best results is given in Table 3; we also ran and averaged ten runs, the results of which are also given in Table 3. The *MAE* for the whole set, training set, and testing set are 3.93, 3.86, and 4.57 K, respectively,

Table 3. The Observed Boiling Point of the Heterogeneous Set of Organic Compounds with the Predicted Boiling Point from an 19:5:1 Artificial Neural Network Along with the Residual Values

					av from 10 runs	
id[a]	compound name	T_b (K)[b]	calc[c]	res[d]	calc[e]	res
1	acrylonitrile	350.5	351.7	−1.2	354.6	−4.1
2	acrylic aldehyde	325.8	325.0	0.8	323.6	2.3
3	ethylene carboxylic acid	414.2	417.9	−3.8	417.8	−3.7
4	3-chloropropene	318.1	310.8	7.3	316.1	2.0
5	propylene	225.4	252.8	−27.4	249.2	−23.8
6	1,2-dichloropropene	361.3	365.5	−4.2	368.5	−7.3
7	allyl alcohol	370.2	364.1	6.2	368.6	1.7
8	methylvinyl ether	278.7	285.5	−6.9	287.3	−8.6
9	*n*-propionaldehyde	321.2	320.6	0.6	319.9	1.3
10	ethylformate	327.5	340.4	−13.0	341.4	−14.0
11	methylacetate	330.1	332.5	−2.4	331.5	−1.5
12	proprionic acid	414.3	422.0	−7.7	418.6	−4.3
13	1-bromopropane	344.1	344.2	−0.1	347.2	−3.1
14	isopropyl chloride	308.9	304.7	4.2	310.2	−1.3
15	*n*-propyl chloride	319.7	315.3	4.4	318.5	1.2
16	allylamine	326.5	327.3	−0.8	325.6	0.9
17	propane	231.1	258.1	−27.0	251.5	−20.4
18	methylethyl ether	280.5	285.4	−4.9	285.6	−5.1
19	*n*-propanol	370.4	366.7	3.7	365.0	5.4
20	methylal	315.0	321.1	−6.1	318.6	−3.6
21	1,2-propylene glycol	460.8	463.2	−2.4	459.4	1.4
22	isopropylamine	305.6	309.3	−3.7	309.2	−3.6
23	*n*-propylamine	321.6	325.7	−4.1	322.0	−0.4
24	trimethylamine	276.0	278.0	−2.0	277.9	−1.9
25	1,3-butadiene	268.7	275.2	−6.5	271.3	−2.6
26	divinyl ether	301.5	307.8	−6.3	310.3	−8.8
27	methacrolein	341.1	353.8	−12.7	354.2	−13.1
28	*tert*-crotonic acid	458.1	457.7	0.4	458.3	−0.2
29	methylmacrylate	353.4	348.7	4.7	350.2	3.2
30	vinyl acetate	345.6	352.7	−7.1	354.4	−8.8
31	butyronitrile	390.8	391.6	−0.8	387.2	3.6
32	1-butene	266.9	274.9	−8.0	272.7	−5.8
33	*cis*-2-butene	276.9	273.8	3.1	276.1	0.8
34	*trans*-2-butene	274.0	273.8	0.2	276.1	−2.1
35	isobutene	266.3	279.3	−13.0	275.2	−8.9
36	*n*-butyraldehyde	348.0	349.0	−1.0	348.2	−0.2
37	isobutyraldehyde	337.3	336.1	1.2	337.5	−0.2
38	methylethyl ketone	352.8	345.5	7.3	344.2	8.6
39	tetrahydrofuran	338.0	328.3	9.7	323.9	14.1
40	*n*-butyric acid	436.4	438.3	−1.9	437.3	−0.9
41	ethyl acetate	350.2	350.0	0.2	349.9	0.3
42	isobutyric acid	427.9	428.1	−0.2	427.9	0.0
43	methyl proprionate	352.6	351.1	1.5	350.1	2.5
44	*n*-propyl formate	354.0	361.1	−7.1	362.5	−8.5
45	1-bromobutane	374.9	375.5	−0.6	371.6	3.3
46	1-chlorobutane	351.6	352.0	−0.4	350.5	1.1
47	*n*-butane	272.6	280.7	−8.1	275.7	−3.1
48	isobutane	261.4	274.6	−13.2	270.5	−9.1
49	butanol	390.8	386.8	4.0	387.6	3.2
50	*sec*-butyl alcohol	372.7	375.5	−2.8	375.3	−2.6
51	*tert*-butyl alcohol	355.6	361.5	−5.9	357.2	−1.6
52	diethyl ether	308.6	310.2	−1.6	310.8	−2.2
53	isobutanol	380.8	377.3	3.5	377.0	3.8
54	1,3-butanediol	480.1	480.4	−0.3	479.1	1.0
55	*n*-butylamine	350.6	351.7	−1.1	348.2	2.4
56	diethylamine	328.6	325.4	3.2	327.7	0.9
57	isobutylamine	340.9	336.0	4.9	337.4	3.5
58	pyridine	388.4	387.9	0.5	387.6	0.8
59	cyclopentene	317.4	316.8	0.6	322.0	−4.6
60	isoprene	307.2	300.6	6.6	299.9	7.3
61	acetylacetone	413.6	413.7	−0.1	418.3	−4.7
62	allyl acetate	377.1	368.1	9.0	370.7	6.4
63	ethyl arylate	372.6	366.7	5.9	367.9	4.7
64	cyclopentane	322.4	326.2	−3.8	326.7	−4.3
65	1-pentene	303.1	300.6	2.5	300.4	2.7
66	diethyl ketone	375.1	369.4	5.7	368.2	6.9
67	2-pentanone	375.5	371.1	4.4	370.1	5.4
68	valeraldehyde	376.1	379.6	−3.5	378.4	−2.3
69	ethylproprionate	372.3	371.4	0.9	371.3	1.0
70	isobutyl formate	371.2	373.5	−2.3	376.5	−5.3
71	*n*-propylacetate	374.6	373.4	1.2	373.5	1.1
72	valeric acid	458.6	457.1	1.5	458.0	0.6

Table 3 (Continued)

					av from 10 runs	
id[a]	compound name	T_b (K)[b]	calc[c]	res[d]	calc[e]	res
73	1-chloropentane	381.5	388.3	−6.8	383.5	−2.0
74	isopentane	301.0	302.6	−1.6	300.7	0.3
76	neopentane	282.6	286.8	−4.2	286.3	−3.7
76	*n*-pentane	309.2	308.2	1.0	306.4	2.8
77	2-methyl-1-butanol	401.9	398.9	3.0	399.3	2.6
78	2-methyl-2-butanol	375.1	377.6	−2.5	375.4	−0.3
79	3-methyl-1-butanol	404.4	399.3	5.1	399.6	4.8
80	3-methyl-2-butanol	384.6	388.5	−3.9	387.8	−3.2
81	1-pentanol	411.0	405.7	5.3	407.5	3.5
82	2-pentanol	392.1	394.9	−2.8	395.3	−3.2
83	3-pentanol	388.5	394.1	−5.6	394.4	−5.9
84	neopentyl glycol	483.0	476.0	7.0	479.5	3.5
85	1,5-pentanediol	512.1	511.9	0.2	512.4	−0.3
86	*o*-dichlorobenzene	453.6	455.3	−1.7	455.6	−2.0
87	bromobenzene	429.2	428.6	0.6	429.9	−0.7
88	chlorobenzene	404.9	412.1	−7.2	408.5	−3.6
89	benzene	353.2	350.3	2.9	351.6	1.6
90	phenol	455.0	462.5	−7.5	461.9	−6.9
91	*p*-hydroquinone	558.1	552.4	5.7	550.5	7.6
92	aniline	457.6	452.0	5.6	457.0	0.6
93	2-methylpyridine	402.6	410.2	−7.6	409.6	−7.0
94	adiponitrile	568.1	566.9	1.2	567.8	0.3
95	cyclohexene	356.1	353.4	2.7	356.9	−0.8
96	cyclohexanone	428.9	422.6	6.3	419.0	9.9
97	mesityl oxide	403.0	408.3	−5.3	411.7	−8.7
98	cyclohexane	353.9	353.4	0.5	356.9	−3.0
99	2,3-dimethyl-1-butene	328.8	329.0	−0.2	330.0	−2.1
100	2,3-dimethyl-2-butene	346.4	344.8	1.6	345.6	0.8
101	2-ethyl-1-butene	337.8	333.7	4.1	335.9	1.9
102	1-hexene	336.6	331.7	4.9	333.7	2.9
103	*cis*-2-hexene	342.0	337.9	4.1	338.6	3.4
104	*trans*-2-hexane	341.0	337.9	3.1	338.6	2.4
105	methylcyclopentane	345.0	346.4	−1.4	347.0	−2.0
106	2-methyl-1-pentane	335.3	333.8	1.5	336.2	−0.9
107	4-methyl-1-pentene	327.0	322.6	4.4	324.2	2.8
108	butylvinyl ether	367.0	366.3	0.7	362.3	4.7
109	cyclohexanol	434.0	435.7	−1.7	439.2	−5.2
110	2-hexanone	400.9	399.1	1.8	397.8	3.1
111	methylisobutyl ketone	389.6	387.6	2.0	388.2	1.4
112	ethyl *n*-butyrate	394.6	391.6	3.0	391.4	3.2
113	2-ethylbutyric acid	467.0	464.7	2.3	465.0	2.0
114	*n*-hexanoic acid	478.9	475.7	3.2	477.9	1.0
115	isobutyl acetate	389.8	381.7	8.1	383.9	5.9
116	cyclohexylamine	407.6	410.0	−2.4	411.1	−3.5
117	2,2-dimethylbutane	322.9	317.4	5.5	318.6	4.3
118	2,3-dimethylbutane	331.1	327.4	3.7	326.1	5.0
119	*n*-hexane	341.9	336.2	5.7	337.3	4.6
120	2-methylpentane	333.4	331.5	1.9	331.5	1.9
121	3-methylpentane	336.4	331.5	4.9	331.5	4.9
122	disopropyl ether	341.5	352.1	−10.6	350.2	−8.7
123	di-*n*-propyl ether	362.8	362.4	0.4	362.3	0.5
124	acetal	376.8	375.6	1.2	374.8	2.0
125	disopropylamine	357.2	360.1	−2.9	358.2	−1.0
126	di-*n*-propylamine	382.0	382.0	0.0	379.1	2.9
127	benzaldehyde	451.9	458.9	−7.0	456.0	−4.1
128	benzoic acid	522.4	519.5	2.9	522.1	0.3
129	toluene	383.8	374.3	9.5	380.8	3.0
130	benzyl alcohol	477.9	479.8	−1.9	480.9	−3.0
131	*m*-cresol	475.4	467.3	8.1	467.7	7.7
132	*o*-cresol	464.1	468.5	−4.4	468.7	−4.6
133	*m*-toluidine	476.6	476.7	−0.1	475.6	1.0
134	*o*-toluidine	473.6	474.1	−0.5	473.3	0.3
135	*n*-butyl acrylate	421.0	419.8	1.2	418.7	2.3
136	isobutyl acrylate	405.1	410.0	−4.9	408.7	−3.6
137	ethylcyclopentane	376.6	376.7	−0.1	377.9	−1.3
138	2,3-dimethylpentane	362.9	358.3	4.6	357.0	5.9
139	*n*-heptane	371.6	365.1	6.5	368.1	3.5
140	2-methylhexane	363.2	360.6	2.6	361.6	1.6
141	3-methylhexane	365.0	360.6	4.4	361.6	3.4
142	styrene	418.3	411.2	7.1	409.1	9.2
143	acetophenone	475.1	476.6	−1.5	471.5	3.6
144	ethylbenzene	409.4	404.7	4.7	407.1	2.3
145	*m*-xylene	412.3	411.9	0.4	410.9	1.4

Table 3 (Continued)

					av from 10 runs	
id[a]	compound name	T_b (K)[b]	calc[c]	res[d]	calc[e]	res
146	*p*-xylene	411.5	413.2	−1.7	412.0	−0.5
147	2,6-xylenol	474.2	475.9	−1.7	473.8	0.4
148	*N*,*N*′-dimethylaniline	466.7	464.7	2.0	463.9	2.8
149	*cis*-1,2-dimethylcyclohexane	402.9	399.8	3.1	396.8	6.1
150	*trans*-1,2-dimethylcyclohexane	396.6	399.8	−3.2	396.8	−0.2
151	*cis*-1,3-dimethylcyclohexane	393.2	398.6	−5.4	395.8	−2.6
152	*trans*-1,3-dimethylcyclohexane	397.6	398.6	−1.0	395.8	1.8
153	*cis*-1,4-dimethylcyclohexane	397.5	398.6	−1.1	395.8	1.7
154	*trans*-1,4-dimethylcyclohexane	392.5	398.6	−6.1	395.8	−3.3
155	ethylcyclohexane	405.0	403.6	1.4	404.5	0.5
156	*n*-propylcyclopentane	404.1	403.6	0.5	404.5	−0.4
157	2,4,4-trimethyl-1-pentene	374.6	374.1	0.5	375.9	−1.3
158	2,4,4-trimethyl-2-pentene	378.1	391.3	−13.2	382.6	−4.5
159	isobutylisobutyrate	420.6	412.7	7.9	415. 6	5.0
160	2,3-dimethylhexane	388.8	387.6	1.2	385.5	
161	2-methyl-3-ethylpentane	388.8	387.6	1.2	385.5	3.3
162	*n*-octane	398.1	393.6	4.5	397.0	1.1
163	2,2,3-trimethylpentane	383.0	383.8	−0.8	383.7	−0.7
164	2,2,4-trimethylpentane	372.4	380.7	−8.3	380.8	−8.4
165	2,3,3-trimethylpentane	387.9	380.7	7.2	380.8	7.1
166	di-*n*-butyl ether	413.4	414.5	−1.1	412.3	1.1
167	2-ethyl-1-hexanol	457.8	460.4	−2.6	455.4	2.4
168	quinoline	510.8	509.6	1.2	510.4	0.4
169	α-methylstyrene	438.6	441.3	−2.7	439.9	−1.3
170	ethyl benzoate	486.6	493.9	−7.3	486.7	−0.1
171	cumene	425.6	422.6	3.0	423.1	2.5
172	*o*-ethyltoluene	438.3	436.6	1.7	434.6	3.7
173	*p*-ethyltoluene	435.2	436.5	−1.3	434.5	0.7
174	mesitylene	437.9	443.4	−5.5	441.3	−3.4
175	*n*-propylbenzene	432.4	431.4	1.0	430.6	1.8
176	1,2,3-trimethylbenzene	449.3	443.7	5.6	441.6	7.7
177	1,2,4-trimethylbenzene	442.5	442.6	−0.1	440.4	2.1
178	isophorone	488.4	484.9	3.5	478.7	9.7
179	*n*-propylcyclohexane	429.9	429.7	0.2	429.4	0.5
180	3,3-diethylpentane	419.3	422.5	−3.2	422.6	−3.3
181	*n*-nonane	424.0	421.2	2.8	423.5	0.5
182	2,2,3,3-tetramethylpentane	413.4	409.5	3.9	405.2	8.2
183	dimethylterephthalate	561.1	558.2	2.9	554.1	7.0
184	1,2,3,4-tetrahydronaphthalene	480.8	474.0	6.8	477.6	3.2
185	*n*-butylbenzene	456.5	459.1	−2.6	455.8	0.7
186	*tert*-butylbenzene	442.3	441.6	0.7	439.7	2.6
187	*p*-cymene	450.3	447.0	3.3	447.7	2.6
188	*m*-diethylbenzene	454.3	456.2	−1.9	455.3	−1.0
189	*o*-diethylbenzene	456.6	456.0	0.6	455.3	1.3
190	isobutylbenzene	445.9	449.5	−3.6	445.7	0.2
191	*n*-butylcyclohexane	454.1	454.6	−0.5	452.7	1.4
192	*n*-decane	447.3	447.7	−0.4	448.0	−0.7
193	2-ethylhexylacrylate	489.1	502.1	−13.0	492.2	−3.1
194	diphenyl ether	531.5	527.1	4.4	530.1	1.4
195	diphenylamine	575.1	576.3	−1.2	575.3	−0.2
196	diethylphthalate	567.1	570.9	−3.8	569.2	−2.1
197	*m*-diisopropylbenzene	476.3	474.4	1.9	475.8	0.5
198	bicyclohexyl	512.2	506.8	5.4	506.5	5.7
199	1-dodecene	486.5	492.5	−6.0	488.7	−2.2
200	*n*-dodecane	489.5	495.1	−5.6	491.0	−1.5
201	di-*n*-hexyl ether	498.9	500.3	−1.4	493.5	5.4
202	diphenylmethane	537.4	543.6	−6.2	535.8	1.6
203	benzylbenzoate	596.6	596.2	0.4	595.5	1.1
204	1,2-diphenylethane	553.6	553.9	−0.3	552.0	1.6
205	1-tetradecene	524.2	531.5	−7.3	526.5	−2.3
206	*n*-tetradecane	526.7	535.1	−8.4	528.7	−2.0
207	dibutylphthalate	613.1	609.0	4.1	611.2	1.9
208	1-hexadecene	558.0	561.1	−3.1	558.3	−0.3
209	dibutylsebacate	622.1	615.2	6.9	621.3	0.8
210	1-octadecene	588.0	584.2	3.8	585.4	2.6
211	stearic acid	648.3	634.9	13.4	637.4	10.9
212	*n*-octadecane	589.9	589.7	0.2	587.8	2.1
213	*n*-nonadecane	603.0	598.9	4.1	599.2	3.8
214	*n*-butylstearate	623.1	633.9	−10.8	633.7	−10.6
215	methylisopropenylketone	371.1	372.3	−1.2	372.4	−1.3
216	1,1-dichloropropane	361.3	365.1	−3.8	363.7	−2.4
217	1,3-dichloropropane	393.6	395.8	−2.2	389.0	4.6
218	1,3-propylene glycol	487.6	490.4	−2.8	482.3	5.3

Table 3 (Continued)

					av from 10 runs	
id[a]	compound name	T_b (K)[b]	calc[c]	res[d]	calc[e]	res
219	isobutyronitrile	376.8	377.2	−0.4	374.5	2.3
220	1,4-dichlorobutane	427.1	427.4	−0.3	421.5	5.6
221	2-bromobutane	364.4	368.8	−4.4	365.4	−1.0
222	*sec*-butylchloride	341.3	344.8	−3.5	344.4	−3.1
223	methylisopropyl ether	323.8	306.1	17.7	306.4	17.4
224	1,4-butanediol	501.1	502.0	−0.9	498.3	2.8
225	2,3-butanediol	453.9	454.7	−0.8	453.9	0.0
226	*sec*-butylamine	336.1	336.2	−0.1	336.2	−0.1
227	*tert*-butylamine	317.6	322.8	−5.2	322.0	−4.4
228	valeronitrile	414.5	418.4	−3.9	415.4	−0.9
229	2-methyl-1-butene	304.3	306.1	−1.8	305.2	−0.9
230	2-methyl-2-butene	311.7	304.2	7.5	307.3	4.4
231	3-methyl-1-butene	293.2	291.9	1.3	292.0	1.2
232	1,5-dichloropentane	453.1	451.3	1.8	450.6	2.5
233	methylisopropyl ketone	367.6	359.3	8.3	360.6	7.0
234	isopropyl acetate	361.6	358.5	3.1	362.0	−0.4
235	2-methylbutyric acid	450.1	444.3	5.8	444.8	5.3
236	methyl-*n*-butyrate	375.9	375.6	0.3	374.7	1.2
237	piperidine	379.6	374.5	5.1	376.9	2.7
238	2,2-dimethyl-1-propanol	386.3	387.5	−1.2	386.0	0.3
239	ethylpropyl ether	337.0	337.2	−0.2	337.6	−0.6
240	methyl-*sec*-butyl ether	332.1	331.7	0.4	331.4	
241	methyl *tert*-butyl ether	328.4	321.2	7.2	323.6	4.8
242	methyl isobutyl ether	331.7	336.0	−4.3	334.2	−2.5
243	*n*-pentylamine	377.6	377.8	−0.2	375.6	2.0
244	3-methylpyridine	417.3	410.9	6.4	410.3	7.0
245	1,3-cyclohexadiene	353.5	354.2	−0.7	352.4	1.1
246	methylcyclopentadiene	345.9	342.8	3.1	344.3	1.6
247	2,3-dimethyl-3-butadiene	341.9	339.7	2.2	338.1	3.8
248	1,5-hexadiene	332.6	329.8	2.8	328.3	4.3
249	hexanenitrile	436.8	434.5	2.3	435.5	1.3
250	3,3-dimethyl-1-butene	314.4	309.9	4.5	311.2	3.2
251	2-methyl-2-pentene	340.5	336.9	3.6	338.8	1.7
252	3-methyl-1-pentene	327.3	322.2	5.1	323.8	3.5
253	ethylisopropyl ketone	386.6	382.7	3.9	382.9	3.7
254	1-hexanal	401.5	407.6	−6.1	405.4	−3.9
255	3-hexanone	396.6	398.9	−2.3	397.8	−1.2
256	*sec*-butyl acetate	385.1	383.1	2.0	386.1	−1.0
257	*tert*-butyl acetate	369.1	373.2	−4.1	368.7	0.4
258	ethyl isobutyrate	383.0	380.6	2.4	383.2	−0.2
259	*n*-propyl propionate	395.6	393.8	1.8	393.7	1.9
260	*n*-butylethyl ether	365.4	364.2	1.2	364.1	1.3
261	2-ethyl-1-butanol	419.6	418.6	1.0	418.2	1.4
262	1-hexanol	430.1	425.3	4.8	426.6	3.5
263	2-hexanol	413.0	415.2	−2.2	414.7	−1.7
264	2-methyl-1-pentanol	421.1	419.9	1.2	419.4	1.7
265	4-methyl-2-pentanol	404.9	409.0	−4.1	407.3	−2.4
266	methyl-*tert*-pentyl ether	359.5	352.5	7.0	355.1	4.4
267	1,6-hexanediol	516.1	522.3	−6.2	525.9	−9.8
268	*n*-hexylamine	404.6	400.8	3.8	400.3	4.3
269	acetone[f]	329.4	319.4	10.0	317.3	12.1
270	2-propanol[f]	355.4	366.7	−11.3	365.0	−9.6
271	ethylvinyl ether[f]	308.7	307.3	1.4	307.9	0.8
272	cyclopentadiene[f]	314.6	313.8	0.8	319.0	−4.4
273	*n*-butyl formate[f]	379.3	387.9	−8.6	389.0	−9.7
274	isovaleric acid[f]	448.3	448.2	0.1	449.0	−0.7
275	*n*-butyl acetate[f]	399.1	394.9	4.2	395.3	3.8
276	hexamethylene imine[f]	404.9	401.2	3.7	403.5	1.4
277	*p*-cresol[f]	475.1	467.9	7.2	468.3	6.8
278	*p*-toluidine[f]	473.4	474.6	−1.2	474.3	−0.9
279	1-heptene[f]	366.8	360.5	6.3	363.6	3.2
280	methylcyclohexane[f]	374.1	375.1	−1.0	376.6	−2.5
281	2,2,3-trimethlbutane[f]	354.0	348.1	5.9	349.1	4.9
282	*o*-xylene[f]	417.6	411.9	5.7	410.9	6.7
283	1-octene[f]	394.4	389.4	5.0	392.4	2.0
284	benzyl acetate[a]	486.6	500.9	−14.3	496.7	−10.1
285	*n*-ethyltoluene[f]	434.5	435.3	−0.8	433.5	1.0
286	dimethyl phthalate[f]	556.8	552.8	4.0	549.5	7.3
287	*sec*-butyl benzene[f]	446.5	447.3	−0.8	443.7	2.8
288	*p*-diethylbenzene[f]	456.9	456.2	0.7	455.3	1.6
289	1-decane[f]	443.8	444.5	−0.7	444.5	−0.7
290	1-decanol[f]	503.4	505.9	−2.5	499.1	4.3
291	*p*-diisopropylbenzene[f]	483.6	474.4	9.2	475.8	7.8

Table 3 (Continued)

id[a]	compound name	T_b (K)[b]	calc[c]	res[d]	calc[e]	res
292	1,1-diphenylethane[f]	545.8	543.3	2.5	537.4	8.4
293	*n*-hexadecane[f]	560.0	566.4	−6.4	561.0	−1.0
294	2-bromopropane[f]	332.6	332.5	0.1	339.7	−7.1
296	*tert*-butyl chloride[f]	323.8	326.5	−2.7	325.4	−1.6
297	4-methylpyridine[f]	418.5	410.2	8.3	409.6	8.9
298	*n*-pentyl formate[f]	406.6	411.3	−4.7	412.5	−5.9

[a] id: The serial number for the compounds in the table. This id is the same as in ref 11 by Jurs. [b] The observed boiling point on the Kelvin scale, T_b (K), as obtained from Jurs (11). [c] calc: The value predicted for boiling point from the 19:5:1 neural network model. See text. [d] res: Observed boiling point − calculated boiling point. [e] The results from ten runs on the same artificial network using a different starting set of network weights for each run. Calc is the average of calculated values from ten runs. Res is observed − calc. See text. [f] Compounds which were used as a set test.

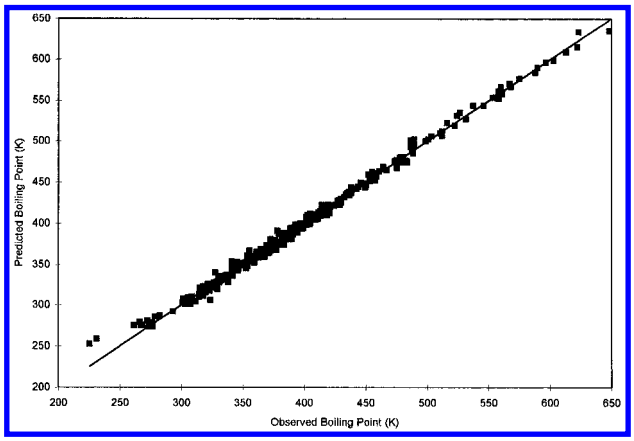


Figure 1. The plot of calculated versus observed boiling point for the set of 298 heterogeneous organic compounds based on a 19:5:1 artificial neural network.

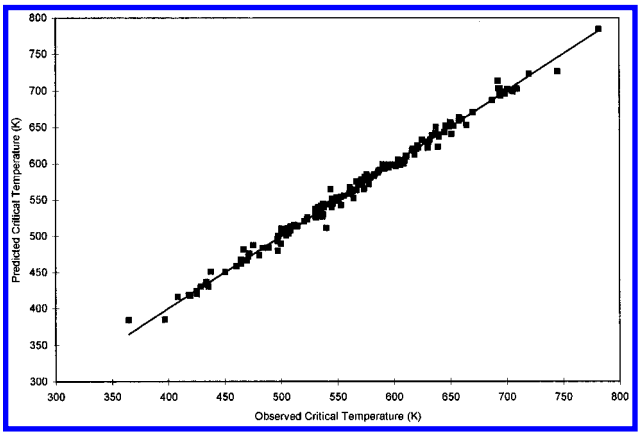


Figure 2. The plot of calculated versus observed critical temperature for the set of 165 heterogeneous organic compounds based on a 19:4:1 artificial neural network.

for the best run. The average relative percent error for ten runs is 0.94% for the total set and 1.12% for the test set. The summary of results for the selected model is given in Table 2a. The results of the best model are given in Table 3, including the compound name, the observed boiling point, calculated boiling point, and the residual. The plot of calculated versus observed boiling point is given in Figure 1.

Critical Temperature. For the studies performed in this investigation for critical temperature, overall relative error for the best run (for the whole data set of 165 compounds) ranged from 0.97% to 1.17%. The *MAE* for the whole set, training set, and testing set are 4.52, 4.39, and 5.59 K, respectively, for the best run. Average relative percent error for five runs is 0.77% for the total set and 0.95% for the test set. The summary for the selected model is given in Table 2b. The results from the run with the best results is given in Table 4; we also ran and averaged five runs, the results of which are also given in Table 4, including the compound name, the observed boiling point, the calculated boiling point and the residual. A plot of calculated versus observed boiling point is given in Figure 2.

DISCUSSION

The data in these studies are two heterogeneous sets of organic compounds, based on 298 organic compounds and their properties taken from the DIPPR collection. The first set contains boiling points for all 298 compounds. The second set is a subset of the first, that is, those compounds of set 1 for which critical temperature values exist in the DIPPR database. The molecules in both sets are characterized by 19 atom types (listed in the METHODS section) including five elements: carbon, nitrogen, oxygen, chlorine, and bromine. In addition to single, double, triple, and aromatic bonds, the following functional groups are present: chloro, bromo, alcohol (including diols), cyano, amino, ether, carboxylic acid, ester, and carbonyl. Nonpolar and polar molecules are included as well as those exhibiting hydrogen bonding. The boiling points range from 225 K (propylene) to 648 K (stearic acid). For critical temperature, the values range from 365 K (propylene) to 782 K (quinoline).

Boiling Point. For the boiling point model, only two compounds (propylene and propane) have residuals greater than 20 K, and ten have residuals between 10 and 20 K. Among the 30 compounds in the test set, only three have residuals greater than 10; none are greater than 13 K. A plot of calculated boiling point versus observed boiling point (Figure 1) shows a very tight arrangement along the line. The plot of residuals (not shown) displays no patterns and appears to be random. It is interesting that the two largest residuals belong to the two smallest molecules in the data set, propylene and propane. In fact, most of the larger residuals belong to the smallest molecules, those for which prediction is of little importance.

For purposes of comparison, a linear model was obtained using multiple linear regression. For the same 19 input variables as used in the neural network, the standard error was found to be much larger, 15.6 K. Further, there were many more large residuals, including 40 with residual larger than 20 K; five residuals were larger than 40 K. Thus, it appears that the nonlinear nature of the neural network model is a better vehicle for modeling boiling point.

The results for this model are excellent. The mean absolute error for the whole data set is under 4 K and for the test set as well. This value is the lowest *MAE* for such

Table 4. The Observed Critical Temperature of the Heterogeneous Set of Organic Compounds with the Predicted Critical Temperature from an 19:4:1 Artificial Neural Network Along with the Residual Values

					av from 5 runs	
id[a]	compound name	T_c (K)[b]	calc[c]	res[d]	calc[e]	res
1	3-chloropropene	514.5	515.0	−0.5	513.8	0.7
2	propylene	364.8	383.7	−18.9	383.5	−18.7
3	allylalcohol	545.1	545.4	−0.4	545.8	−0.7
4	ethylformate	508.4	509.6	−1.2	511.1	−2.7
5	methylacetate	506.8	505.2	1.6	503.3	3.5
6	proprionicacid	604.0	606.3	−2.3	606.5	−2.5
7	*n*-propylchloride	503.2	502.5	0.7	503.6	−0.5
8	propane	396.8	385.6	11.2	386.0	10.9
9	methyl ethyl ether	437.8	442.4	−4.6	442.3	−4.5
10	*n*-propanol	536.7	522.7	14.0	528.2	8.5
11	methylal	480.6	480.3	0.3	480.3	0.3
12	isopropylamine	471.9	474.4	−2.5	474.4	−2.5
13	*n*-propylamine	497.0	491.7	5.3	493.8	3.1
14	trimethylamine	433.3	433.2	0.0	433.2	0.0
15	1,3-butadiene	425.4	418.6	6.8	418.5	6.8
16	butyronitrile	582.3	582.1	0.1	581.9	0.4
17	1-butene	419.6	418.5	1.1	418.9	0.7
18	*cis*-2-butene	435.6	432.1	3.4	430.7	4.9
19	*trans*-2-butene	428.6	432.5	−3.8	431.0	−2.4
20	isobutene	417.9	421.2	−3.3	418.4	−0.5
21	methylethylketone	535.5	536.9	−1.4	533.7	1.8
22	tetrahydrofuran	540.2	527.3	12.8	526.0	14.2
23	*n*-butyric acid	628.0	630.4	−2.4	630.5	−2.5
24	ethylacetate	523.3	527.4	−4.1	527.9	−4.6
25	isobutyricacid	609.2	605.5	3.6	606.7	2.5
26	methylproprionate	530.6	530.5	0.1	528.6	2.0
27	*n*-propylformate	538.0	533.4	4.6	539.2	−1.2
28	*n*-butane	425.2	426.4	−1.2	426.9	−1.7
29	isobutane	408.1	414.6	−6.4	414.7	−6.6
30	butanol	562.9	563.8	−0.9	562.4	0.6
31	*sec*-butyl alcohol	536.0	539.8	−3.8	538.1	−2.1
32	*tert*-butyl alcohol	506.2	514.2	−8.0	515.1	−8.9
33	diethyl ether	466.7	478.7	−12.0	476.7	−10.0
34	isobutanol	547.7	547.7	0.1	547.6	0.1
35	*n*-butylamine	531.9	530.8	0.1	547.6	0.1
35	diethylamine	496.6	494.2	2.4	495.0	1.6
37	pyridine	620.0	620.2	−0.3	619.2	0.7
38	cyclopentene	511.8	510.5	1.3	513.5	−1.8
39	1-pentene	464.8	461.9	2.8	462.3	2.5
40	diethylketone	561.0	560.4	0.5	559.0	2.0
41	2-pentanone	561.1	562.6	−1.5	561.8	−0.7
42	ethylproprionate	546.0	547.0	−1.0	545.4	0.6
43	isobutyrlformate	551.4	549.0	2.3	550.9	0.5
44	*N*-propylacetate	549.4	547.8	1.6	549.9	−0.5
45	valeric acid	651.0	648.9	2.1	648.1	2.9
46	isopentane	460.4	457.5	2.9	457.9	2.6
47	neopentane	433.8	437.8	−4.0	436.6	−2.9
48	*N*-pentane	469.7	469.3	0.3	469.4	0.3
49	2-methyl-2-butanol	545.2	536.0	9.1	536.2	9.0
50	3-methyl-1-butanol	579.5	581.8	−2.3	580.1	−0.6
51	1-pentanol	586.2	590.2	−4.1	588.0	−1.9
52	bromobenzene	670.2	670.1	0.0	670.1	0.0
53	chlorobenzene	632.4	632.2	0.1	632.4	−0.1
54	benzene	562.2	563.0	−0.9	564.4	−2.3
55	phenol	694.3	696.5	−2.3	694.9	−0.7
56	aniline	699.0	699.2	−0.2	697.1	1.9
57	2-methylpyridine	621.0	627.2	−6.2	626.7	−5.7
58	cyclohexene	560.4	561.7	−1.3	559.0	1.4
59	cyclohexanone	629.2	627.1	2.0	628.1	1.0
60	cyclohexane	535.5	543.8	−8.2	542.1	−6.5
61	1-hexene	504.0	505.8	−1.8	506.4	−2.3
62	methylcyclopentane	532.8	536.9	−4.1	535.6	−2.8
63	cyclohexanol	625.2	635.9	−10.7	634.4	−9.3
64	2-hexanone	587.1	585.2	1.9	586.8	0.3
65	methylisobutylketone	571.4	576.9	−5.5	577.3	−5.9
66	ethyl-*n*-butyrate	571.0	566.1	4.9	566.1	4.9
67	isobutylacetate	561.0	565.8	−4.8	565.9	−4.9
68	2,2-dimethylbutane	488.8	482.4	6.4	485.1	3.7
69	2,3-dimethylbutane	500.0	488.0	12.0	488.9	11.1
70	*n*-hexane	507.4	507.3	0.2	507.9	−0.5
71	2-methylpentane	497.5	499.5	−2.0	500.4	−2.9
72	3-methylpentane	504.4	500.0	4.4	500.6	3.8

Table 4 (Continued)

					av from 5 runs	
id[a]	compound name	T_c (K)[b]	calc[c]	res[d]	calc[e]	res
73	disopropylether	500.1	503.9	−3.8	507.0	−6.9
74	di-*n*-propylether	530.6	532.0	−1.4	532.9	−2.3
75	di-isopropylamine	523.1	529.2	−6.1	526.8	−3.7
76	di-*n*-propylamine	555.8	552.5	3.3	552.4	3.4
77	benzaldehyde	695.0	695.7	−0.7	695.2	−0.2
78	toluene	591.8	592.1	−0.3	590.5	1.3
79	*m*-cresol	705.9	698.2	7.7	698.7	7.2
80	*o*-cresol	697.6	697.2	0.4	698.1	
81	*m*-toluidine	709.2	701.7	7.5	702.1	7.0
82	*o*-toluidine	694.2	702.5	−8.3	703.0	−8.8
83	ethylcyclopentane	569.5	574.6	−5.1	573.6	−4.1
84	2,3-dimethylpentane	537.4	532.6	4.7	531.6	5.8
85	*n*-heptane	540.3	542.3	−2.0	542.1	−1.9
86	2-methylhexane	530.4	538.2	−7.8	538.2	−7.8
87	3-methylhexane	535.3	538.1	−2.8	537.8	−2.6
88	ethylbenzene	617.2	619.1	−1.9	616.5	0.7
89	*m*-xylene	617.1	619.0	−2.0	620.1	−3.1
90	*p*-xylene	616.3	618.4	−2.1	619.5	−3.3
91	2,6-xylenol	701.1	700.8	0.2	702.3	−1.3
92	*N*,*N*′-dimethylaniline	687.2	687.1	0.0	687.1	0.0
93	*cis*-1,2-dimethylcyclohexane	606.2	597.2	9.0	596.1	10.1
94	*trans*-1,2-dimethylcyclohexane	596.2	597.9	−1.7	597.1	−1.0
95	*cis*-1,3-dimethylcyclohexane	591.2	598.8	−7.7	598.4	−7.3
96	*trans*-1,3-dimethylcyclohexane	598.0	598.3	−0.3	597.9	0.1
97	*trans*-1,4-dimethylcyclohexane	590.2	598.4	−8.3	598.1	−7.9
98	ethylcyclohexane	609.2	598.7	10.4	599.7	9.4
99	isobutylisobutyrate	602.0	597.7	4.3	596.5	5.5
100	2,3-dimethylhexane	563.4	564.5	−1.1	563.5	−0.1
101	2-methyl-3-ethylpentane	567.0	565.4	1.6	564.2	2.8
102	*n*-octane	568.8	566.9	1.9	567.8	1.0
103	2,2,3-trimethylpentane	563.5	565.2	−1.7	565.8	−2.3
104	2,2,4-trimethylpentane	544.0	565.9	−21.9	566.1	−22.1
105	2,3,3-trimethylpentane	573.5	563.9	9.6	564.1	9.4
106	2-ethyl-1-hexanol	640.3	630.2	10.1	636.9	3.4
107	quinoline	782.2	777.4	4.8	781.5	0.7
108	cumene	631.2	626.5	4.7	627.2	3.9
109	*o*-ethyltoluene	651.2	641.7	9.5	641.9	9.2
110	*p*-ethyltoluene	640.2	640.7	−0.6	641.0	−0.8
111	mesitylene	637.4	650.2	−12.8	649.3	−12.0
112	*n*-propylbenzene	638.4	639.6	−1.2	637.5	0.9
113	1,2,3-trimethylbenzene	664.5	651.9	12.6	650.7	13.8
114	1,2,4-trimethylbenzene	649.1	650.8	−1.7	650.2	−1.0
115	*n*-propylcyclohexane	639.2	624.1	15.0	627.7	11.5
116	3,3-diethylpentane	610.1	603.8	6.3	604.0	6.0
117	*n*-nonane	595.7	591.5	4.2	592.3	3.3
118	2,2,3,3-tetramethylpentane	660.6	658.9	1.7	659.7	0.8
119	1,2,3,4-tetrahydronaphthalene	720.2	727.3	−7.1	724.3	−4.1
120	*n*-butylbenzene	660.6	658.9	1.7	659.7	0.8
121	*p*-cymene	653.2	653.1	0.1	656.6	−3.4
122	isobutylbenzene	650.2	652.6	−2.5	655.1	−5.0
123	*n*-decane	618.5	610.9	7.6	611.9	6.5
124	*n*-tetradecane	658.2	666.7	−8.5	669.2	−11.0
125	*n*-octadecane	692.4	714.7	−22.3	714.0	−21.6
126	*n*-nonadecane	745.3	724.6	20.7	722.9	22.3
127	*sec*-butylchloride	520.6	520.8	−0.2	520.7	−0.1
128	methylisopropyl ether	464.5	464.7	−0.2	462.8	1.7
129	*sec*-butylamine	514.3	513.1	1.2	512.8	1.5
130	*tert*-butylamine	483.9	488.9	−5.0	485.2	−1.3
131	2-methyl-1-butene	460.0	457.3	2.7	458.9	1.1
132	2-methyl-2-butene	471.0	470.9	0.1	472.1	−1.1
133	3-methyl-1-butene	450.4	447.7	2.7	449.8	0.5
134	methylisopropylketone	553.0	540.7	12.3	547.3	5.7
135	methyl-*n*-butyrate	545.5	554.8	−9.3	552.0	−6.5
136	piperidine	594.1	594.0	0.1	595.1	−1.0
137	ethylpropyl ether	500.2	510.0	−9.8	507.7	−7.4
138	methyl *tert*-butyl ether	497.1	479.4	17.7	481.7	15.4
139	3-methylpyridine	645.0	639.0	6.0	639.8	5.2
140	1,5-hexadiene	507.0	509.2	−2.2	508.3	−1.3
141	hexanenitrile	622.1	622.2	−0.1	622.4	−0.3
142	3-hexanone	582.8	590.6	−7.7	586.1	−3.3
143	ethylisobutyrate	553.2	561.9	−8.7	561.2	−8.1
144	*n*-propylproprionate	578.0	568.9	9.1	568.1	9.9
145	1-hexanol	611.4	611.6	−0.3	609.8	1.6

Table 4 (Continued)

					av from 5 runs	
id[a]	compound name	T_c (K)[b]	calc[c]	res[d]	calc[e]	res
146	2-hexanol	586.2	588.9	−2.7	585.6	0.6
147	4-methyl-2-pentanol	574.4	581.9	−7.5	579.3	−4.9
148	acetone[f]	508.2	508.7	−0.5	501.2	7.0
149	2-propanol[f]	508.3	503.2	5.1	507.2	1.1
150	ethyl vinyl ether[f]	475.2	483.7	−8.6	482.4	−7.3
151	isovaleric acid[f]	634.0	642.8	−8.8	637.6	−3.6
152	*n*-butylacetate[f]	579.2	570.5	8.7	575.0	4.2
153	*p*-cresol	704.7	702.6	2.0	701.8	2.9
154	*p*-toluidine[f]	693.2	704.4	−11.2	703.7	−10.6
155	1-heptene[f]	537.3	545.1	−7.8	544.4	−7.1
156	methylcyclohexane[f]	572.2	575.2	−3.0	573.0	−0.8
157	2,2,3-trimethylbutane[f]	531.2	527.2	4.0	526.8	4.4
158	*o*-xylene	630.4	625.0	5.3	624.4	6.0
159	1-octene[f]	566.6	575.3	−8.7	574.8	−8.2
160	*m*-ethyltoluene[f]	637.2	645.1	−7.9	644.3	
161	*p*-diethylbenzene[f]	658.0	662.0	−4.0	661.5	−3.6
162	1-decene[f]	617.1	618.4	−1.3	620.4	−3.3
163	propionitrile[f]	564.4	552.7	11.7	554.8	9.6
164	4-methylpyridine[f]	646.2	646.7	−0.5	7649.7	−3.6
165	*n*-pentylformate[f]	576.0	573.8	2.2	585.1	−9.1

[a] id: The serial number for the compounds in the table. This id is the same as in ref 11 by Jurs. [b] The observed critical temperature on the Kelvin scale, T_c (K), as obtained from Jurs (11). [c] calc: The value predicted for critical temperature from the 19:4:1 neural network model. See text. [d] res: Observed critical temperature − calculated critical temperature. [e] The results from five runs on the same artificial network using a different starting set of network weights for each run. Calc is the average calculated value for five runs. Res is observed − calc See text. [f] Compounds which were used as a test set.

a large, heterogeneous data set published to date (as far as we know). The average relative error for the whole data set is 1.05% and 1.12% for the test set. These values compare very favorably with the estimated experimental errors.[11] The slope of the line in the plot of calculated versus observed (Figure 2) is 1.00 for the whole set and 0.98 for the test set; this is a very gratifying result. The value of unity for the slope indicates that over the range of input variables, there is a strong trend between calculated and observed boiling point.

In the work of Jurs et al. the objective was development of a minimal set of descriptors for modeling in addition to prediction. For that reason, direct comparison may not be useful. However, in the boiling point model the measures of error obtained by Jurs are significantly larger than obtained here, and the number of outliers with residuals above 20 K is much larger. There appears to be systematic problems with compounds containing bromine and with formate esters. Such systematic errors are not present in the atom type E-state model.

Critical Temperature. For the critical temperature model, only three compounds have residuals greater than 20 K: *n*-nonadecane, 2,2,4-trimethylpentane, and *n*-octadecane; 16 residuals lie between 10 and 20 K. In the test set, only two have residuals greater than 10 K; none are greater than 12 K. The plot of calculated versus observed critical temperature (Figure 2) represents an excellent relationship, with the points located very close to the line. The plot of the residuals versus observed critical temperature (not shown) appears to be random, showing no pattern. There seems to be no particular pattern among the types of compounds found with the residuals greater than 10 K. When the calculated critical temperature values are compared to the observed values, the plot yields a slope of 1.00 for the test set as well as for the training set. These results show that, over the range of the input values, there is a strong relationship between the calculated and observed values.

The summary results for this model are also excellent. The *MAE* for the whole data set is 4.5 and 5.6 K for the test set. The average relative error is 0.94% for the whole set and 0.97% for the test set, which compare favorably with the estimated experimental errors.[11] These averages are excellent for this data which has not been widely modeled. For comparison, Jurs et al. (who used fewer variables) found a standard deviation of 11.88 K for the training data set (147 compounds) as compared to 7.1 K found in the atom type E-state model (modeling results were not given for the test set).

For purposes of comparison, a linear model was obtained using multiple linear regression. For the same 19 input variables as used in the neural network, the standard error was found to be much larger, 23.7 K. Further, there were many more large residuals, including 49 with residual larger than 20 K; eight residuals were larger than 40 K. Thus, it appears that the nonlinear nature of the neural network model is a better vehicle for modeling boiling point. Further, the improvement over the linear model is even greater for critical temperature than for boiling point.

Another investigation of the neural network modeling was performed. For each data set, several runs were made, and the calculated values were averaged; ten runs for boiling point and five runs for the critical temperature. The average predicted values from these runs are given in Tables 3 and 4 together with the residual computed as observed value − average predicted value, along the right hand side of each table. In both cases the residuals from the average predicted values tend to be consistent with the residual values from the best run. In general, the average relative error is somewhat smaller for the average of the several (ten or five) runs.

CONCLUSIONS

The atom type E-state indices are readily computed with available software. Each index represents the electrotopo-

logical state of a particular atom type in the molecular structure. The E-state value indicates the electron accessibility at that atom in the molecular structure. The atom type E-state indices combine the power of the structure representation inherent in the electrotopological state with the general applicability of the atom typing approach.

For properties such as the boiling point and critical temperature, the electronic structure information of the E-state is related to the intermolecular forces present. For this data set, three types of intermolecular forces are considered present: dispersion, dipolar, and hydrogen bonding. Further, there are many structure types present in this heterogeneous data set, related to more than a dozen functional groups. The high quality of the QSAR models for both boiling point and critical temperature clearly demonstrate the strong modeling ability of the atom type E-state indices and confirms that the E-state is a powerful representation of molecular structure. The artificial neural network models developed here also seem to be very useful in modeling properties such as boiling point and critical temperature. We are continuing to explore the usefulness of artificial neural networks for modeling and the atom type electrotopological state indices as structure descriptors for heterogeneous data sets on physicochemical properties and for biological data sets as well.

ACKNOWLEDGMENT

We wish to give grateful acknowledgment to the DuPont Chemical Corporation for a grant which provided partial support for this research investigation.

REFERENCES AND NOTES

(1) Kier, L. B.; Hall, L. H. An Atom-Centered Index for Drug QSAR Models. In *Advances in Drug Design*; Testa, B., Ed.; Academic Press: 1992; Vol. 22.

(2) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.

(3) Kier, L. B.; Hall, L. H. An Electrotopological State Index for Atoms in Molecules. *Pharm. Res.* **1990**, *7*, 801−807.

(4) Hall, L. H.; Mohney, B. K.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76−82.

(5) Hall, L. H.; Kier, L. B. An Index of Electrotopological State for Atoms in Molecules. *J. Math Chem.* **1991**, *7*, 229−241.

(6) Hall, L. H.; Mohney, B. K.; Kier, L. B. The Electrotopological State: An Atom Index for QSAR. *Quant. Struc.−Act. Relat.* **1991**, *10*, 43−51.

(7) Kier, L. B.; Hall, L. H. An Index of Atom Electrotopological State. In *QSAR in Design of Bioactive Compounds, A Telesymposium*; Biaggi, A.; Ed.; J. R. Prous Publishers: 1992.

(8) Hall, L. H.; Mohney, B. K.; Kier, L. B. Comparison of Electrotopological State Indices with Molecular Orbital Parameters: Inhibition of MAO by Hydrazides. *Quant. Struct.-Act. Relat.* **1993**, *12*, 44−48.

(9) Hall, L. H.; Kier, L. B. Binding of Salicylamides: QSAR Analysis with Electrotopological State Indexes. *Med. Res. Rev.* **1992**, *2*, 497−502.

(10) Hall, L. H.; Kier, L. B.; Brown, B. B. Molecular Similarity Based on Novel Atom Type Electrotopological State Indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1074−1080.

(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) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH, Weinheim, 1993.

(13) Feuilleaubois, E.; Fabart, V.; Doucet, J. P. Implementation of the Three-Dimensional-Pattern Search Problem on Hopfield-like Neural Networks. *SAR QSAR Environ. Res.* **1993**, *1*, 97−114.

(14) Gakh, A.; Gakh, E.; Sumpter, B.; Noid, D. Neural Network-Graph Theory Approach to the Prediction of the Physical Properties of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 832−839.

(15) Xu, L.; Ball, J.; Dixon, S.; Jurs, P. Quantitative Structure-Activity Relationships for Toxicity of Phenols Using Regression Analysis and Computational Networks. *Environ. Toxicol. Chem.* **1994**, *13*, 841−851.

(16) Molconn-Z (Molconn-X ver 3.0); Hall Associates: 2 Davis Street, Quincy, MA 02170.

(17) Neuralyst; Cheshire Engineering Corp.: 650 Sierra Madre Villa Avenue, Pasadena, CA 91107.

CI960375X