Estimation of Critical Properties with Group Contribution Methods

Currently available group-contribution methods for T_c, P_c, and V_c were evaluated using the Ambrose (1980) critical property data compilation. The Ambrose estimation methods were found to be the most accurate. Linear regression methods were also employed to develop alternate estimation methods which were found to have an accuracy comparable to those of Ambrose.

K. M. KLINCEWICZ and
R. C. REID
Department of Chemical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

SCOPE

Critical properties (T_c, P_c, V_c) are often required when using generalized property estimation methods. Due to the paucity of experimental values, one must often estimate these properties, and a group-contribution approach is generally used. The method developed by Lydersen (1955) is the most widely employed. However, many new experimental critical properties have become available since the Lydersen report and several alternate estimation schemes have been proposed. In the first portion of this paper, a few methods have been briefly reviewed and evaluated for accuracy using the currently available data.

With the rapid advances in linear regression programming, it was felt to be of value to use this methodology to optimize the coefficients and group contributions utilizing a number of objective functions involving T_c, P_c, or V_c. In this manner, one can rapidly test a number of objective functions, selected group-contribution substructures, auxiliary variables, etc. to obtain the least error method of estimation.

CONCLUSIONS AND SIGNIFICANCE

Evaluating all available estimation methods for T_c, P_c, and V_c, it was shown that the methods of Ambrose (1979, 1980b) were the most accurate. These methods employ $T_b/(T_c - T_b)$, $(MW/P_c)^{1/2}$, and V_c as objective functions (Q) with the equation forms given in Table 1 and the necessary group contributions in Table 3. The statistics relating to expected errors are shown in Table 5.

Linear regression techniques also provided group estimation methods close in accuracy to those of Ambrose but with a simpler set of structural subgroups. In this type of analysis, multicollinearity was identified as a serious problem and the final choice of subgroups minimized the difficulties encountered. The new proposed estimation equations are given as Eqs. 10 through 12 with group contributions in Table 6. Expected errors are shown in Table 7.

Models not involving the concept of group contributions were able to explain a large portion of the variance in the critical property data. Reasonably accurate estimation methods could be found where T_c, $(MW/P_c)^{1/2}$, and V_c were correlated only with T_b, molecular weight and the number of atoms in the compound. These preliminary results suggest that factor analysis may provide a viable alternate route to group-contributions in the estimation of critical properties.

The properties of one material are often related to properties of another by assuming the respective intermolecular potentials of both are similar if the energy and separation variables are scaled appropriately. Materials that behave in this manner are called *conformal substances* (Rowlinson, 1969). Utilizing relationships between intermolecular potentials, the equation of state for two conformal substances may then be expressed as

$$\mathcal{F}(P_\alpha/P^*, T_\alpha/T^*, V_\alpha/V^*) = \mathcal{F}(P_\beta, T_\beta, V_\beta) \tag{1}$$

where subscripts α and β refer to two different conformal substances and P^*, T^*, and V^* are appropriate scaling factors. Equation 1 is the *principle of corresponding states*. As Rowlinson notes, given the P-V-T surface for one member (β) of a conformal pair, the P-V-T surface for the other member (α) is geometrically similar with the P-V-T axes multiplied by P^*, T^*, and V^* respectively. "It follows that all singular points on the surface, such as the solid, liquid, and gas at the triple point and the fluid at the critical point have values of P, T, and V in the ratio of these scale factors."

The scale factors most commonly employed by engineers relate to the critical points, thus with $P^* = P_{c_\alpha}/P_{c_\beta}$, $T^* = T_{c_\alpha}/T_{c_\beta}$, and $V^* = V_{c_\alpha}/V_{c_\beta}$, Eq. 1 could be rewritten as

$$\mathcal{F}[(P/P_c)_\alpha, (T/T_c)_\alpha, (V/V_c)_\alpha] = \mathcal{F}[(P/P_c)_\beta, (T/T_c)_\beta, (V/V_c)_\beta] \tag{2}$$

Equations of the form of Eq. 2 may also be obtained from taking simple equations of state such as the Redlich-Kwong (1949) and applying the critical criteria

$$(\partial P/\partial V)_T = (\partial^2 P/\partial V^2)_T = 0 \qquad \text{at } T_c, P_c, V_c \tag{3}$$

While additional characterizing parameters can be introduced into Eq. 2 to account for *non*conformal materials (e.g., the critical compressibility factor or the Pitzer acentric factor), it is still a prerequisite to know the critical properties of a material to employ such generalized equations of state. Also, for mixtures, it is common to relate *mixture* parameters to composition, pure component critical properties (and other characterization factors), as well as binary interaction terms.

Despite their significance in generalized property correlations, it is a discouraging fact that critical properties have been experimentally determined for relatively few pure substances. The most complete compilation of known critical properties is that developed by Ambrose (1980a).

If one is not fortunate enough to find reliable experimental values

TABLE 1. SUMMARY OF PREVIOUS KEY GROUP ESTIMATION TECHNIQUES FOR CRITICAL PROPERTIES†

$$Q = a + \sum_{j=1}^{M} n_j \Delta_j + b \left[\sum_{j=1}^{M} n_j \Delta_j \right]^2$$

Author	$\hat{Q}$ Function	a	b	Comments
Critical Temperature				
Riedel (1952)	T_b/T_c	0.574	0	
Eduljee (Gambill, 1959)	T_b/T_c	0	0	
Fedors (1982)	$10^{T_c/535}$	0	0	Table 4
Lydersen (1955)	T_b/T_c	0.567	−1	Table 2
Ambrose (1980b)†††	$T_b/(T_c - T_b)$	1.242††	0	Table 3
Critical Pressure				
Riedel (1949)	$(MW/P_c)^{1/2}$	0.32	0	
Eduljee (Gambill, 1959)	$(MW/P_c)^{1/2}$	0	0	
Lydersen (1955)	$(MW/P_c)^{1/2}$	0.33	0	Table 2
Ambrose (1979)	$(MW/P_c)^{1/2}$	0.339††	0	Table 3
Critical Volume				
Fedors (1979)	V_c	26.6	0	
Vetere (1976)	$(V_c - 33)^{0.9728}$	0	0	
Lydersen (1955)	V_c	40	0	Table 2
Ambrose (1979)	V_c	40	0	Table 3

† All correlations have been changed so that temperatures are in K, pressures in bar, and volumes in cm^3/mol. MW is the molecular weight.

†† For perfluorinated compounds or with monohalogenated fluorocarbons, change a to 1.570 for T_c and to 1.000 for P_c.

††† See also Ambrose et al. (1974).

of critical properties for the fluids of interest, recourse is made to *estimation* methods. In almost every instance, such methods employ a group contribution approach.

In this technique, it is assumed that some property of a substance, Q, is a function of the *structure* of the molecule, e.g., the number and types of chosen molecular substructures each of which is assigned a numerical value. This approach has been widely used for *ideal-gas* properties such as heat capacities, enthalpies of formation, etc. (Reid et al., 1977). For such properties, intermolecular forces play no role. The concept has also been used to estimate molar volumes of pure liquids at their normal boiling point and, of particular interest here, the critical properties T_c, P_c, and V_c.

TABLE 2. LYDERSEN GROUP CONTRIBUTIONS†

	Δ_j Values for		
	T_c	P_c	V_c
$-CH_3$, $-CH_2-$	0.020	0.226	55
$-CH_2-$ (ring)	0.013	0.183	44.5
$>CH-$	0.012	0.209	51
$>CH-$ (ring)	0.012	0.191	46
$>C<$	0.000	0.209	41
$>C<$ (ring)	−0.007	0.153	31
$=CH_2$, $=CH-$	0.018	0.197	45
$=CH-$ (ring)	0.011	0.153	37
$>C=$, $=C=$	0.000	0.197	36
$>C=$ (ring)	0.011	0.153	36
$\equiv CH$, $\equiv C-$	0.005	0.152	36
$-F$	0.018	0.223	18
$-Cl$	0.017	0.328	49
$-Br$	0.010	0.50	70
$-I$	0.012	0.83	95
$-O-$	0.021	0.16	20
$-O-$ (ring)	0.014	0.12	8
$-OH$ (alcohols)	0.082	0.06	18
$-OH$ (phenols)	0.035	−0.02	3
$>CO$	0.040	0.29	60
$>CO$ (ring)	0.033	0.2	50
$-CHO$	0.048	0.33	73
$-CO-O-$	0.047	0.47	80
$-COOH$	0.085	0.4	80
$-NH_2$	0.031	0.094	28
$>NH$	0.031	0.134	37
$>NH$ (ring)	0.024	0.09	27
$>N-$	0.014	0.17	42
$>N-$ (ring)	0.007	0.13	32
$>CN$	0.060	0.36	80
$-SH$, $-S-$	0.015	0.27	55
$-S-$ (ring)	0.008	0.24	45
$=S$	0.003	0.24	47
$-NO_2$	0.005	0.42	78
$>Si<$	0.03	0.54	—
$>B-$	0.03		

† Free bonds are attached to nonhydrogen atoms.

The assignation of group values may be as simple as atomic weights if the property Q is the molecular weight. On the other hand, group contributions may relate to a particular grouping of atoms in a specified environment of other atoms. The group-estimation methods developed by Benson and his coworkers would be of this latter type where, for example, for each carbon atom, the type and number of adjoining atoms must be specified (Reid et al., 1977).

For critical property estimation, however, rather simple molecular groups are normally proposed, e.g., $-CH_3$, $-NH-$, etc., but correction terms are often included to allow for conjugation, ortho substitutions, and other special structural entities.

REVIEW OF GROUP CONTRIBUTION METHODS FOR T_c, P_c, V_c

There have been many techniques proposed to estimate critical properties; only a few of the more important are noted here and, of these, only a subset is evaluated to delineate the accuracy.

In all cases, the estimated critical property—or some function of the property, is denoted as $\hat{Q}$. Group contribution methods may then be expressed as:

$$\hat{Q} = a + \sum_{j=1}^{M} n_j \Delta_j + b \left[\sum_{j=1}^{M} n_j \Delta_j \right]^2 \tag{4}$$

Δ_j is the contribution for group j, n_j is the number of such groups and M is the total number of different types of groups. $\hat{Q}$, a, and b are shown for several well-known methods in Table 1. The Lydersen and Ambrose techniques are the most accurate and group contributions for these are shown in Tables 2 and 3. The Fedors (1982) method for critical temperature requires no boiling-point temperature. As this may be a real advantage in some instances, his group contributions are given in Table 4.

These group contribution methods were tested using the data in the Ambrose (1980a) compilation and the results shown in Table 5. The Ambrose approach gives significantly more accurate predictions than Lydersen's. Fedor's shows a high deviation, but this might have been expected since no normal boiling point was included.

TABLE 3. AMBROSE GROUP CONTRIBUTIONS[†]

	Δ_j Values for		
	T_c	P_c	V_c
Carbon atoms in alkyl groups	0.138	0.226	55.1
corrections:			
>CH— (each)	−0.043	−0.006	−8
>C< (each)	−0.120	−0.030	−17
double bonds (nonaromatic)	−0.050	−0.065	−20
triple bonds	−0.200	−0.170	−40
delta Platt number,[1] multiply by	−0.023	−0.026	—
Aliphatic functional groups			
—O—	0.138	0.160	20
>CO	0.220	0.282	60
—CHO	0.220	0.220	55
—COOH	0.578	0.450	80
—CO—O—OC—	1.156	0.900	160
—CO—O—	0.330	0.470	80
—NO$_2$	0.370	0.420	78
—NH$_2$	0.208	0.095	30
—NH—	0.208	0.135	30
>N—	0.088	0.170	30
—CN	0.423	0.360	80
—S—	0.105	0.270	55
—SH	0.090	0.270	55
—SiH$_3$	0.200	0.460	119
—O—Si(CH$_3$)$_2$	0.496	—	—
—F	0.055	0.223	14
—Cl	0.055	0.318	45
—Br	0.055	0.500	67
—I	0.055	—	90
Halogen correction in aliphatic compounds:			
F is present	0.125	—	—
F is absent, but Cl, Br, I present	0.055	—	—
Aliphatic alcohols[2]	Footnote 3	Footnote 4	15
Ring compound increments (only listed when different from aliphatic values)			
—CH$_2$—	0.090	0.182	44.5
>CH in fused ring	0.030	0.182	44.5
double bond	−0.030	—	−15
—O—	0.090	—	10
—NH—	0.090	—	—
—S—	0.090	—	30
Aromatic compounds			
benzene	0.448	0.924	Footnote 5
pyridine	0.448	0.850	
C$_4$H$_4$ (fused as in naphthalene)	0.220	0.515	
—F	0.080	0.183	
—Cl	0.080	0.318	
—Br	0.080	0.600	
—I	0.080	0.850	
—OH	0.198	−0.025	
corrections for non-halogenated substituents			
first	0.010	0	
each subsequent	0.030	0.020	
ortho pairs containing —OH	−0.080	−0.050	
ortho pairs with no —OH	−0.040	−0.050	
Highly fluorinated aliphatic compounds			
—CF$_3$, —CF$_2$—, >CF—	0.200	0.550	
—CF$_2$—, >CF— (ring)	0.140	0.420	
>CF— (in fused ring)	0.030	—	
—H (monosubstitution)	−0.050	−0.350	
double bond (nonring)	−0.150	−0.500	
double bond (ring)	−0.030	—	
(other increments as in nonfluorinated compounds).			

[†] Free bonds are attached to nonhydrogen atoms.

[1] The delta Platt number is defined as Platt number (isomer)-Platt number (n-alkane) where the Platt number is the number of pairs of carbon atoms three bond apart (Platt, 1947, 1952).

[2] Includes naphthenic alcohols and glycols but not aromatic alcohols such as xylenol.

[3] First determine the hydrogen homomorph, i.e., substitute —CH$_3$ for each —OH. Calculate $\hat{Q}$ for this compound. Then, subtract 0.138 for each —OH substituted. Next, add

$$(0.87 - 0.11\,n + 0.003\,n^2)$$

where

$$n = [T_b(\text{alcohol}, K) - 314.1]/19.2$$

Exceptions include methanol ($\hat{Q} = 2.142$), ethanol ($\hat{Q} = 2.181$), and any alcohol whose value of n exceeds 10.

[4] Determine the hydrocarbon homomorph as in footnote 3. Calculate $\hat{Q}$. Subtract 0.226 for each —OH substituted. Add $0.100 - 0.013\,n$ where n is calculated from T_b as in footnote 3.

[5] When estimating critical volumes of aromatic substances, use alkyl group values.

TABLE 4. FEDORS GROUP CONTRIBUTIONS FOR T$_c$[†]

	Δ_j Values
—CH$_3$	1.79
—CH$_2$—	1.34
>CH—	0.45
(except for adjacent pairs of >CH—, then, for each)	0.76
>C<	−0.22
=CH$_2$	1.59
=CH—	1.40
>C=	0.89
≡CH—	1.79
≡C—	2.46
=C=	1.03
—CO—OH	10.72
—CO—O—OC— (anhydride)	7.95
—CO—O—	5.32
—O—OC—CO—O— (oxalate)	6.25
—CO—	5.36
—O—	1.56
—O— (aromatic)	2.68
—OH	5.63
—OH (aromatic)	9.65
—CHO	5.49
—C≡N	8.49
—C≡N (aromatic)	9.38
—NH$_2$	4.56
—NH$_2$ (aromatic)	9.20
—NH—	3.04
—NH— (aromatic)	7.64
>N—	0.89
>N— (aromatic)	4.74
—N=	4.51
—S—S—	9.83
—S—	4.91
—SH	5.36
—F	2.10
—F (aromatic)	0.45
—F (perfluoro)	0.54
—Cl	4.20
—Cl (disubstituted)	3.71
—Cl (trisubstituted)	3.17
—Br	5.58
—I	8.04
—I (aromatic)	10.77
3-membered ring	0.45
5-membered ring	2.23
6-membered ring	2.68
heteroatom in ring	0.45
substitution on carbon in a double bond (nonaromatic)	0.58
ortho substitution in a benzene ring	1.16
conjugation, per double bond	0.13

[†] Free bonds refer to nonhydrogen atoms.

PROPOSED ESTIMATION METHODS

Following an evaluation of current group contribution methods, we felt it to be worthwhile to try and improve on the methods using linear regression techniques to aid in selecting the optimum contributions for atomic groups and, also, to delineate the dependence of a given property on other available data such as molecular weights, number of atoms, etc.

In our work, we first selected a subset of the critical property compilation by Ambrose (1980a) which contained 199 different organic compounds. This subset was chosen because, for each

TABLE 5. COMPARISON OF PREVIOUS GROUP CONTRIBUTION METHODS

Author	Property	Table No.	Deviation[†] Mean	Deviation[†] 95%[††]	% Error Mean	% Error 95%	No. of Compounds Tested
Lydersen	T_c	2	8.2	31	1.4	5.1	396
Fedors		4	15	—	4	—	199
Ambrose		3	4.3	19	0.7	3.7	398
Lydersen	P_c	2	3.3	16	8.9	38	288
Ambrose		3	1.8	6	4.7	13	290
Lydersen	V_c	2	10	34	3.1	10	205
Ambrose		3	8.5	32	2.8	9.6	207

[†] Deviations are given in K for T_c, bar for P_c, and cm^3/mol for V_c, Eq. 8.
[††] 95% signifies the range within which there is a 95% probability of finding the true value, Eq. 9.

TABLE 6. PROPOSED GROUP CONTRIBUTIONS FROM REGRESSION ANALYSIS[†]

	Δ_j Values for T_c Eq. 10	Δ_j Values for P_c Eq. 11	Δ_j Values for V_c Eq. 12
—CH$_3$	−2.433	0.026	16.2
—CH$_2$—	0.353	−0.015	16.1
—CH$_2$— (ring)	4.253	−0.046	8.2
>CH—	6.266	−0.083	12.1
>CH— (ring)	−0.335	−0.027	7.4
>C<	16.416	−0.136	8.95
>C< (ring)	12.435	−0.111	−6.6
=CH$_2$	−0.991	−0.015	13.9
=CH—	3.786	−0.050	9.8
=CH— (ring)	3.373	−0.066	5.1
>C=, =C=	7.169	−0.067	2.7
>C= (ring)	5.623	−0.089	0.2
≡CH	−4.561	−0.056	7.5
≡C—	7.341	−0.112	3.0
—OH	−28.930	−0.190	−24.0
—O—	5.389	−0.143	−26.1
—O— (ring)	7.127	−0.116	−36.6
>CO, —CHO	4.332	−0.196	−6.7
—COOH	−25.085	−0.251	−37.0
—CO—O—	8.890	−0.277	−28.2
—NH$_2$	−4.153	−0.127	−0.1
>NH	2.005	−0.180	53.7
>NH (ring)	2.773	−0.172	−8.0
>N—	12.253	−0.163	−0.7
=N— (ring)	8.239	−0.104	−18.4
—CN	−10.381	−0.064	12.0
—SH	28.529	−0.303	−27.7
—S—	23.905	−0.311	−27.3
—S— (ring)	31.537	−0.208	−61.9
—F	5.191	−0.067	−34.1
—Cl	18.353	−0.244	−47.4
—Br	53.456	−0.692	−148.1
—I	94.186	−1.051	−270.6
—XCX (X = halogen)[††]	−1.770	0.032	0.8
—NO$_2$	11.709	−0.325	−39.2

[†] Free bonds are attached to nonhydrogen atoms.
[††] The number of possible pairwise interactions of halogen atoms attached to the same carbon, e.g., —CHF$_2$ contains one XCX while —CCl$_3$ contains three XCX values (Cramer, 1980).

compound, an experimental value was available for all three critical properties, T_c, P_c, and V_c. From this subset, we obtained our group contributions. After selecting the most accurate, error analyses were then extended to cover all the data given by Ambrose. The results could then be compared with existing techniques such as those of Ambrose or Lydersen.

Our first task involved the selection of atomic groups to be used. We benefited greatly from the experience of earlier workers, but we also desired to choose groups which were simple in nature and led to unambiguous molecular structure representations. Many listings were prepared and tested.

One problem found in certain choices related to multicollinearity. That is, if two or more groups are highly correlated with each other, a regression may lead to a result which fits the data set well, but the coefficients of the correlation will have large standard deviations. The accuracy of the result is then reduced when predicting properties not included in the original data set (Freund and Minton, 1979). As a simple example, consider a group contribution scheme for a set of alkanes which are structurally represented by —CH$_3$, —CH$_2$—, >CH—, and >C<. The number of quaternary carbon atoms is highly correlated with the number of chain ends in the alkane, represented by —CH$_3$. Though the groups relate to real structures, they are not random variables because all combinations are not possible. Each molecule must have two ends. Additional ends indicate tertiary or quaternary carbons are present. The variables —CH$_3$ and >C< become less correlated when additional classes of molecules are included because —CH$_3$ is no longer the only chain end and >CH— and >C< no longer are the only types of branching structures. For alkanes, more accurate group contribution methods are usually obtained when only three group structures are used.

Another example is the differentiation between aromatic and non-aromatic substituents. Structures such as —Br (aromatic) are highly correlated with structures such as (=CH—)$_{\text{ring}}$ or (=C<)$_{\text{ring}}$ which comprise the aromatic entity. A more general variable such as the sum of all bromine containing structures might have yielded better results, but the additional complexity was not warranted.

As a final example, consider structures containing the group ≡N. There are few cases where this group is not attached to ≡C. If a regression sample contains a reasonable ratio of nitriles to alkynes, the variable ≡N will be highly correlated with ≡C and the alkynes will not be fairly represented. It would be preferable to restructure to provide two groups, —C≡N and ≡C— (nonnitrile). Similarly it is better to define an *ester* group as —COO— rather than the sum of =C, =O, —O— because the former relates to a particular set of compounds with similar chemical structures.

In devising group contributions, it is good practice to examine changes in the variance explained by each structure in the model. Structures which explain only a small fraction of the total variance are not good choices. Another measure of a group's significance is the t-ratio, i.e., the ratio of the structure's regression coefficient divided by the standard deviation of the coefficient. t-ratios less than the t-value for the degrees of freedom of the residual indicate that the coefficient is not statistically different from zero. In most cases considered in this work, the number of degrees of freedom was sufficiently large that the test t-value was about two.

The groups finally selected are given in Table 6.

In developing a correlation equation, we began with the general form:

$$\hat{Q} = a_1 + a_2 \times (MW) + a_3 \times A + a_4 \times T_b + a_5 \times (MW)^{-1} + a_6 \sum_{j=1}^{M} n_j \Delta_j \quad (5)$$

where $\hat{Q}$ is the estimated critical property (or some function of this property). In Eq. 5, a_4 and a_5 were set equal to zero for all tests where $\hat{Q}$ related to critical pressure and volume. (They could be, and were, set equal to zero in many of the tests of critical temperature estimations.)

TABLE 7. TEST RESULTS FOR PROPOSED METHOD

Condition	Property	Eq. No.	Deviation[†] Mean	Deviation[†] 95%[††]	% Error Mean	% Error 95%	R^2 (adj.)	No. of Compounds
Base Set[†††]	T_c	10	3.5	10	0.7	2.1	99.7	199
	P_c	11	2.2	6.9	5.1	14	99.1	199
	V_c	12	8.9	26	3.0	8.6	98.6	199
All Compounds	T_c	10	7.5	26	1.2	4.3	—	398
	P_c	11	3.0	9.5	7.8	25	—	290
	V_c	12	8.9	26	2.9	8.5	—	207

[†] Deviations are given in K for T_c, bar for P_c, and cm^3/mol for V_c, Eq. 8.

[††] 95% signifies the range within which there is a 95% probability of finding the true value, Eq. 9.

[†††] The base set of compounds consisted of those wherein experimental values were available for T_c, P_c, and V_c.

The property function $\hat{Q}$ was varied widely to test many suggestions of earlier workers. For example, when studying critical temperature, the various $\hat{Q}$ functions included T_c, $T_c^{1/2}$, $T_c^{-1/2}$, T_b/T_c, $(T_b/T_c)^{1/2}$, $(T_b/T_c)^2$, $(T_c \times MW)$, $(T_c \times MW)^{1/2}$, MW/T_c, $(T_c/MW)^{1/2}$, $(T_c/MW)^2$.

The coefficient a_6 was set equal to zero or unity. The choice of zero simply eliminated any contribution to $\hat{Q}$ from individual group contributions.

Once the choice of $\hat{Q}$ had been made as well as the selection of any zero coefficients, linear regression programs were carried out to determine the values of the nonzero coefficients and group contributions which minimized the sum of squares of the differences between the *true* values and *predicted* values of $\hat{Q}$.

A measure of how well any selected format of Eq. 5 fitted the data was the squared multiple-regression coefficient, R^2 (Ryan, 1980). ($R^2 = 100$ indicated a perfect fit while $R^2 = 0$ would mean there was no linear correlation).

$$R^2 = 100\left[1 - \left(\frac{\sum_{j=1}^{k}(Q-\hat{Q})^2}{\sum_{j=1}^{k}(Q-\overline{Q})^2}\right)\right] \tag{6}$$

$\hat{Q}$ is the estimated Q-function, Q is the true value, and $\overline{Q}$ the mean value. The index k indicates the number of compounds tested. Unfortunately, R^2 is a function of the number of groups, M, employed in the test procedure. (For example, with the final choice of groups as shown in Table 6, $M = 35$.)

Increasing M may cause R^2 to be larger even if the additional variables are not significant. Another statistic, R^2 (adj.), eliminates this bias and was employed in this study.

$$R^2(\text{adj.}) = 100\left(1 - \frac{(k-M)\sum_{j=1}^{k}(Q-\hat{Q})^2}{(k-1)\sum_{j=1}^{k}(Q-\overline{Q})^2}\right) \tag{7}$$

$(k - M)$ equals the degrees of freedom of the residual. R^2(adj.) provides a valuable criterion when evaluating two group contribution schemes with the same true data (Q) but different values of M.

The other statistics used to compare the methods were the absolute values of both the residuals $[|r| = |Q - \hat{Q}|]$ and the percent deviation $[|d| = 100 \times |Q - \hat{Q}|/Q]$. The mean value of $|r|$ and the 95% confidence limits were found by

$$|\bar{r}| = (1/k)\sum_{j=1}^{k}|r_j| \tag{8}$$

$$|r|_{95\%} = 1.96\left[\frac{1}{k-1}\sum_{j=1}^{k} r_j^2\right]^{1/2} \tag{9}$$

with similar expressions used for $|\bar{d}|$ and $|d|_{95\%}$.

The results of many tests are given elsewhere (Klincewicz, 1982). The most accurate correlations are shown below in Eqs. 10 through 12.

$$T_c = 45.40 - 0.77 \times MW + 1.55 \times T_b + \sum_{j=1}^{35} n_j\Delta_j \tag{10}$$

$$(MW/P_c)^{1/2} = 0.335 + 0.010 \times MW + \sum_{j=1}^{35} n_j\Delta_j \tag{11}$$

$$V_c = 25.2 + 2.80 \times MW + \sum_{j=1}^{35} n_j\Delta_j \tag{12}$$

where the units are Kelvin, bar, cm^3/mol, and, for MW, g/mol. The contributions for Δ_j are shown in Table 6.

The summary of the test results of the proposed methods is given in Table 7; these may be compared with similar test results given in Table 5 for earlier studies. The proposed methods are more accurate than those given by Lydersen but less accurate (except for V_c) than those of Ambrose. The differences are in most instances small. Note that the residuals $|\bar{r}|$ and $|r|_{95\%}$ are expressed in Kelvin, bar, and cm^3/mol for T_c, P_c and V_c.

CONCLUDING REMARKS

In this study we employed group contribution methods to estimate the critical properties T_c, P_c, and V_c. A wide range of property functions ($\hat{Q}$) was tested and linear regression methods allowed us to determine optimum coefficients and group contribution values. The results were encouraging but, in most instances, the method of Ambrose, though slightly more complicated, led to less error than the best of the correlations developed here.

One interesting result of the linear regression procedure was the finding that some correlations which neglected entirely any group contributions still provided quite reasonable estimations. For example, in considering critical temperature, using Eq. 5 with $a_6 = 0$, the best correlation was found to be

$$T_c = 50.2 - 0.16 \times MW + 1.41 \times T_b \tag{13}$$

This simple form gave an R^2 (adj.) of 97.8 and a $|r|_{95\%}$ of 24.5 K for the 199 test compounds. The best equation using group contributions was Eq. 10 with an R^2 (adj.) of 99.7. Thus by adding 35 more variables ($k = 35$), an increase of only 1.9 was obtained for R^2 (adj.). In Eq. 10 only ten of the 35 variables gave t-ratios greater than 2 and only six of the 35 variables added more than 0.1 to R^2 (adj.). In other words, the inclusion of specific group contributions increased the overall predictive accuracy only slightly. In a similar manner for critical pressures and volumes, we found

$$(MW/P_c)^{1/2} = 0.335 + 0.009 \times MW + 0.019\,A \tag{14}$$

$$V_c = 20 + 0.088 \times \text{MW} + 13.4 \times A \tag{15}$$

gave R^2 (adj.) values of 95.1 and 95.3 and $|r|_{95\%}$ values of 14.6 bar and 47 cm^3/mol respectively. Again, the inclusion of 35 more variables in the regression led to an increase of only 2 to 3 in R^2 (adj.).

It will also be noted that molecular weight (MW) was used as a variable in all of the better correlations we tested. Elimination of this variable always led to lower R^2 (adj.) values. Inclusion of the number of atoms (A) in a molecule as a variable usually increased the accuracy of a method. Inclusion of group contributions also improved the correlations, but in only a marginal manner.

The importance of variables such as MW, A, T_b suggests that a viable alternate approach to the estimation of critical properties would be through *factor analysis* in which critical properties are related directly to other measured properties. This approach will be examined in a later paper.

NOTATION

a = constant in Eq. 4

$a_1, \ldots, a_6$ = constants in Eq. 5

A = number of atoms in the molecule

b = constant in Eq. 4

$|\bar{d}|$ = absolute mean % deviation between experimental and estimated value

$|d|_{95\%}$ = absolute value of 95% confidence limit of $|d|$

k = number of compounds tested in any estimation method

M = number of structural groups used in an estimation method

MW = molecular weight

n = parameter in Ambrose's methods, Table 3

n_j = number of structural groups of type j in molecule

P = pressure, bar; P_c, critical pressure, bar; P^*, scaling pressure, dimensionless

Q = experimentally determined variable

$\hat{Q}$ = estimated value

$\bar{Q}$ = mean value of Q

$|r|$ = residual value of estimate, $|Q - \hat{Q}|$

$|\bar{r}|$ = absolute value of mean residual, Eq. 8

$|r|_{95\%}$ = absolute value of 95% confidence limits of residual, Eq. 9

R^2 = squared multiple-regression coefficient, Eq. 4

R^2(adj.) = adjusted value of R^2, Eq. 7

T = temperature, K; T_c, critical temperature, K; T_b, normal boiling temperature, K; T^*, scaling temperature, dimensionless

V = specific volume, cm^3/mol; V_c, critical volume, cm^3/mol; V^*, scaling volume, dimensionless

X = halogen atom

Greek Letters

Δ_j = contribution for group j

LITERATURE CITED

Ambrose, D., B. E. Broderick, and R. Townsend, "The Critical Temperatures and Pressures of Thirty Organic Compounds," *J. Appl. Chem. Biotechnol.*, **24,** 359 (1974).

Ambrose, D., "Correlation and Estimation of Vapour-Liquid Critical Properties. I. Critical Temperatures of Organic Compounds," National Physical Laboratory, Teddington, NPL Report Chem 92 (Sept., 1978/Corrected Mar., 1980b).

Ambrose, D., "Vapour-Liquid Critical Properties," National Physical Laboratory, Teddington, NPL Report Chem 107 (Feb., 1980a).

Ambrose, D., "Correlation and Estimation of Vapor-Liquid Critical Properties. I. Critical Temperatures of Organic Compounds," National Physical Laboratory, Teddington, NPL Report 92 (Sept., 1978/Corrected Mar., 1980b).

Cramer, R. D., "BC(DEF) Parameters. 1. The Intrinsic Dimensionality of Intermolecular Interactions in the Liquid State;" "2. An Empirical Structure-Based Scheme for the Prediction of Some Physical Properties," *J. Am. Chem. Soc.*, **102,** 1837, 1849 (1980).

Fedors, R. F., "A Method to Estimate Critical Volumes," *AIChE J.*, **25,** 202 (1979).

Fedors, R. F., "A Relationship Between Chemical Structure and the Critical Temperatures," *Chem. Eng. Comm.*, **16,** 149 (1982).

Freund, R. J., and P. D. Minton, "Regression Methods—A Tool for Data Analysis," Marcel Dekker Inc., New York, 118, 185 (1979).

Gambill, W. R., "Predict Critical Temperature," *Chem. Eng.* **66,** (12), 181 (1959); *ibid*, "Predict Critical Pressure," *Chem. Eng.*, **66,** (14), 157 (1959).

Klincewicz, K. M., "Prediction of Critical Temperatures, Pressures, and Volumes of Organic Compounds from Molecular Structure," S. M. Thesis, M.I.T. (1982).

Lydersen, A. L., "Estimation of Critical Properties of Organic Compounds," Univ. of Wisconsin, *Eng. Exp. Sta.* Report 3, Madison (April, 1955).

Platt, J. R., "Influence of Neighbor Bonds on Additive Bond Properties in Paraffins," *J. Chem. Phys.*, **15,** 419 (1947); *ibid.*, "Prediction of Isomeric Differences in Paraffin Properties," *J. Phys. Chem.*, **56,** 328 (1952).

Redlich, O., and J. N. S. Kwong, "On the Thermodynamics of Solutions, V.," *Chem. Rev.*, **44,** 233 (1949).

Reid, R. C., J. M. Prausnitz, and T. K. Sherwood, "The Properties of Gases and Liquids," McGraw-Hill Book Co., 3rd Ed., Ch. 7 (1977).

Riedel, L., "Ein neus Verfahren zur Abschätzung unbekannter kritischer Drucke von organischen Verbindungen," *Z. Electrochem.*, **53,** 222 (1949).

Riedel, L., "Additives Verfahren zur Abschützung der kritischen Temperatur aus dem normalen Siedepunkt," *Chem. Ing. Tech.*, **24,** 353 (1952).

Rowlinson, J. S., "Liquids and Liquid Mixtures," Butterworth, London, 2nd Ed., 265 (1969).

Ryan, T. A., B. L. Joiner, and B. F. Ryan, *Minitab Reference Manual*, Duxbury Press, North Scituate, MA, 66 (1980).

Vetere, A., "An Empirical Method for Evaluating Critical Molar Volumes," *AIChE J.*, **22,** 950 (1976); Errata, *ibid.*, **23,** 406 (1977).

Manuscript received December 10, 1982; revision received March 11, and accepted March 16, 1983.