Kirk-Othmer Encyclopedia of Chemical Technology.

ENGINEERING, CHEMICAL DATA CORRELATION

It is a familiar scenario in the chemical sciences: a straightforward set of calculations becomes bogged down for lack of one or more pieces of data, particularly physical property data. Whereas the use of experimentally obtained data is usually preferred over values obtained from data prediction methods, such data are often unavailable or are of dubious quality. In such cases, correlations to obtain the required values must be employed. Because it is not sufficient to provide just a solution, but the best solution given the constraints in time, funding, and information, a large portion of scientific and engineering investigation is devoted to the assessment of the degree to which mathematical constructs faithfully mimic real systems. It is perhaps equally important to be able to determine the level of confidence to which a solution is ascribed.

Theoretically based correlations (or semitheoretical extensions of them), rooted in thermodynamics or other fundamentals are ordinarily preferred. However, rigorous theoretical understanding of real systems is far from complete, and purely empirical correlations typically have strict limits on applicability. Many correlations result from curve-fitting the desired parameter to an appropriate independent variable. Some fitting exercises are rooted in theory, eg, Antoine's equation for vapor pressure; others can be described as being semitheoretical. These distinctions usually do not refer to adherence to the observations of natural systems, but rather to the agreement in form to mathematical models of idealized systems. The advent of readily available computers has revolutionized the development and use of correlation techniques (see Chemometrics; Computer technology; Dimensional analysis).

It is often advisable to begin a project not by attempting to solve the real problem, but rather by constructing an idealized model which can be rigorously solved, then extrapolating based on assumptions and sensitivity analysis, to an estimation of the real solution. The field of classical thermodynamics, developed to relate physical properties to one another, is such an idealized model, although it cannot be used directly to predict physical properties. Empirical modifications to theoretically based models are often reliable because these tend to apply over a broad range of conditions and qualitatively exhibit the correct behavior near physical boundaries and limits. Other techniques such as molecular thermodynamics and kinetic theory might allow direct predictions of physical properties, but such schemes are usually complex and difficult to apply. Molecular descriptors such as shape, size, the chemical moieties present, etc, provide more information for basing correlations. For engineering and physical property data correlations it is important to: (*1*) determine the accuracy of the experimental data to which the correlation is to be applied; (*2*) understand the limitations of the correlation technique employed; (*3*) minimize mathematical complexity (additional fitting parameters may improve data regressions, but may also generate local minima/maxima which do not exist beyond the mathematical model); (*4*) perform sufficient sensitivity analyses to be able to assess the confidence level of final results; and (*5*) present solutions within the context of confidence levels and assumptions.

An overview of some basic mathematical techniques for data correlation is to be found herein together with background on several types of physical property correlating techniques and a road map for the use of selected methods. Methods are presented for the correlation of observed experimental data to physical properties such as critical properties, normal boiling point, molar volume, vapor pressure, heats of vaporization and fusion, heat capacity, surface tension, viscosity, thermal conductivity, acentric factor, flammability limits, enthalpy of formation, Gibbs energy, entropy, activity coefficients, Henry's constant, octanol–water partition coefficients, diffusion coefficients, virial coefficients, chemical reactivity, and toxicological parameters. Correlation methods discussed include basic mathematical and numerical techniques, and approaches based on reference substances, empirical equations, nomographs, group contributions, linear solvation energy relationships, molecular connectivity indexes, and graph theory. Chemical data correlation foundations in classical, molecular, and statistical thermodynamics are introduced.

1. Thermodynamic Correlations for Property Estimation

In the broadest sense, thermodynamics is concerned with mathematical relationships that describe equilibrium conditions as well as transformations of energy from one form to another. Many chemical properties and parameters of engineering significance have origins in the mathematical expressions of the first and second laws and accompanying definitions. Particularly important are those fundamental equations which connect thermodynamic state functions to real-world, measurable properties such as pressure, volume, temperature, and heat capacity (1–3) (see also Thermodynamic properties).

The phase rule specifies the number of intensive properties of a system that must be set to establish all other intensive properties at fixed values (3), without providing information about how to calculate values for these properties. The field of applied engineering thermodynamics has grown out of the need to assign numerical values to thermodynamic properties within the constraints of the phase rule and fundamental laws. In the engineering disciplines there is a particular demand for physical properties, both for pure fluids and mixtures, and for phase equilibrium data (4, 5).

Data compilations, the first recourse for an engineering calculation requiring physical property or parameter data, are often incomplete or do not contain data within the appropriate range of temperature or pressure (6–9). For this reason, correlation and estimation methods play an important role in applied thermodynamics.

The common theme in the evolution of methods for property and parameter prediction is the development of equations, either theoretical or empirical, containing quantities that can be calculated from theoretical considerations or experimental data. Mathematical expressions for correlating thermodynamic data may take several forms.

1.0.1. Limiting Laws

Simple laws that tend to describe a narrow range of behavior of real fluids and substances, and which contain few, if any, adjustable parameters are called limiting laws. Models of this type include the ideal gas law equation of state and the Lewis-Randall fugacity rule (10).

1.0.2. Equations of State

Equations of state having adjustable parameters are often used to model the pressure–volume–temperature (PVT) behavior of pure fluids and mixtures (1, 2). Equations that are cubic in specific volume, such as a van der Waals equation having two adjustable parameters, are the mathematically simplest forms capable of representing the two real volume roots associated with phase equilibrium, or the three roots (vapor, liquid, solid) characteristic of the triple point.

1.0.3. Fundamental Property Relation

The fundamental property relation, which embodies the first and second laws of thermodynamics, can be expressed as a semiempirical equation containing physical parameters and one or more constants of integration. All of these may be adjusted to fit experimental data. The Clausius-Clapeyron equation is an example of this type of relation (1–3).

1.0.4. Semiempirical Relationships

Exact thermodynamic relationships can be approximated, and the unknown parameters then adjusted or estimated empirically. The virial equation of state, truncated after the second term, is an example of such a correlation (3).

1.0.5. Mathematical Consistency

Consistency requirements based on the property of exact differentials can be applied to smooth and extrapolate experimental data (2, 3). An example is the use of the Gibbs-Duhem coexistence equation to estimate vapor mole fractions from total pressure versus liquid mole fraction data for a binary mixture.

1.0.6. Chemical Potential

Equilibrium calculations are based on the equality of individual chemical potentials (and fugacities) between phases in contact (10). In gas–solid adsorption, the equilibrium state can be defined in terms of an adsorption potential, which is an extension of the chemical potential concept to pore-filling (physisorption) onto microporous solids (11–16).

1.0.7. Generalized Correlations

Generalized correlations are often the only recourse when a property value cannot be determined from empirical correlations or by other means. Several powerful correlating techniques fall under this category, including the principle of corresponding states (3, 17), reduced property models (1), and the Polanyi-type characteristic curve for microporous adsorbents (14).

This summary of correlating approaches is far from complete. The focus herein is primarily on pure components, although the phase equilibrium models presented do account for mixture interactions (10). The literature should be consulted for electrolyte systems or transport parameters such as diffusion coefficients (1–3, 18).

1.1. Theoretical and Empirical Correlating Forms

1.1.1. Fundamental Property Relation

For homogeneous, single-phase systems the fundamental property relation (3), is a combination of the first and second laws of thermodynamics that may be written as

$$d(nU) = Td(nS) - Pd(nV) + \Sigma(\mu_i dn_i) \qquad (1)$$

where U = internal energy, T = absolute temperature, S = entropy, P = pressure, μ_i = chemical potential, V = system volume, n_i = moles of i, and n = total moles in the system.

The definitions of enthalpy, H, Helmholtz free energy, A, and Gibbs free energy, G, also give equivalent forms of the fundamental relation (3) which apply to changes between equilibrium states in any homogeneous fluid system:

$$d(nH) = Td(nS) + (nV)dP + \Sigma(\mu_i dn_i) \qquad (2)$$

$$d(nA) = -(nS)dT - Pd(nV) + \Sigma(\mu_i dn_i) \qquad (3)$$

$$d(nG) = -(nS)dT + (nV)dP + \Sigma(\mu_i dn_i) \qquad (4)$$

The chemical potential is a partial molar quantity defined by the last term in equatin 4:

$$\mu_i = \left[\frac{\partial(nG)}{\partial n_i}\right]_{T,P,nj} \qquad (5)$$

where the subscripted term indicates that the moles of all chemical species other than n_{i} are held constant, and ∂ is a partial derivative.

1.1.2. Clausius-Clapeyron Equation

Derived from equation 1, the Clapeyron equation is a fundamental relationship between the latent heat accompanying a phase change and pressure–volume–temperature (PVT) data for the system (1):

$$\Delta H = T\Delta V\frac{dP^{\mathrm{sat}}}{dT} \qquad (6)$$

where ΔV = the volume change accompanying the phase change at temperature T; ΔH = the latent heat of the phase change; and dP^{sat}/dT = the differential change of the saturation pressure with temperature.

The Clapeyron equation is most often used to represent the relationship between the temperature dependence of a pure liquid's vapor pressure curve and its latent heat of vaporization. In this case, dP^{sat}/dT is the slope of the vapor pressure–temperature curve, ΔV is the difference between the volume of the saturated vapor, V_{g}, and the saturated liquid, V_{l}, and ΔH^{vap} is the latent heat of vaporization. Commonly, V_{l} is small in comparison to V_{g} and the ideal gas law is assumed for the vapor phase.

If the latent heat of vaporization is then assumed to be constant over the temperature range of interest, equation 6 can be integrated to give the Clausius-Clapeyron expression:

$$\ln P^{\mathrm{sat}} = \frac{-\Delta H^{\mathrm{vap}}}{RT} + I \qquad (7)$$

Two empirical parameters are evident in equation 7, the heat of vaporization and the integration constant, I. Experimental data indicate that the linear relationship suggested by Clausius-Clapeyron may not be followed over a large temperature range (4); therefore additional adjustable parameters have been added to equation 7 to improve its correlating ability. The most prominent of these is the Antoine equation:

$$\log_{10} P^{\mathrm{sat}} = A - \frac{B}{T+C} \qquad (8)$$

where the constant A is related to the constant of integration, B is analogous to the latent heat term, and C is a correction factor added to the temperature T, usually in $^{\circ}$C.

Extensive tabulations of Antoine parameters are available for many chemicals of importance to engineers, chemists, and environmental scientists (9, 19, 20). Caution is in order when using tabulated Antoine constants because several forms of the correlating equation are found in the literature. In particular, there are variations in the sign before the second term, the units of temperature, and the use of natural or decimal logarithms of the vapor pressure.

The equation developed by the Design Institute for Physical Property Data (DIPPR) is another successful correlating tool for vapor pressure (4). It is an empirical extension of the Antoine equation and has two additional constants, D and E:

$$\ln P^{\mathrm{sat}} = A + \frac{B}{T} + C\ln T + DT^{E} \qquad (9)$$

1.1.3. Equations of State

An equation of state can be an exceptional tool for property prediction and phase equilibrium modeling. The term equation of state refers to the equilibrium relation among pressure, volume, temperature, and composition of a substance (2). This substance can be a pure chemical or a uniform mixture of chemicals in gaseous or liquid form.

Theoretical equation forms may be derived from either kinetic theory or statistical mechanics. However, empirical and semitheoretical equations of state have had the greatest success in representing data with high precision over a wide range of conditions (1). At present, theoretical equations are more limited in range of application than empirical equations. There are several excellent references available on the application and development of equations of state (2, 3, 18, 21).

1.1.3.1. The Virial Expansion. Many equations of state have been proposed for gases, but the virial equation is the only one having a firm basis in theory (1, 3). The pressure-explicit form of the virial expansion is

$$Z \equiv \frac{PV}{RT} = 1 + B'P + C'P^2 + D'P^3 + \cdots \qquad (10)$$

where the ratio PV/RT is called the compressibility factor and is given the symbol Z. An equivalent expression for Z, explicit in volume, is

$$Z = 1 + \frac{B}{V} + \frac{C}{V^2} + \frac{D}{V^3} + \cdots \qquad (11)$$

Equations 10 and 11 are known as virial expansions, and the coefficients B', C', D', . . . and B, C, D, . . . are called virial coefficients. For a given substance, these coefficients are functions of temperature only (1–3).

Statistical mechanics provides physical significance to the virial coefficients (18). For the expansion in $1/V$, the term B/V arises because of interactions between pairs of molecules (eq. 11), the term C/V^2, because of three-molecule interactions, etc. Because two-body interactions are much more common than higher order interactions, truncated forms of the virial expansion are typically used. If no interactions existed, the virial coefficients would be zero and the virial expansion would reduce to the ideal gas law ($Z = 1$).

1.1.3.2. Liquid-Phase Models. Theoretical models of the liquid state are not as well established as those for gases; consequently, the development of general equations for the description of liquid-phase equilibrium behavior is not far advanced. Cubic equations of state give a qualitative description of liquid-phase equilibrium behavior, but do not generally yield good quantitative results (3). For engineering calculations, equations and estimation techniques developed specifically for liquids must normally be used.

The Tait equation for the pressure–volume behavior of liquids (1) correlates data accurately, and is expressed mathematically as

$$V = V_0 - D \ln\left(\frac{P + E}{P_0 + E}\right) \qquad (12)$$

where the parameters D and E are empirical constants for a given temperature, and V_0 and P_0 are molar volume and pressure values at some reference state of the liquid. If sufficient data are available to allow determination of D and E, the Tait equation can represent the isothermal pressure–volume relation up to very high pressures (1).

Another empirical model for liquid pressure–volume behavior is the generalized equation for the molar volumes of saturated liquids given by the Rackett equation:

$$V^{\text{sat}} = V_c Z_c^{(1-T_r)^{0.2857}} \qquad (13)$$

which employs critical volume, V_c, the compressibility factor at the critical point, Z_c, and reduced temperature, T_r. The Rackett equation produces results accurate to about 2% (1).

1.1.4. Mixing Rules

Kay's method is a mixing rule for determining the pseudocritical temperature, T_{cm}, and pressure, P_{cm}, of a mixture of gases (1, 3). In this method, pseudocritical values for mixtures of gases are calculated on the assumption that each component in the mixture contributes to the pseudocritical value in the same proportion as the number of moles of that component. The resulting values are referred to as linearly weighted mole-average pseudocritical properties of pressure, temperature, and volume. Kay's method is classified as a two-parameter rule because only P_c and T_c for each component are involved in the calculation of the compressibility factor, Z. Pseudoreduced properties of the mixture are calculated from the pseudocritical values from Kay's method, allowing the mixture to be treated as a pure fluid. Pseudoreduced properties then feed into an appropriate generalized correlation. If mixture data are scarce, Kay's rule is a good starting point. Other empirical mixing formulas may also be found in the literature (3).

1.1.5. Correlations for Enthalpy of Vaporization

Enthalpy or heat of vaporization, which is an important engineering parameter for liquids, can be predicted by a variety of methods which focus on either prediction of the heat of vaporization at the normal boiling point, or estimation of the heat of vaporization at any temperature from a known value at a reference temperature (5).

Enthalpies of vaporization, $\Delta H^{\text{vap}}{}_n$, at the normal boiling point, T_b, can be estimated roughly with the simple rule of Pictet and Trouton (19) which state that the ratio of $\Delta H^{\text{vap}}{}_n$ to T_b should have the same value for all liquids:

$$\Delta H_n^{\text{vap}}/T_b = \text{constant} \approx 88\ \text{J}/(\text{K·mol})\,(21\ \text{cal}/(\text{K·mol})) \qquad (14)$$

It is equivalent to say that entropy of vaporization is a constant value for non-associating liquids. Associating liquids, eg, ammonia, water, methanol, and ethanol, do not obey the rule of Pictet and Trouton. Despite its simplicity, the Pictet-Trouton view of liquid vaporization (19) is an excellent example of the many rules of thumb that have been useful aids in engineering calculations for decades (5, 7–9, 21). However, proper application requires an understanding of the physical reasoning behind each rule.

An equation that yields values of $\Delta H^{\text{vap}}{}_n$ in units of kJ/mol to within about 2% is Chen's equation (1):

$$\Delta H_n^{\text{vap}} = \frac{T_b\left[0.0331\,(T_b/T_c) + 0.0297\log_{10}P_c - 0.0923\right]}{1.07 - T_b/T_c} \qquad (15)$$

where T_b is the normal boiling point of the liquid in Kelvin, T_c is the critical temperature in Kelvin, and P_c is the critical pressure in kPa.

Another useful method for predicting the heat of vaporization at the normal boiling point is Riedel's equation (1):

$$\frac{\Delta H_n^{\text{vap}}}{T_n} = \frac{\left(9.079\times10^{-3}\ln P_c - 0.051\right)}{0.930 - T_{rn}} \qquad (16)$$

where $\Delta H^{\text{vap}}{}_n$ is the molar latent heat of vaporization at the normal boiling point in kJ/mol, T_{rn} is the reduced temperature at the normal boiling point, and P_c is the critical pressure in kPa (1).

Estimates of the latent heat of vaporization of a pure liquid at any temperature from a known value at a single temperature are possible by several methods. The known value may be experimental, or it may be estimated by equation 16. Of the methods proposed, the Watson correlation (1), which is both simple and reliable, has found the greatest acceptance:

$$\frac{\Delta H_2}{\Delta H_1} = \left(\frac{1 - T_{r2}}{1 - T_{r1}}\right)^{0.38} \qquad (17)$$

where ΔH_2 is the heat of vaporization of a pure liquid at reduced temperature, T_{r2}, and ΔH_1 is the heat of vaporization of the same liquid at T_{r1}. The exponent is sometimes treated as an adjustable parameter in the Watson correlation (7). For example, various other values of the exponent are available (22).

1.2. Limiting Laws

1.2.1. Ideal Gas Behavior

In 1787 it was demonstrated that the volume of a gas varies directly with temperature if the pressure remains constant. Other investigations determined complementary correlating relations from which the perfect or ideal gas law was drawn (1–3). Expressed mathematically, the ideal gas law is

$$pV = nRT \qquad (18)$$

where p is the pressure exerted by the gas, V is the volume of the gas, n is the number of moles of gas, R is the universal gas constant, and T is the absolute temperature of the gas. Real gases closely obey this pressure–volume–temperature relationship at low pressures and high temperatures.

No real gas obeys equation 18 precisely over wide ranges of temperature and pressure. Under ordinary circumstances, lighter gases such as hydrogen, oxygen, and air obey the ideal gas law closely. Gases such as sulfur dioxide, carbon dioxide, and large hydrocarbons, however, deviate considerably from ideal gas behavior, particularly at high pressures and low temperatures.

The ideal gas law is an example of a correlating expression that comes directly from experimental observations, but has theoretical significance. Despite its simplicity, the ideal gas law is an excellent estimation tool. Often, it is the first approximation in systems involving real gases of all types. Unfortunately for liquids and solids no laws of such general utility are available.

1.2.2. Ideal Liquid Solutions

Two limiting laws of solution thermodynamics that are widely employed are Henry's law and Raoult's law, which represent vapor–liquid partitioning behavior in the concentration extremes. These laws are used frequently in equilibrium problems and apply to a variety of real systems (10).

The equilibrium partitioning of a chemical solute between a liquid and vapor phase is governed by Henry's law when the liquid mixture is very dilute in the solute. Henry's law generally is valid at concentrations below 0.01 mol/L of solution, although the upper limit can sometimes extend to 0.1 mol/L or higher (10). Over this concentration range, a direct proportionality, ie, Henry's constant, is observed between the partial pressure of the chemical in the gas phase and its mole fraction in the liquid phase. Henry's constant, when expressed in this way, has units of pressure (3).

In general, equality of component fugacities, ie, chemical potentials, in the vapor and liquid phases yields the following relation for vapor–liquid equilibrium:

$$y_i \phi_i P_T = x_i \gamma_i P_i^{\text{sat}} \phi_{ii}^{\text{sat}} \pi_i \qquad (19)$$

in which the Poynting correction factor, π_i, is defined as:

$$\pi_i = \exp\left(\frac{V_i^l\left(P_T - P_i^{\mathrm{sat}}\right)}{RT}\right) \qquad (20)$$

and where y_i and x_i are the equilibrium mole fractions of i in the vapor and liquid phases, respectively, $V^l{}_i$ is the liquid molar volume of pure i, φ_i is the vapor-phase fugacity coefficient, and P_T is the total pressure.

At system pressures up to several tens of MPa, the fugacity coefficients, φ_i and $\varphi^{\mathrm{sat}}{}_i$, and the Poynting factor, π_i, are usually near unity. A simplified version of equation 19 can therefore be used for the majority of vapor–liquid equilibrium problems:

$$y_i P_T = x_i \gamma_i P_i^{\mathrm{sat}} \qquad (21)$$

In equation 21 the vapor phase is considered to be ideal, and all nonideality effects are attributed to the liquid-phase activity coefficient, γ_i. For an ideal solution ($\gamma_i = 1$), equation 21 becomes Raoult's law for the partial pressure, p_i, exerted by the liquid mixture:

$$p_i = y_i P_T = x_i P_i^{\mathrm{sat}} \qquad (22)$$

Equation 22 is a special application of the general Lewis-Randall ideal solution model (3, 10) that is typically used for near-ambient pressures and concentrated nonpolar solutes.

1.2.3. Partial Molar Properties

The properties of individual components in a mixture or solution play an important role in solution thermodynamics. These properties, which represent molar derivatives of such extensive quantities as Gibbs free energy and entropy, are called partial molar properties. For example, in a liquid mixture of ethanol and water, the partial molar volume of ethanol and the partial molar volume of water have values that are, in general, quite different from the volumes of pure ethanol and pure water at the same temperature and pressure (21). If the mixture is an ideal solution, the partial molar volume of a component in solution is the same as the molar volume of the pure material at the same temperature and pressure.

Perhaps the most significant of the partial molar properties, because of its application to equilibrium thermodynamics, is the chemical potential, μ_i. This fundamental property, and related properties such as fugacity and activity, are essential to mathematical solutions of phase equilibrium problems. The natural logarithm of the liquid-phase activity coefficient, $\ln\gamma_i$, is also defined as a partial molar quantity. For liquid mixtures, the activity coefficient, γ_i, describes nonideal liquid-phase behavior.

1.2.4. Heat Capacities

The heat capacities of real gases are functions of temperature and pressure, and this functionality must be known to calculate other thermodynamic properties such as internal energy and enthalpy. The heat capacity in the ideal-gas state is different for each gas. Constant pressure heat capacities, C'_p, for the ideal-gas state are independent of pressure and depend only on temperature. An accurate temperature correlation is often an empirical equation of the form:

$$C'_p = \alpha + \beta T + \gamma T^2 \qquad (23)$$

where α, β, and γ are constants characteristic of the gas considered and T is temperature (3). The prime superscript on C'_p denotes the ideal-gas state.

For the ideal-gas state there is an exact relation between the constant pressure heat capacity and the constant volume heat capacity, C'_v, via the ideal-gas constant, R.

$$C'_p - R = C'_v \qquad (24)$$

From this equation, the temperature dependence of C'_p is known, and vice versa (21). The ideal-gas state at a pressure of 101.3 kPa (1 atm) is often regarded as a standard state, for which the heat capacities are denoted by $C^0{}_p$ and $C^0{}_v$. Real gases rarely depart significantly from ideality at near-ambient pressures (3); therefore, C'_p and C'_v usually represent good estimates of the heat capacities of real gases at low to moderate, eg, up to several hundred kPa, pressures. Otherwise thermodynamic excess functions are used to correct for deviations from ideal behavior when such situations occur (3).

When experimental data are not available, methods of estimation based on statistical mechanics are employed (7, 19). Classical kinetic theory suggests a contribution to $C^0{}_p$ of $\frac{1}{2}$ R for each translational degree of freedom in the molecule, a contribution of $\frac{1}{2}$ R for each axis of rotation, and of R for each vibrational degree of freedom. A crude estimate of $C^0{}_p$ for small molecules can be obtained which neglects vibrational degrees of freedom:

$$C_p^0 = R(1 + g/2) \qquad (25)$$

where g is the sum of the translational and rotational modes in the molecule. Equation 25 gives only rough estimates, so empirical correlations such as equation 23 are recommended whenever possible (2). Ideal-gas heat capacities increase smoothly with increasing temperature toward an upper limit, which is reached when all translational, rotational, and vibrational modes of molecular motion are fully excited.

1.3. Mathematical Consistency Requirements

Theoretical equations provide a method by which a data set's internal consistency can be tested or missing data can be derived from known values of related properties. The ability of data to fit a proven model may also provide insight into whether that data behaves correctly and follows expected trends. For example, poor fit of vapor pressure versus temperature data to a generally accepted correlating equation could indicate systematic data error or bias. A simple semilogarithmic form, (eg, the Antoine equation, eq. 8), has been shown to apply to most organic liquids, so substantial deviation from this model might indicate a problem. Many other simple thermodynamics relations can provide useful data tests (1–5, 18, 21).

From the definition of a partial molar quantity and some thermodynamic substitutions involving exact differentials, it is possible to derive the simple, yet powerful, Duhem data testing relation (2, 3, 18). Stated in words, the Duhem equation is a mole-fraction-weighted summation of the partial derivatives of a set of partial molar quantities, with respect to the composition of one of the components (2, 3). For example, in an n-component system, there are n partial molar quantities, M_{i}, representing any extensive molar property. At a specified temperature and pressure, only $(n - 1)$ of these properties are independent. Many experiments, however, measure quantities for every chemical in a multicomponent system. It is this redundance in reported data that makes thermodynamic consistency tests possible.

The well-known Gibbs-Duhem equation (2, 3, 18) is a special mathematical redundance test which is expressed in terms of the chemical potential (3, 18). The general Duhem test procedure can be applied to any set of partial molar quantities. It is also possible to perform an overall consistency test over a composition range with the integrated form of the Duhem equation (2).

In some cases, reported data do not satisfy a consistency check, but these may be the only available data. In that case, it may be possible to smooth the data in order to obtain a set of partial molar quantities that is thermodynamically consistent. The procedure is simply to reconstruct the total molar property by a weighted mole fraction average of the n measured partial molar values and then recalculate normalized partial molar quantities. The new set should always be consistent.

For interrelated properties, a method of assessing data quality is to see if data for those properties are consistent with theoretical equations. One example of such a test could be how well the slope of a semilogarithmic vapor pressure plot agrees with that chemical's measured heat of vaporization. Recall that the integrated Clausius-Clapeyron form (eq. 7) suggests that a semilogarithmic plot of vapor pressure versus inverse temperature should be a straight line having a slope equal to the heat of vaporization.

Many additional consistency tests can be derived from phase equilibrium constraints. From thermodynamics, the activity coefficient is known to be the fundamental basis of many properties and parameters of engineering interest. Therefore, data for such quantities as Henry's constant, octanol–water partition coefficient, aqueous solubility, and solubility of water in chemicals are related to solution activity coefficients and other properties through fundamental equilibrium relationships (10, 23, 24). Accurate, consistent data should be expected to satisfy these and other thermodynamic requirements. Furthermore, equilibrium models may permit a missing property value to be calculated from those values that are known (2).

1.4. Phase Equilibria and Chemical Potential

1.4.1. Models of Phase Nonideality

The correlation and prediction of phase equilibria at low (near-vacuum) to moderate (several hundred kPa) pressures is typically based on a form of equation 19. The key is the degree to which the various parameters for phase nonideality are known or can be estimated.

1.4.1.1. Fugacity Coefficients. An exact equation that is widely used for the calculation of fugacity coefficients and fugacities from experimental pressure–volume–temperature (PVT) data is

$$\ln\hat{\phi}_i = \int_0^P \left(\bar{Z}_i - 1\right)\frac{dP}{P}\,(\text{constant } T, x) \qquad (26)$$

where ϕ_i = fugacity coefficient of i in a vapor mixture, and

$$\bar{Z}_i = P\bar{V}_i/RT \qquad (27)$$

which leads to another form of equation 26:

$$\ln\hat{\phi}_i = \frac{-1}{RT}\int_0^P \left(\frac{RT}{P} - \bar{V}_i\right) dP\,(\text{constant } T, x) \qquad (28)$$

Here a suitable equation of state is required to provide a mathematical expression for the mixture molar volume, $\overline{V}_i$. For some equations of state, it is better to use a form of equation 28 in which the integral is volume explicit (3). Note also that for an ideal gas $Z_i = \overline{Z}_i = 1$, and $\phi_i = \hat{\phi} = 1$.

1.4.1.2. Activity Coefficients. Activity coefficients in liquid mixtures are directly related to the molar excess Gibbs energy of mixing, ΔG^{E}, which is defined as the difference in the molar Gibbs energy of mixing between the real and ideal mixtures. It is typically an assumed function. Various functional forms of ΔG^{E} give rise to many of the different activity coefficient models found in the literature (1–3, 18). Typically, the liquid-phase activity coefficient is a function of temperature and composition; explicit pressure dependence is rarely included.

There are many simple two-parameter equations for liquid mixture constituents, including the Wilson (25), Margules (2, 3, 18), van Laar (3, 26), nonrandom two-liquid (NRTL) (27), and universal quasichemical (UNIQUAC) (28) equations. In the case of the NRTL model, one of the three adjustable parameters has been found to be relatively constant within some homologous series, so NRTL is essentially a two-parameter equation. The third parameter is usually treated as a constant which is set according to the type of chemical system (27). A third parameter for Wilson's equation has also been suggested for use with partially miscible systems (29–31). These equations all require experimental data to fit the adjustable constants. Simple equations of this type have the additional attraction of being useful for hand calculations.

Other algorithms are much more complex and require computer software, such as group contribution models for estimating liquid-phase activity coefficients. These models include analytical solution of groups (ASOG) (32), modified separation of cohesive energy density (MOSCED) (33, 34), and UNIQUAC functional-group activity coefficient (UNIFAC) (28, 35–40). Each model is applicable for some, but not all, chemical systems and conditions, and the choice of the functional form for the Gibbs excess energy leads to specialization and limitation.

The solvophobic model of liquid-phase nonideality takes into account solute–solvent interactions on the molecular level. In this view, all dissolved molecules expose microsurface area to the surrounding solvent and are acted on by the so-called solvophobic forces (41). These forces, which involve both enthalpy and entropy effects, are described generally by a branch of solution thermodynamics known as solvophobic theory. This general solution interaction approach takes into account the effect of the solvent on partitioning by considering two hypothetical steps. First, cavities in the solvent must be created to contain the partitioned species. Second, the partitioned species is placed in the cavities, where interactions can occur with the surrounding solvent. The idea of solvophobic forces has been used to estimate such diverse physical properties as absorbability, Henry's constant, and aqueous solubility (41–44). A principal drawback is calculational complexity and difficulty of finding values for the model input parameters.

The solvophobic model has been used to deduce a functional form for a Henry's constant correlation based on molecular connectivity index and polarizability (42). Accurate predictions are reported over a span of seven log units in Henry's constant. A reliable solvophobic model of aqueous solubility has also been reported (45, 46).

1.4.1.3. Regular Solution Theory. The key assumption in regular-solution theory is that the excess entropy, S^{E}, is zero when mixing occurs at constant volume (3, 18). This idea of a regular solution (26) leads to the equations:

$$G^E = \sum_k \left[x_k V_k \left(\delta_k - \bar{\delta}\right)^2 \right] \qquad (29)$$

and

$$\ln\gamma_i = \frac{V_i}{RT}\left(\delta_i - \bar{\delta}\right)^2 \qquad (30)$$

where

$$\bar{\delta} = \frac{\Sigma_k \left(x_k V_k \delta_k\right)}{\Sigma_k \left(x_k V_k\right)} \qquad (31)$$

In these equations, V_{k} is the molar volume and δ_k is the solubility parameter of the pure liquid k. The solubility parameter is the square root of an energy density, defined as

$$\delta_i = \left(\frac{\Delta U_i^{\mathrm{vap}}}{V_i}\right)^{1/2} \qquad (32)$$

where $\Delta U^{\mathrm{vap}}{}_i$ is the internal energy change of vaporization of pure liquid i. Values are determined from heats of vaporization.

The solubility parameter, δ_i, is a function of temperature, but the difference $(\delta_i - \overline{\delta})$ is only weakly dependent on temperature. By convention, both δ_i and V_i are evaluated at 25°C and are treated as constants independent of both T and P. The activity coefficients given by equation 30 are therefore functions of liquid composition and temperature, but not of pressure.

A key feature of this model is that no data for mixtures are required to apply the regular-solution equations because the solubility parameters are evaluated from pure-component data. Results based on these equations should be treated as only qualitative. However, mixtures of nonpolar or slightly polar, nonassociating chemicals, can sometimes be modeled adequately (1, 3, 18). Applications of this model have been limited to hydrocarbons (qv) and a few gases associated with petroleum (qv) and natural gas (see Gas, natural) processing, such as N_2, H_2, CO_2, and H_2S. Values for δ_i and V_i can be found in many references (1–3, 7).

1.4.2. Thermodynamics of Vapor–Liquid Equilibrium

Assuming ideal vapor and choosing the Lewis-Randall standard state, a chemical distributed between a vapor and liquid in equilibrium often behaves according to:

$$y_i P_T = p_i = x_i \gamma_i P^{\text{sat}_i} \qquad (33)$$

where y_i is the mole fraction of i in the vapor phase; x_i is the mole fraction of i in the liquid phase; P_T is the total system pressure; p_i is the partial pressure of i in the vapor mixture; γ_i is the liquid-phase activity coefficient of i (Lewis-Randall basis); and $P^{\text{sat}}{}_i$ is the saturation pressure of pure liquid i at the system temperature. Note that equation 33, which is based on equal fugacities, ie, equal chemical potentials, of a partitioned component in the vapor and liquid, neglects the Poynting and saturation fugacity coefficient corrections presented earlier (eq. 19). This equilibrium model, despite its simplicity, is suitable for many engineering calculations. Setting $\gamma_i = 1$ reduces equation 33 to Raoult's law (1–3, 18).

1.4.2.1. Henry's Law Constant. Henry's law for dilute concentrations of contaminants in water is often appropriate for modeling vapor–liquid equilibrium (VLE) behavior (47). At very low concentrations, a chemical's Henry's constant is equal to the product of its activity coefficient and vapor pressure (3, 10, 48). Activity coefficient models can provide estimated values of infinite dilution activity coefficients for calculating Henry's constants as a function of temperature (35–39, 49).

The short-cut technique frequently used to estimate the Henry's constant of a volatile substance in water is to calculate the ratio of the pure compound's vapor pressure to its aqueous saturation limit (23):

$$H_i = \frac{P_i^{\text{sat}}}{x_s} \qquad (34)$$

where x_s is the ultimate solubility of the substance in water in mole fraction, $P^{\text{sat}}{}_i$ is the vapor pressure of the pure liquid substance in kPa, and H_i is the Henry's constant in kPa. This assumes the miscibility of water in the volatile substance is negligible. A more rigorous solution, however, yields

$$H_i = \frac{(1 - x_w) P_i^{\text{sat}}}{x_s} \qquad (35)$$

where x_w is the saturation limit of water in the substance in mole fraction.

As indicated by equation 35, the vapor pressure–solubility ratio (eq. 34), only provides accurate results when (*1*) reliable vapor pressure and aqueous solubility data are employed, (*2*) the quantity x_w is small enough to be neglected, and (*3*) Henry's law applies for the dissolved substance in water up to the saturation limit. If water miscibility in the organic is expected to be significant, equation 35 is much preferred, assuming miscibility data are available (42). This correction can amount to several percent even when the aqueous solubility is low.

For example, water solubility in *n*-octanol is about 5% whereas the *n*-octanol solubility in water is in the parts per thousand range.

From fundamental thermodynamic relations, the temperature and pressure dependence of Henry's constant can be shown (18, 50, 51) to be:

$$d(\ln H_i) = \left(\frac{(h_i - h_i^{\infty})}{RT^2}\right) dT + \left(\frac{V_i^{\infty}}{RT}\right) dP \qquad (36)$$

where h_i is the molar enthalpy of i in the ideal gas state, h_i^{∞} is the partial molar enthalpy of i at infinite dilution, and V_i^{∞} is the partial molar volume of i at infinite dilution.

For experiments conducted at constant pressure, the second term in equation 36 disappears. The expression for the temperature dependence is then obtained by performing an indefinite integration on the remainder of the equation after assuming that the enthalpy change of volatilization, $(h_i - h_i^{\infty})$, is constant with respect to temperature. The resulting equation is

$$\ln H_i = \frac{-(h_i - h_i^{\infty})}{RT} + I \qquad (37)$$

where I is an integration constant.

By treating the quantities $(h_i - h_i^{\infty})/R$ and I as fitting parameters, H_i data obtained at various temperatures can be correlated to equation 37 using linear regression. The final equation for H_i has the form:

$$H_i = \exp\left[B - \left(\frac{A}{T}\right)\right] \qquad (38)$$

where A and B are the "best-fit" constants for the experimental data. In environmental systems, where there are temperature gradients, the temperature in equation 38 can be taken to be the air–water interface temperature (48).

1.4.3. Thermodynamics of Liquid–Liquid Equilibrium

Phase splitting of a liquid mixture into two liquid phases (I and II) occurs when a single liquid phase is thermodynamically unstable. The equilibrium condition of equal fugacities (and chemical potentials) for each component in the two phases allows the fugacities f'_i and f''_i in phases I and II to be equated and expressed as:

$$x'_i \gamma'_i = x''_i \gamma''_i \qquad (39)$$

The same reference (standard) state, $f^0{}_i$, is chosen for the two phases, so that it cancels on both sides of equation 39. The products $x'_i \gamma'_i$ and $x''_i \gamma''_i$ are referred to as activities. Because equation 39 holds for each component of a liquid–liquid system, it is possible to predict liquid–liquid phase splitting when the activity coefficients of the individual components in a multicomponent system are known. These values can come from vapor–liquid equilibrium experiments or from prediction methods developed for phase-equilibrium problems (4, 5, 10). Some binary systems can be modeled satisfactorily in this manner, but only rough estimations appear to be possible for multicomponent systems because activity coefficient models are not yet sufficiently developed in this area.

1.4.3.1. Aqueous Solubility. Solubility of a chemical in water can be calculated rigorously from equilibrium thermodynamic equations. Because activity coefficient data are often not available from the literature or direct experiments, models such as UNIFAC can be used for structure–activity estimations (24). Phase-equilibrium relationships can then be applied to predict miscibility. Simplified calculations are possible for low miscibility; however, when there is a high degree of miscibility, the phase-equilibrium relationships must be solved rigorously.

For low miscibility, the solubility of a substance in water is often estimated as the inverse of the chemical's infinite dilution activity coefficient:

$$x_s \cong \frac{1}{\gamma_i^{\infty}} \qquad (40)$$

where x_s = the solubility limit of the chemical in water in mole fraction, and γ_i^{∞} = the infinite dilution activity coefficient, which is either known from experiments or estimated. Equation 40 is valid only when Henry's law applies to the dissolved chemical up to its aqueous solubility limit and the solubility of water in the chemical can be neglected.

If miscibility is significant in a binary chemical–solvent system, the calculations become more complex because the coupled nonlinear phase equilibrium expressions must be solved (10):

$$x_s \cdot \gamma_s^{\mathrm{II}} = (1 - x_w) \cdot \gamma_s^{\mathrm{I}} \qquad (41)$$

$$x_w \cdot \gamma_w^{\mathrm{I}} = (1 - x_s) \cdot \gamma_w^{\mathrm{II}} \qquad (42)$$

and

$$\gamma_i = \mathrm{f}\,(x_s, x_w, T) \qquad (43)$$

In the case of water as the solvent, I designates the chemical-rich phase, II corresponds to the water-rich phase, x_{s} is the aqueous solubility of the chemical, and x_{w} is the solubility of the water in the chemical. The activity coefficients are nonlinear functions of concentration, requiring a numerical solution to find the values of x_{s} and x_{w} that satisfy equations 41 and 42. In principle, this method can be extended to multiple components, and general-purpose "flash" procedures are available in the form of programmable algorithms (3, 10). High quality organic-water mutual miscibility data are scarce (52–60), underscoring the need for predictive methods.

1.4.3.2. Octanol–Water Partition Coefficient. In environmental calculations, the octanol–water partition coefficient, K_{ow}, is related to a chemical's lipophilicity and correlates well with many properties. K_{ow} is defined as the ratio of the concentrations of a third component distributed between octanol-rich and water-rich phases in equilibrium. As the mole fraction of the distributed chemical in the two phases approaches zero, the K_{ow} concentration ratio approaches a constant value (23).

Based on liquid–liquid equilibrium principles, a general model of octanol–water partitioning is possible if accurate activity coefficients can be determined. First, phase equilibrium relationships based on activity coefficients permit liquid–liquid equilibrium calculations for the binary octanol–water system. Because the two components are almost immiscible in each other, two phases form: an octanol-rich phase containing dissolved water, and a water-rich phase containing dissolved octanol.

Once the composition of each equilibrium phase is known, infinite dilution activity coefficients for a third component in each phase can then be calculated. The octanol–water partition coefficient is directly proportional to the ratio of the infinite dilution activity coefficients for a third component distributed between the water-rich and octanol-rich phases (5, 24). The primary drawback to the activity coefficient approach to K_{ow} estimation is the difficulty of the calculations involved, particularly when the activity coefficient model is complex.

2. Generalized and Reduced Equations

Reduced conditions are corrected, or normalized, conditions of temperature T, pressure p, and specific volume $\hat{V}$, and are expressed mathematically as

$$T_r = T/T_c \qquad (44)$$

$$p_r = p/p_c \qquad (45)$$

$$\hat{V}_r = \hat{V}/\hat{V}_c \qquad (46)$$

where the subscript c denotes the critical state, and r the reduced state.

The idea of using reduced variables to correlate the pressure–volume–temperature properties of gases, was suggested by van der Waals in 1873. The underlying principle is that all substances behave alike in the reduced, ie, corrected, states. That is, any substance would have the same reduced volume at the same reduced temperature and reduced pressure. Certainly, the simplest form of such an equation would be a variation of the ideal gas law:

$$p_{\mathrm{r}}\hat{V}_{\mathrm{r}} = \gamma R T_r \qquad (47)$$

where γ is some universal constant and $\hat{V}_{\mathrm{r}}$ is the reduced specific molar volume. Although equation 47 works for many gases at high temperature and pressure, it fails to predict gas properties adequately in the region where the ideal gas law applies, ie, at low pressure somewhat below $p_r = 1.0$.

2.1. Principle of Corresponding States

2.1.1. Generalized Correlations

A simple and reliable method for the prediction of vapor–liquid behavior has been sought for many years to avoid experimentally measuring the thermodynamic and physical properties of every substance involved in a process. Whereas the complexity of fluids makes universal behavior prediction an elusive task, methods based on the theory of corresponding states have proven extremely useful and accurate while still retaining computational simplicity. Methods derived from corresponding states theory are commonly used in process and equipment design.

A generalized correlation is a functional relationship between dimensionless or reduced variables. Most generalized correlations may be expressed by the following function:

$$z = \psi|T_r, P_r| \qquad (48)$$

which is the keystone to the theory of corresponding states. This form readily lends itself to the use of a generalized chart, where the compressibility factor (or any other property) is plotted as a function of the reduced pressure for each reduced isotherm. A great advantage of the generalized chart is its ease of use and fair accuracy (within 10%) obtained from limited information, ie, usually the critical pressure and critical temperature. Drawbacks to the generalized chart include the inability to predict properties with high precision ($<1\%$ error) and the inconvenience of the form to computer application.

Generalized charts are applicable to a wide range of industrially important chemicals. Properties for which charts are available include all thermodynamic properties, eg, enthalpy, entropy, Gibbs energy; and PVT data, compressibility factors, liquid densities, fugacity coefficients, surface tensions, diffusivities, transport properties, and rate constants for chemical reactions. Charts and tables of compressibility factors vs reduced pressure and reduced temperature have been produced. Data is available in both tabular and graphical form (61–72).

An alternative to the use of generalized charts is an analytical equation of state. Equations of state which are expressed as a function of reduced properties and nondimensional variables are said to be generalized. The term generalization is in reference to the wide applicability to the estimation of fluid properties for many substances.

Numerous other methods have been used to predict properties of gases and liquids. These include group contribution, reference substance, approaches, and many others. However, corresponding states theory has been one of the most thoroughly investigated methods and has become an important basis for the development of correlation and property estimation techniques. The methods derived from the corresponding states theory for liquid and gas property estimation have proved invaluable for work such as process and equipment design.

2.1.2. Reduced Properties

One of the first attempts at achieving an accurate analytical model to describe fluid behavior was the van der Waals equation, in which corrections to the ideal gas law take the form of constants a and b to account for molecular interactions and the finite volume of gas molecules, respectively.

$$\left(P + \frac{a}{\hat{V}^2}\right)(\hat{V} - b) = RT \qquad (49)$$

This equation bridged the gap between gases and liquids by showing each could be described by a single cubic equation. Although the van der Waals equation's accuracy fails at the critical point and in the liquid region, it has had a great effect on the understanding of fluids. It is from this simple equation that the theory of corresponding states was proposed.

All fluids, when compared at the same reduced temperature and reduced pressure have approximately the same compressibility factor and deviate from ideal gas behavior to the same extent, giving

$$\left(P_r + \frac{3}{\hat{V}_r^2}\right)(3\hat{V}_r - 1) = 8T_r \qquad (50)$$

which is theoretically equivalent for all substances.

In 1893, it was shown that corresponding states are not unique to van der Waals' equation of state (73). Rather, for any equation of state having not more than three constants, corresponding states are only a mathematical consequence.

The concept of corresponding states was based on kinetic molecular theory, which describes molecules as discrete, rapidly moving particles that together constitute a fluid or solid. Therefore, the theory of corresponding states was a macroscopic concept based on empirical observations. In 1939, the theory of corresponding states was derived from an inverse sixth power molecular potential model (74). Four basic assumptions were made: (*1*) classical statistical mechanics apply, (**2**) the molecules must be spherical either by actual shape or by virtue of rapid and free rotation, (*3*) the intramolecular vibrations are considered identical for molecules in either the gas or liquid phases, and (*4*) the potential energy of a collection of molecules is a function of only the various intermolecular distances.

Using these assumptions and a classical partitioning function, Q, integrated over the coordinates and momenta of all molecules, a universal function was defined:

$$Q = F\left(\frac{\hat{V}}{\hat{V}_c}, \frac{T}{T_c}\right) \qquad (51)$$

Whereas this two-parameter equation states the same conclusion as the van der Waals equation, this derivation extends the theory beyond just PVT behavior. Because the partition function, Q, can also be used to derive all the thermodynamic functions, the functional form, F, can be changed to describe this data as well. Corresponding states equations are typically written with respect to temperature and pressure because of the ambiguities of measuring volume at the critical point.

The theory of corresponding states only works for certain classes of fluids, thus fluids have been placed into generalized groups based on behavior (75, 76). These groupings include the following.

2.1.2.1. Simple Fluids. Spherical compounds having little molecular interaction, eg, argon, krypton, xenon, and methane, are known as simple fluids and obey the theory of corresponding states.

2.1.2.2. Normal Fluids. Asymmetrical compounds having little molecular interaction, eg, carbon monoxide, n-butane, and n-hexane, deviate slightly from the theory of corresponding states and are considered to be normal fluids.

2.1.2.3. Slightly Polar Fluids. Compounds having weak polarity such as benzene and toluene cannot be accurately predicted by corresponding states and are known as slightly polar fluids.

2.1.2.4. Hydrogen-Bonded, Associating, and Polar Fluids. Compounds having strong hydrogen-bonding and polar behavior, eg, water and alcohols, as well as compounds having strong ionic characteristics, such as weak acids and bases, are not applicable to simple corresponding states equations. Only complex equations can begin to describe the behavior of these compounds.

2.1.2.5. Quantum Fluids. Light compounds which exhibit behavior resulting from quantum effects, eg, hydrogen, helium, and neon, are called quantum fluids.

2.1.3. Reduced Equations of State

A simple modification to the cubic van der Waals equation, developed in 1946 (72), uses a term called the ideal or pseudocritical volume, $V_c^{\text{ideal}} = RT_c/P_c$, to avoid the uncertainty in the measurement of volume at the critical point.

$$P_r = \frac{T_r}{0.125 - \phi} - \frac{0.422}{\phi^2} \qquad (52)$$

where

$$\phi = \frac{\hat{V}}{V_c^{\text{ideal}}}$$

This expression, which has an average error of 2% or less, is valid up to the critical density. It also allows a corrective factor to be added to the critical pressure and temperature so that hydrogen and helium behavior can be accurately predicted (3, 21).

A generalized cubic equation is the Redlich-Kwong equation (77):

$$Z = \frac{1}{1-h} - \frac{4.9340}{T_r^{1.5}}\left(\frac{h}{1+h}\right) \qquad (53)$$

where

$$h = \frac{0.08664P_r}{ZT_r}$$

This equation is useful for gases above the critical point. Only reduced pressure, P_r, and reduced temperature, T_r, are needed. In the form represented by equation 53, iteration quickly gives accurate values for the compressibility factor, Z. However, this two-parameter equation only gives accurate values for simple and nonpolar fluids. Unless the Redlich-Kwong equation (eq. 53) is explicitly solved for pressure in nonreduced variables, it does not give accurate liquid volumes.

An example of a more complex equation that exhibits better accuracy is the Benedict-Webb-Rubin equation:

$$P = \frac{RT}{\hat{V}} + \frac{B_0RT - A_0 - C_0/T^2}{\hat{V}^2} + \frac{bRT - a}{\hat{V}^3} + \frac{a\alpha}{\hat{V}^6} + \frac{c}{\hat{V}^3T^2}\left(1 + \frac{\gamma}{\hat{V}^2}\right)e^{-\left(\frac{\gamma}{\hat{V}^2}\right)} \qquad (54)$$

where A_0, B_0, C_0, a, b, c, α, and γ are all empirical constants for specific substances. The complexity and lack of constant values for an equation of this sort is a weakness of some generalized equation correlations.

The equations given predict vapor behavior to high degrees of accuracy but tend to give poor results near and within the liquid region. The compressibility factor can be used to accurately determine gas volumes when used in conjunction with a virial expansion or an equation such as equation 53 (77). However, the prediction of saturated liquid volume and density requires another technique. A correlation was found in 1958 between the critical compressibility factor and reduced density, based on inert gases. From this correlation an equation for normal and polar substances was developed (78):

$$\log\left(\frac{V_l^{\mathrm{sat}}}{\hat{V}_c}\right) = (1 - T_r)^{2/7}\log Z_c \qquad (55)$$

which can predict saturated liquid volumes, $V^{\mathrm{sat}}{}_l$, to within 1.5%.

Corresponding states have been used in other equations. For example, the Peng-Robinson equation is a modified Redlich-Kwong equation formulated to better correlate vapor–liquid equilibrium (VLE) vapor pressure data. This equation, however, is not useful in reduced form because it is specifically designed to calculate accurate pressure data. Reduced equations generally presuppose knowledge of the pressure.

2.1.4. Three Parameter Models

Most fluids deviate from the predicted corresponding states values. Thus the acentric factor, ω, was introduced to account for asymmetry in molecular structure (79). The acentric factor is defined as the deviation of reduced vapor pressure from 0.1, measured at a reduced temperature of 0.7. In equation form this becomes:

$$\omega = -\log P_r - 1.000 \qquad (56)$$

where P_r is evaluated at $T_r = 0.7$. This definition was chosen because of the abundance and accuracy of vapor pressure data near the normal boiling point. Acentric factors for various pure compounds are given in many references (3).

The acentric factor, ω, was the third parameter used (20) in an equation based on the second virial coefficient. This equation was further modified and is suitable for reduced temperatures above 0.5.

$$\left(\frac{BP_c}{RT_c}\right) = \left(\frac{B^{(0)}P_c}{RT_c}\right) + \omega\left(\frac{B^{(1)}P_c}{RT_c}\right) \qquad (57)$$

where

$$\left(\frac{B^{(0)}P_c}{RT_c}\right) = 0.1445 - \frac{0.330}{T_r} - \frac{0.1385}{T_r^2} - \frac{0.0121}{T_r^3} - \left(\frac{0.000607}{T_r^8}\right)$$

$$\left(\frac{B^{(1)}P_c}{RT_c}\right) = 0.073 + \frac{0.46}{T_r} - \frac{0.50}{T_r^2} - \frac{0.097}{T_r^3} - \frac{0.0073}{T_r^8}$$

The second virial coefficient is related to the compressibility factor:

$$Z = 1 + B^{(0)}\frac{P_r}{T_r} + \omega B^{(1)}\frac{P_r}{T_r} \qquad (58)$$

2.1.5. Four Parameter Models

Two- and three-parameter theories are only accurate for simple, normal, and some slightly polar fluids. In order to accurately predict polar fluid behavior a fourth parameter is needed (80). The Stiel polarity factor, χ, is one such fourth parameter and follows from the equation

$$\log P_r = (\log P_r)^{(0)} + \omega(\log P_r)^{(1)} + \chi(\log P_r)^{(2)} \qquad (59)$$

where $(\log P_r)^{(0)}$ and $(\log P_r)^{(1)}$ are the same as those obtained for a normal fluid by equation 57.

To correlate χ to the acentric factor a quadratic Taylor series in terms of the compressibility factor was formulated. This equation is represented as

$$Z = z^{(0)} + \omega z^{(1)} + \chi z^{(2)} + \chi\omega z^{(3)} + \chi^2 z^{(4)} \qquad (60)$$

for $0 < \omega < 0.65$; $-0.06 < \chi < 0.04$ where the values of $z^{(0)}$ and $z^{(1)}$ are the same as those for normal fluids. The values of $z^{(2)}$, $z^{(3)}$, and $z^{(4)}$ have been determined for the gaseous and liquid regions for reduced temperatures from 0.8 to 1.15 and reduced pressures from 0.2 to 6.0 (81). The average percentage error for polar fluids using this method is below 1% compared to over 10% error if the Pitzer correlation is used.

2.2. Application to Other Properties

2.2.1. Extension of Generalized Charts

In 1975, the usefulness of generalized charts was extended upon the publication of extensive tables of residual enthalpy, entropy, and heat capacity (82). This tabular data has also been converted into graphical form (3). The corresponding equations incorporate the acentric factor:*Residual enthalpy*:

$$\left[\frac{H^* - H}{RT_c}\right] = \left[\frac{H^* - H}{RT_c}\right]^{(0)} + \omega\left[\frac{H^* - H}{RT_c}\right]^{(1)} \qquad (61)$$

Residual entropy:

$$\left[\frac{S^* - S}{R}\right] = \left[\frac{S^* - S}{R}\right]^{(0)} + \omega\left[\frac{S^* - S}{R}\right]^{(1)} \qquad (62)$$

Heat capacity:

$$\left[\frac{C_p - C_p^*}{R}\right] = \left[\frac{C_p - C_p^*}{R}\right]^{(0)} + \omega\left[\frac{C_p - C_p^*}{R}\right]^{(1)} \qquad (63)$$

where * denotes an ideal gas value. Combining these equations with ideal-gas property values gives accurate results for any system obeying the corresponding states theory.

Another property which can be represented by generalized charts is fugacity, φ. The fugacity of a substance can be regarded as a corrected vapor pressure. At low pressures (below atmospheric) the use of pressure in the place of fugacity leads to little error in calculations. The fugacity coefficient is defined by

$$\phi = \frac{f}{P} \qquad (64)$$

where φ is the fugacity coefficient, f is the fugacity, and P is the pressure. In the limit as the pressure approaches zero the fugacity coefficient is unity. If the equation

$$\ln f_{\mathrm{pure}i} = \int_0^P Z_{\mathrm{pure}i}\,(d\ln P) \qquad (65)$$

where pure i denotes a pure component i and $Z_{\mathrm{pure}i}$ is its compressibility factor, is evaluated by graphical integration and the results are plotted as the fugacity coefficient vs the reduced pressure for constant reduced temperatures, a generalized chart can be constructed.

A generalized chart for the prediction of the fugacity coefficients for 23 compounds using actual PVT data and averaging or smoothing the curves has been constructed (83). The predicted values from the generalized chart have been compared to the experimental values, and agreement within 4% obtained. No limitations were given, but in general the charts should not be used for highly polar substances, quantum fluids, or associating substances.

One of the most versatile and accurate generalized correlations for the prediction of the fugacity coefficient (3) involves a three-parameter generalized correlation which takes advantage of the acentric factor. The correlation breaks the fugacity coefficient into two parts: φ^0 and φ^1.

$$\phi = \left(\phi^0\right)\left(\phi^1\right)^{\omega} \qquad (66)$$

Generalized charts are available for the calculation of φ^0 and φ^1 based on experimental data (61, 65, 66). A generalized chart for methane and pentane has also been made (84).

2.2.2. Generalized Surface Tension Correlations

Use of the principle of corresponding states has provided a practical and accurate method for the estimation of surface tensions. The functional relationship for the surface tension of a pure substance (85) is

$$\sigma_r = \sigma_r\,(T_r; \alpha_k) \qquad (67)$$

where

$$\sigma_r = \frac{\sigma}{P_c^{2/3}\,(xT_c)^{1/3}} \qquad (68)$$

x is a constant that depends on the units of P_c and T_c (86) and

$$\sigma = 4.6 \times 10^{-4}\left[P_c^{2/3} T_c^{1/3} \alpha_k (1 - T_r)^{11/9}\right] \qquad (69)$$

$$\alpha_k = 0.1207\left[1 + \frac{T_b}{T_c}\,\frac{(\ln P_c - 11.526)}{\left(1 - \dfrac{T_b}{T_c}\right)}\right] - 0.281$$

For 84 simple inorganic substances and a variety of organic compounds, equation 69 produced an average error of 3.0% (87). The above correlations should not be applied to the quantum substances, associating substances, or highly polar substances.

In 1971 the Stiel polar factor was incorporated in an equation that accounts for polar fluids (88)

$$\sigma = P_c^{2/3} T_c^{1/3} \left(\sigma_r \,\Big|_{T_r = 0.6}\right)\left(\frac{1 - T_r}{0.4}\right)^m \qquad (70)$$

where

$$\sigma_r \,\Big|_{T_r = 0.6} = 0.1574 + 0.359\omega - 1.769\chi - 13.69\chi^2 - 0.510\omega^2 + 1.298$$

and

$$m = 1.210 + 0.5385\omega - 14.61\chi - 32.07\chi^2 - 1.65\omega^2 + 22.03\omega\chi$$

This equation was empirically derived from 16 polar fluids and has an average error of 2.9%. A technique for estimating surface tension using nonretarded Hamaker constants (89) has also been presented.

2.2.3. Generalized Correlations for Viscosity

Gas viscosity has also been predicted by corresponding states theory (90) using

$$\eta\xi = \left[0.807 T_r^{0.618} - 0.357 e^{(-0.449 T_r)} + 0.340 e^{(-4.058 T_r)} + 0.018\right] F_P^0 F_Q^0 \qquad (71)$$

where ξ is the reduced, inverse gas viscosity and is defined as

$$\xi = 0.176\left(\frac{T_c}{M^3 P_c^4}\right)^{1/6} \qquad (72)$$

where M is the molecular weight of the gas. For nonpolar, nonquantum gases the values of the low pressure polarity and quantum factors, $F^0{}_P$ and $F^0{}_Q$, are equal to 1.

2.2.4. Miscellaneous Generalized Correlations

Generalized charts and corresponding states equations have been published for many other properties in addition to those presented. Most produce accurate results over a wide range of conditions. Some of these properties include: (*1*) transport properties (64, 91); (*2*) second virial coefficients (80, 92); (*3*) third virial coefficients (72); (*4*) liquid mixture activity coefficients (93); (*5*) Henry's constant (94); and (*6*) diffusivity (95).

2.3. Reference Substances

Use of a reference substance has its origins in the work of Clausius-Clapeyron and equation 73, a form of equation 7:

$$\ln P^{\text{sat}} = -\frac{\Delta H^{\text{vap}}}{RT} + I \qquad (73)$$

where I is an integration constant. If the natural log of the vapor pressure, P^{sat}, is plotted vs the inverse of absolute temperature, T, then the slope of the line becomes the heat of vaporization, ΔH^{vap}, divided by the universal gas constant. A plot called a Cox chart can be constructed by plotting the ln P^{sat} of an unknown compound vs the ln P^{sat} of a well-documented (3) substance at equal temperatures. By recording only the temperature at which the reference vapor pressure was measured on the x axis, the Cox chart becomes a valuable tool for estimating vapor pressures at temperatures for which no experimental values are available.

In another type of reference plot the temperature of a property of compound A is plotted vs the temperature of the reference substance at equal vapor pressures values.

The values for a single property of two compounds, A and B, are useless unless these values are compared at equal temperature or pressure. Then a deviation from some intermediate value can be determined. If this intermediate value is chosen to be the value of one particular substance, ie, A, the reference substance, both A and B can then be expressed as functions of the reference substance. One very simplistic example is specific gravity where the density of a compound is expressed as the actual density divided by the density of water at 4°C and water is the reference substance.

By use of Othmer plots of reference substances, large tables of thermodynamic data can be expressed as simple correlations which are extremely accurate and easy to use. The real power of these correlations is the ability to interpolate and extrapolate the correlations beyond the experimental values with considerable accuracy. Mathematically stated

$$\log(\textit{property}) = m\log(\textit{property}_{\text{ref}}) + C \qquad (74)$$

where the coefficient m and C are determined by graphing experimental data for the chemical of interest at the same critical difference temperature, T_{D}, where $T_D = T_c - T$, vs reference values.

Correlations based on reference substances are limited to compounds which have experimentally determined values, but the number of data points needed to produce a correlation is relatively small. A reference substance should be as chemically and physically compatible to the chemical with which it is being compared as possible. Use of reference substances in the ideal assumes no property deviations, thus the smaller the deviations, the lower the absolute error in the correlations.

Values for many properties can be determined using reference substances, including density, surface tension, viscosity, partition coefficient, solubility, diffusion coefficient, vapor pressure, latent heat, critical properties, entropies of vaporization, heats of solution, colligative properties, and activity coefficients. Table 1 lists the equations needed for determining these properties.

Compared to the theory of corresponding states, the reference substance method gives highly accurate results for compounds having sparse experimental data. The corresponding states method gives moderate accuracy for numerous compounds even without actual data.

Table 1. Othmer Plot Equations

Property	Equation[a]	Comments
surface tension, γ	$\log\gamma = m\log\gamma' + C$	m = ratio of surface activation energies
liquid density, d	$\log d = m\log d' + C$	$m = 1$ for most organic liquids
viscosity, η	$\log\eta = m\log P' + C$	P' is the vapor pressure of the reference substance
solubility in liquids, x	$\log x = \frac{(h_l - h_s)}{L'}\log P' + C$	L' is the molal latent heat of reference substance, h_s and h_l are the molal heat content of the solid and saturated solution, respectively
partition coefficients, K_D	$\log K_D = \frac{\Delta H}{L'}\log P' + C$	ΔH is molal heat of transfer of solute between the two layers
diffusion coefficients, D	$\log D = \left(\frac{E_d}{L'}\right)\log P' + C$	E_d is the energy of activation for diffusion
vapor pressure, P	$\log P = m_T\log P' + C$	m_T is the ratio of the latent heats per mole

[a] The superscript prime denotes the reference substance.

Table 2. Curve Equations and Linear Reduction Methods

Curve	Equation	Method of reduction
linear	$y = a + bx$	plot y vs x; $a = y -$ intercept; b = slope
parabolic	$y = a + cx^n$	first obtain $a = y -$ intercept of y vs x plot; then plot log $(y - a)$ vs log x; c = antilog of new y-intercept at $x = 1$; n = slope
polynomial quadratic[a]	$y = a + bx + cx^2$	first obtain $a = y -$ intercept on y vs x plot; then plot $y - y_n/x - x_n$ vs x where $y_\mathrm{n}, x_\mathrm{n}$ are the coordinates of any point on a smooth curve through the experimental points; c = slope; $b + cx_n = y -$ intercept
exponential	$y = ab^x$	plot log y vs x; a = antilog of y-intercept; b = antilog of slope
	$y = ae^{bx}$	plot ln y vs x; a = antilog of y-intercept; b = slope
hyperbolic	$y = a + b/x$	plot y vs $1/x$; $a = y -$ intercept; b = slope
logarithmic	$y = a\log x$	plot y vs log x; a = slope
logistic	$1/y = a + be^x$	plot $1/y$ vs e^x; $a = y -$ intercept and b = slope or plot $1/y$ vs x; $a = y -$ intercept; then plot $\log\left(\frac{1}{y} - a\right)$ vs log x; b = antilog of y-intercept

[a] For the general polynomial quadratic equation, $y = z + bc + cx^2 + dx^3 + \cdots$, graphical procedures are almost impossible.

3. Empirical Representations of Data

3.1. Fitting Simple Functions

In many engineering applications there may be no apparent theoretical basis for the relationship of two variables or the relationship may be too complex to apply. Thus the search for a correlating equation form may at first be along empirical lines. A simple plot of the data in ordinary Cartesian coordinates gives an immediate indication of the essential form of the data.

Each of the graphs shown in Figure 1 suggests an equation given in Table 2 by which a data set may be correlated. For most of these equations, the method of reduction to a linear form may be apparent, possibly through a transformation of variables. Typically, it is desirable to reduce the information to a linear form because the constants of the resulting linear equation can be found either directly from the slope and intercept of the straight line graph, or by means of analytical techniques.

Once the form of the correlation is selected, the values of the constants in the equation must be determined so that the differences between calculated and observed values are within the range of assumed experimental error for the original data. However, when there is some scatter in a plot of the data, the best line that can be drawn representing the data must be determined. If it is assumed that all experimental errors (ϵ) are in the y values and the x values are known exactly, the least-squares technique may be applied. In this method the

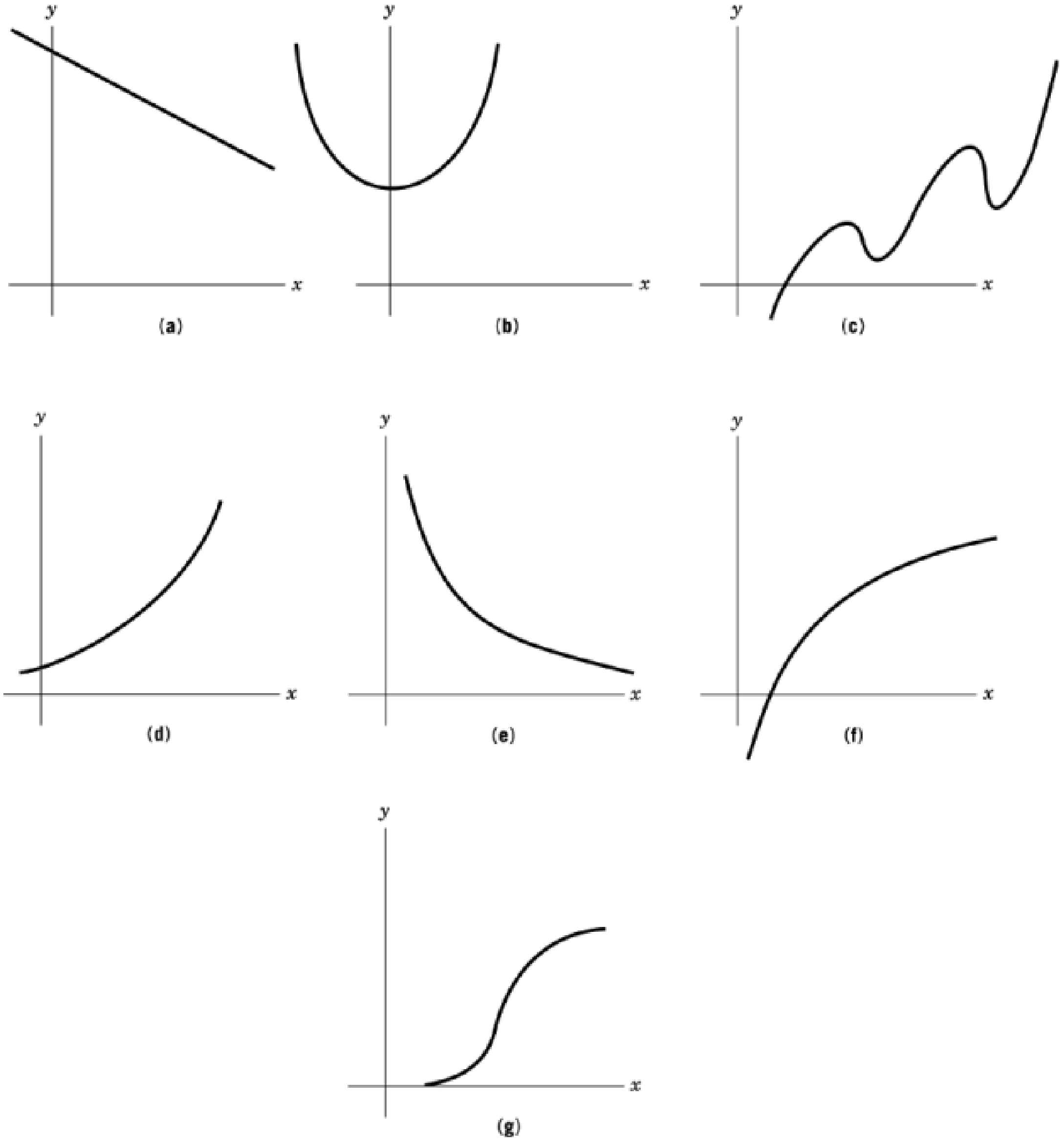

Fig. 1. Some common curves: (**a**) linear; (**b**) parabolic; (**c**) polynomial; (**d**) exponential; (**e**) hyperbolic; (**f**) logarithmic; and (**g**) logistic. See Table 2 for corresponding equations.

constants of the best line are those that minimize the sum of the squares of the residuals, ie, the difference, α, between the observed values, y, and the calculated values, Y. In general, this sum of the squares of the residuals, R, is represented by

$$R = \sum_{i=1}^{n} \alpha_i^2 = \sum_{i=1}^{n} (y_i - Y_i)^2 \qquad (75)$$

For a linear correlation in Y with one independent variable, x, and constants a and b:

$$Y = a + bx \qquad (76)$$

Substitution for Y in equation 75 yields

$$R = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = 0 \qquad (77)$$

where n is the number of data points for which a correlation is required. Values of the two coefficients that minimize R in this equation are found by taking its partial derivative with respect to each of the coefficients and setting these derivatives equal to zero:

$$\frac{\partial R}{\partial a} = -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \qquad (78)$$

$$\frac{\partial R}{\partial b} = -2\sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0 \qquad (79)$$

Rearrangement yields a set of linear equations with two unknowns:

$$\sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} y_i \qquad (80a)$$

$$\sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} bx_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (80b)$$

Additional rearrangement gives the so-called normal equations:

$$na + b\Sigma x_i - \Sigma y_i = 0 \qquad (81)$$

$$a\Sigma x_i + b\Sigma x_i^2 - \Sigma x_i y_i = 0 \qquad (82)$$

which can be readily solved for a and b as follows:

$$b = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad (83)$$

$$a = \frac{\Sigma y_i - b\Sigma x_i}{n} \qquad (84)$$

Thus finding the best values for the slope, b, and intercept, a, of a straight line simply involves developing the sums indicated and combining them in equations 83 and 84.

The least-squares procedure can be applied to the transformed variables of any of the equations in Table 2, where a simple transformation of one or both of the variables results in a linearized expression. The sums for equations 83 and 84 must be formed from the transformed variables rather than from the original data.

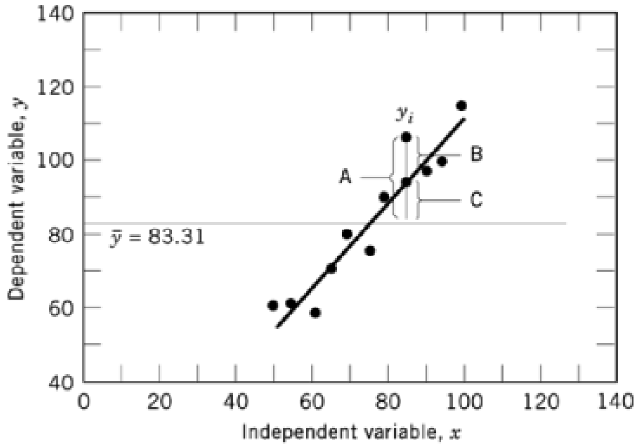

Fig. 2. Least-squares plot showing the determinants for the coefficient of determination. A, total deviation $(y_i - \bar{y})$; B, unexplained deviation $(y_i - y)$; and C, explained deviation $(y - \bar{y})$.

Prior to correlation of the data, all that can be considered is the scatter of all the $y_{\rm i}$ values about the mean $\bar{y}$:

$$\sum (y_i - \bar{y})^2 \qquad (85)$$

After obtaining a correlation, it is more meaningful to speak of the scatter of the $y_{\rm i}$ values around the correlating function:

$$\sum (y_i - Y)^2 \qquad (86)$$

Subtracting $(y_i - \bar{y})$ from $(y_i - Y)$ gives $(Y - \bar{y})$, which is that portion of the deviation of any data point from the mean which is explained by the correlation. The coefficient of determination, γ^2, a statistical parameter which varies from 0.0 to 1.0, is defined as:

$$\gamma^2 = \frac{\Sigma\,(Y - \bar{y})^2}{\Sigma\,(y_i - \bar{y})^2} \qquad (87)$$

How well the developed equation fits the data is given by the correlation coefficient, γ (0.0 = poor fit, ±1.0 = excellent fit), which is the square root of the coefficient of determination. The physical significance of this ratio is apparent from the graph shown in Figure 2. Simple algebraic transformation gives the following more useful form:

$$\gamma^2 = \frac{\left[\Sigma x_i y_i - \dfrac{\Sigma x_i y_i}{n}\right]}{\left[\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}\right]\left[\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}\right]} \qquad (88)$$

This involves the same summations of data values which were required in the evaluation of the constants a and b and can be determined at the same time.

The least-squares technique can be extended to any number of variables as long as the equation is linear in its coefficients. The linear correlation of y vs x can be extended to the correlation of y vs multiple independent variables generating an equation of the form:

$$Y = a + bx_1 + cx_2 + dx_3 + \cdots \Omega x_n \qquad (89)$$

where $x_1, x_2, \ldots x_n$ are n different variables. For example, y might be the yield of a reaction product, whereas x_1 might be the reaction temperature, x_2 the reaction pressure, x_3 the concentration of a catalyst, x_4 the space velocity, and so on.

The coefficients of equation 89 can be determined by an extension of the procedure outlined earlier, ie, by minimizing the sum of the squares of the deviations of observed y values from those that would be predicted by the model. Substitution of equation 89 into equation 75 yields:

$$R = \sum_{i=1}^{n} (y_i - a - bx_{1i} - cx_{2i} - dx_{3i} - \cdots \Omega x_{ni}) \qquad (90)$$

Setting

$$\frac{\partial R}{\partial a} = 0, \qquad \frac{\partial R}{\partial b} = 0, \qquad \text{etc} \qquad (91)$$

results in a set of simultaneous equations that determine simple sums of terms involving the experimentally determined data values.

$$ma + b\Sigma x_1 + c\Sigma x_2 + d\Sigma x_3 + \cdots \Omega\Sigma x_n = \Sigma y \qquad (92)$$

$$a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2 + d\Sigma x_1 x_3 + \cdots \Omega\Sigma x_1 x_n = \Sigma x_1 y \qquad (93)$$

$$a\Sigma x_2 + b\Sigma x_1 x_2 + c\Sigma x_2^2 + d\Sigma x_2 x_3 + \cdots \Omega\Sigma x_2 x_n = \Sigma x_2 y \qquad (94)$$

$$a\Sigma x_n + b\Sigma x_1 x_n + c\Sigma x_2 x_n + d\Sigma x_3 x_n + \cdots \Omega\Sigma x_n^2 = \Sigma x_n y \qquad (95)$$

In principle, this set of equations can be solved for the various constants, a through Ω, just as a and b were obtained previously. In practice, however, the actual numerical evaluation involves considerable computation in all but the simplest examples. Computer solution by matrix techniques designed specifically to handle this type of data correlation problem is usually required.

Curve fitting to data is most successful when the form of the equation used is based on a known theoretical relationship between the variables associated with the data points, eg, use of the Clausius-Clapeyron equation for vapor pressure. In the absence of known theoretical relationships, polynomials are one of the most useful forms to describe a curve. Polynomials are easy to evaluate; the coefficients are linear; and the degree, ie, the highest power appearing in the equation, is a convenient measure of smoothness. Lower orders yield smoother fits.

Fitting polynomials to data points involves essentially the same technique as for correlation of multiple variables. Replacing x_1 through x_n in equation 89 by x^1 through x^n results in the polynomial:

$$Y = a + bx + cx^2 + dx^3 + \cdots \Omega x^n \qquad (96)$$

Continuing this substitution through the resulting set of simultaneous equations and solving them generates the coefficients of this polynomial equation. This process is termed polynomial regression or curvilinear regression. It should be noted that equation 96 is still linear in the coefficients a through Ω.

A nonlinear model would involve one or more of the coefficients in nonlinear form such as the Antoine equation (eq. 8) or

$$Y = a + be^{-cx} \qquad (97)$$

or

$$Y = ae^{bx} + cx^2 \qquad (98)$$

The nonlinear constant c in these equations cannot be evaluated directly by the methods previously described. Even forms such as these can be handled, however. For example, subtracting a trial value of a from y and taking logarithms transforms equation 97 into the linear form:

$$\ln(Y - a) = \ln b - cx \qquad (99)$$

from which the constants b and c as well as the correlation coefficient can be determined. Iterations on the trial values of the constant a establishes the value which results in the best fit. Following well-known techniques, the familiar Antoine vapor pressure correlation (eq. 8) can be evaluated using the substitutions $a = A$; $b = AC - B$; and $c = -C$; yielding

$$aT + b - c\log_{10}P^{\text{sat}} - T\log_{10}P^{\text{sat}} = 0 \qquad (100)$$

which can be handled quite easily by standard techniques. Upon substitution of sufficient data to yield the constants a, b, and c one can find the Antoine constants from $A = a$; $B = -(ac + b)$; and $C = -c$. Equation 100 can represent the vapor pressure of many pure substances from the triple point to the critical point.

Some formulas, such as equation 98 or the van der Waals equation, are not readily linearized. In these cases a nonlinear regression technique, usually computational in nature, must be applied. For such nonlinear equations it is necessary to use an iterative or trial-and-error computational procedure to obtain roots to the set of resultant equations (96). Most of these techniques are well developed and include methods such as successive substitution (97, 98), variations of Newton's rule (99–101), and continuation methods (96, 102).

The usual practice in these applications is to concentrate on model development and computation rather than on statistical aspects. In general, nonlinear regression should be applied only to problems in which there is a well-defined, clear association between the independent and dependent variables. The generalization of statistics to the associated confidence intervals for nonlinear coefficients is not well developed.

3.2. Nomographs

The word *nomograph*, from the Greek, means the image of the law, so literally a nomograph is any diagram that represents a mathematical function. The function is evaluated by drawing lines through axes. These lines intersect a result curve. There are three types of nomographs, the Cartesian graph, slide rule, and alignment chart.

Alignment charts, developed in 1899 (103), are sometimes difficult to interpret. Whereas the Cartesian graph of a line looks like a line, the alignment diagram of a line is a point between two axes. By extension, a three-axis Cartesian surface may be represented by a line straddling three axes. This ability to condense an immense amount of data to simple graphical correlations is the strength of alignment charts and nomographs.

Nomographs were first introduced in the form of the Cartesian graph. Characterized by perpendicular axes and reference lines, two-axis Cartesian graphs can be designed to show the solution to any two-variable equation. Expression of three-variable equations is possible with the use of families of curves or three-axis surfaces. The former involves considerable interpolation, because a desired value often lies somewhere between two curves on the graph. Unlike a Cartesian graph, the axes of a nomograph are not necessarily perpendicular, nor parallel, nor even linear. The main feature of an alignment chart is the reference line, which is not necessarily perpendicular to any axis. All nomography is based on geometrical construction and there are great differences in the appearance of various forms of nomographs.

3.2.1. Comparison of Alignment Charts and Cartesian Graphs

There are typically fewer lines on an alignment chart as compared to Cartesian plots. This reduces error introduced by interpolation and inconsistency between scales. For example, to find a point (x,y) on a Cartesian graph one draws two lines, one perpendicular to each axis, and these reference lines intersect at the point (x,y). This point (x,y) may correspond to some finite value found by reading a contour map represented by a family of curves corresponding to different values of the function.

An alignment chart is used by drawing one reference line through the two axes. This reference line, which need not be perpendicular to either axis, crosses a result axis at a unique finite value. This result axis represents the contour map on a Cartesian graph. Thus each line on an alignment chart represents a point on a Cartesian graph.

3.2.2. Construction of Alignment Charts

Of the ways to construct alignment charts, the brute force method, which requires some idea of the geometry for the chart, is the easiest method to use. The mathematical method, which uses parametric equations of scale to determine the placement and scale of each axis, is the most accurate, but the most difficult to apply.

3.2.2.1. Brute Force Method. An example of this method is demonstrated in Figure 3 for the three-variable equation, $x + 3y = z$, which requires two known points in order to find the third. The first step is to determine two axes and the appropriate scales, drawing two parallel lines and marking the lines with evenly spaced graduations as shown in Figure 3. This is the basic form of the parallel alignment chart. The scales should be evenly spaced. The third axis is determined by evaluating the equation at several points. The position of the third axis is found by evaluating the equation for several sets of points which share the same value for the third axis. This can be done for two values, in order to find two points which lie on the unknown axis. A straight line is used to define the axis. The scales are marked similarly by evaluating the equation for useful values on the third axis. Because any value of z is found only on one point on each of the other axes, the placement of the z-axis can be determined by choosing an arbitrary value for z and solving for values of x and y which satisfy the equation. For example, if $z = 13$, then if $y = 1$, then $x = 13 - 3(1) = 10$, giving $(x, y, z) = (10, 1, 13)$; if $y = 3$, then $x = 13 - 3(3) = 4$, giving $(x, y, z) = (4, 3, 13)$. The lines drawn connecting each of these pairs of points intersect at the point on the z-axis, where $z = 13$. Another point on the axis can be found by choosing a new value for z. Setting $z = 40$, if $y = 10$, then $x = 40 - 3(10) = 10$, giving $(x, y, z) = (10, 10, 40)$; if $y = 9$, then $x = 40 - 3(9) = 13$, giving $(x, y, z) = (13, 9, 40)$. A vertical line connects these intersection points to define the z-axis.

The scale can be determined in much the same way, by choosing the values to mark on the axis and evaluating the equation for sets of points. The limits of this axis can be found by evaluating the extremes. For $x = 1$, $y = 1$, $z = 4$; for $x = 13$, $y = 13$, $z = 65$. So values of z can be chosen anywhere between 4 and 65. If increments of 10 are desired, then values of x and y which, when evaluated with the equation, give the desired values of z incremented by 10, should be found. A comparison between the Cartesian representation and the alignment chart equivalent is shown in Figure 3. The alignment chart is simpler because of the smaller number of lines.

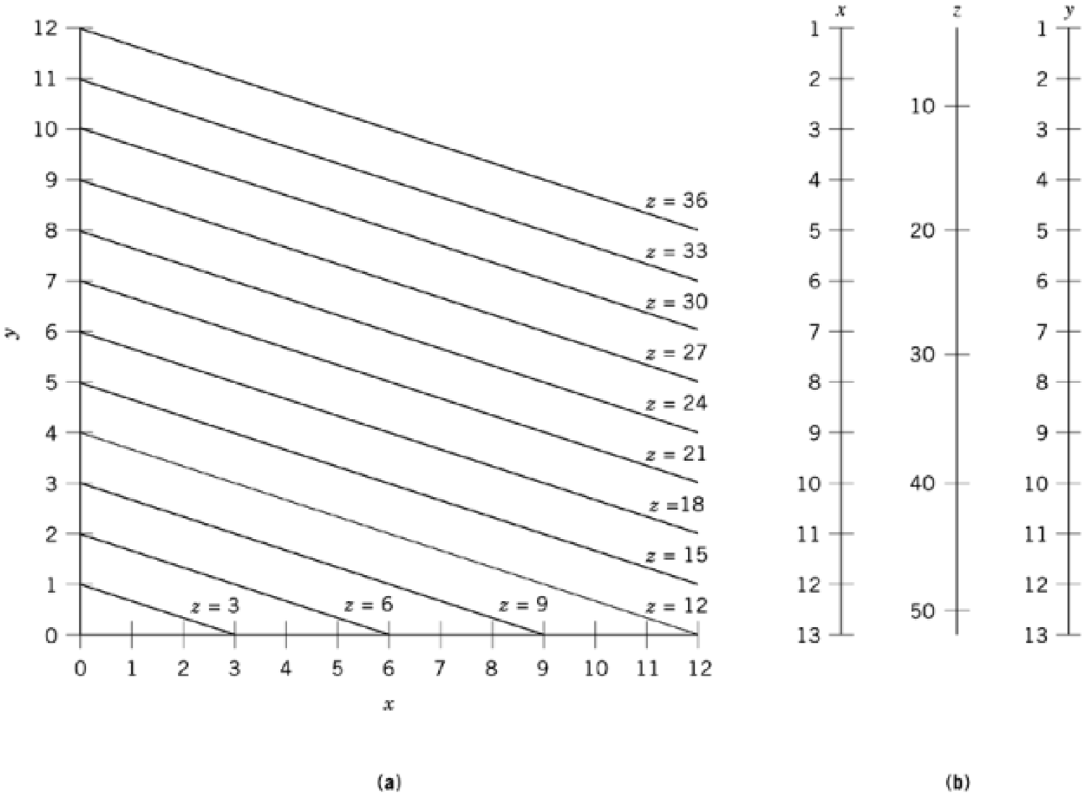

Fig. 3. Comparison for the equation $x+3y=z$ of (**a**) Cartesian and (**b**) nomographic formats. See text.

3.2.2.2. Mathematical Derivation of Scales. The brute force method for determining scale is not always the most convenient. Scale direction and axis placement can be difficult to guess. In addition, there may be some advantage to nonlinear axes. An accurate method, based on the use of parametric equations for determining the scale has been developed using determinants (104–107). For complex equations, it is possible to break the problem into several parts that can be simultaneously evaluated. Because any function can be broken into a set of parametric functions, the task is to graphically interpret the functions. The parametric equations can be converted into equations of scale which differ from chart to chart, depending on whether the function involved is linear (eq. 101x), exponential (eq. 101y), or inverse (eq. 101z):

$$L\,[t] = mf\,(t) + n$$

$$L\,[t] = m\log f\,(t) + n \qquad (101y)$$

$$L\,[t] = mf\,(t)^{-1} + n$$

where L is the linear placement of the scale value; $f(t)$ is the value of the function; t is to be marked on the scale; m is the modulus of the scale and determines the length of the axis; and n determines the placement of the scale. These parametric equations of scale can be expanded to be used on a Cartesian plane with both

x- and y-axis coordinates

$$L\,[t] = \left\{ m_x f\ (t) + n_x,\, m_y f\ (t) + n_y \right\} \qquad (102)$$

giving an x- and a y-coordinate for each mark of the t-axis. A full axis can be derived with this equation. A set of parametric equations can be derived that produce simultaneous solutions to a function. The challenge is to align them properly.

Each of the parametric equations that can be formed from an expression represents an axis on the chart. Each set of parametric equations must simultaneously agree with the equation they represent. In other words, on a line drawn through any two variables, a third variable can be found which satisfies the parametric equations. So an evaluation of the chart requires that values produced by each parametric equation be on the chart as a line. A determinant can be used to determine whether or not points are collinear. The parametric equations must be evaluated so they always produce values which lie on a line. By replacing the x and y points with parametric equations of scale for the chart, it is possible to create any diagram. This method can be used to determine the placement of the axes, because the parametric equations can be transformed into equations of scale.

3.2.3. Various Nomographs

The alignment chart is restricted neither to addition operations, nor to three-variable problems. Alignment charts can be used to solve most mathematical problems, from linear ones having any number of variables, to ratiometric, exponential, or any combination of problems. A very useful property of these alignment diagrams is the fact that they can be combined to evaluate a more complex formula. Nomographs for complex arithmetical expressions have been developed (108).

3.3. Computer Techniques

Rapid development of a wide range of data regression techniques has occurred in the latter 1900s. Basic to most engineering work stations are databases (qv) for efficient data retrieval and management, and spreadsheets (109) in which numerous routine calculations can be quickly correlated, regressed, or otherwise manipulated (see also Computer-aided design and manufacture (CAD/CAM); Computer-aided engineering (CAE)).

The ability to perform complex regressions and matrix inversions are subjects in numerous undergraduate and graduate courses (110). Problems requiring estimated solutions to sparse matrices, eg, group contribution parameter estimations, have been found to be mathematically similar to annealing of metals and mathematical Boltzmann machines (111, 112). Numerous software packages have been developed for specific applications, such as kinetics (113) and chemical formulations (114), and as general tools for modeling and reformulation (115) and expert systems (qv) (116).

4. Molecular-Based Methods

In the most general sense, a molecular-based property prediction method can be thought of as any method in which the physical features of a molecule (or molecules for multicomponent properties) are correlated to a property. Some methods, such as molecular thermodynamics, generalize characteristics of whole molecules to correlate them to bulk properties consistent with classical thermodynamics. Other methods correlate molecular fragments, bonds, atomic compositions, molecular topological features, and solution interactions.

4.1. Molecular Thermodynamics, the Kinetic Theory of Gases, and Statistical Thermodynamics

Molecular thermodynamics attempts to explain observed physical properties via individual molecular properties (117). It uses classical thermodynamics, as well as concepts from statistical thermodynamics and chemical physics (118). The molecular view of thermodynamics can be traced back to 1738, when Bernoulli proposed that a gas consisted of individual molecules in constant motion, colliding with each other and with their containment vessels. During the nineteenth century, Maxwell and Boltzmann developed this into the kinetic theory of gases (119–121), which applies the laws of mechanics to microsystems. From this theory expressions for the pressure of a gas, its internal energy, and its specific heat capacity can be derived (122). Statistical thermodynamics, or equilibrium statistical mechanics, provides the mathematical glue between the kinetic theory and macrosystems, ie, systems containing large numbers of molecules. It takes advantage of the fact that molecules are very numerous and that average properties can be estimated through probability and statistical analysis (122).

For example, the measured pressure exerted by an enclosed gas can be thought of as a time-averaged manifestation of the individual molecules' random motions. When one considers an individual molecule, however, statistical thermodynamics would propose its random motion or pressure could be quite different from that measured by even the most sensitive gauge which acts to average a distribution of individual molecule pressures. The particulate nature of matter is fundamental to statistical thermodynamics as opposed to classical thermodynamics, which assumes matter is continuous. Further, these elementary particles and their complex substructures exhibit wave properties even though intra- and interparticle energy transfers are quantized, ie, not continuous. Statistical thermodynamics holds that the impression of continuity of properties, and even the solidity of matter is an effect of scale.

The kinetic theory attempts to describe the individual molecules' energies and interactions; statistical thermodynamics attempts to fundamentally develop the equation of state from considerations of groupings of molecules. These approaches are complementary in many ways (3, 123, 124). A well-referenced text covering molecular thermodynamics is also available (125).

4.1.1. Basic Problem in Statistical Thermodynamics and Quantum States

All problems in statistical thermodynamics have been postulated to be essentially forms of the same problem (126), that is, given a constant energy level, E, for a closed set of identical systems, eg, molecules, determine the distribution of these systems over all possible states in which the set can exist. This set is often referred to as an assembly or ensemble. Implicit in this description is that whereas there is interaction between the systems, the interaction is so weak that it can be disregarded. Thus, from a mathematical point of view, each system is thought of as containing its own distinct energy and E is obtained by summing all distinct system energies.

Distinct molecular energies are found at discrete, or quantized, levels. One method to determine the energy associated with a quantum state of a system is by describing the spins, or angular-momentums of the components and subcomponents. The wave functions which are used in these descriptions, eg, Schrödinger's wave equation (126, 127), may be symmetric or antisymmetric depending on the spins. An antisymmetric wave function is one in which the sign of the wave function is dependent on component positions within the system. Components having even numbers of subcomponents have integral spins, and therefore symmetric wave functions, ie, sign not dependent on component position. The Pauli exclusion principle (128) indicates that within a system of components having antisymmetric wave functions, no two particles can have the same quantum state.

Systems containing symmetric wave function components are called Bose-Einstein systems (129); those having antisymmetric wave functions are called Fermi-Dirac systems (130, 131). Systems in which all components are at a single quantum state are called Maxwell-Boltzmann systems (122). Further, a boson is a particle obeying Bose-Einstein statistics, a fermion is one obeying Fermi-Dirac statistics (132).

Maxwell-Boltzmann particles are distinguishable, and a partition function, or distribution, of these particles can be derived from classical considerations. Real systems exist in which individual particles are indistinguishable. For example, individual electrons in a solid metal do not maintain positional proximity to specific atoms. These electrons obey Fermi-Dirac statistics (133). In contrast, the quantum effects observed for most normal gases can be correlated with Bose-Einstein statistics (117). The approach to statistical thermodynamics described thus far is referred to as wave mechanics. An equivalent quantum theory is referred to as matrix mechanics (134–136).

4.1.2. Ensembles and Postulates

Once estimations of time-averaged energies for specific quantum states of identical systems have been made, a large number of identical systems called an ensemble is defined. An ensemble is typically defined according to a set of boundary conditions encountered in a real system. The ensemble is a mathematical construct that facilitates summation of all possible states. The typical method is to assign a state to each system in an ensemble. If the number of systems is great enough, summing an ensemble of systems with fixed states arrives at the same total as averaging all possible states of a single system and multiplying by the number of systems. The microcanonical and canonical ensembles are the two most frequently encountered in chemical thermodynamics (118). A microcanonical ensemble represents an isolated system having independent parameters, E; volume, V; and N molecules. A canonical ensemble represents a closed isothermal system with independent parameters N, V, and temperature T. An ensemble encountering fewer mathematical difficulties is the grand canonical ensemble in which restrictions on E and N are relaxed. T, V, and the chemical potentials μ_1, μ_2, ... of components 1,2, ... are the usual independent parameters.

Mechanical–thermodynamic property values (137) for each system are assigned based on quantum theory and the following postulates: (*1*) the most probable distribution of systems among all possible quantum states is analogous to the equilibrium property value from classical thermodynamics, and (*2*) the most probable distribution of systems among all possible quantum states is that distribution for which the number of quantum states of the systems and their surroundings is a maximum. Mechanical–thermodynamic properties are those for which a system in a quantum state would have a well-defined value, eg, pressure, volume, energy, and mass. Nonmechanical–thermodynamic properties, eg, temperature and entropy, do not have well-defined values for a system, and therefore can only be evaluated through a comparison with classical thermodynamic relationships. For example, the mass of an ensemble may be found by adding the masses of the constituent systems. However, the same cannot be said of temperature. The term classical refers to property values derived from macroscopic measurements, eg, mass, pressure, volume, and temperature, or calculated from these derived values, eg, enthalpy, entropy, and internal energy. A concise review of the mathematics necessary to calculate properties from these postulates based on Maxwell-Boltzmann particles is available (118).

4.1.3. Applications

Statistical thermodynamics holds great promise as a means to characterize molecules, bonds, reactions, and energy fields in ways which are consistent with both observed and calculated classical thermodynamic properties. However, it is exceedingly difficult, if not impossible, to solve analytically for the trajectories of just three mutually interacting bodies from quantum or classical mechanics (118), and the problem at hand involves on the order of 10^{23} bodies. Few physical property prediction methods are available that are based solely on statistical thermodynamics. However, its methodology is fundamental to group contribution and most other molecular-based physical property prediction methods (122, 137–141). Some insight into the applications of statistical thermodynamics to nonequilibrium problems is available (142, 143). An alternative approach in which field theory is applied to replace volume and pressure with strain and stress tensors, respectively, involves continuum mechanics (144). Areas in which molecular thermodynamics goes beyond classical thermodynamics is of interest. The second and third coefficients of the virial equation-of-state for real gases have been shown to correspond to binary and ternary molecular interactions through statistical thermodynamics (118, 121). From this information, and a basic understanding of intermolecular forces, eg, polar attractions, dipole and quadrupole moments, and dispersion forces (118), these coefficients have been predicted for simple gases (118, 138, 145). Equations-of-state for a series of low to medium pressure interacting gases have been developed, as well as thermodynamic properties of crystalline solids (138). Spectroscopic properties of various energy radiations related to electron and atomic transitions have been explained largely through statistical thermodynamics (146–148). Mathematical descriptions of radiation processes including gamma and x-rays, radiant heating, the propagation of light and the laser effect are demonstrated through statistical considerations (138, 149, 150). Other equilibrium statistical thermodynamic demonstrations can be found in prediction of magnetic properties (138, 151, 152). Nonequilibrium statistical thermodynamics are used to explain and predict Brownian motion (153, 154) and other physical phenomena based on particle fluctuations (138, 142, 143). *Ab initio* techniques are beginning to be developed to directly correlate atomic wave functions to physical properties (155–158); however, this area does not extend far beyond descriptions of simple species such as methane.

4.2. Group Contribution Methods

It has been shown that many macroscopic physical properties, ie, those derived from experimental measurements of bulk solutions or substances, can be related to specific constituents of individual molecules. These constituents, or functional groups, are usually composed of commonly found combinations of atoms. One procedure for correlating functional groups to a property is as follows. (*1*) A set of substances is selected for which the property has been determined experimentally. (*2*) Groups are determined for each substance based on a previously determined set of rules. (*3*) Property contributions for each group are determined based on a previously determined summation equation. This is done with a variety of computer optimization algorithms. (*4*) Property values are calculated based on the group contributions and compared to the experimental results. As a further test, property values are sometimes estimated for a selection of substances not included in the optimization.

Group contribution methods routinely predict properties of substances, both real and imagined, for which the only information known is structural. For example, if groups are selected such that each group has one carbon atom and its associated hydrogen atoms, then the groups CH_2 and CH_3 would be adequate to describe almost all normal alkanes, methane being the exception. Regression of a given physical property by an appropriate equation using these two groups would allow the prediction of the property of any normal alkane for which the structure could be written.

Aside from the rather trivial matter of predicting molecular weight by summing the mass contribution of constituent atoms, group contribution methods have numerous limitations. (*1*) Predictions for molecules of drastically different sizes than those in the optimization set usually exhibit large errors. (*2*) Short of having an experimental determination of the property for the compound of interest, there is no rigorous method for determining the accuracy of a group contribution-based prediction. Without experimental data, prediction of accuracy is inherently extrapolative and there is no means to validate that the compound of interest falls inside the range of application of the correlating equation. (*3*) Although attempts have been extensive, group contribution methods are predominantly empirical and involve numerous assumptions. Researchers are far from understanding the behavior of groups in a general way, such as the understanding of PVT relationships afforded by classical thermodynamics. (*4*) Seemingly similar substances are sometimes observed to be exceptions to the rule for many group contribution methods, (*5*) Percentage error in group contribution methodology can be as much as an order of magnitude or greater than that deemed acceptable for many applications, such as industrial plant design.

Recognizing that there is presently a need for property values for tens of thousands of substances, but experimental data for only a small percentage of these substances, group contribution methods are viewed as the only choices for many problems such as newly or yet-to-be-synthesized compounds, situations where available data are well outside the conditions of interest, and reaction kinetics studies involving unknown intermediates.

A relatively simple example of a group contribution technique is the method for estimating liquid and solid heat capacities (159). This method is a modification of Kopp's rule (160, 161) which was originally proposed in 1864. Kopp's rule states that, at room temperature, the heat capacity of a solid compound is approximately equal to a stoichiometric summation of the heat capacities of its atoms (elements). The Hurst-Harrison modified equation is as follows:

$$C_p = \sum_{i=1}^{n} (C_i \times N_i) + C_{\text{misc}} \times N_{\text{misc}} \qquad (103)$$

where C_p is specific heat in J/(K·mol) at 298 K; C_i is a constant associated with element *i;* N_i is the number of occurrences of element *i;* C_{misc} is a constant associated with elements not having a specific C_{i}; N_{misc} is the number of occurrences of elements not having a specific C_{i}; n is the number of different elements in the compound for which there are specific constants.

Values for 32 elements and the C_{misc} constant have been tabulated (159). An average absolute error of 9.6% for 721 solid compounds (13% standard deviation) and 8.8% for 477 liquid compounds (11.5% standard deviation) has been reported.

4.2.1. QSAR Methods for Fluid Solubility Prediction

Several group contribution methods for predicting liquid solubilities have been developed. These methods as well as other similar methods are often called quantitative structure-activity relationships (QSARs). This field is experiencing rapid development.

4.2.1.1. UNIFAC and ASOG Development. Pertinent equations of the UNIQUAC functional-group activity coefficient (UNIFAC) model for prediction of activity coefficients including example calculations are available (162). Much of the background of UNIFAC involves another QSAR technique, the analytical solution of groups (ASOG) method (163).

Predecessors to QSAR correlations include the Cox chart families (164), the local composition concept (165), and the development of mixing rules and solution theory by many researchers (118). During the 1960s a set of group selection rules was developed (25, 32, 49) for a broad range of organic compounds and a molecular solution model. Combination of this set of rules, the model, and emerging numerical computational capabilities resulted in ASOG, which allowed for the prediction of activity coefficients using no species-specific data beyond structure. With each successive modification or refinement, modeling of the interactions between molecules in solution improved.

The nonrandom, two-liquid (NRTL) equation (166) was developed along similar lines, and in 1975 the universal quasichemical (UNIQUAC) equation and UNIFAC models were published (28, 35). The latter was a group contribution method which was claimed to have comparable accuracy in activity-coefficient estimation to ASOG for vapor–liquid systems. This greatly improved accuracy for liquid–liquid systems. A more rigorous mathematical description of the original method (167) and several revisions have been published (28, 35–39, 168). Available groups and interactions through the 1991 revision (168) are provided in Table 3 and Figure 4, respectively. In addition, application-specific modifications have been proposed. ASOG development has continued at a slower pace.

The basic equational form of UNIFAC and many other QSARs is

$$GC = GC_{\text{combinatorial}} + GC_{\text{residual}} \qquad (104)$$

where GC is the group contribution, $GC_{\text{combinatorial}}$ is based only on composition and size and shape effects, and GC_{residual} depends on these effects and intermolecular forces. GCs are summed to estimate the activity

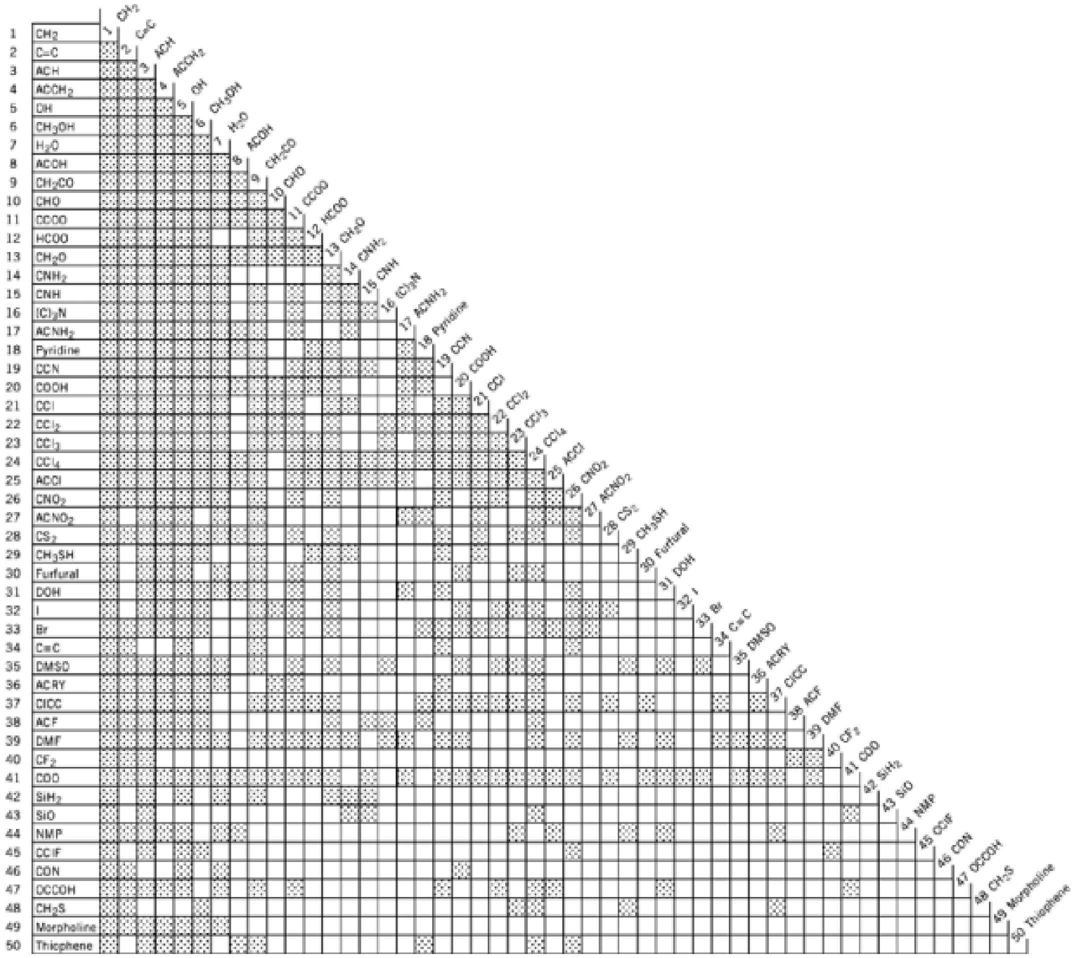

Fig. 4. UNIFAC group interaction parameter matrix, the ⊡ represents parameters fit and □ parameters not available (168). A represents an aromatic moiety. See Table 3 for additional terms.

coefficient or excess Gibbs energy. These *GC* parts are of logarithmic form consistent with the relationship between excess Gibbs energy and activity coefficients.

Table 3. UNIFAC Functional Groups[a, b]

Main group	Subgroup	Component	Constituent groups
C—H	C	2,2-dimethylbutane	4 CH_3, 1 CH_2, 1 C
	CH	2-methylpropane	3 CH_3, 1 CH
	CH_2	pentane	2 CH_3, 3 CH_2
	CH_3		
	CH_4	methane	1 CH_4
CH=CH	C=C	2,3-dimethyl-2-butene	4 CH_3, 1 C=C
	CH=C	2-methyl-2-butene	3 CH_3, 1 CH=C
	CH_2=C	2-methyl-1-butene	2 CH_3, 1 CH_2, 1 CH_2=C
	CH=CH	2-pentene	2 CH_3, 1 CH_2, 1 CH=CH
	CH_2=CH	1-butene	1 CH_3, 1 CH_2, 1 CH_2=CH

Table 3. ***Continued***

Main group	Subgroup	Component	Constituent groups
	$CH_2{=}CH_2$	ethylene	1 $CH_2{=}CH_2$
ACH	AC	styrene	1 $CH_2{=}CH$, 5 ACH, 1 AC
	ACH	benzene	6 ACH
ACCH	ACCH	cumene	2 CH_3, 5 ACH, 1 ACCH
	$ACCH_2$	*n*-propylbenzene	1 CH_3, 1 CH_2, 5 ACH, 1 $ACCH_2$
	$ACCH_3$	toluene	5 ACH, 1 $ACCH_3$
OH	OH	2-butanol	2 CH_3, 1 CH_2, 1 CH, 1 OH
CH_3OH		methanol	1 CH_3OH
H_2O		water	1 H_2O
ACOH	ACOH	phenol	5 ACH, 1 ACOH
CH_nCO	CH_2CO	(ketone group as any group except second carbon)	
		3-hexanone	2 CH_3, 2 CH_2, 1 CH_2CO
	CH_3CO	(ketone group as second carbon)	
		2-pentanone	1 CH_3, 2 CH_2, 1 CH_3CO
CHO	CHO	acetaldehyde	1 CH_3, 1 CHO
CH_nCOO	CH_2COO	pentyl propanoate	2 CH_3, 4CH_2, 1 CH_2COO
	CH_3COO	butyl acetate	1 CH_3, 3CH_2, 1 CH_3COO
HCOO	HCOO	methyl formate	1 CH_3, 1 HCOO
CH_nO	FCH_2O	tetrahydrofuran	3 CH_2, 1 FCH_2O
	CHO	diisopropyl ether	4 CH_3, 1 CH, 1 CHO
	CH_2O	diethyl ether	2 CH_3, 1 CH_2, 1 CH_2O
	CH_3O	dimethyl ether	1 CH_3, 1 CH_3O
CH_nNH_2	$CHNH_2$	isobutylamine	2 CH_3, 1 CH_2, 1 CHNH
	CH_2NH_2	propylamine	1 CH_3, 1 CH_2, 1 CH_2NH_2
	CH_3NH_2	methylamine	1 CH_3NH_2
CH_nNH	CHNH	diisopropylamine	4 CH_3, 1 CH, 1 CHNH
	CH_2NH	diethylamine	2 CH_3, 1 CH_2, 1 CH_2NH
	CH_3NH	methylethylamine	1 CH_3, 1 CH_2, 1 CH_3NH
CH_nN	CH_2N	triethylamine	3 CH_3, 2 CH_2, 1 CH_2N
	CH_3N	trimethylamine	1 CH_3, 1 CH_3NH
$ACNH_2$	$ACHN_2$	aniline	5 ACH, 1 $ACNH_2$
C_5H_nN	C_5H_3N	2,3-dimethylpyridine	2 CH_3, 1 C_5H_3N
	C_5H_4N	3-methylpyridine	1 CH_3, 1 C_5H_4N
	C_5H_5N	pyridine	1 C_5H_5N
CH_nCN	CH_2CN	propionitrile	1 CH_3, 1 CH_2CN
	CH_3CN	acetonitrile	1 CH_3CN
CH_nOOH	HCOOH	formic acid	1 HCOOH
	COOH	acetic acid	1 CH_3, 1 COOH
CH_nCl	CCl	2-chloro-2-methylpropane	3 CH_3, 1 CCl
	CHCl	2-chlorobutane	2 CH_3, 1 CH_2, 1 CHCl
	CH_2Cl	1-chlorobutane	1 CH_3, 2 CH_2, 1 CH_2Cl
CH_nCl_2	CCl_2	2,2-dichloropropane	2 CH_3, 1 CCl_2
	$CHCl_2$	1,1-dichloropropane	1 CH_3, 1 CH_2, 1 $CHCl_2$
	CH_2Cl_2	dichloromethane	1 CH_2Cl_2
CH_nCl_3	CCl_3	1,1,1-trichloroethane	1 CH_3, 1 CCl_3
	$CHCl_3$	chloroform	1 $CHCl_3$
CCl_4		tetrachloromethane	1 CCl_4
ACCl		chlorobenzene	5 ACH, 1 ACCl
CH_nNO_2	$CHNO_2$	2-nitrobutane	2 CH_3, 1 CH_2, 1 $CHNO_2$
	CH_2NO_2	1-nitropropane	1 CH_3, 1 CH_2, 1 CH_2NO_2
	CH_3NO_2	nitromethane	1 CH_3NO_2
$ACNO_2$		nitrobenzene	5 ACH, 1 $ACNO_2$

Table 3. *Continued*

Main group	Subgroup	Component	Constituent groups
CS_2		carbon disulfide	1 CS_2
CH_nSH	CH_2SH	ethanethiol	1 CH_3, 1 CH_2SH
	CH_3SH	methanethiol	1 CH_3SH
furfural		furfural	1 furfural
$(CH_2OH)_2$		1,2-ethanediol	1 $(CH_2OH)_2$
I	I	1-iodoethane	1 CH_3, 1 CH_2, 1 I
Br	Br	1-bromoethane	1 CH_3, 1 CH_2, 1 Br
CH_nC	$C{\equiv}C$	2-pentyne	2 CH_3, 1 CH_2, 1 $C{\equiv}C$
	$CH{\equiv}C$	1-butyne	1 CH_3, 1 CH_2, 1 $CH{\equiv}C$
	$CH{\equiv}CH$	ethyne	1 $CH{\equiv}CH$
DMSO	$(CH_3)_2SO$	dimethylsulfoxide	1 $(CH_3)_2SO$
$CH_2{\equiv}CHCN$		acrylonitrile	1 $CH_2{=}CHCN$
$Cl(C{=}C)$	$Cl(C{=}C)$	tricholorethylene	1 $CH{=}C$, 3 $Cl(C{=}C)$
ACF	ACF	hexafluorobenzene	6 ACF
DMF	DMF-1	dimethylformamide	1 DMF-1
	DMF-2	diethylformamide	2 CH_3, 1 DMF-2
CF_n	CF	perfluoromethylcyclohexane	1 CH_3, 5 CH_2, 1 CF
	CF_2		
		perfluorohexane	2 CF_3, 4 CF_2
	CF_3		
COO	COO	dimethyl oxalate	2 CH_3, 2 COO
SiH_2	Si	hexamethyldisiloxane	6 CH_3, 1 SiO, 1 Si
	SiH	heptamethyltrisiloxane	7 CH_3, 2 SiO, 1 SiH
	SiH_2	diethylsilane	2 CH_3, 2 CH_2, 1 SiH_2
	SiH_3	methylsilane	1 CH_3, 1 SiH_3
SiO	SiO	octamethylcyclotetrasiloxane	8 CH3, 4 SiO
	SiHO	1,1,3,3-tetramethyldisiloxane	4 CH_3, 1 SiHO, 1 SiH
	SiH_2O	1,3-dimethyldisiloxane	2 CH_3, 1 SiH_2O, 1 SlH_2
NMP	NMP	*N*-methylpyrrolidine	1 NMP
CClF	(8 subgroups)	represents freons	
CON	(6 subgroups)	represents amides	
OCCOH	(2 subgroups)	represents glycol ethers	
CH_2S	(3 subgroups)	represents organic sulfides	
morpholine		morpholine	1 morpholine
thiophene	(3 subgroups)	represents thiophenes	

[a] Ref. 168.

[b] A represents an aromatic moiety.

UNIQUAC and UNIFAC embody a generalized form of Guggenheim's lattice solution model (28, 169) and aspects of Derr and Deal's solution-of-groups theory (32). The former is the basis for the equations governing molecular interactions in UNIFAC; the latter provides the method for assessing group contributions once the groups have been selected. A key assumption in UNIQUAC and UNIFAC is that perturbations in the solution lattice can be equated to excess Gibbs energy and therefore to the activity coefficient by use of appropriate correction factors. Several textbooks on the subject suggest these assumptions are reasonable at ordinary (<500 kPa (5 atm)) pressures and over moderate (300–425 K) temperature ranges for nonelectrolyte substances well below their critical points. A description of the lattice theory of solutions is available (118).

UNIQUAC generalizes Guggenheim's model to mixtures containing molecules of varying sizes by using Wilson's local-composition theory (165) which proposes that when a solution is viewed microscopically, the mixture is not homogeneous with regard to composition. Whereas this assumption may be of little consequence to some bulk fluid properties, it is salient to liquid-mixture modeling. In UNIQUAC, a molecule is divided into equal segments where each segment occupies a lattice intersection, or cell. Guggenheim's ratio of like and unlike neighbors (each neighbor is a molecule occupying one cell) is replaced with local area fractions (ratio of like and unlike segments) to correct for the differences in proximity effects of end vs interior groups. Because all segments in a given molecule are assumed equal (even though chemical differences exist in real molecules), energy parameters for the method represent averages over the individual molecular topologies. Statistical aspects of the lattice structure for UNIQUAC were left the same as Guggenheim's model (169).

Linking this molecular model to observed bulk fluid PVT-composition behavior requires a calculation of the number of possible configurations (microstructures) of a mixture. There is no exact method available to solve this combinatorial problem (28). ASOG assumes the athermal (no heat of mixing) Flory-Huggins equation for this purpose (118, 170, 171). UNIQUAC claims to have a formula that avoids this assumption, although some aspects of athermal mixing are still present in the model.

UNIQUAC is significant because it provides a means to estimate multicomponent interactions using no more than binary interaction experimental data, bond angles, and bond distances. There is an implicit assumption that the combinatorial portion of the model, ie, the size and shape effects, can be averaged over a molecule and that these can be directly related to molecular surface area and volume. This assumption can be found in many QSAR methods and probably makes a significant contribution to the generally low accuracy of many QSAR prediction techniques.

UNIFAC assumes that the size, shape, and intermolecular interactions can be calculated by stoichiometrically and linearly summing group contributions of these parameters for a given system. Size and shape are related to van der Waals group volume and surface areas (172) and are subsequently summed and averaged over each molecule type. Any stearic or relative positional effects are lost in this averaging, but it does allow for the segments to be different from one molecule to the next. This feature is not accommodated by ASOG's Flory-Huggins assumption (170, 171). UNIFAC's interaction parameter summation differs from its volume and surface area summations only in that it is formulated as a difference to a reference substance containing only molecules of the same group distribution as the substance from which the interaction parameter data is derived. In this way, activity coefficients become unity for pure substances. Although the theoretical validity of these linear summations may be challenged, this technique does provide a mathematically simple expression.

Most of the assumptions are based on idealized models, indicating the limitations of the mathematical methods employed and the quantity and type of experimental data available. For example, the details of the combinatorial entropy of a binary mixture may be well understood, but modeling requires, in large measure, uniformity so the statistical relationships can be determined. This uniformity is manifested in mixing rules and a minimum number of adjustable parameters so as to avoid problems related to the mathematics, eg, local minima and multiple solutions.

Updates for the UNIFAC method primarily expand the list of groups and group interactions and make small adjustments to some of the experimental constants in the equations (39, 173, 174). Modifications to mixing rules and the like are not common in the more recent literature, although correction factors and similar features have been proposed for most of the underlying theories. For example, an adjustment to the Flory-Huggins theory correlating molecular surface area-to-volume ratio to combinatorial entropy, making the theory more accurate for nonpolymeric substances has been proposed (175). A degree of mixing correction factor to Wilson's equation has also been provided (176, 177), but this correction is not included as a variable parameter in ASOG or UNIFAC, probably in part because of the mathematical and data acquisition difficulties it would create. A strategy for selection of this correction factor is also available (29). Presentations of the Wilson equation (118) often do not show this correction factor, which is commonly called the liquid's coordination number.

Both UNIFAC and ASOG are typically generalized to multicomponent systems in commercial software packages. An important feature of these methods is that only binary interaction information is used to generate multicomponent predictions.

4.2.1.2. Assessment of UNIFAC. UNIFAC is a method to predict the activity of binary liquid solutions in the absence of all data except structural information. Because state-of-the-art real fluid estimation methods are empirical or semi-empirical, the use of more data results in improved activity estimation. UNIFAC was developed for situations where experimental data are scarce, and its application should generally be that of last resort (3, 7, 162, 178). That is, UNIFAC is not a method for comparative testing of methods based on experimental data even though the method is sometimes listed in commercial computer simulator menus without indication of its limitations.

Numerous assessments of the reliability of UNIFAC for various applications can be found in the literature. Extrapolating a confidence level for some new problem is ill-advised because accuracy is estimated by comparing UNIFAC predictions to experimental data. In some cases, the data are the same as that used to generate the UNIFAC interaction parameters in the first place. Extrapolating a confidence level for a new problem requires an assumption that the nature of the new problem is similar to that of the UNIFAC test systems previously considered. With no more than structural information, such an assumption may not be valid.

A general review of the UNIFAC and ASOG (32) methods, through the 1985 updates for the American Institute of Chemical Engineers (AIChE) under the Design Institute for Physical Property Data (DIPPR Project 802), is available (179). For temperatures in the range 300–425 K and pressures below 500 kPa (5 atm), UNIFAC is recommended as is the ASOG method (163). Investigation of the availability of interaction parameters is suggested as is comparing errors for the two methods to make a selection for a given problem. A general approach to method selection for specific problems and a comparison of binary and ternary solubility predictions to experimental results for a wide range of functional groups are available (162). In making a method selection, it is important to confirm that the method version matches the selection procedure. This confirmation is not always easy to make, based on commercial software documentation.

UNIFAC's combinatorial term is based on the Stavermann-Guggenheim term (180, 181) which is intended to account for size and shape effects of groups. The original ASOG method uses a Flory-Huggins type of approach (170, 171). Both methods neglect temperature effects. The ASOG residual term includes temperature dependence, whereas UNIFAC's does not. However, no evidence that this resulted in improved predictions by ASOG has been found (179). Where the residual term is zero, ie, there are no molecular interactions, the Stavermann-Guggenheim term (UNIFAC) gives exaggerated predictions of nonideality (40). No preference was indicated for nonathermal mixtures. The expression for excess Gibbs energy in UNIFAC is not flexible enough to produce reliable predictions for strongly associating or solvating systems in dilute regions (179). Infinite dilution activity coefficients usually exceed 10 for such systems. Predictions for systems containing symmetrical molecules were found to be no better than those by Raoult's law. The use of a simple equation of state having zero interaction coefficients for such situations has been recommended (179). All methods tested, eg, NRTL and UNIQUAC, which use more data, were recommended over UNIFAC for problems where such data were available.

Modified UNIFAC parameters applicable to both vapor–liquid and liquid–liquid environmental problems have been developed (23, 24, 182, 183), extending the UNIFAC range for selected dilute, nonideal systems. Several authors suggest addition of corrector groups to account for steric features, eg, a nonaromatic cyclic group (34). Inclusion of numerous groups containing first members of homologous series (5) suggests the UNIFAC developers were also aware of this problem.

The use of UNIFAC for estimating activity coefficients in binary and multicomponent organic and organic–water systems is recommended for those systems composed of nonelectrolyte, nonpolymer substances for which only structural information is known. UNIFAC is not recommended for systems for which some reliable experimental data are available. The method, including revisions through 1987 (39), is available in commercial software packages such as AspenPlus (174).

4.2.2. Group Contribution Method Survey

Group contribution methods to estimate physical properties vary in reliability and usage. Almost all are restricted to organic compounds. Before a method is used in studies requiring more than rough estimation, the literature should be reviewed. Concise descriptions of many of these methods including comparisons to nongroup contribution methods are available (5, 7, 184).

4.2.3. Critical Properties

Several methods have been developed to estimate critical pressure, P_c; temperature, T_c; and volume, V_c. Many other properties can be estimated from these properties. Error propagation can be large for physical property estimations based on critical properties from group contribution methods. Thus sensitivity analyses are recommended. The Ambrose method (185) was found to be more accurate (186) than the Lyderson (187) method, although it is computationally more complex. The Joback and Reid method (188) is only slightly less accurate overall than the Ambrose method, and is more accurate for some specific substances. Other methods of lesser overall accuracy are also available (189, 190) (T_c, V_c), (191, 192) (T_c, P_c), (193) (T_c, and (194) (V_c). A second-order method using Benson-type groups (195, 196) and normal boiling point for predicting critical temperature and pressure of hydrocarbons has been developed (197). This was extended to include organic compounds containing oxygen, nitrogen, sulfur, and halogens (198).

Parachor is the name (199) of a temperature-independent parameter to be used in calculating physical properties. Parachor is a function of liquid density, vapor density, and surface tension, and can be estimated from structural information. Critical constants for about 100 organic substances have been correlated to a set of equations involving parachors and molar refraction (200).

4.2.3.1. Normal Boiling Point–Freezing Point. One method's deviation for boiling and freezing point data is reported to be less than 2 K for 80% of a 600 compound set and less than 5 K for an additional 18% (201). However, this method is limited to organic compounds having only one functional group and a short base group such as methyl, vinyl, and phenyl. In contrast, the Joback method (188) has much wider applicability and a reported average error of 12.9 K. The Joback method also includes correlating parameters for freezing point at atmospheric pressure. A 1.29% average absolute error for predictions of a diverse set of 1169 organic species based on a nonlinear group contribution model has been reported (202). The parachor method also estimates normal boiling point (200), and a group contribution based correction factor for critical temperature to estimate boiling point has been developed (187). Reviews of other methods are available (5, 19).

4.2.3.2. Liquid Molar Volume. An average absolute error of approximately 1% for a range of liquid molar volumes of organic solvents and oligomers between their melting and normal boiling points has been reported (203). In addition, a 2.7% average error for selected amorphous polymers was estimated. The Le Bas method (204) estimates liquid molar volume at the normal boiling point. An average error of 3.9% for 32 diverse chemical species has been reported for this method (7). For many species, the Tyn and Calus equation (205), which relates normal boiling point molar volume to V_c, results in more accurate estimations than Le Bas (204) when the Ambrose (185) or Joback (188) method is used to estimate V_c.

4.2.3.3. Vapor Pressure. Empirical vapor pressure equations have been generalized, based on structure through Kamlet-Taft solvatochromic parameters (206). This correlation falls under the linear solvation energy relationship class of correlations. However, it has been extended to derive vapor pressures from UNIFAC activity coefficient estimations by an imaginary inert methane matrix solvent. An indirect means of vapor pressure prediction has been proposed (207) using, for example, a modified set of UNIFAC groups fit to the Antoine equation constants (208). Halogenated aromatic hydrocarbon vapor pressures have been fit (209) to groups (210, 211). Perfluorinated saturated hydrocarbons have been fit (212) to a modified set of Lawson's groups (213), as have 48 substituted benzenes (214).

4.2.3.4. Heats of Vaporization and Fusion. A simple linear summation of most of the Lyderson groups (187) has been proposed for heat of vaporization at the normal boiling point and heat of fusion at atmospheric pressure for a wide variety of organic compounds (188). Average errors of 1.2 and 4.3% for group contribution-based estimations of heats of vaporization for selected *n*- and iso-alkanes, respectively, have been reported (215).

4.2.3.5. Heat Capacity. The multiple property estimation methods for constant pressure ideal-gas heat capacities cover a broad range of organic compounds (188, 216, 217). Joback's method (188) is the easiest to use; however, usage of all these methods has been recommended only over the range 280–1100 K (7). An accurate method for ideal-gas heat capacities (constant pressure), limited to hydrocarbons, has been presented (218) that involves a fit of seven variables, and includes steric, ring, branching, alkene, and even allene corrections.

Constant volume heat capacities for liquid organic compounds were estimated with a four parameter fit (219). A 1.3% average absolute error for 31 selected species was reported. A group contribution method for heat capacities of pure solids and liquids based on elemental composition has also been provided (159).

4.2.3.6. Surface Tension. The relationship between surface tension and liquid molar volume (220), and the group contribution methods for liquid molar volume can be utilized to estimate surface tension.

4.2.3.7. Viscosity. A corresponding states method that requires critical pressure, temperature, and dipole moment has been developed for low pressure gas viscosity (221). This method, which includes a group contribution parameter, is also applicable to gas mixtures. Whereas a group contribution method is not available for dipole moment, the influence this parameter has can be neglected for many species.

Pure, low temperature organic liquid viscosities can be estimated by a group contribution method (7) and a method combining aspects of group contribution and connectivity indexes theories (222). Caution is recommended in the use of these methods because the calculated absolute errors are as high as 100% for individual species in a 150-compound, 10-family test set (223). A new method based on a second-order fit of Benson-type groups with numerous steric correctors is suggested as an alternative. Lower errors are claimed for the same test set.

4.2.3.8. Thermal Conductivity. A method for determining the thermal conductivity of polyatomic organic gases based on critical properties and a series of correction factors based on functional groups has been developed (224). Organic liquid thermal conductivity has also been estimated utilizing a second-order fit of Benson-type groups (225). An average deviation of 5.9% was reported for 266 compounds.

4.2.3.9. Acentric Factor. A method restricted to alkanes has been proposed for estimating acentric factors (226). A 2.2% average absolute error for 59 species was reported.

4.2.3.10. Flammability Limits. Some 1358 compounds selected from the DIPPR Compilation File (Pennsylvania State University, 1991; Ref. 4) have been fit for upper and lower flammability limits (227). Average errors reported were 0.266% (volume) and 0.06% (volume) for upper and lower flammability limits, respectively. A detailed analysis by functional group classification is included that identifies classifications with high error for several methods.

4.2.3.11. Enthalpy of Formation. The methods of several workers (188, 216–218) include parameters for estimating enthalpies of formation.

4.2.3.12. Gibbs (Free) Energy. The methods of several workers (188, 218) include parameters for estimating Gibbs energy of formation. In addition, based on its thermodynamic relationship to the activity coefficient, excess Gibbs energy can be estimated through the UNIFAC and ASOG methods.

4.2.3.13. Entropy. Standard entropies can be estimated by available methods (216–218).

4.2.3.14. Activity Coefficients. Most activity coefficient property estimation methods are generally applicable only to pure substances. Methods for properties of multicomponent systems are more complex and parameter fits usually rely on less experimental data. The primary group contribution methods of activity coefficient estimation are ASOG and UNIFAC. Of the two, UNIFAC has been fit to more combinations of groups and therefore can be applied to a wider variety of compounds. Both methods are restricted to organic compounds and water.

4.2.3.15. Henry's Constant. UNIFAC interaction parameters have been refit (227–229) to estimate Henry's law constants for dilute organic-in-water systems. Because this method is based on UNIFAC, it can be used for organic species with multiple function groups. A work following similar lines to the method of Reference 230 was presented for octanol–water partitioning coefficient and water solubility (183). Another UNIFAC-based method applied to methane, nitrogen, and oxygen gas phases has been developed (230). A method not based on UNIFAC applies to hydrocarbons and compounds having single functional groups (231, 232). A bond contribution method for a diverse set of 59 organic compounds is also available (233).

4.2.3.16. Octanol–Water Partition Coefficient. The f-Fragment approach (234–236) has been reviewed (227) and another method similar to the UNIFAC refit for Henry's constant has been proposed. Improved accuracy for many species and the ability to correct for temperature effects have been claimed for the newer method.

4.2.3.17. Diffusion Coefficient. The method of Reference 237 has been recommended for many low pressure binary gases (238). Other methods use solvent and solute parachors to calculate diffusion coefficients of dissolved organic gases in liquid solvents (239, 240). Molar volume and viscosity are also required and may be estimated by the methods previously discussed. Caution should be exercised because errors are multiplicative by these methods.

4.2.3.18. Second Virial Coefficient. A group contribution method including polar and nonpolar contributions has been proposed for second virial coefficients (241). This method has been applied to both pure components and mixtures, the latter through prediction of cross-second virial coefficients.

4.2.3.19. Chemical Reactivity. Among other things, reactivity plays a vital role in determining fate of pollutants in the environment. A prototype computer system to estimate reactivity of systems based on structure, SPARC (SPARC Performs Automated Reasoning in Chemistry), is described (242). As of this writing, procedures for hydrolysis rate constants, ionization pK_a uv–vis light absorption spectra, and several physical properties have been developed for SPARC.

4.2.3.20. Toxicology and Other Biological System Studies. Quantitative structure–relationship (QSAR) studied in the field of toxicology are extensive. These relationships range from functional group correlations to conventional group contributions to solvation energy relationships to molecular connectivity indexes. The U.S. Fish and Wildlife Service has developed an expert system that generates linear solvation energy relationship (LSER) parameters based on contaminant chemical structure and uses them for predicting aquatic toxicity (243). Overviews of and specific QSAR uses in environmental toxicology were presented in 1983 (244–247). An overview of QSAR studies related to drugs acting on the central system has also been compiled (248). A historical review of research into the relationship of biological activity and aqueous solubility to partition coefficient as well as discussions of related QSAR-based models is available (249) as is a strategy for ranking chemical hazards based on QSAR techniques (250). Specific property studies on acute fish toxicity relationships can be found in the literature (251, 252). Other studies propose QSARs for specific chemical classes, such as the organochlorine and synthetic pyrethroid insecticide study (253).

4.3. Linear Free Energy–Linear Solvation Energy Relationships

Linear free energy (LFER) and linear solvation energy (LSER) relationships are used to develop correlations between selected properties of similar compounds. These are fundamentally a collection of techniques whereby properties can be predicted from other properties for which linear dependency has been observed. Linear relationships include not only simple $y = mx + b$ relationships, but also more complicated expressions such as the Hammett equation (254) which correlates equilibrium constants for benzenes,

$$\log(K_x/K_H) = \rho\sigma_x$$

where K_x and K_H are equilibrium constants for substituted and unsubstituted benzenes, respectively, ρ is a reaction parameter, and ς_x is a substituent parameter.

The literature is not particularly clear on the distinction between LFERs and LSERs. In some instances, the terms are used interchangeably. For the present, consider LFERs as those relationships for which properties are linearly related to free energy. These include equilibrium constants, reaction rates, various spectral properties, polarographic half-wave lengths, dipole moments, and redox behavior (255). These have been used extensively to investigate the relationship between structure and chemical, physical, and biological properties of organic compounds, in particular. LSERs can be thought of as that subset of LFERs concerned with properties relating to solubility and partitioning. An active area of LSER research involves relating solubility-related parameters, eg, 1-octanol–water partitioning coefficient, K_{ow}, to the evaluation of the hazards of environmental contaminants (see Eenvironmental impact).

The distinction between LFERs and group contribution methods is also confusing. Many group contribution methods can be considered to be simply extensions of LFERs for which the independent parameters are structural. From this perspective, both group contribution techniques and LSERs are subsets of LFERs. Another viewpoint holds that the differences are largely semantical, owing to the historical fact that many group contribution techniques originate in the works of thermodynamicists and engineers, whereas many LFER–LSER methods are first found in the writings of chemists, biologists, and environmental scientists.

An excellent historical review of LFERs is available (255). An example of an LFER study, relating partition coefficients to aqueous solubility of organic liquids, may be found in Reference 256.

4.3.1. Kamlet-Taft Linear Solvation Energy Relationships

Most recent works on LSERs are based on a powerful predictive model, known as the Kamlet-Taft model (257), which has provided a framework for numerous studies into specific molecular thermodynamic properties of solvent–solute systems. This model is based on an equation having three conceptually explicit terms (258).

$$Prop = cavity\ term + dipolar\ term + hydrogen-bonding\ terms \qquad (105)$$

where *Prop* is the physical property of interest, the *cavity term* describes the energy associated with creation and stabilization of a hole of correct size and shape in a continuous solvent phase large enough to accept a solute molecule, and the remaining terms relate electrostatic and hydrogen-bonding interactions between the molecules to stabilize the solute molecule in the hole. Such an equation is readily transformed into a form amenable to a quantitative structure–activity approach such as,

$$Prop = Prop_0 + mV_i/100 + s\pi^* + b\beta_m + a\alpha_m \qquad (106)$$

where *m, s, b*, and *a* are constants for a particular property and set of conditions. The normally endoergic cavity term, $mV_i/100$, contains $V_i/100$,, the van der Waals molecular volume which is scaled to units consistent with the other terms. The dipolar term, $s\pi^*$, is a measure of the largely exoergic solute–solvent dipole–dipole and dipole-induced dipole effects interactions, and π^* is a measure of the ability of nonspecific dielectric effects to stabilize neighboring charges and dipoles. The remaining terms incorporate the exoergic effects of the formation of hydrogen bonds. The β_m parameter is used for conditions in which the solvent is a hydrogen bond donor acid (HBD) and the solute is a hydrogen bond acceptor base (HBA), whereas α_m involves an HBA solvent and an HBD solute. The subscript *m* indicates that only the monomeric solute is considered, irrespective of the fact that it may self-associate. These last terms form the framework for expanded definitions of acids and bases which include nonelectrolyte systems. This approach has been used to examine solvent effects on equilibria and reaction rates, the influence of solute effects by water, blood, etc, solubilities, phase distribution, and carbon adsorption (259).

Owing to the original determination from uv–vis spectral solvatochromic shifts, π^*, β_m, and α_m are called solvatochromic parameters. General rules for estimation of these variables have been proposed (258). Examples of individual parameter investigations are available (260, 261). As previously mentioned, individual LFER–LSER studies are performed on related materials. A common method to link these individual studies to group contribution methods, and thereby expand the applicability, is by expansion of solvatochromic parameters to log–linear relationships, such as

$$\pi_x = \log_{10} P_x - \log_{10} P_H$$

where P_x is a coefficient related to a derivative compound, P_H is related to its parent molecule, and π_x is the resultant solvatochromic parameter or solvation property. As defined (262), the subscript x refers to the π value for a substitute group. This so-called parent–derivative method has been used for estimation of octanol–water partition coefficients for sets of similar compounds.

A sampling of applications of Kamlet-Taft LSERs include the following. (*1*) The Solvatochromic Parameters for Activity Coefficient Estimation (SPACE) method for infinite dilution activity coefficients where improved predictions over UNIFAC for a database of 1879 critically evaluated experimental data points has been claimed (263). (*2*) Observation of inverse linear relationship between log 1-octanol–water partition coefficient and liquid (or subcooled liquid) solubility (264). (*3*) Correlation of fish bioconcentration factors with solvatochromic parameters (265). (*4*) The distribution of organic solutes between water and immiscible organic solvents (266). (*5*) Carbon adsorption of organic compounds (267). A similar, alternative approach is one based on solvophobic theory (244). (*7*) Correlation of octanol–water partition coefficients of organic nonelectrolytes (including strong HBD solutes) (268).

4.4. Molecular Connectivity Indexes and Graph Theory

Perhaps the chief obstacle to developing a general theory for quantification of physical properties is not so much in the understanding of the underlying physical laws, but rather the inability to solve the requisite equations. The plethora of assumptions and simplifications in the statistical mechanics and group contribution sections of this article provide examples of this. Computational procedures are simplified when the number of parameters used to describe the salient features of a problem is reduced. Because many properties of molecules correlate well with structures, parameters have been developed which grossly quantify molecular structural characteristics. These parameters, or connectivity indexes, are usually based on the numbers and orientations of atoms and bonds in the molecule.

A forerunner to recent connectivity indexes is found in work on the properties of saturated hydrocarbons (269, 270). The Wiener index, ω, can be calculated by taking the product of the number of carbon atoms on the two sides of each C—C bond and then summing these products for all C—C bonds in the molecule (271). For example, 2-methylpentane, 2,2-dimethyl butane, and 2,3-dimethyl-butane have Wiener indexes of 32, 28, and 26, respectively. It can be readily seen that ω combines both molecular bulk and branching in a single parameter. As is to be expected, a single parameter does not generally provide sufficient information to do accurate property estimation. By adding a second index, p, which is equal to the number of pairs of carbon atoms separated by three bonds and normalizing ω by the square of the number of carbon atoms in the molecule (N^2), a general formula for estimating several properties was generated (269, 270):

$$P_i = a\left(\omega/N^2\right) + bp + c \qquad (107)$$

where a, b, and c are constants appropriate to property P_i.

A series of connectivity indexes based on this same type of information, but having mathematical form more closely aligned with potential functions, has also been developed (272–274). This approach allows for a series of indexes to be calculated based on assemblages of increasing numbers of contiguous bonds and has been successfully applied to a variety of biological and physical properties (275), including: boiling point (274, 276), density (274), heats of formation (273), molar refractive index (275), partition coefficients (277) and polarity of solvents (278), aqueous solubility (45), surface tension (279), and critical volume and acentric factor (280).

As computing capability has improved, the need for automated methods of determining connectivity indexes, as well as group compositions and other structural parameters, for existing databases of chemical species has increased in importance. New naming techniques, such as SMILES, have been proposed which can be easily translated to these indexes and parameters by computer algorithms. Discussions of the more recent work in this area are available (281, 282). SMILES has been used to input Contaminant structures into an expert system for aquatic toxicity prediction by generating LSER parameter values (243, 258).

Perhaps a more flexible approach to incorporating structural information into a series of rule-based parameters is found in the incorporation of graph theoryinto chemical representation studies. Graph theory, introduced in 1736, is based on ideas from topology and combinatorics (271). Introductions to this branch of mathematics are available (283, 284). The enhanced flexibility of graph theory stems from the fact that it retains molecular structural information in a matrix format which is essentially a two- (or more) dimensional representation of a molecule's spatial orientation. Through the use of large computer capacity and sophisticated regression techniques, parameters of optimum correlation to the property of interest may be obtained. An excellent introduction to the application of graph theory to structure–property relationships is available (285). An example of pertinent calculations related to alkanes may be found in Reference 286.

Whereas graph theory and other computer-intensive approaches can provide insights into the area of physical property correlation previously unobtainable, the general guideline should be to use no more parameters than are necessary to achieve the desired level of accuracy in the results. Many so-called ad hoc descriptors, which are usually based on few fitting parameters, have been successfully used (287–291). For example, it has been suggested (271) that two indexes, the number of carbon atoms and the number of terminal methyl groups, can be combined with three fitting parameters in a linear equation for alkane normal boiling points. A correlation coefficient of 0.987 for 39 species has been reported. LSER (243, 258) uses only four variables, and a routine correlation coefficient range of 0.93–0.98 has been reported for numerous investigations. Concise introductions to the basic concepts of predicting chemical and physical properties from molecular topology may be found in the literature (271, 292).

BIBLIOGRAPHY

"Data Correlation" in *ECT* 1st ed., Vol. 4, pp. 846–873, by D. F. Othmer and E. S. Roszkowski, Polytechnic Institute of Brooklyn; "Data—Interpretation and Correlation Correlation" in *ECT* 2nd ed., Vol. 6, pp. 705–755, by D. F. Othmer, Polytechnic Institute of Brooklyn; "Engineering and Chemical Data Correlation" in *ECT* 3rd ed., Vol. 9, pp. 45–117, by D. F. Othmer, Polytechnic Institute of New York.

Cited Publications

1. D. M. Himmelblau, *Basic Principles and Calculations in Chemical Engineering*, 5th ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1989.
2. M. Modell and R. C. Reid, *Thermodynamics and Its Applications*, 5th ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1974.
3. J. M. Smith and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 4th ed., McGraw-Hill Book Co., Inc., New York, 1987.
4. *Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (DIPPR), AIChE, 1983 (updated in 1986).
5. W. J. Lyman, W. F. Reehl, and D. H. Rosenblatt, *Handbook of Chemical Property Estimation Methods*, McGraw-Hill Book Co., Inc., New York, 1982.
6. *EPA Superfund Public Health Evaluation Manual*, EPS/540/1-86/060, U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, Washington, D.C., October 1986.

7. R. C. Reid, J. M. Prausnitz, and B. E. Poling, *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill Book Co., Inc., New York, 1987.
8. R. H. Perry and C. H. Chilton, *Chemical Engineer's Handbook*, 5th Ed., McGraw-Hill Book Co., Inc., New York, 1973.
9. D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 71st ed. CRC Press, Boca Raton, 1990.
10. S. M. Walas, *Phase Equilibria in Chemical Engineering*, Butterworth Publishers, Boston, 1985.
11. L. Berenyi, *Z. Phys. Chem.* **94**, 628 (1920).
12. L. Berenyi, *Z. Phys. Chem.* **105**, 55 (1923).
13. M. M. Dubinin, *Chem. Rev.* **60**, 235–241 (1960).
14. M. M. Dubinin and L. V. Radushkevich, *Dokl. Akad. Nauk. SSSR* **55**, 327, (1947).
15. M. Polanyi, *Verh. Dtsch. Phys. Ges.* **16**, 1012 (1914).
16. M. Polanyi, *Z. Elektrochem.* **26**, 371 (1920).
17. A. L. Myers and S. Sircar, in M. J. McGuire and I. H. Suffet, eds., *Treatment of Water by Granular Activated Carbon*, American Chemical Society, Washington, D.C., 1983, 63–76.
18. J. M. Prausnitz, *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.
19. R. C. Reid and T. K. Sherwood, *The Properties of Gases and Liquids: Their Estimation and Correlation*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1966.
20. K. S. Pitzer and R. F. Curl, *J. Amer. Chem. Soc.*, 79 (May 20, 1957).
21. P. W. Atkins, *Physical Chemistry*, W. H. Freeman and Co., San Francisco, 1978, 204–208.
22. C. L. Yaws, *Thermodynamic and Physical Property Data*, Gulf Publishing Co., Houston, Tex., 1992.
23. W. B. Arbuckle, *Env Sci Tech* **17**(9), 537–542 (1983).
24. W. B. Arbuckle, *Env Sci Tech.* **20**(10), 1060–1064 (1986).
25. G. M. Wilson and C. H. Deal, *Ind. Eng. Chem. Fund.* **1**(1), 20–23 (1962).
26. J. H. Hildebrand, J. M. Prausnitz, and R. L. Scott, *Regular and Related Solutions*, Van Nostrand Reinhold, Co., Inc., New York, 1970.
27. F. A. Mato, R. B. Mato, and F. Mato, *Ind. Eng. Chem. Res.* **28**(9), 1441–1446 (1989).
28. D. S. Abrams and J. M. Prausnitz, *AIChE J.* **21**, 116 (1975).
29. M. Hiranuma, *J. Chem. Eng. Japan* **8**(2), 162–163 (1975).
30. I. Nagata, M. Nagashima, and M. Ogura, *J. Chem. Eng. Japan* **8**(5), 406–408 (1975).
31. T. Tsuboka and T. Katayama, *J. Chem. Eng. Japan* **8**(3), 181–187 (1975).
32. E. L. Derr and C. H. Deal, *I. Chem. E. Symposium Series* **32**, 3:40–3:51 (1969).
33. J. H. Park and P. W. Carr, *Anal. Chem.* **59**(21), 2596–2602 (1987).
34. E. R. Thomas and C. A. Eckert, *Ind. Eng. Chem. Process Des. Dev.* **23**(2), 194–209 (1984).
35. Aa. Fredenslund, R. L. Jones, and J. M. Prausnitz, *AIChE. J.* **21**, 1086–1099 (1975).
36. S. Skjold-Jorgensen, B. Kolbe, J. Gmehling, and Rasmussen, *Ind. Eng. Chem. Process Des. Dev.* **18**(4) 714–722 (1979).
37. J. Gmehling, P. Rasmussen, and Aa. Fredenslund, *Ind. Eng. Chem., Process Des. Dev.* **21**(1), 118–127 (1982).
38. E. A. Macedo, U. Weidlich, J. Gmehling, and P. Rasmussen, *Ind. Eng. Chem. Process Des. Dev.* **22**(4), 676–678 (1983).
39. D. Tiegs and co-workers, *Ind. Eng. Chem. Res.* **26**(1), 159–161 (1987).
40. I. Kikic, P. Alessi, P. Rasmussen, and Aa. Fredenslund, *Can. J. Chem. Eng.* **58**, 253–258 (Apr. 1980).
41. O. Sinanoglu in B. Pullman, ed., *Molecular Associations in Biology*, Academic Press, New York, 1968, p. 427.
42. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **22**(11), 1349–1357 (1988).
43. G. Belfort, G. L. Altshuler, K. K. Thallam, C. P. Feerick, Jr., and K. L. Woodfield, *AIChE. J.*, **30**(2), 197 (Mar. 1984).
44. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **24**(4), 575 (1990).
45. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.*, **22**(3), 328–338 (1988).
46. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **3**, 708–713 (1989).
47. J. J. Carroll, *Chem. Eng. Prog.* (Sept. 1991).
48. R. A. Ashworth, G. B. Howe, M. E. Mullins, and T. N. Rogers, *J. Hazd. Matls.* **18**, 25–36 (1988).
49. G. J. Pierotti, C. H. Deal, and E. L. Derr, *Ind. Eng. Chem.* **51**(1), 95–102 (1959).
50. J. M. Gossett and A. H. Lincoff, *Solute Gas Equilibriums in Multi-organic Aqueous Systems*, report 1981, AFOSR-TR-81-0858; order no. AD-A109082, NTIS, 1981.
51. J. M. Gossett, *Env. Sci. Tech.* **21**(2), 202–208 (1987).
52. C. Black, G. G. Joris, and H. S. Taylor, *J. Chem. Phys.* **16**(5), 537–543 (1948).

53. R. Stephenson, J. Stuart, and M. Tabak, *J. Chem. Eng. Data* **29**, 287–290 (1984).
54. R. Stephenson and J. Stuart, *J. Chem. Eng. Data* **31**, 56–70 (1986).
55. K. Kinoshita and H. Ishikawa, *Bull. Chem. Soc. Japan* **31**(9), 1081–1082 (1958).
56. W. Y. Shiu and D. Mackay, *J. Phys. Chem. Ref. Data* **15**(2), 911–929.
57. S. Banerjee, S. H. Yalkowsky, and S. C. Valvani, *Env. Sci. Tech.* **14**(10), 1227–1229 (1980).
58. M. M. Miller, S. Ghodbane, S. P. Wasik, Y. B. Tewari, and D. E. Martire, *J. Chem. Eng. Data* **29**, 184 (1984).
59. V. N. Kabadi and R. P. Danner, *Hydrocarbon Proc.* **58**(5), 245–246 (1979).
60. C. L. Yaws and X. Pan, *Oil & Gas J.* 79–82, (Apr. 8, 1991).
61. B. I. Lee and M. G. Kesler, *AIChE. J.* **21**, 3 (May 1975).
62. J. R. Brock and R. B. Bird, *AIChE. J.* **1**, 2 (June 1955).
63. B. F. Dodge, *Inc. Eng. Chem.* **24**, 12 (Dec. 1932).
64. W. K. Lewis, *Ind. Eng. Chem.*, **28**, 257–262 (Feb. 1936).
65. W. K. Lewis and C. D. Luke, *Trans. Am. Soc. Mech. Eng.* **PME-54**, 8 (1932).
66. W. K. Lewis and C. D. Luke, *Ind. Eng. Chem.* **25**, 7 (July 1933).
67. F. D. Maslan and T. M. Littman, *Ind. Eng. Chem.* **45**, 7 (July 1953).
68. H. P. Meissner and R. Seferian, *Chem. Eng. Prog.* **47**, 11 (Nov. 1951).
69. L. C. Nelson and E. F. Obert, *Trans. Am. Soc. Mech. Eng.* 76 (Oct. 1954).
70. L. Nelson and E. F. Obert, *AIChE. J.* **1**, 1 (Mar. 1955).
71. L. Obert, *Ind. Eng. Chem.* **40**, 11 (Nov. 1948).
72. G. J. Su and C. H. Chang, *Ind. Eng. Chem.* 38 (1946).
73. Meslin, *Comptes Rendus cxvi* (1893).
74. K. S. Pitzer, *J. Chem. Phys.* 7 (Aug.1939).
75. K. S. Pitzer, *J. Amer. Chem. Soc.* **77**, 13 (July 16, 1955).
76. T. W. Leland, Jr. and P. S. Chappelear, *Ind. Engr. Chem.* **60**, 7 (July 1968).
77. O. Redlich and J. N. Kwong, *Chem. Rev.* **44**, 233–244 (1949).
78. H. G. Rackett, *J. Chem. Eng. Data* **15**, 4 (1970).
79. K. S. Pitzer, *J. Amer. Chem. Soc.* **77**, 13 (July 16, 1955).
80. R. L. Halm and L. Stiel, *AIChE. J.*, **3**, 2 (Mar. 1967).
81. G. K. Stipp, S. D. Bai, and L. I. Stiel, *AIChE. J.*, **19**, 6 (Nov. 1973).
82. B. I. Lee and M. G. Kesler, *AIChE. J.* **21**, 3 (May 1975).
83. R. H. Newton, *Ind. Eng. Chem.* **27**, 3 (Mar. 1935).
84. C. W. Selheimer and co-workers, *Ind. Eng. Chem.* **24**, 5 (May 1932).
85. L. Raidel, *Chemie Ingenieur Technik*, (in German), VCH, Weinheim, 1956, p. 28.
86. R. Danner and T. Daubert, *DIPPR 802 Data Prediction Manual, AIChE*. University Park, Pa., 1986, pp. 7A1–7B2.
87. J. R. Brock, and R. B. Bird, *AIChE. J.* **1**, 174 (1955).
88. D. I. Hakim, D. Steinberg, and L. I. Stiel, *Ind. Eng. Chem. Fundam.* **10**(1), 174–175 (1971).
89. M. D. Croucher and M. L. Hair, *J. Phys. Chem.* **81**, 17 (1977).
90. K. Lucas, *Phase Equilibrium and Fluid Properties in the Chemical Industry*, Dechema, Frankfurt, 1980.
91. L. I. Stiel and G. Thodos, *Progress in International Research on Thermodynamics and Transport Properties*, American Society of Mechanical Engineers, Academic Press, Inc., New York, 1962.
92. C. Tsonopoulos, *AIChE. J.* **20**, 2 (Mar. 1974).
93. S. Chidambaram and G. Narismhan, *Chem. Eng.* 135, (Nov. 23, 1964).
94. R. D. Gunn, T. Yamada, and D. Whitman, *AIChE. J.* **20**(5), 906 (Sept. 1974).
95. H. C. Chen and S. H. Chen, *Ind. Eng. Chem. Fund.* **24**(2), 183 (1985).
96. W. C. Rheinbolt, *Numerical analysis of Parameterized Nonlinear Equations*, Wiley-Interscience, New York, 1986.
97. J. H. Wegenstien, *Commun. ACM.* 1 (1958).
98. O. Orbach and C. M. Crowe, *Can. J. Chem. Eng.* **68**(7–8), 1033–1042 (1971).
99. D. Marquard, *SIAM J. App. Math*, 11 (1963).
100. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Non-Linear Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
101. M. J. D. Powell, *Numerical Methods for Non-Linear Algebraic Equations*, Gordon and Breach, New York, 1970.
102. W. J. Lin, J. D. Seader, and T. L. Wayburn, *AIChE. J.* **33**(6), 886–897 (1987).

103. D. S. Davis, *Empirical Equations and Nomography*, McGraw-Hill Book Co., Inc., New York, 1943, 91–160.
104. M. Kraitchik, *Alignment Charts Construction and Use*, D. Van Nostrand Co., Inc., New York, 1944.
105. J. Lipka, *Graphical and Mechanical Computation, Part I. Alignment Charts*, John Wiley & Sons, Inc., New York, 1921.
106. W. C. Marshall, *Graphical Methods for Schools, Colleges, Statisticians, Engineers and Executives*, McGraw-Hill Book Co., Inc., New York, 1921, 167–185.
107. J. B. Peddle, *The Construction of Graphical Charts*, McGraw-Hill, London, 1910.
108. A. S. Levins, *Nomography*, 2nd ed., Fearon Publishers, Belmont, Calif., 1971.
109. J. Angel, *Chem. Eng. Progress* **86**, 85–87 (1990).
110. M. E. Davis, *Numerical Methods for Chemical Engineers*, John Wiley & Sons, Inc., New York, 1984.
111. E. Aarts, and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, Inc., New York, 1989.
112. J. A. Nelder and R. Mead, *Comput. J.* **7**, 308 (1965).
113. J. A. Solodar, *Chemtech.* **16**, 372–377 (1986).
114. P. K. Mookerjee, *Chemtech.* **15**, 606–610 (1985).
115. R. J. Amarger, L. T. Biegler, and I. E. Grossman, *Computers Chem. Eng.* **16**, 623–636 (1992).
116. J. J. Ferrada and J. M. Holmes, *Chem. Eng. Progress* **86**, 34–41 (1990).
117. J. A. Fay, *Molecular Thermodynamics*, Addison-Wesley, Reading, Mass. 1965, Chapt. 1, p. 83.
118. J. M. Prausnitz, R. N. Lichtenthaler, and E. G. de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1986, Chapts. 4, 5, 7; App. III, pp. xv, 6–7, 306–327, 360–364, 509–520.
119. L. Boltzmann, *Lectures on Gas Theory*, English translation University of California Press, (1964); German original published by J. A. Barth, Leipzig, 1896, Part I; 1898, Part II.
120. L. Loeb, *Kinetic Theory of Gases*, 2nd ed., Dover, 1934.
121. J. O. Hirschfelder, C. F. Curtis, and R. B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York, 1954.
122. F. W. Sears and G. L. Salinger, *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*, 3rd ed., Addison-Wesley, Reading, Mass. 1975, Chapts. 9 and 11; 320–322.
123. I. Müller, *Thermodynamics*, Pittman Advanced Publishing Program, Boston, 1985, Chapt. 5.
124. J. D. Lewins, ed., *Teaching Thermodynamics*, Plenum Press, New York, 1985.
125. J. W. Whalen, *Molecular Thermodynamics: A Statistical Approach*, John Wiley & Sons, Inc., New York, 1991.
126. E. Schrödinger, *Statistical Thermodynamics*, 2nd ed., Cambridge University Press, Cambridge, UK, 1964.
127. F. Daniels and R. A. Alberty, *Physical Chemistry*, 4th ed., John Wiley & Sons, Inc., New York, 1975, 396–422.
128. W. Pauli, Jr., *Z. Physik* **31**, 765–783 (1925).
129. A. Einstein, *Akademie der Wissenschaften*, Sitzungsberichte, Berlin, 1924, p. 261; 1925, p. 3.
130. E. Fermi, *Z. Physik* **36**, 902 (1926).
131. P. A. M. Dirac, *Proceedings of the Royal Society of London*, **A112**, 661 (1926).
132. C. Kittel, *Thermal Physics*, John Wiley & Sons, Inc., New York, 1969, Chapt. 9.
133. D. A. McQuarrie, *Statistical Thermodynamics*, Harper & Row, New York, 1973, 164–169.
134. W. Heisenberg, *Z. Physik* **33**, 879–893 (1925).
135. M. Born and P. Jordan, *Z. Physik* **34**, 858–858 (1925).
136. W. Pauli, Jr., *Z. Physik* **36**, 336–363 (1926).
137. E. L. Knuth, *Introduction to Statistical Thermodynamics*, McGraw-Hill, New York, 1966, Chapt. 2.
138. J. Kestin, and J. R. Dorfman, *A Course in Statistical Thermodynamics*, Academic Press, New York, 1971, Chapts. 7–11, 14.
139. F. C. Andrews, *Equilibrium Statistical Mechanics*, John Wiley & Sons, Inc., New York, 1975.
140. N. O. Smith, *Elementary Statistical Thermodynamics—A Problems Approach*, Plenum Press, New York, 1982.
141. R. J. Finfelstein, *Thermodynamics and Statistical Physics—A Short Introduction*, W. H. Freeman, San Francisco, 1969.
142. B. H. Lavenda, *Nonequilibrium Statistical Thermodynamics*, John Wiley & Sons, Inc., Chichester, England, 1985.
143. H. J. Kruezer, *Nonequilibrium Thermodynamics and its Statistical Foundations*, Clarendon Press, Oxford, 1981.
144. H. Ziegler, *An Introduction to Thermomechanics*, 2nd ed. North-Holland Publishing, Amsterdam, distributed in the United States by Elsevier Science Publishing, New York, 1983.
145. E. A. Mason and T. H. Spurling, in E. A. Mason and T. H. Spurling, *The International Encyclopedia of Physical Chemistry and Chemical Physics*, Vol. **2**, topic 10, Pergammon Press, New York, 1969.

146. G. Herzberg, *Atomic Spectra and Atomic Structure*, Prentice Hall, Englewood Cliffs, N.J., 1937, reprinted by Dover, New York, 1944.
147. G. Herzberg, *Molecular Spectra and Molecular Structure, I. Spectra of Diatomic Molecules*, 2nd ed., Van Nostrand, Princeton, N.J., 1950.
148. G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, Princeton, N.J., 1945.
149. E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer*, Brooks/Cole Publishing, Belmont, Calif., 1966.
150. D. Q. Kern, *Process Heat Transfer*, McGraw-Hill, New York, 1950.
151. C. Kittel, *Solid State Physics*, 3rd ed., John Wiley & Sons, Inc., New York, 1967, Chapt. 14.
152. G. H. Wannier, *Statistical Physics*, John Wiley & Sons, Inc., New York, 1966, Chapt. 15.
153. A. Einstein, *Investigations of the Theory of the Brownian Movement*, Dover, New York, 1956.
154. P. C. Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, New York, 1977, 106–115.
155. F. Penotti, J. Gerratt, D. L. Cooper, and M. Raimondi, *THEOCHEM* **46**, 421–436 (1988).
156. D. E. Williams, *J. Comput. Chem.* **9**, 745–763 (1988).
157. J. Niwa, *Bull. Chem. Soc. Jpn.* **62**, 226–233 (1989).
158. N. Koga and K. Morokuma, *J. Phys. Chem.* **94**, 5454–5462 (1990).
159. J. E. Hurst, Jr. and B. K. Harrison, *Chem. Eng. Comm.* **112**, 21 (1992).
160. R. M. Felder and R. W. Rousseau, *Elementary Principles of Chemical Processes*, 2nd ed., John Wiley & Sons, Inc., New York, 1986, p. 352.
161. D. M. Himmelblau, *Basic Principles and Calculations in Chemical Engineering*, 3rd ed., Prentice-Hall, Englewood Cliffs, N.J., 1974, p. 270.
162. R. P. Danner, and T. E. Daubert, *Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (AIChE), 1986, Chapt. 6.
163. K. Kojima and K. Tochigi, *Prediction of Vapor-Liquid Equilibria by the ASOG Method*, Elsevier and Kodansha Ltd., New York, 1979.
164. R. R. Dreisbach, *Pressure–Volume–Temperature Relationships of Organic Compounds*, Handbook Publishers, Sandusky, Ohio, 1952, pp. 294.
165. G. M. Wilson, *J. Am. Chem. Soc.* **86**, 127 (1964).
166. H. Renon and J. M. Prausnitz, *AIChE. J.* **14**, 135 (1968).
167. A. Fredenslund, J. Gmehling, and P. Rasmussen, *Vapor–Liquid Using UNIFAC, a Group Contribution Method*, Elsevier Scientific Publishing, Amsterdam, 1977.
168. H. K. Hansen, P. Rasmussen, and A. Fredenslund, *Ind. Eng. Chem. Res.* **30**, 2355–2358 (1991).
169. E. A. Guggenheim, *Mixtures*, Oxford University Press, 1952.
170. M. L. Huggins, *J. Phys. Chem.* **9**, 440 (1941); *Ann. N.Y. Acad. Sci.* **43**, 1 (1942).
171. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, N.Y., 1953.
172. A. Bondi, *Physical Properties of Molecular Liquids, Crystals and Glasses*, John Wiley & Sons, Inc., New York, 1968.
173. T. Magnussen, P. Rasmussen, and A. Fredenslund, *Ind. Eng. Process Dev.* **20**, 331–339 (1980).
174. Aspen Technology, Inc., *AspenPlus Users Guide*, Cambridge, Mass., 1988, 5–46, appendix pp. C25–C34.
175. R. N. Lichtenthaler, D. S. Abrams and J. M. Prausnitz, *Can. J. Chem.* **51**, 3071 (1973).
176. G. Scatchard and G. M. Wilson, *J. Am. Chem. Soc.* **86**, 133 (1964).
177. H. Renon and J. M. Prausnitz, *AIChE. J.* **15**, 785 (1969).
178. R. F. Wilcox, *Chem. Eng.* 66 (Oct 17, 1983).
179. P. A. Gupte, M. Nagvekar, R. P. Danner, and T. E. Daubert, *Documentation of the Basis for Selection of the Contents of Chapter 6 Phase Equilibrium in Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (AIChE), (1987).
180. E. A. Guggenheim, *Applications of Statistical Mechanics*, Oxford University Press, London, 1966.
181. A. J. Staverman, *Recl. Trav. Chim. Pays-Bas* **69**, 163 (1950).
182. T. N. Rogers and G. Howe, *A UNIFAC Model for Predicting Henry's Law Constants*, Research Triangle Institute, unpublished report, 1988.
183. G. Wienke and J. Gmehling, *Prediction of Properties of Environmental Interest Using UNIFAC, Documentation from AIChE Annual Meeting presentation*, Miami Beach, Fla., Nov. 1–6, 1992 (available from author at Universität Oldenburg, Technische Chemie, Postfach 25 03, D-2900 Oldenburg, Germany).

184. C. Jochum, M. G. Hicks, and J. Sunkel, eds., *Physical Property Prediction in Organic Chemistry, Proceedings of the Beilstein Workshop, May 16–20, 1988, Schloss Korb, Italy*, Springer-Verlag, Berlin and New York, 1988.
185. D. Ambrose, *Correlation and Estimation of Vapor-Liquid Critical Properties. I. Critical Temperatures of Organic Compounds*, National Physical Laboratory, Teddington, UK, NPL Report 92 (1978, corrected 1980).
186. K. M. Klincewicz, and R. C. Reid, *AIChE J.* **30**, 137 (1984).
187. A. L. Lyderson, *Estimation of Critical Properties of Organic Compounds*, University of Wisconsin, Eng. Exp. Sta. Report 3, Madison (1955).
188. K. G. Joback, and R. C. Reid, *Chem. Eng. Comm.* **57**, 233 (1987).
189. R. F. Fedors, *AIChE. J.* **25**, 202 (1979).
190. R. F. Fedors, *Chem. Eng. Comm.* **16**, 149 (1982).
191. L. Reidel, *Z. Electrochem.* **53**, 222 (1949).
192. L. Reidel, *Chem. Ing. Tech.* **24**, 353 (1952).
193. W. R. Gambill, *Chem. Eng.*, **66**(12), 181 (1959).
194. A. Vetere, *AIChE J.* **22**, 950 (1976); Errata, **23**, 406 (1977).
195. S. W. Benson, and J. H. Buss, *J. Chem. Phys.* **29**, 546 (1958).
196. S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, and R. Walsh, *Chem. Rev.* **69**(3), 279–324 (1969).
197. J. W. Jalowka, and T. E. Daubert, *Ind. Eng. Chem. Process Des. Dev.* **25**, 139 (1986).
198. T. E. Daubert, and R. Bartakovits, *Ind. Eng. Chem Res.* **28**, 5 (1989).
199. S. Sugden, *The Parachor and Valency*, Routledge & Sons, London, 1930.
200. H. P. Meissner, *Chem. Eng. Prog.* **45**, 2, 149 (1949).
201. Y. Ogata, and M. Tsuchida, *Ind. Eng. Chem.* **49**, 415 (1957).
202. W. Y. Lai, D. H. Chen, and R. N. Maddox, *Ind. Eng. Chem. Res.* **26**, 1072 (1987).
203. H. S. Elbro, A. Fredenslund, and P. Rasmussen, *Ind. Eng. Chem. Res.* **30**, 2576 (1991).
204. G. Le Bas, *The Molecular Volumes of Liquid Chemical Compounds*, Longmans, Green, New York, 1915.
205. M. T. Tyn, and W. F. Calus, *J. Chem. Eng. Data* **20**, 106 (1975).
206. S. Banarjee, P. H. Howard, and S. Lande, *Chemosphere* **21**, 11–12, 1173–1180, (1990).
207. P. P. Radecki, T. N. Rogers, and M. E. Mullins, *Initial Development of a Method to Predict Antoine Constant Group Contributions by Simultaneous Optimization*, AIChE-DIPPR Project 912 Meeting, Michigan Technological University (June 8, 1992).
208. C. Antoine, *Compt. Rend.* **107**, 681, 836 (1888).
209. L. P. Burkhard, *Ind. Eng. Chem. Fundam.* **24**, 119 (1985).
210. A. B. Macknick, and J. M. Prausnitz, *Ind. Eng. Chem. Fundam.* **18**, 348 (1979).
211. A. Bondi, *Physical Properties of Molecular Crystals, Liquids and Gases*, John Wiley & Sons, Inc., New York, 1968, Chapt. 14.
212. C. M. Kelly, P. M. Mathias, and F. K. Schweighardt, *Ind. Eng. Chem. Res.* **27**, 1732 (1988).
213. D. D. Lawson, *Appl. Energy* **6**, 241 (1980).
214. D. Hoshino, X. Zhu, K. Nagahama, and M. Hirata, *Ind. Eng. Chem. Fundam.* **24**, 112 (1985).
215. D. Hoshino, K. Nagahama, and M. Hirata, *J. Chem. Eng. Japan* **11**, 403 (1978).
216. S. W. Benson, *Thermochemical Kinetics*, John Wiley & Sons, Inc., New York, 1968, Chapt. 2.
217. K. Ogiwara, Y. Arai, and S. Saito, *J. Chem. Eng. Japan* **14**, 156 (1981); Y. Yoneda, *Bull. Chem. Soc. Japan* **52**, 1297 (1979).
218. T-P Thinh, J-L Duran, and R. S. Ramalho, *Ind. Eng. Chem. Process Des. Dev.* **10**, 576 (1971).
219. K. Ogiwara, Y. Arai, and S. Saito, *J. Chem. Eng. (Japan)* **14**, 156 (1981).
220. D. B. MacLeod, *Trans. Faraday Soc.* **19**, 38 (1923).
221. D. Reichenberg, *Symp. Transp. Prop. Fluids and Fluid Mixtures*, Natl. Eng. Lab., East Kilbride, Glasgow, Scotland, 1979; also *AIChE. J.* **19**, 854 (1973); **21**, 181 (1975).
222. D. van Velzen, R. L. Cardozo, and H. Langenkamp, *Ind. Eng. Chem. Fundam.* **11**, 20 (1972).
223. T. E. Daubert, private communication, Department of Chemical Engineering, The Pennsylvania State University, University Park, Pa., 1992.
224. D. Roy, and G. Thodos, *Ind. Eng. Chem. Fundam.* **7**, 529 (1968); **9**, 71 (1970).
225. M. Nagvekar, and T. E. Daubert, *Ind. Eng. Chem. Res.* **26**, 1362 (1987).

226. D. Hoshino, K. Nagahama, and M. Hirata, *J. Chem. Eng. Japan* **15**, 153 (1982).
227. M. E. Mullins, T. N. Rogers, A. J. Pintar and J. R. Mihelcic, *AIChE Design Institute for Physical Property Data, Research Project 912, Environmental, Safety and Health Data Estimation Manual*, Year End Progress Report, Michigan Technological University, Houghton, Mich. March, 1992, 91–114.
228. T. N. Rogers, *Henry's Law Constants for Environmental Applications*, presented at the Second International Conference on Thermodynamics of Aqueous Systems with Industrial Applications, Warrenton, Virginia (May 10–14, 1987).
229. W. B. Arbuckle, *Environ. Sci. Technol.* **17**, 537 (1983).
230. C. Antunes and D. Tassios, *Ind. Eng. Chem. Process Des. Dev.* **22**, 457 (1983).
231. J. Hine and P. K. Mookerjee, *J. Org. Chem.* **40**, 292 (1975).
232. W. J. Lyman, Estimation of Physical Properties in W. J. Lyman, Estimation of Physical Properties *Environmental Exposure from Chemicals*, Vol **1**, CRC Press, Boca Raton, Fla., 1987, 13–47.
233. W. M. Meylan, and P. H. Howard, *Env. Toxicology and Chem.* **10**, 1283 (1991).
234. G. G. Nys, and R. F. Rekker, *Eur. J. Med. Chem. Chim. Ther.* **8**, 521 (1973); **9**, 361 (1975).
235. A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, *J. Med. Chem.* **18**, 865 (1975).
236. J. T. Chou, and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **19**, 172 (1979).
237. E. N. Fuller, P. D. Schettler, and J. C. Giddings, *Ind. Eng. Chem.* **58**, 19 (1966).
238. D. P. Northup, A. J. Engel, R. P. Danner, and T. E. Daubert, *Design Institute for Physical Property Data Documentation Report No. 802-10-86*, American Institute of Chemical Engineers, New York, 1987.
239. M. T. Tyn and W. F. Calus, *J. Chem. Eng. Data* **20**, 106 (1975).
240. W. Hayduk and B. S. Minhas, *Can. J. Chem. Eng.* **60**, 295 (1982).
241. J. A. Abusleme and J. H. Vera, *AIChE. J.* **35**, 481 (1989).
242. S. W. Karickhoff, V. K. McDaniel, C. Melton, A. N. Vellino, D. E. Nute, and L. A. Carriera, *Env. Toxicology and Chem.* **10**, 1405 (1991).
243. J. P. Hickey, A. Aldridge, D. R. May Passino, A. M. Frank, *Expert System Predicts Aquatic Toxicity from Contaminant Chemical Structure*, National Fisheries Research Center-Great Lakes, U.S. Fish and Wildlife Service, Ann Arbor, Mich., 1991; *Ibid., Drug Information Journal* **26**, 487 (1992).
244. K. L. E. Kaiser, ed., *Proceedings of the Workshop on Quantitative Structure–Activity Relationships (QSAR) in Environmental Toxicology held at McMaster University, Hamilton, Ontario, Canada, August 16–18, 1983*, D. Reidel Publishing Co., Dordrecht, Germany, 1983.
245. J. D. McKinney, *Environmental Health Perspectives* **61**, 5 (1985).
246. C. M. Auer, J. V. Nabholz, and K. P. Baetke, *Environmental Health Perspectives* **87**, 183 (1990).
247. P. N. Craig and K. Enslein, *Structure-Activity in Hazard Assessment in Hazard Assessment of Chemicals Current Developments*, Vol. **1**, Academic Press, New York, 1981, p. 381.
248. S. P. Gupte, *Chem. Rev.* **89**, 1765 (1989).
249. J. C. Dearden, *Environmental Health Perspectives* **61**, 203–228 (1985).
250. M. L. Tosato, L. Vigano, B. Skagerberg, and S. Clementi, *Environ. Sci. Technol.* **25**(4), 695 (1991).
251. J. M. McKim, S. P. Bradbury, and G. J. Niemi, *Environmental Health Perspectives* **71**, 171 (1987).
252. J. L. M. Hermans, *Environmental Health Perspectives* **87**, 219 (1990).
253. J. R. Coats, *Environmental Health Perspectives* **87**, 255 (1990).
254. L. P. Hammett, *Physical Organic Chemistry*, 1st Ed., McGraw-Hill, New York, 1940, p. 184.
255. M. Charton, *Chemtech* **4**, 502 (1974).
256. C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.* **33**, 347 (1968).
257. R. W. Taft, J-L. M. Abboud, M. L. Kamlet, and M. H. Abraham, *J. of Solution Chem.* **14**, 3, 153 (1985). An excellent review source including an extensive list of properties correlated with solvatochromic parameters.
258. J. P. Hickey, and D. R. Passino-Reader, *Environ. Sci. Technol.* **25**, 1754 (1991).
259. M. H. Abraham, R. M. Doherty, M. J. Kamlet, and R. W. Taft, *Chemistry in Britain*, 551 (June, 1986); M. J. Kamlet, R. M. Doherty, J-L. M. Abboud, M. H. Abraham, and R. W. Taft, *Chemtech*, 566 (Sept 1986).
260. D. C. Leuhrs, R. E. Brown, and K. A. Godbole, *J. of Solution Chem.* **18**(5), 463 (1989).
261. D. C. Leuhrs, and K. A. Godbole, *J. of Solution Chem.* **192**, 187 (1990).

262. A. Leo, C. Hansch, and D. Elkins, *Chemical Reviews*, **71**, 6, 525–616 (1971).
263. M. J. Hait, C. L. Liotta, C. A. Eckert, D. L. Bergmann, A. M. Karachewski, A. J. Dallas, D. I. Eikens, J. J. Li, P. W. Carr, R. B. Poe, and S. C. Rutan, *The SPACE Predictor for Infinite Dilution Activity Coefficients*, unpublished, presented at AIChE 1992 Annual Meeting, Nov. 3, Miami Beach, Fla. For information, write to Charles. A. Eckert, School of Chemical Engineering, Georgia Institute of Technology, Atlanta, Ga. 30332-0100.
264. D. Hawker, *Chemosphere* **19**, 1585–1593 (1989).
265. D. Hawker, *Chemosphere* **20**, 467–477 (1990).
266. Y. Marcus, *J. Phys. Chem.* **95**, 8886–8891 (1991).
267. M. J. Kamlet, R. M. Doherty, M. H. Abraham, and R. W. Taft, *Carbon* **23**(5), 549 (1985).
268. M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus, and R. W. Taft, *J. Phys. Chem.* **92**, 5244–5255 (1988).
269. H. Wiener, *J. Am. Chem. Soc.* **69**, 17 (1947).
270. H. Weiner, *J. Chem. Phys.* **15**, 766 (1947).
271. P. G. Seybold, M. May, and U. A. Bagal, *J. of Chemical Education* **64**, 575 (1987).
272. M. Randic, *J. Am. Chem. Soc.* **97**, 6609 (1975).
273. L. B. Kier, and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
274. L. B. Kier, W. J. Murray, M. Randic, and L. H. Hall, *J. Pharm. Sci.* **65**, 1226 (1976).
275. L. B. Kier and L. H. Hall, *J. Pharm. Sci.* **65**, 1806 (1975).
276. L. H. Hall, L. B. Kier, and W. J. Murray, *J. Pharm. Sci.* **64**, 1974 (1975).
277. W. J. Murray, L. H. Hall, and L. B. Kier, *J. Pharm. Sci.* **64**, 1978 (1975).
278. L. B. Kier, *J. Pharm. Sci.* **70**, 930 (1981).
279. D. T. Stanton and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **32**, 109 (1992).
280. M. B. Perry and C. M. White, *AIChE. J.* **33**, 146 (1987).
281. D. Weininger and J. L. Weininger, in C. Hansch, P. G. Sammes, and J. B. Taylor, eds., *Comprehensive Medicinal Chemistry, Vol. 4*, Pergamon Press, New York, 1990.
282. S. Borman, *C. & E. News*, 28 (Sept. 18, 1989).
283. H. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass., 1969.
284. R. J. Wilson, *Introduction to Graph Theory*, Academic Press, New York, 1972.
285. Z. Mihalic and N. Trinajstic, *J. of Chemical Education* **69**, 701 (1992).
286. H. P. Schultz, E. B. Schultz, and T. P. Schultz, *J. Chem. Inf. Comput. Sci.* **32**, 69 (1992).
287. K. Yuta, and P. C. Jurs, *J. Med. Chem.* **24**, 241 (1981).
288. M. N. Hasan, and P. C. Jurs, *Anal. Chem.* **55**, 263 (1983).
289. J. S. Wishnok In R. A. Scanlan, and S. R. Tannenbaum, Eds., *N-Nitroso Compounds*, ACS Symposium Series, No. 174, Am. Chem. Soc., Washington, D.C., 1981, 77–88.
290. R. Dagani, *Chem. Eng. News* **26**, (Mar 9, 1981).
291. K. Enslein, and P. N. Craig, *J. Toxicol. Environ. Health* **10**, 521 (1982).
292. D. H. Rouvray, *Am. Sci.* **61**, 729 (1973).

Michael E. Mullins
Tony N. Rogers
Michigan Technological University
Peter P. Radecki
Center for Clean Industrial and Treatment Technologies (CenCITT)

Related Articles

Chemometrics; Computer technology; Dimensional analysis; Thermodynamic properties; Computer-aided engineering