

## Thermodynamics in the treatment of vapor/liquid equilibrium (VLE) data

Hendrick C. Van Ness

Howard P. Isermann Department of Chemical Engineering,  
Rensselaer Polytechnic Institute, Troy, New York 12180, USA

**Abstract:** The Rossini lecture affords the lecturer an opportunity to review his contributions to a subject and to reflect on them from the perspective of an extended career. The subject here is vapor/liquid equilibrium, and the topics reviewed include property relations, experimental measurements, consistency tests, and data reduction and correlation.

In the first of these lectures Dr. Rossini remarked that the sponsoring group was named the Commission on Thermodynamics rather than the Commission on *Chemical* Thermodynamics so that engineering thermodynamics could be included. I am, however, the first engineer to give this lecture, a possible exception being Dr. Rossini himself. His B.S. degree was in fact in chemical engineering; however, he immediately turned to graduate work in chemistry, and certainly considered himself a chemist during his professional career. So it would seem that all of my predecessors have been chemists, very distinguished chemists, who have set a high standard indeed for this lecture. As the tenth Rossini lecturer I am expected to draw upon contributions I have made to thermodynamics. Of greatest practical interest no doubt are those relating to vapor/liquid equilibrium (VLE). I propose therefore to expound on the thermodynamic treatment of VLE data, with consideration limited to *binary* systems at low to moderate pressures, a scope suited to the occasion. This topic is more complex and less widely understood than might be supposed.

### PROPERTY RELATIONS

Central to the correlation of VLE data are the thermodynamic *excess properties*, which describe the behavior of the liquid phase. These properties, introduced during the mid-1930's, express the differences between actual property values of a solution and the values which would be exhibited by an ideal solution at the same temperature, pressure, and composition. The equation interrelating these properties was derived in one of my earliest papers on thermodynamics (1). Now known as the *fundamental excess-property relation*, it is an integral part of the thermodynamic structure that we lay before undergraduate students (2):

$$d(\text{ng}) = \frac{nV^E}{RT} dP - \frac{nH^E}{RT^2} dT + \sum \ln \gamma_i dn_i \quad (1)$$

where  $\gamma_i$  is the activity coefficient of species  $i$  in solution and

$$g \equiv G^E/RT$$

Implicit in Eq. (1) is the relation,

$$\left[ \frac{\partial(\text{ng})}{\partial n_i} \right]_{T,P,n_j} = \ln \gamma_i \quad (2)$$

which identifies  $\ln \gamma_i$  as a partial property with respect to  $g$ . These quantities therefore obey the *summability relation* characteristic of partial properties:

$$ng \equiv \frac{nG^E}{RT} = \sum n_i \ln \gamma_i \quad (3)$$

General differentiation gives

$$d(ng) = \sum n_i d \ln \gamma_i + \sum \ln \gamma_i dn_i$$

Comparison with Eq. (1) yields the *Gibbs/Duhem equation* for excess properties:

$$\frac{nV^E}{RT} dP - \frac{nH^E}{RT^2} dT = \sum n_i d \ln \gamma_i \quad (4)$$

Applied to one mole of a liquid phase containing species 1 and 2, Eqs. (1) and (4) become:

$$dg = \frac{V^E}{RT} dP - \frac{H^E}{RT^2} dT + \ln \frac{\gamma_1}{\gamma_2} dx_1 \quad (5)$$

and

$$\frac{V^E}{RT} dP - \frac{H^E}{RT^2} dT = x_1 d \ln \gamma_1 + x_2 d \ln \gamma_2 \quad (6)$$

For simplicity we invoke the following definitions:

$$\varepsilon_P = \frac{V^E}{RT} \frac{dP}{dx_1} \quad (7)$$

$$\varepsilon_T = \frac{-H^E}{RT^2} \frac{dT}{dx_1} \quad (8)$$

The equations which describe binary VLE are written either for constant  $T$  or constant  $P$ . For the isobaric case,  $\varepsilon_P = 0$ ; for the isothermal case,  $\varepsilon_T = 0$ . Either way, only one  $\varepsilon$  term is needed; thus for application to binary VLE we rewrite Eqs. (5) and (6) as

$$\frac{dg}{dx_1} = \ln \frac{\gamma_1}{\gamma_2} + \varepsilon \quad (9)$$

and

$$x_1 \frac{d \ln \gamma_1}{dx_1} + x_2 \frac{d \ln \gamma_2}{dx_1} - \varepsilon = 0 \quad (10)$$

where  $\varepsilon$  is evaluated by Eq. (7) or Eq. (8) as appropriate. Equation (10) is the Gibbs/Duhem equation pertinent to this work.

Equation (3) here becomes

$$g = x_1 \ln \gamma_1 + x_2 \ln \gamma_2 \quad (11)$$

Thus the treatment of binary VLE is based on three very simple equations [Equations (9), (10), and (11)], of which any pair may be chosen as independent.

## EXPERIMENTAL MEASUREMENTS

The directly measured variables in binary VLE are  $x_1, y_1, T$ , and  $P$ . Experimental values of the activity coefficient of species  $i$  in the liquid are related to these variables by:

$$\gamma_i^* = \frac{y_i P}{x_i P_i^{\text{sat}}} \Phi_i \quad (i = 1, 2) \quad (12)$$

where

$$\Phi_i \equiv \frac{\hat{\phi}_i}{\phi_i^{\text{sat}}} \exp \left\{ \frac{-V_i^l(P - P_i^{\text{sat}})}{RT} \right\} \quad (13)$$

In Eq. (12) the asterisk denotes an experimental value, determined primarily by the Raoult's-law factor  $y_i P/x_i P_i^{\text{sat}}$ , in which the vapor pressure  $P_i^{\text{sat}}$  is a strong function of temperature. For data at low to moderate pressure, the other factor  $\Phi_i$  is of relatively minor importance, incorporating quantities extraneous to the VLE data set. The fugacity coefficients  $\hat{\phi}_i$  and  $\phi_i^{\text{sat}}$  account for vapor-phase nonidealities, and are readily evaluated from second virial coefficients (2). The exponential is the Poynting factor.

We write a simple summability relation analogous to Eq. (11) to define  $g^*$  as an experimental value:

$$g^* \equiv x_1 \ln \gamma_1^* + x_2 \ln \gamma_2^* \quad (14)$$

Because experimental measurements are subject to systematic error, sets of values of  $\ln \gamma_1^*$  and  $\ln \gamma_2^*$  determined by experiment may not exhibit partial-property behavior with respect to  $g^*$ . Although partial properties obey the summability relation, summability does not confer partial-property characteristics, and Eq. (14) does not make  $\ln \gamma_1^*$  and  $\ln \gamma_2^*$  values into partial properties. Moreover, when we apply the analogs of Eqs. (9) and (10) to sets of experimental values, they become tests of the thermodynamic consistency of the data, rather than valid general relationships.

#### AREA TEST FOR THERMODYNAMIC CONSISTENCY

Once thermodynamics came to be applied in the correlation of VLE data, interest developed in testing the experimental values  $\ln \gamma_1^*$  and  $\ln \gamma_2^*$  for conformity with Eq. (10), the Gibbs/Duhem equation. Implicit in this equation is a *slope test* for thermodynamic consistency that compares slopes of curves drawn to fit data points on plots of  $\ln \gamma_1^*$  vs.  $x_1$  and  $\ln \gamma_2^*$  vs.  $x_1$ . This test proved to be a tedious and uncertain business, and it never found serious application. Consistent data should also conform to Eq. (9), but again direct application requires measurement of slopes. A practical means to test for consistency is provided by the *area test*.

Differentiation of Eq. (14) gives

$$\frac{dg^*}{dx_1} = x_1 \frac{d \ln \gamma_1^*}{dx_1} + \ln \gamma_1^* + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \ln \gamma_2^*$$

or

$$dg^* = \left( \ln \frac{\gamma_1^*}{\gamma_2^*} + \varepsilon \right) dx_1 + \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right) dx_1 \quad (15)$$

This equation may be integrated over the full composition range:

$$\int_{x_1=0}^{x_1=1} dg^* = \int_0^1 \left( \ln \frac{\gamma_1^*}{\gamma_2^*} + \varepsilon \right) dx_1 + \int_0^1 \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right) dx_1$$

But  $g^* = 0$  at both  $x_1 = 0$  and  $x_1 = 1$ , making the left-hand integral zero. Therefore

$$\int_0^1 \left( \ln \frac{\gamma_1^*}{\gamma_2^*} + \varepsilon \right) dx_1 + \int_0^1 \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right) dx_1 = 0 \quad (16)$$

If in this equation the integrand of the second integral is everywhere zero as required by Eq. (10), the Gibbs/Duhem equation, then both integrals are zero, and we may write:

$$\boxed{\int_0^1 \left( \ln \frac{\gamma_1^*}{\gamma_2^*} + \varepsilon \right) dx_1 = 0} \quad (17)$$

The presumption in this derivation is that the data are consistent. Given values of  $\gamma_1^*$ ,  $\gamma_2^*$ , and  $\varepsilon$ , this equation then represents a test of consistency that is very easy to implement. One merely plots values

of the quantity in parentheses vs.  $x_1$ , and the test requires the net area be zero, i.e., the positive area (above the  $x_1$  axis) should equal the negative area (below the  $x_1$  axis). The test is almost always applied with  $\varepsilon$  set equal to zero, an approximation acceptable for isothermal data, but less likely to be realistic for isobaric data.

This *area test* was proposed in 1947/48 by independent publication in England by Herington (3) and in the United States by Redlich and Kister (4). It was immediately accepted because of its simplicity, and quickly became widely applied. And so it is until the present day. The original publications have been routinely referenced for 45 years, and are among the most cited of all papers on experimental thermodynamics. The *area test* is a remarkable success; perhaps it is folly to inveigh against it, but that is my intent. Fortunately I can offer a new and far superior alternative.

An immediate problem with the *area test* is to decide just how close to zero the net area must be in order to justify a claim of consistency for the data. The criterion in general use is that the *net* area, i.e., the area determined with due regard to sign, should be no more than 10% of the *total* area, i.e., the area determined without regard to sign. This is far from a stringent requirement, and has been happily accepted by experimentalists eager to sanctify their data.

The *area test* is a gross test applied to the data set as a whole, and as I pointed out many years ago (5), it is merely a necessary, not a sufficient condition for consistency. The integrand of the second integral in Eq. (16) can exhibit positive values over part of the range of the integral and compensating negative values over the remaining part that reduce the value of the integral, even to the point of making it zero. It is entirely possible, indeed it is common, for data to meet the criterion of consistency adopted for the *area test* when they are in fact seriously inconsistent with the Gibbs/Duhem equation itself.

The final problem attending use of the *area test* depends on whether the VLE data are isobaric or isothermal. For data taken at constant pressure,  $\varepsilon$  in Eq. (15) is

$$\varepsilon \equiv \varepsilon_T = \frac{-H^E dT}{RT^2 dx_1}$$

a quantity that cannot in general be regarded as negligible. However, the data required for its evaluation are often missing, and even when they are accessible the need to evaluate  $\varepsilon$  is often dismissed. The result is that in virtually all applications of the *area test* to isobaric data,  $\varepsilon$  is treated as making a negligible contribution, and is omitted.

For the far more useful case of data taken at constant temperature,  $\varepsilon$  is given by

$$\varepsilon = \varepsilon_P = \frac{V^E dP}{RT dx_1}$$

Here the assumption that  $\varepsilon$  makes a negligible contribution is fully justified. The problem associated with isothermal data is of quite a different nature.

The possibility that sets of values of  $\gamma_1^*$  and  $\gamma_2^*$  may fail to satisfy the Gibbs/Duhem equation arises only when the two sets are determined independently from experimental measurement of all the variables— $x_1$ ,  $y_1$ ,  $T$ , and  $P$ . However, the equilibrium state is fixed for given values of  $T$  and  $x_1$  when *either*  $P$  or  $y_1$  is known. Thus, measurement of *both*  $P$  and  $y_1$  is an overdetermination of the equilibrium state, which introduces the potential for inconsistency. Indeed, the real question is whether the measurements of  $P$  are consistent with those of  $y_1$ . Without both measurements, no experimental values of  $\gamma_1^*$  and  $\gamma_2^*$  can be calculated, and no test by the Gibbs/Duhem equation is possible.

When the ratio  $\gamma_1^*/\gamma_2^*$  is formed,  $P$  cancels:

$$\frac{\gamma_1^*}{\gamma_2^*} = \frac{y_1 P \Phi_1 / x_1 P_1^{\text{sat}}}{y_2 P \Phi_2 / x_2 P_2^{\text{sat}}} = \frac{y_1 \Phi_1 x_2 P_2^{\text{sat}}}{y_2 \Phi_2 x_1 P_1^{\text{sat}}}$$

The pressures need not even be measured; they can be assigned any values at all, and the *area test* is unchanged. The minor factors  $\Phi_1$  and  $\Phi_2$  do depend on  $P$ , but an appropriate value is readily found by a simple iterative procedure. The *area test* can therefore be applied to the  $y_1 - x_1$  subset of VLE data taken at constant temperature. So what does it test? Only whether the fixed value of the vapor-pressure ratio  $P_2^{\text{sat}}/P_1^{\text{sat}}$  is appropriate to the  $y_1 - x_1$  data subset; it is otherwise worthless.

## DATA REDUCTION

Though the reason is hardly obvious, the subject of data reduction is relevant to further treatment of consistency. Data reduction is the process of finding a suitable analytic relation for  $g$  ( $\equiv G^E/RT$ ) as a function of its independent variables  $T$  and  $x_1$ , thus producing a correlation of VLE data. Although  $g$  is in principle also a function of  $P$ , the dependence is so weak as to be universally and properly neglected.

When a binary liquid phase is in equilibrium with its vapor, there are but two degrees of freedom. Thus when  $T$  is fixed, then for any value of  $x_1$  we can calculate values for  $y_1$  and  $P$ . These are BUBL P calculations. Similarly, when  $P$  is fixed and  $x_1$  is given, values for  $y_1$  and  $T$  are determined by BUBL T calculations. Both of these routine engineering calculations (2) are based on an ability to evaluate  $g$ ,  $P_1^{\text{sat}}$ ,  $P_2^{\text{sat}}$ ,  $\Phi_1$ , and  $\Phi_2$ .

We assume the availability of an expression for  $g$  as a function of  $x_1$  and  $T$  that is inherently capable of correlating the data to within their experimental precision. An example is the 3-parameter Margules equation, an expression often useful for this purpose:

$$g \equiv \frac{G^E}{RT} = (A_{21}x_1 + A_{12}x_2 - Cx_1x_2)x_1x_2 \quad (18)$$

where the parameters  $A_{21}$ ,  $A_{12}$ , and  $C$  are functions of temperature. Equation (18) then represents *derived* values of  $g$ , and application of Eq. (2) provides *derived* values of the partial properties  $\ln \gamma_1$  and  $\ln \gamma_2$ .

Data reduction is particularly simple for the isothermal case, where the parameters, here  $A_{21}$ ,  $A_{12}$ , and  $C$ , are constants, and the objective is their evaluation. Given a computer, a nonlinear regression program, and starting values for the parameters, initial values for  $\ln \gamma_1$  and  $\ln \gamma_2$  are found from Eq. (2). Written for *derived* values, Eq. (12) yields:

$$y_1P = \frac{x_1\gamma_1P_1^{\text{sat}}}{\Phi_1} \quad \text{and} \quad y_2P = \frac{x_2\gamma_2P_2^{\text{sat}}}{\Phi_2}$$

By addition

$$P = \frac{x_1\gamma_1P_1^{\text{sat}}}{\Phi_1} + \frac{x_2\gamma_2P_2^{\text{sat}}}{\Phi_2}$$

whence

$$y_1 = \frac{x_1\gamma_1P_1^{\text{sat}}}{\Phi_1P}$$

Thus for given values of the independent variables  $T$  and  $x_1$  these BUBL P calculations provide *derived* values for  $P$  and  $y_1$ , in addition to the derived values of  $g$ ,  $\ln \gamma_1$ ,  $\ln \gamma_2$ , and  $\ln(\gamma_1/\gamma_2)$ . The corresponding set of *experimental* values includes measured values of the dependent variables  $P^*$  and  $y_1^*$ , values of  $\gamma_1^*$  and  $\gamma_2^*$  calculated by Eq. (12) written as

$$\gamma_i^* = \frac{y_i^*P^*}{x_iP_i^{\text{sat}}}\Phi_i$$

values of  $g^*$  from Eq. (14), and values of

$$\ln(\gamma_1^*/\gamma_2^*) \equiv \ln \gamma_1^* - \ln \gamma_2^*$$

We define a *residual* as the difference between a derived value and the corresponding experimental value. Sets of values for residuals  $\delta P$ ,  $\delta y_1$ ,  $\delta g$ ,  $\delta \ln \gamma_1$ ,  $\delta \ln \gamma_2$ , and  $\delta \ln(\gamma_1/\gamma_2)$  are readily calculated. The computer program can operate to adjust the parameters in the expression for  $g$  so as to minimize an *objective function* formed from the sum of squares of any one of these sets of residuals— $\sum(\delta P)^2$ ,  $\sum(\delta y_1)^2$ ,  $\sum(\delta g)^2$ , etc. The result is an expression for  $g = g(x_1)$  with optimized parameters. Unless the data are perfectly consistent, each choice of objective function leads to a different set of optimized parameters.

Dealing with *isobaric* data is more difficult, because both  $T$  and  $x_1$ , upon which  $g$  depends, are variables. Done properly, data reduction requires that  $g$  be expressed as a function of both  $T$  and  $x_1$ . The process is similar to that used for isothermal data, though the required BUBL T calculations are more difficult than BUBL P calculations. Nevertheless, they are readily carried out, given appropriate data, and they yield the same sets of residuals as for isothermal data, except that  $\delta T$  replaces  $\delta P$ .

### DIRECT TEST OF THERMODYNAMIC CONSISTENCY

Although the quantity  $\ln(\gamma_1^*/\gamma_2^*)$  itself provides an *area test* of very limited value, the *residuals*  $\delta \ln(\gamma_1/\gamma_2)$  offer a unique opportunity for consistency testing. We write Eq. (15) as

$$\frac{dg^*}{dx_1} = \ln \frac{\gamma_1^*}{\gamma_2^*} + \varepsilon + x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon$$

and subtract it from Eq. (9):

$$\frac{dg}{dx_1} - \frac{dg^*}{dx_1} = \ln \frac{\gamma_1}{\gamma_2} - \ln \frac{\gamma_1^*}{\gamma_2^*} - \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right)$$

In terms of residuals this becomes

$$\frac{d(\delta g)}{dx_1} = \delta \ln \frac{\gamma_1}{\gamma_2} - \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right) \quad (19)$$

If a data set, either isothermal or isobaric, is reduced with  $\sum(\delta g)^2$  as objective function so as to make the  $\delta g$  residuals scatter about zero, then  $d(\delta g)/dx_1$  is effectively zero, and

$$\boxed{\delta \ln \frac{\gamma_1}{\gamma_2} = x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon} \quad (20)$$

The right-hand side of this equation is exactly the quantity that Eq. (10), the Gibbs/Duhem equation, requires to be zero for consistent data. The residual on the left is therefore a direct measure of deviations from the Gibbs/Duhem equation. The extent to which values of this residual fail to scatter about zero measures the departure of the data from thermodynamic consistency.

*Thus is realized a long-sought goal—a simple direct test of thermodynamic consistency for each point of a VLE data set with respect to the Gibbs/Duhem equation itself.*

Although the *area test* fades into insignificance by comparison, its relation to the *direct test* is of interest. We return to Eq. (19), multiply it by  $dx_1$ , and integrate:

$$\int_{x_1=0}^{x_1=1} d(\delta g) = \int_0^1 \delta \ln \frac{\gamma_1}{\gamma_2} dx_1 - \int_0^1 \left( x_1 \frac{d \ln \gamma_1^*}{dx_1} + x_2 \frac{d \ln \gamma_2^*}{dx_1} - \varepsilon \right) dx_1$$

Since  $g$  and  $g^*$  are both zero at both  $x_1 = 0$  and  $x_1 = 1$ , then  $\delta g \equiv g - g^*$  is also zero at both conditions, regardless of the objective function chosen for data reduction. Thus the left-most integral is zero, and comparison of the resulting equation with Eq. (16) shows that:

$$\int_0^1 \delta \ln \frac{\gamma_1}{\gamma_2} dx_1 = - \int_0^1 \left( \ln \frac{\gamma_1^*}{\gamma_2^*} - \varepsilon \right) dx_1 \quad (21)$$

The integral on the right represents the *area test* of Eq. (17); the left-hand integral therefore provides an alternative formulation of this test. Since the right-hand integral includes only experimental quantities, its value is independent of how the data are fit. It follows that the value of the left-hand integral is also independent of fit, independent of the choice of objective function, even though values of  $\delta \ln(\gamma_1/\gamma_2)$  do themselves depend on this choice. Thus a plot of these residuals vs.  $x_1$ , made to represent the *direct test*, also displays the *area test*. However, for this specific purpose a more suitable objective function is  $\sum[\delta \ln(\gamma_1/\gamma_2)]^2$ , because this causes the residuals to scatter about a horizontal line. Its ordinate is zero when the test is satisfied.

## EXAMPLES

The data of Mato et al. (6) for methyl tert-butyl ether[MTBE](1)/chloroform(2) at 313.15 K provide the first example. This is a superb set of data, atypical of what is found in the literature. One gets an immediate impression of the quality of a data set, though not necessarily of its consistency, by looking at a graph on which are shown the experimental values  $\ln \gamma_1^*$ ,  $\ln \gamma_2^*$ , and  $g^*/x_1x_2$  plotted vs.  $x_1$ , as shown by Fig. 1. We note that the points easily define smooth curves, and only slightly erratic behavior is exhibited by values at very high dilution. Such values are inherently uncertain, because limiting values at infinite dilution are indeterminate. However, with this data set extrapolation to reasonable infinite-dilution values is a simple matter. The values used for vapor pressures  $P_i^{\text{sat}}$  are here critical, and an advantage is often gained for high-quality data when these values are found by regression in the manner described by Abbott and Van Ness (7). This has been done here; hence the  $P_i^{\text{sat}}$  values used are slightly different from those reported by the authors.

Figure 2 illustrates the graph traditionally required for the *area test*, where the experimental values  $\ln(\gamma_1^*/\gamma_2^*)$  are plotted vs.  $x_1$ . The smoothness of the data is again evident, and the positive and negative areas clearly appear to be about equal.

Data reduction depends on an equation for  $g$  as a function of  $x_1$  by which the experimental measurements are correlated. Equation (18), the 3-parameter Margules equation, is here entirely adequate, and may be applied to the data to yield sets of optimized parameters. With  $\sum(\delta g)^2$  as objective function, the  $\delta g$  residuals scatter closely about zero, as required for the *direct test*. However, we see from Fig. 3 that this fit produces  $\delta \ln(\gamma_1/\gamma_2)$  residuals that fail to scatter about zero, indicating that the data are not perfectly consistent. Nevertheless, the residuals are very small, and a true indication of how close the data come to being consistent is given by the RMS value of these residuals: 0.011. This set of data is therefore in the category of highest quality. The *area test* is also represented on this plot by the net area under a smoothing curve drawn to correlated the points. However, if out of nostalgia one is seriously interested in the *area test*, then it is best to plot these same residuals as determined from a fit with  $[\delta \ln(\gamma_1/\gamma_2)]^2$  as objective function. As is evident from Fig. 4 these residuals tend to scatter about a horizontal line, which in this case lies slightly above zero.

Next we examine a set of data for the system methanol(1)/water(2) at atmospheric pressure reported by Kurihara et al. (8). Again we first examine a graph, shown by Fig. 5, of the experimental values  $\ln \gamma_1^*$ ,  $\ln \gamma_2^*$ , and  $g^*/x_1x_2$  plotted vs.  $x_1$ . This data set is evidently not of the highest quality, but is representative of many sets appearing in the literature. These data too are well fit by the 3-parameter Margules equation, in this case with temperature-dependent parameters. Excess enthalpy (heat-of-mixing) data for this system allow the temperature dependence of the parameters to be determined independent of the VLE data, and fitting to the VLE data then establishes the correlation. The *direct test* requires a fit of the  $\delta g$  residuals, and the resulting  $\delta \ln(\gamma_1/\gamma_2)$  residuals are shown by Fig. 6. The data clearly exhibit significant departures from thermodynamic consistency; the RMS value of the residual is 0.064, suggesting that the data are only of acceptable quality. Figure 6 also clearly shows that the *area test* is not satisfied, as is confirmed by Fig. 7.

An older set of data for the methanol(1)/water(2) system at atmospheric pressure is that of Uchida and Kato (9). It has been treated exactly as the preceding set, and the corresponding plots are shown by Figs. 8 through 10. It is truly a dreadful set of data. This is immediately evident from the plots of experimental values shown in Fig. 8, and is further confirmed by the *direct test* as shown by Fig. 9, which produces the RMS value for the residual of 0.277, a totally unacceptable result. Astonishingly, the *area test* for this set of data is satisfied. This is suggested by Fig. 9, where the positive and negative areas appear roughly in balance, and is fully confirmed by Fig. 10, where the residuals are seen to scatter about zero. The information provided here by this test is clearly delusive.

These results and other considerations suggest that a scale be established to indicate the quality of a data set as judged by its departure from thermodynamic consistency. The appropriate measure is the RMS value of  $\delta \ln(\gamma_1/\gamma_2)$  as determined from the *direct test*. TABLE 1 shows a proposed *consistency index* which starts at 1 for highly consistent data and goes to 10 for data of very poor quality. It replaces the totally inadequate +/- or yes/no designation presently employed to characterize the consistency of a data set.

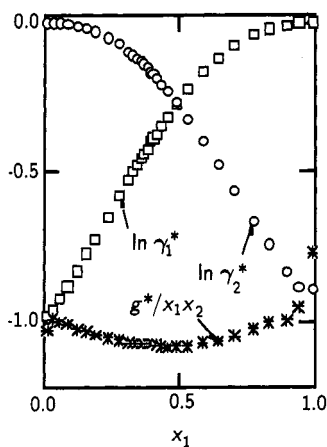


Fig. 1 Experimental values.  
MTBE(1)/chloroform(2), 313.15 K(6).

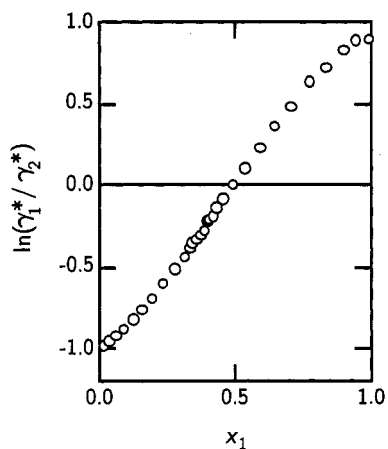


Fig. 2 Traditional area test.  
MTBE(1)/chloroform(2), 313.15 K(6).

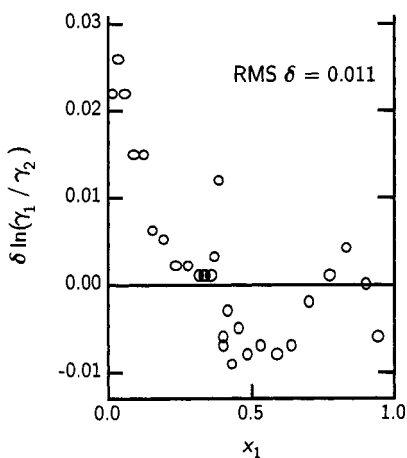


Fig. 3 Data of Fig. 1, fit by Eq. (18).  
Direct test of consistency.

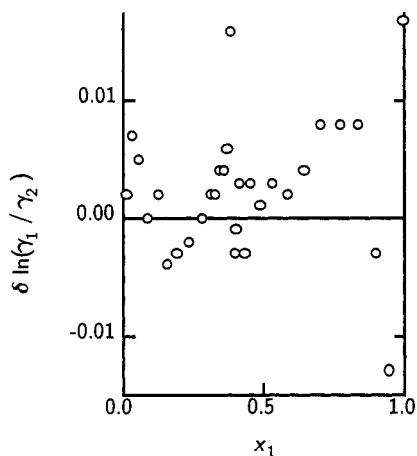


Fig. 4 Data of Fig. 1, fit by Eq. (18).  
Alternative area test.

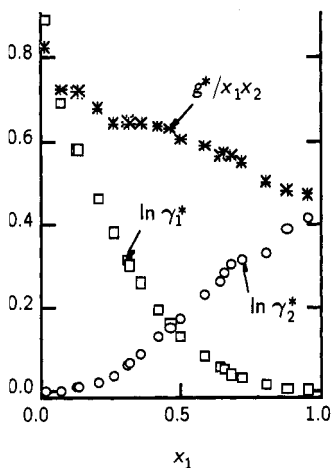


Fig. 5 Experimental values.  
Methanol(1)/Water(2) at 101.325 kPa(8).

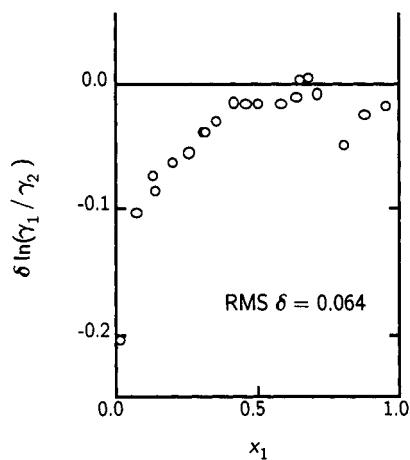


Fig. 6 Data of Fig. 5, fit by Eq. (18).  
Direct test of consistency.



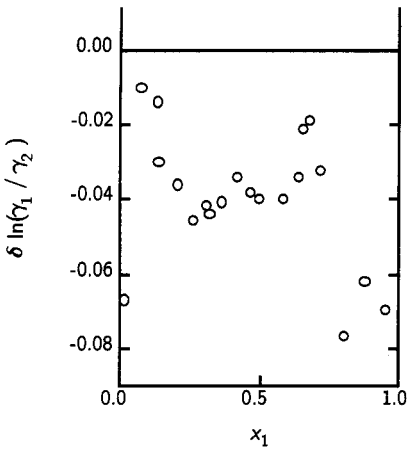


Fig. 7 Data of Fig. 5, fit by Eq. (18). Alternative area test.

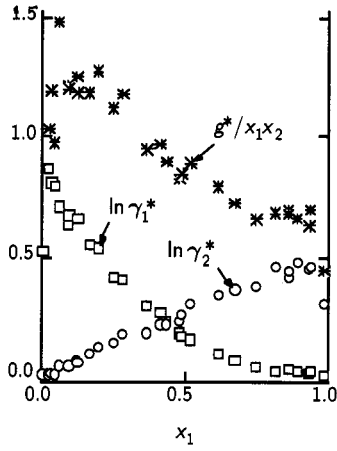


Fig. 8 Experimental values. Methanol(1)/Water(2) at 101.325 kPa(9).

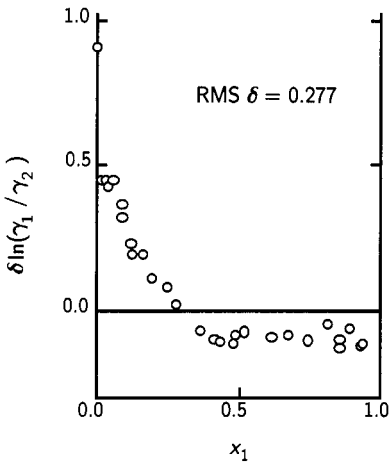


Fig. 9 Data of Fig. 8, fit by Eq. (18). Direct test of consistency.

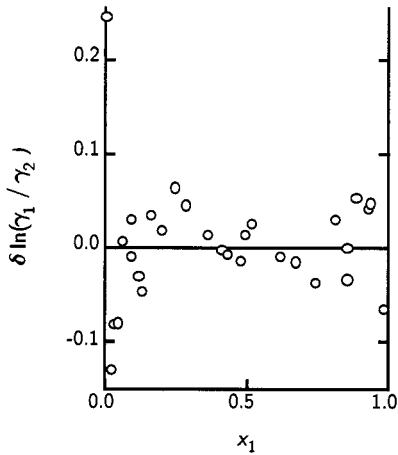


Fig. 10 Data of Fig. 8, fit by Eq. (18). Alternative area test.

TABLE 1. Consistency index for VLE data.

Index	RMS $\delta \ln(\gamma_1/\gamma_2)$	
1	> 0	$\leq 0.025$
2	> 0.025	$\leq 0.050$
3	> 0.050	$\leq 0.075$
4	> 0.075	$\leq 0.100$
5	> 0.100	$\leq 0.125$
6	> 0.125	$\leq 0.150$
7	> 0.150	$\leq 0.175$
8	> 0.175	$\leq 0.200$
9	> 0.200	$\leq 0.225$
10	> 0.225	

## CORRELATION OF VLE DATA

Although the objective functions employed in testing VLE data for thermodynamic consistency certainly produce correlations of the data, they are usually not the best choices for the specific purpose of correlation. Pertinent to this topic are the relations among residuals. Variations in  $g$  for isothermal data and fixed liquid-phase composition are provided by differentiation of Eq. (11):

$$dg = x_1 d \ln \gamma_1 + x_2 d \ln \gamma_2$$

or

$$dg = x_1 \frac{d\gamma_1}{\gamma_1} + x_2 \frac{d\gamma_2}{\gamma_2} \quad (22)$$

Moreover, as a mathematical identity, we have:

$$d \ln \frac{\gamma_1}{\gamma_2} \equiv \frac{d\gamma_1}{\gamma_1} - \frac{d\gamma_2}{\gamma_2} \quad (23)$$

Variations in  $\gamma_i$  in relation to variations in the experimental variables for isothermal data at fixed  $x_i$  are found by differentiation of Eq. (12):

$$d\gamma_i = \frac{P dy_i + y_i dP}{x_i P_i^{\text{sat}}} \Phi_i \quad (i = 1, 2)$$

Variations in the minor factor  $\Phi_i$  are here neglected. Division by Eq. (12) gives:

$$\frac{d\gamma_i}{\gamma_i} = \frac{dy_i}{y_i} + \frac{dP}{P} \quad (i = 1, 2)$$

Writing this equation for  $i = 1$  and for  $i = 2$  and combining the resulting expressions with Eqs. (22) and (23) gives after reduction:

$$\delta g = \frac{x_1 - y_1}{y_1 y_2} \delta y_1 + \frac{\delta P}{P} \quad (24)$$

and

$$\delta \ln \frac{\gamma_1}{\gamma_2} = \frac{\delta y_1}{y_1 y_2} \quad (25)$$

where we have replaced differentials by residuals. Provided the residuals are small, this is an excellent approximation, and Eqs. (24) and (25) are in fact observed to apply regardless of the objective function used for data reduction.

Equations (24) and (25) explain the diversity of results obtained when different objective functions are employed in fitting the same data set. If one adopts as objective function the sum of squares of the  $\delta g$  residuals, then a satisfactory fit makes these residuals scatter about zero. This presumes that the equation representing  $g$  is capable of fitting the data, a premise fundamental to proper data reduction. For fully consistent data, the residuals  $\delta y_1$  and  $\delta P$  also scatter about zero; for inconsistent data, they do not. Rather, they are forced to assume values such that the two terms on the right of Eq. (24) compensate each other, thus making the  $\delta g$  residuals scatter about zero. This often results in inflated values of both the  $\delta y_1$  and  $\delta P$  residuals. However, the behavior of these residuals depends on the relative magnitudes of the two terms on the right of Eq. (24). The first of these terms contains the difference  $x_1 - y_1$ , a quantity that is small over the whole composition range for many systems of interest; this is true in particular for systems that form azeotropes. In this event the second term on the right of Eq. (24) contributes far more than the first to the residuals  $\delta g$ .

If the goal of data reduction is replication of measured values, then use of  $\sum(\delta P)^2$  or  $\sum(\delta y_1)^2$  as objective function is suggested. In the former case the  $\delta P$  residuals are forced to scatter about zero. The measured  $y_1^*$  values are not used in the calculations, but are needed to form the  $\delta y_1$  residuals, which then provide another consistency test (10). When they scatter about zero, then the data are consistent; when they do not, the data are inconsistent. As we shall see, the opposite procedure, based on  $\sum(\delta y_1)^2$  as objective function, has little to recommend it. In any event, when a set of data is

TABLE 2. Residuals from data reduction by Eq. (18).  
Data of Mato et al. (6)  
Methyl tert-Butyl Ether(1)/Chloroform(2) at 313.15 K.

Objective function	RMS $\delta P$ kPa	RMS $\delta y_1$
$\sum(\delta g)^2$	0.028	0.0011
$\sum(\delta P)^2$	0.026	0.0011
$\sum(\delta y_1)^2$	0.072	0.0008

inconsistent, no parameters *exist* for a suitable equation for  $g$  that make *both* the  $\delta y_1$  and  $\delta P$  residuals scatter about zero.

The methyl tert-butyl ether(1)/chloroform(2) system considered earlier does indeed form an azeotrope, and the first term on the right of Eq. (24) contributes little in comparison with the second. Thus we obtain virtually the same fit of the data by Eq. (18) whether the objective function is formed from the  $\delta P$  residuals or from the  $\delta g$  residuals. When the objective function is formed from the  $\delta y_1$  residuals, the fit forces the  $\delta y_1$  residuals to be very small, but there is no constraint on the  $\delta P$  and  $\delta g$  residuals. Although they must be very nearly equal, they are free to inflate together. Thus TABLE 2 shows that the fit based on the  $\delta y_1$  residuals is inferior to that based on either the  $\delta P$  or  $\delta g$  residuals, even for this very excellent set of data.

Equation (25) shows a direct relation between the  $\delta y_1$  residuals and the  $\delta \ln(\gamma_1/\gamma_2)$  residuals. Clearly, Eq. (20) and hence the *direct test* for consistency have alternative forms based on the  $\delta y_1$  residuals. Furthermore, the *area test* represented by the integral on the left side of Eq. (21) also has an alternative form. The remarkable thing about this is that when the *area test* is not satisfied, it is not possible to make the residuals  $\delta \ln(\gamma_1/\gamma_2)$  and hence the quantities  $\delta y_1/y_1y_2$  scatter about zero. In this event it follows that the  $\delta y_1$  residuals themselves cannot be made to scatter about zero, even when the objective function is  $\sum(\delta y_1)^2$ . Surprising as this may be, it is a further reason why  $\sum(\delta y_1)^2$  is a poor choice of objective function.

A number of objective functions, including those discussed here, are treated in detail by Van Ness et al. (11). The conclusion reached there is still valid—although the reduction of isothermal VLE data can be accomplished equally well by several methods, the unweighted least-squares technique based on the objective function  $\sum(\delta P)^2$ , called Barker's method, is at least as good as any, and is certainly the simplest and most direct.

For isobaric data at fixed  $x_i$ , differentiation of Eq. (12) gives:

$$d\gamma_i = \frac{dy_i - y_i d \ln P_i^{\text{sat}}}{x_i P_i^{\text{sat}}} P\Phi_i \quad (i = 1, 2)$$

which may be divided by Eq. (12):

$$\frac{d\gamma_i}{\gamma_i} = \frac{dy_i}{y_i} - d \ln P_i^{\text{sat}} \quad (i = 1, 2)$$

In combination with Eqs. (22) and (23) this ultimately yields:

$$\delta g = \frac{x_1 - y_1}{y_1 y_2} \delta y_1 - (x_1 \delta \ln P_1^{\text{sat}} + x_2 \delta \ln P_2^{\text{sat}}) \quad (26)$$

and

$$\delta \ln \frac{\gamma_1}{\gamma_2} = \frac{\delta y_1}{y_1 y_2} - \delta \ln \frac{P_1^{\text{sat}}}{P_2^{\text{sat}}} \quad (27)$$

Although  $P_i^{\text{sat}}$  depends strongly on temperature for each individual species, the *ratio*  $P_1^{\text{sat}}/P_2^{\text{sat}}$  is insensitive to temperature, and Eq. (27) to an excellent approximation reduces to Eq. (25), exactly as for the isothermal case.

TABLE 3. Residuals from data reduction by Eq. (18).  
Data of Kurihara et al. (8)  
Methanol(1)/Water(2) at 101.325 kPa.

Objective function	RMS $\delta T$ K	RMS $\delta y_1$
$\sum(\delta g)^2$	0.44	0.012
$\sum(\delta T)^2$	0.07	0.007
$\sum(\delta y_1)^2$	0.57	0.003

Equations (25) and (26) apply to the isobaric case in the same way that Eqs. (25) and (24) do for the isothermal case. Much the same discussion applies, except that  $\delta T$  replaces  $\delta P$  as a primary residual, and  $\sum(\delta T)^2$  becomes the preferred objective function. The data set of Kurihara et al. (8) for methanol(1)/water(2) at atmospheric pressure serves as an example. This system does not form an azeotrope, and the difference  $x_1 - y_1$  is large enough that the two terms on the right in Eq. (26) both contribute importantly. Thus when the data are fit with  $\sum(\delta g)^2$  as objective function, the residuals that control these terms must either scatter about zero (consistent data) or must offset each other (inconsistent data). We have already seen that the data exhibit significant inconsistency, and in Fig. 11a we observe offsetting residuals  $\delta y_1$  and  $\delta T$ . When the data are fit with  $\sum(\delta T)^2$  as objective function, the temperature residuals are forced to scatter about zero, as is evident in Fig. 11b. The  $\delta y_1$  residuals clearly do not scatter about zero, confirming the inconsistency already noted for this data set. When  $\sum(\delta y_1)^2$  serves as objective function, *neither* set of residuals scatters about zero, as Fig. 11c shows. The failure of the  $\delta y_1$  residuals to do so reflects the fact that this data set satisfies neither the *direct test* nor the *area test* for consistency. TABLE 3 shows the superiority of the fit wherein the objective function is formed from temperature residuals.

## REFLECTIONS

In the early unsophisticated days of chemical engineering VLE data were taken at constant pressure for direct application in the design of distillation columns, which were treated as though they operated at uniform pressure. There is no longer any excuse for taking isobaric data, but regrettably the practice persists, and we have examined two such data sets. Rigorous thermodynamic treatment of isobaric data presents problems that do not arise with isothermal data. Their origin is the need to take into account not only the composition dependence of the excess Gibbs energy but also its temperature dependence.

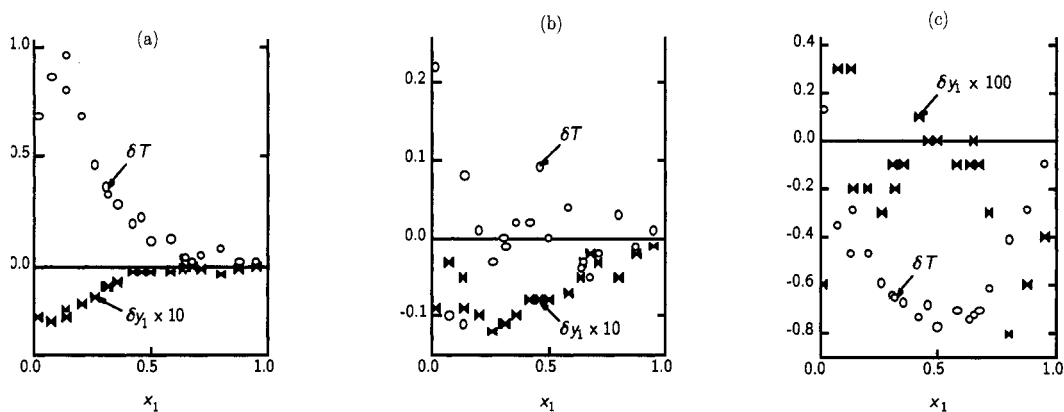


Fig. 11 Data of Kurihara et al.(8), Methanol(1)/Water(2) at 101.325 kPa, fit by Eq. (18). Objective function: (a)  $\sum(\delta g)^2$ , (b)  $\sum(\delta T)^2$ , (c)  $\sum(\delta y_1)^2$ .

Equations (9) and (11) may be combined to eliminate either  $\ln \gamma_1$  or  $\ln \gamma_2$ , yielding

$$\ln \gamma_1 = g + x_2 \left( \frac{dg}{dx_1} - \varepsilon \right) \quad (28a)$$

$$\ln \gamma_2 = g - x_1 \left( \frac{dg}{dx_1} - \varepsilon \right) \quad (28b)$$

These equations are valid for both isothermal and isobaric data. However,  $\varepsilon$  is quite negligible for the isothermal case; our interest here is in data taken at constant pressure, for which  $\varepsilon$  is given by Eq. (8). When pressure is fixed in binary VLE, there is but a single remaining degree of freedom, and  $x_1$  may be treated as the only independent variable. Thus in Eqs. (28)  $dg/dx_1$  is a total derivative. Moreover, it is entirely proper to write an equation that expresses  $g$  as a function of the single variable  $x_1$ . This does not mean that  $g$  is independent of  $T$ ; only that the temperature dependence has been absorbed into the  $x_1$  dependence. It is of course not possible to extract the temperature dependence of  $g$  from such an equation, which is no more than a correlating expression for the constant- $P$  data set. As such it is entirely suitable for use in the *direct test* for thermodynamic consistency.

One may also write an equation that expresses  $g$  as a function of both temperature and composition. When combined with Eqs. (28),  $T$  in such an expression must be treated as a function of  $x_1$ . Applied to the reduction of a constant- $P$  data set, this leads to an equation for  $g$  that evidently shows both the temperature and composition dependence of  $g$ . However, such a result is not to be taken seriously. The temperature range of the data set is too small; moreover,  $T$  and  $x_1$  are strongly correlated, and no single set of constant- $P$  data is adequate to the task of separating one influence from the other.

For isobaric VLE data Eqs. (28) require evaluation of  $\varepsilon$ , but this is, in fact, rarely done. The rationalization is that the temperature range of the data is small and therefore the temperature dependence of the parameters can be ignored. Nevertheless, when  $\varepsilon$  is set equal to zero for isobaric VLE, an approximation is introduced in the calculation of activity coefficients by Eqs. (28). This may not degrade the quality of fit obtained in data reduction for any particular objective function, but quantities derived from the resulting equation for  $g$  reflect the approximation. At best it should be regarded as a data-smoothing procedure.

An alternative to Eqs. (28) is developed as follows. For isobaric VLE,  $g$  is a function of both  $T$  and  $x_1$ , but  $T$  is also a function of  $x_1$ . We may therefore write

$$\frac{dg}{dx_1} = \left( \frac{\partial g}{\partial x_1} \right)_{T,P} + \left( \frac{\partial g}{\partial T} \right)_{P,x_1} \frac{dT}{dx_1} \quad (29)$$

As a consequence of Eq. (5) we have

$$\left( \frac{\partial g}{\partial T} \right)_{P,x_1} = \frac{-H^E}{RT^2} \quad (30)$$

Combining this equation and Eq. (8) with Eq. (29) leads to

$$\frac{dg}{dx_1} - \varepsilon = \left( \frac{\partial g}{\partial x_1} \right)_{P,T} \quad (31)$$

Substitution into Eqs. (28) gives

$$\ln \gamma_1 = g + x_2 \left( \frac{\partial g}{\partial x_1} \right)_{T,P} \quad (32a)$$

$$\ln \gamma_2 = g - x_1 \left( \frac{\partial g}{\partial x_1} \right)_{T,P} \quad (32b)$$

The derivative  $(\partial g/\partial x_1)_{T,P}$  of course has no physical meaning within the context of binary vapor/liquid equilibrium, because  $x_1$  cannot vary at constant  $T$  and  $P$ . However, as applied to an equation giving  $g$  as a function of  $T$  and  $x_1$  it certainly has mathematical meaning. Equations (32) are far more

convenient for such application than are Eqs. (28). Although  $\varepsilon$  need not be evaluated,  $H^E$  data are required for proper evaluation of the temperature dependence of  $g$  through application of Eq. (30), the Gibbs/Helmholtz equation.

First, however, we must have an equation for  $g$  that adequately represents both its  $x_1$  dependence and its  $T$  dependence. The Margules equation lends itself admirably to this application, as described in detail by Van Ness and Abbott (12). However, it is not suitable for some highly nonideal systems, and no adequate alternative is yet known.

Another disadvantage of taking isobaric data is that the temperature dependence of the vapor pressures  $P_i^{\text{sat}}$  must be taken into account. These are the most important thermodynamic quantities in VLE, and correct values are essential to proper data reduction. They should in fact be measured over the temperature range of the data and carefully fit to adequate vapor-pressure equations. This is rarely done; rather correlating equations are taken from the literature and assumed valid, a practice that ignores the high sensitivity of the vapor pressures of chemical species to their purity.

Given the problems that attend reduction of isobaric VLE data, the continued accumulation of such data makes little sense. An equivalent effort spent on taking isothermal VLE data would be far more useful. Once having decided on such a course, we face yet another problem: Just which variables do we measure? One can readily evaluate the parameters in an appropriate expression for  $g$  at fixed temperature from  $P - x_1$ ,  $y_1 - x_1$ , or  $P - y_1$  data subsets. Accurate measurement of  $y_1$  is by far the most difficult, and therefore experiments which produce  $P - x_1$  data subsets are favored. Since such measurements are sufficient to the purpose of data reduction, why would one also make the more difficult measurements of  $y_1$ ? The only logical reason is so as to subject the measurements to a test for thermodynamic consistency. Only when a full set of measurements is available is this possible, because such a set represents an overdetermination of the equilibrium state. If the data satisfy the *direct test* for consistency, then there is no problem; all data-reduction procedures produce essentially the same results, and Barker's method is probably as reasonable a choice as any. Since the  $y_1$  values are not used in this procedure, they may as well not have been measured, except to show consistency. The catch is that such measurements are rarely made with an accuracy sufficient to avoid *introduction* of inconsistency. In this event the various procedures of data reduction produce discordant results, and one must choose among them. Since the greatest uncertainty lies with the  $y_1$  measurements, Barker's method is again the logical choice. The moral is that measurements of  $y_1$  are pointless unless they are made with sufficient care to be very accurate, and then their use is to validate the data set by the demonstration of consistency, not to improve the correlation.

The thermodynamic mill grinds slowly, but it grinds exceedingly fine.

## REFERENCES

1. H. C. Van Ness, *Chem. Eng. Sci.*, **10**, 225 (1959).
2. J. M. Smith and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 4th ed., McGraw-Hill Book Co., New York, 1987.
3. E. F. G. Herington, *Nature*, **160**, 610 (1947).
4. O. Redlich and A. T. Kister, *Ind. Eng. Chem.*, **40**, 345 (1948).
5. H. C. Van Ness, *Chemical Engineering Science*, **11**, 118 (1959).
6. F. A. Mato, Charles Berro, and André Péneloux, *J. Chem. Eng. Data*, **36**, 259 (1991).
7. M. M. Abbott and H. C. Van Ness, *Fluid Phase Equilibria*, **1**, 3 (1977).
8. K. Kurihara, M. Nakamichi, and K. Kojima, *J. Chem. Eng. Data*, **38**, 446 (1993).
9. S. Uchida and S. Kato, *J. Soc. Chem. Ind. Japan*, **37**, suppl. 525 (1934) as reported by J. Gmehling, U. Onkin, and W. Arlt, *Vapor-Liquid Equilibrium Data Collection*, Chemistry Data Series, Vol. 1, Part 1a, DECHEMA, Frankfurt/Main, 1981.
10. H. C. Van Ness, S. M. Byer, and R. E. Gibbs, *AIChE J.*, **19**, 238 (1973).
11. H. C. Van Ness, F. Pedersen, and P. Rasmussen, *AIChE J.*, **24**, 1055 (1978).
12. H. C. Van Ness and M. M. Abbott, *Classical Thermodynamics of Nonelectrolyte Solutions: With Applications to Phase Equilibria*, McGraw-Hill Book Co., New York, 1982.