Vapor-Liquid Equilibrium:

H. C. VAN NESS
Department of Chemical and
Environmental Engineering
Rensselaer Polytechnic Institute
Troy, New York 12181
FINN PEDERSEN
and
PETER RASMUSSEN
Instituttet for Kemiteknik
Danmarks tekniske Hojskole
2800 Lyngby, Danmark

Part V. Data Reduction by Maximum Likelihood

A systematic study is presented of the objective functions applicable to the reduction of vapor-liquid equilibrium data by regression procedures based on the principle of maximum likelihood. Numerical results are based on a set of artificial x-y-P data constructed so as to conform to the assumptions that underlie proper application of this principle.

SCOPE

Experimental data for x, y, and P, measured at constant temperature for binary systems in vapor-liquid equilibrium (VLE), may be reduced by a variety of regression procedures based on weighted least-squares techniques. When errors in the data set are small, independent, and normally distributed, appropriate weighting factors are provided by application of the principle of maximum likelihood. Even then, one must choose from a diversity of objective functions. We present here the results of a systematic study of the characteristics of a number of such objective functions.

CONCLUSIONS AND SIGNIFICANCE

Although several objective functions with weighting factors based on the principle of maximum likelihood are found suitable for reduction of accurate isothermal VLE data, none proves superior to the method proposed by Barker (1953). Since this method is based on the x-P data subset, measured values of y serve no purpose beyond providing the means for a test of thermodynamic consistency. In any event, accurate measurements of P are essential to any acceptable set of VLE data.

Earlier parts of this series (Abbott and Van Ness, 1975; Van Ness et al., 1973) described various methods which permit the reduction of vapor-liquid equilibrium (VLE) data so as to provide thermodynamic properties of liquid phases. One cannot compare these methods on the basis of their correctness because all are mathematically proper. Questions which arise because of imperfections in the data and inadequacies of the fitting equations must be answered from a pragmatic point of view. We report here some practical results of a study of certain regression procedures.

We consider VLE data for binary systems taken at fixed temperature and at pressures low enough that gas-phase nonidealities and liquid-phase Poynting effects do not enter significantly into the calculations. We presume that a data point is made up of values for x_1, y_1, and P taken at a fixed T for which values of P_1^{sat} and P_2^{sat} are established. We define as experimental values those quantities calculated directly from the experimental data by

$$\gamma_i = y_i P / x_i P_i^{sat} \quad (i = 1, 2) \tag{1}$$

$$\gamma_1/\gamma_2 = y_1 x_2 P_2^{sat} / y_2 x_1 P_1^{sat} \tag{2}$$

$$G^E/RT = x_1 \ln \gamma_1 + x_2 \ln \gamma_2 \tag{3}$$

$$\mathcal{G} \equiv G^E/x_1x_2RT = (\ln \gamma_1)/x_2 + (\ln \gamma_2)/x_1 \tag{4}$$

These simple equations reflect the stated assumptions and exclude second-order effects that merely add complexity with respect to this work.

0001-1541-78-1733-1055-$01.05.

The *object* of data reduction is to find an analytical expression for $\mathcal{G}$ as a function of x_1 (a mathematical model) that leads through the equations of thermodynamics to predicted values of the variables in as close agreement as possible with their experimental values. Given a model equation applicable to the liquid phase at constant T

$$\mathcal{G}^* = \mathcal{G}(x_1, \alpha, \beta, \ldots) \tag{5}$$

where α, β, ... are parameters and the asterisk denotes a predicted value, then we can also predict

$$\ln \gamma_1^* = x_2^2 \left(\mathcal{G}^* + x_1 \frac{d\mathcal{G}^*}{dx_1} \right) \tag{6a}$$

$$\ln \gamma_2^* = x_1^2 \left(\mathcal{G}^* - x_2 \frac{d\mathcal{G}^*}{dx_1} \right) \tag{6b}$$

$$\ln \frac{\gamma_1^*}{\gamma_2^*} = \frac{d(G^E/RT)^*}{dx_1} = (x_2 - x_1)\mathcal{G}^* + x_1 x_2 \frac{d\mathcal{G}^*}{dx_1} \tag{7}$$

$$P^* = x_1 P_1^{\text{sat}} \gamma_1^* + x_2 P_2^{\text{sat}} \gamma_2^* \tag{8}$$

$$y_i^* = \frac{x_i P_i^{\text{sat}} \gamma_i^*}{P^*} \quad (i = 1, 2) \tag{9}$$

With known values for the parameters α, β, ... and for given values of x_1, various comparisons can be made between values predicted by these equations and the corresponding experimental values reported as data or calculated from the data by Equations (1) through (4). Moreover, regression techniques allow determination of best values for the parameters α, β, ... by minimization of various functions of the differences between like quantities as predicted from the model and calculated from the data. For a set of x_1-y_1-P data, one has a wide variety of functions from which to choose, and, unless the data are perfect, different choices lead to different best values of the parameters α, β, ... for a given model. We therefore have the problem of determining a best procedure for producing best values of parameters. Probably no truly best procedure exists for all circumstances, given the diverse factors that influence the choice:

1. The completeness, extent, and quality of the data.
2. The nature of the mathematical model, its flexibility, and intended use.
3. The subjective judgment which inevitably conditions the selection of a regression technique.

The thermodynamic properties of a liquid are functions of its composition x_1, the temperature T, and the pressure P. However, outside the critical region, pressure has little influence, and the effect of pressure on the excess Gibbs function and activity coefficients is treated through the Poynting correction. Thus, pressure does not enter as a variable in mathematical expressions which model the excess Gibbs function, and only x_1 and T appear as independent variables. In VLE measurements, P and y_1 are therefore considered dependent variables.

FUNDAMENTALS OF DATA REDUCTION

Several possibilities for reduction of isothermal VLE data are evident from the equations already given. If we define a residual r as the difference between a predicted value and the corresponding experimental value, then parameter evaluation can be accomplished by any procedure which leads to a set of parameters that makes the set of r values very small. A least-squares technique is almost always used. The point here is that several different residuals are available from the data, for example:

1. $r_A = \mathcal{G}^* - \mathcal{G}$, with $\mathcal{G}^*$ given by Equation (5) and $\mathcal{G}$ by Equation (4).
2. $r_B = \gamma_1^* - \gamma_1$, with γ_1^* given by Equation (6*a*) and γ_1 by Equation (1).
3. $r_C = \gamma_2^* - \gamma_2$, with γ_2^* given by Equation (6*b*) and γ_2 by Equation (1).

For a given set of x_1-y_1-P data points, parameter evaluation can be based on each of these residuals, perhaps with discordant results. The problem is that for internal consistency the experimental activity coefficients must conform to the Gibbs-Duhem equation:

$$x_1 \frac{d \ln \gamma_1}{dx_1} + x_2 \frac{d \ln \gamma_2}{dx_1} = 0 \tag{10}$$

In method 2 only activity coefficients for component 1 enter the calculations, and in method 3 only those for component 2. If the experimental values are inconsistent with Equation (10), then methods 2 and 3 yield different sets of parameters, because each method treats the activity coefficients of just a single component.

Method 1, which is based on activity coefficients for both components, yields a third set of parameters for inconsistent data and is best described with reference to the schematic diagram of Figure 1. The circles represent experimental values of $\ln \gamma_1$ and $\ln \gamma_2$ calculated by Equation (1), and the triangles are experimental values of $\mathcal{G}$ from Equation (4). The line labeled $\mathcal{G}^*$ represents values predicted by a suitable model as a result of a fit by method 1. Predicted values of $\ln \gamma_1^*$ and $\ln \gamma_2^*$, represented by the dashed lines, come from Equations (6*a*) and (6*b*). Since the Gibbs-Duhem equation is implicit in these equations, all predicted values must conform to it. Thus, experimental activity coefficients not consistent with this equation cannot be in agreement with predicted values. To put it another way, values of $\ln \gamma_1$ and $\ln \gamma_2$ combined through addition [Equation

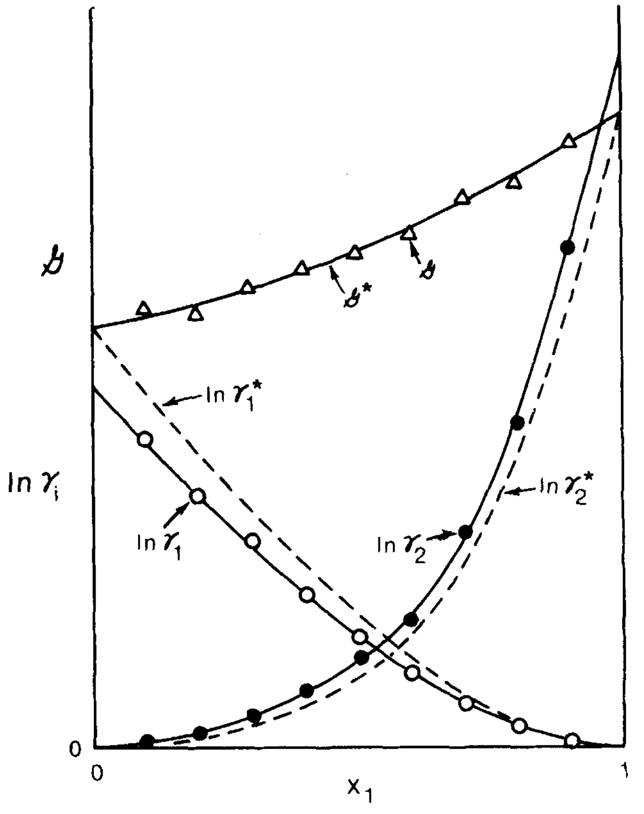

Fig. 1. Schematic diagram showing the relation between experimental and predicted values for inconsistent data.

(4)] to give the correlation $\mathcal{G}^*$ agree with the values of $\ln \gamma_1^*$ and $\ln \gamma_2^*$ obtained through differentiation of $\mathcal{G}^*$ [Equations (6*a*) and (6*b*)] if and only if the former values satisfy Equation (10). This test of thermodynamic consistency was the first of the modern comparison tests to be proposed (Van Ness, 1959).

The fundamental reason that consistency can be tested and that inconsistency can be demonstrated is that data points comprising x_1-y_1-P values contain more information than is necessary with respect to the equations that must be satisfied. Subsets of x_1-P data or x_1-y_1 data are entirely adequate to allow determination of sets of parameters α, β, . . . Since Equation (8) does not require y_1 values for the calculation of P^*, a residual $r = P^* - P$ provides the basis for a data reduction procedure with x_1-P data only, a method originally proposed by Barker (1953). Similarly, the ratio γ_1^*/γ_2^* of Equation (7) does not contain the pressure

$$\frac{\gamma_1^*}{\gamma_2^*} = \frac{y_1^* x_2 P_2^{\text{sat}}}{(1 - y_1^*) x_1 P_1^{\text{sat}}}$$

Thus, a residual $r = y_1^* - y_1$ is the basis for reduction of a x_1-y_1 data subset. Two sets of parameters α, β, . . . determined from x_1-P and x_1-y_1 data subsets cannot possibly be in agreement unless the full x_1-y_1-P data set is thermodynamically consistent. If one has available only a subset of data, then no question of consistency arises, because each contains alone exactly the information necessary for data reduction.

REGRESSION PROCEDURES

Best values for a set of parameters α, β, . . . are usually obtained through regression by a weighted least-squares technique which converges on values that minimize the sum

$$S = \sum_{j=1}^{n} \frac{r_j^2}{w_j} \tag{11}$$

where j is an index running over all data points. As already indicated, one has a choice of residuals r_j. Similarly, a choice may also be made of the weighting factors w_j, which are determined in some reasonable way. Many authors have considered weighting factors of unity for all data points to be reasonable. Recently, Fabries and Renon (1975) and Péneloux et al. (1976) have recommended weighting factors based on the principle of maximum likelihood as being more reasonable. We now consider in a systematic way various residuals and their maximum likelihood weighting factors for use in the objective function of Equation (11).

Equation (1) represents the fundamental relation which must be satisfied for vapor-liquid equilibrium in binary systems under the simplifying assumptions we have imposed. We rewrite it as

$$0 = \gamma_i - \frac{y_i P}{x_i P_i^{\text{sat}}} \quad (i = 1, 2)$$

Note that all symbols stand for experimental values. If we replace γ_i by γ_i^*, the difference between terms on the right is not, in general, zero because we must presume imperfect agreement between the model and the experimental data. Therefore, we write

$$F_i(x_i, y_i, P, T, \alpha, \beta, \ldots) = \gamma_i^* - \frac{y_i P}{x_i P_i^{\text{sat}}} = \gamma_i^* - \gamma_i \equiv r_i$$

We distinguish between the function on the left, which is considered continuous and differentiable, and the residuals on the right, which are discrete values corresponding to data points. The functional relationship is more simply stated as

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

The principle of maximum likelihood leads to the result that the weighting factors are the variances of the residuals, given by

$$w_j = \sigma_j^2 = \left(\frac{\partial F}{\partial x}\right)_j^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)_j^2 \sigma_y^2 + \left(\frac{\partial F}{\partial P}\right)_j^2 \sigma_P^2 + \left(\frac{\partial F}{\partial T}\right)_j^2 \sigma_T^2 \tag{13}$$

For simplicity, we omit the subscript i, which merely identifies a constituent. The quantities σ_x, σ_y, σ_P, and σ_T are estimates of the standard deviations in the measurements of x, y, P, and T. One assigns values to these quantities from an evaluation of the experimental data set. Different values could be assigned for each data point, but in practice assignments for the whole data set are more realistic. We therefore put no index j on these quantities. The partial derivatives are taken at constant values of α, β, . . . but are evaluated anew for each data point.

The principle of maximum likelihood as applied here is based on several fundamental assumptions:

1. Experimental errors are small.
2. Experimental errors in the measured variables are independent of one another, both for a given data point and from point to point.
3. Experimental errors display a Gaussian distribution with zero mean deviation from correct values.
4. The fitting equation (mathematical model) is inherently capable of representing the correct values with deviations of an order less than the order of the experimental errors.

These assumptions must be satisfied if one is to assert that the principle of maximum likelihood provides weighting factors for Equation (11) more reasonable, more rigorous, or statistically more meaningful than some other method.

Expressions for the derivatives in Equation (13) that follow from Equation (12) are obtained by partial differentiation:

$$\frac{\partial F_i}{\partial x_i} = \frac{d\gamma_i^*}{dx_i} + \frac{y_i P}{x_i^2 P_i^{\text{sat}}} = \gamma_i^* \frac{d \ln \gamma_i^*}{dx_i} + \frac{\gamma_i}{x_i}$$

Since errors are assumed small, γ_i^* and γ_i are for practical purposes the same and can here be used interchangeably. Therefore

$$\frac{\partial F_i}{\partial x_i} = \frac{\gamma_i^*}{x_i}\left(1 + x_i \frac{d \ln \gamma_i^*}{dx_i}\right)$$

For the first chemical species, $i = 1$, and

$$\frac{\partial F_1}{\partial x_1} = \frac{\gamma_1^*}{x_1}\left(1 + x_1 \frac{d \ln \gamma_1^*}{dx_1}\right)$$

When $i = 2$

$$\frac{\partial F_2}{\partial x_2} = \frac{\gamma_2^*}{x_2}\left(1 + x_2 \frac{d \ln \gamma_2^*}{dx_2}\right) = \frac{\gamma_2^*}{x_2}\left(1 - x_2 \frac{d \ln \gamma_2^*}{dx_1}\right)$$

As a result of Equation (10), this becomes

$$\frac{\partial F_2}{\partial x_2} = \frac{\gamma_2^*}{x_2}\left(1 + x_1 \frac{d \ln \gamma_1^*}{dx_1}\right)$$

Letting

$$\Gamma \equiv 1 + x_1 \frac{d \ln \gamma_1^*}{dx_1}$$

Table 1. Interchangeable Functions and Residuals for Maximum Likelihood

Case	Function, F_i	Residual, r_i
1	$\gamma_i^* - \dfrac{y_iP}{x_iP_i^{\text{sat}}}$	$\gamma_i^* - \gamma_i$
2	$\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{P} - y_i$	$y_i^* - y_i = y_i\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
3	$\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{y_i} - P$	$P^* - P = P\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
4	$x_iP_i^{\text{sat}}\,\gamma_i^* - y_iP$	$p_i^* - p_i = p_i\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
5	$\dfrac{y_iP}{P_i^{\text{sat}}\,\gamma_i^*} - x_i$	$x_i^* - x_i = -x_i\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
6	$\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{y_iP} - 1$	$\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
7†	$\ln\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{y_iP}$	$\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
8†	$x_i\ln\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{y_iP}$	$x_i\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$
9†	$\dfrac{1}{(1-x_i)}\ln\dfrac{x_iP_i^{\text{sat}}\,\gamma_i^*}{y_iP}$	$\dfrac{1}{(1-x_i)}\dfrac{(\gamma_i^* - \gamma_i)}{\gamma_i}$

† The expression for r_i follows from the approximation

$$\ln\left(1 + \frac{\gamma_i^* - \gamma_i}{\gamma_i}\right) \simeq \frac{\gamma_i^* - \gamma_i}{\gamma_i}$$

we have

$$\partial F_i/\partial x_i = \gamma_i^*\Gamma/x_i$$

Similarly

$$\partial F_i/\partial y_i = -\gamma_i^*/y_i$$

$$\partial F_i/\partial P = -\gamma_i^*/P$$

$$\partial F_i/\partial T = \gamma_i^*\pi_i$$

where

$$\pi_i \equiv \frac{d\ln P_i^{\text{sat}}}{dT}$$

For a particular data point j, Equation (13) becomes

$$\sigma_i^2 = (\gamma_i^*\Gamma/x_i)^2\sigma_x^2 + (\gamma_i^*/y_i)^2\sigma_y^2 + (\gamma_i^*/P)^2\sigma_P^2 + (\gamma_i^*\pi_i)^2\sigma_T^2 \quad (i = 1, 2)$$

and each term of the sum given by Equation (11) becomes

$$\left(\frac{r_i}{\sigma_i}\right)^2 = \frac{[(\gamma_i^* - \gamma_i)/\gamma_i^*]^2}{(\Gamma\sigma_x/x_i)^2 + (\sigma_y/y_i)^2 + (\sigma_P/P)^2 + (\sigma_T\pi_i)^2} \quad (i = 1, 2) \quad (14)$$

Since Equation (1) can be written in various equivalent ways, a number of residual functions alternative to that of Equation (12) may be written. Each represents a specific residual, and all are interrelated. This is indicated by Table 1, where case 1 shows the function given by Equation (12). Cases 6 and 7 are clearly equivalent, as are cases 5 and 8. If we go through the derivations for each case, exactly as was done for case 1, we find that Equation (14) is regenerated in every instance. All cases are therefore equivalent, and the two fundamental relations of equilibrium, Equation (1) with $i = 1$ and with $i = 2$, lead to just two objective functions for Equation (11), regardless of how the residual is expressed. This is a property of the maximum likelihood method.

In addition, however, we can treat the two equilibrium criteria jointly, combining the two objective functions ($i = 1$ and $i = 2$) and minimizing them simultaneously:

$$S = \sum_{j=1}^{n}\left[\left(\frac{r_1}{\sigma_1}\right)_j^2 + \left(\frac{r_2}{\sigma_2}\right)_j^2\right] \quad (15)$$

Alternatively, we can adopt the method of Fabries and Renon (1975), which takes account of the covariance of the residuals r_1 and r_2, and minimize

$$S = \sum_{j=1}^{n}\left[\frac{r_1^2\sigma_2^2 + r_2^2\sigma_1^2 - 2r_1r_2\sigma_{12}^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^4}\right] \quad (16)$$

In both cases, $r_1 = \gamma_1^* - \gamma_1$ and $r_2 = \gamma_2^* - \gamma_2$. The covariance σ_{12} is given by

$$\sigma_{12}^2 = \left(\frac{\partial r_1}{\partial x_1}\right)\left(\frac{\partial r_2}{\partial x_1}\right)\sigma_x^2 + \left(\frac{\partial r_1}{\partial y_1}\right)\left(\frac{\partial r_2}{\partial y_1}\right)\sigma_y^2 + \left(\frac{\partial r_1}{\partial P}\right)\left(\frac{\partial r_2}{\partial P}\right)\sigma_P^2 + \left(\frac{\partial r_1}{\partial T}\right)\left(\frac{\partial r_2}{\partial T}\right)\sigma_T^2 \quad (17)$$

If $\sigma_{12} = 0$, Equation (16) reduces to Equation (15). In these methods, the Gibbs-Duhem equation automatically enters as a constraint.

A still greater variety of objective functions results when we generate new functions F and residuals r by combining the F_i's and the r_i's; for example, by addition

Table 2. Composite Functions and Residuals for Maximum Likelihood. Case Numbers Correspond to Those of Table 1

Case	Function, F	Residual, r
1	$F_1 + F_2 = \gamma_1^* - \dfrac{y_1P}{x_1P_1^{\text{sat}}} + \gamma_2^* - \dfrac{y_2P}{x_2P_2^{\text{sat}}}$	$(\gamma_1^* - \gamma_1) + (\gamma_2^* - \gamma_2)$
1A	$F_1 - F_2 = \gamma_1^* - \dfrac{y_1P}{x_1P_1^{\text{sat}}} - \gamma_2^* + \dfrac{y_2P}{x_2P_2^{\text{sat}}}$	$(\gamma_1^* - \gamma_1) - (\gamma_2^* - \gamma_2)$
4	$F_1 + F_2 = x_1P_1^{\text{sat}}\,\gamma_1^* + x_2P_2^{\text{sat}}\,\gamma_2^* - P$	$p_1^* + p_2^* - P = P^* - P$
7	$F_1 - F_2 = \ln\dfrac{\gamma_1^*}{\gamma_2^*} - \ln\dfrac{x_2P_2^{\text{sat}}\,y_1}{x_1P_1^{\text{sat}}\,y_2}$	$\dfrac{(\gamma_1^* - \gamma_1)}{\gamma_1} - \dfrac{(\gamma_2^* - \gamma_2)}{\gamma_2}$
8	$F_1 + F_2 = x_1\ln\dfrac{x_1P_1^{\text{sat}}\,\gamma_1^*}{y_1P} + x_2\ln\dfrac{x_2P_2^{\text{sat}}\,\gamma_2^*}{y_2P}$	$\left(\dfrac{G^E}{RT}\right)^* - \dfrac{G^E}{RT}$
9	$F_1 + F_2 = \dfrac{1}{x_2}\ln\dfrac{x_1P_1^{\text{sat}}\,\gamma_1^*}{y_1P} + \dfrac{1}{x_1}\ln\dfrac{x_2P_2^{\text{sat}}\,\gamma_2^*}{y_2P}$	$\mathcal{G}^* - \mathcal{G}$

$$F = F_1 + F_2$$

$$r = r_1 + r_2$$

Those of greatest apparent interest are shown by Table 2. In each case, an appropriate variance σ^2 is determined for each function F for use as the weighting factor w with the corresponding residual r to form an objective function for minimization in accord with Equation (11). For example, the derivatives required in Equation (13) for the function of case 4 (Table 2) are

$$\partial F/\partial x_1 = (P_1^{\text{sat}}\gamma_1^* - P_2^{\text{sat}}\gamma_2^*)\Gamma$$

$$\partial F/\partial y_1 = 0$$

$$\partial F/\partial P = -1$$

$$\partial F/\partial T = x_1 P_1^{\text{sat}}\gamma_1^*\pi_1 + x_2 P_2^{\text{sat}}\gamma_2^*\pi_2$$

From these results and from the function itself, we see that the y_1 values of the data do not enter into the data reduction process. Only the x_1-P data subset is required in this case, which is the maximum likelihood version of Barker's method.

Similarly, application of case 7 does not require P values, and the x_1-y_1 data subset suffices for data reduction. In this case, the derivatives are

$$\partial F/\partial x_1 = \Gamma/x_1 x_2$$

$$\partial F/\partial y_1 = -1/y_1 y_2$$

$$\partial F/\partial P = 0$$

$$\partial F/\partial T = \pi_1 - \pi_2$$

Case 8 is particularly interesting in relation to much of the work done in the past. Here, the necessary derivatives are

$$\partial F/\partial x_1 = 0$$

$$\partial F/\partial y_1 = (y_1 - x_1)/y_1 y_2$$

$$\partial F/\partial P = -1/P$$

$$\partial F/\partial T = x_1\pi_1 + x_2\pi_2$$

Surprisingly, this function is insensitive to errors in x_1. Since x_1 is the independent variable, values for x_1 are, of course, essential. Nevertheless, errors in x_1, provided they are small, do not propagate into F. This same result is also found for case 9, which in maximum likelihood is identical to case 8. Thus, when residuals in the thermodynamic functions G^E and $\mathcal{G}$ are the basis for reduction of VLE data, effective use is not made of measured values of x_1, and the quality of fit depends solely on the accuracy of the P and y_1 measurements. However, the independent variable x_1 is invariably measured, and experimental methods are designed to provide accurate values. Moreover, calculations show that these functions yield results inferior to the others; we therefore exclude them from further consideration.

In addition to the cases listed in Table 2, each of which leads to a distinct objective function, we can combine objective functions and minimize them simultaneously. Perhaps the most promising combination is of case 4, which depends on the x_1-P data subset, and case 7, which depends on the x_1-y_1 data subset. Then we minimize

$$S = \sum_{j=1}^{n}\left[\left(\frac{r}{\sigma}\right)_j^2 + \left(\frac{r'}{\sigma'}\right)_j^2\right] \tag{18}$$

where r and r' appear in Table 2 as cases 4 and 7. We note that the ratios r/σ and r'/σ' as provided by the method of maximum likelihood are automatically scaled so as to be suitable for direct combination.

RESULTS OF DATA REDUCTION

A diversity of objective functions is clearly available within the context of the maximum likelihood method. They lead to diverse results, as we will illustrate. However, we must be certain that the data set conforms to the assumptions which condition proper application of the method. To this end, we generate by Equations (6), (8), and (9) an artificial set of perfect data points, representing approximately the data of Fried et al. (1967) for pyridine-tetrachloroethylene at 60°C. The mathematical model is a three-parameter Margules equation:

$$\mathcal{G}^* = A_{21}x_1 + A_{12}x_2 - Dx_1x_2 \tag{19}$$

with

$$A_{21} = 0.83379 \quad A_{12} = 0.87395 \quad D = 0.26813$$

The assigned vapor pressures are

$$P_1^{\text{sat}} = 110.95 \text{ mm Hg} \quad P_2^{\text{sat}} = 94.88 \text{ mm Hg}$$

No corrections are made for gas-phase imperfections or for the Poynting effect, either here or subsequently. Data points generated at 99 values of x_1 are distributed uniformly in the interval $0 < x_1 < 1.0$. These together with the given P_i^{sat} values represent a perfect set of data, perfectly correlated by Equation (19).

To the y_1 and P values of this perfect set we add errors picked randomly from two infinite sets of random errors having a Gaussian distribution with zero mean value and with standard deviations of $\sigma_y = 0.003$ and $\sigma_P = 0.3$ mm Hg. This collection of 99 x_1-y_1-P data points represents the experimental data base. We could also introduce error into the x_1 values, but this seems pointless and merely gives a set of slightly different values to the independent variable. The data base as generated certainly satisfies the assumptions upon which the maximum likelihood method is predicated. The errors are small, independent, and normally distributed; the fitting equation is inherently capable of representing the correct values without error. It is not possible with a set of real data to know that these conditions are fulfilled.

With this data base as a source, we study the effectiveness of various objective functions in data reduction by the maximum likelihood method. The steps followed are:

1. Select ten data points at random from the data base. These, with the perfect P_i^{sat} values, form a data set.
2. Reduce this data set to determine A_{21}, A_{12}, and D in Equation (19), with nine different objective functions.
3. Calculate y_1^* and P^* for each x_1 by Equations (8) and (9) for comparison with the perfect values. This comparison yields values for RMS Δy_1 and RMS ΔP for each objective function.
4. Repeat this process twenty times. From the complete set of results, find for each objective function mean values of the parameters $\overline{A_{21}}$, $\overline{A_{12}}$, and $\overline{D}$ and their standard deviations $\sigma(A_{21})$, $\sigma(A_{12})$, $\sigma(D)$, and mean values of the RMS errors $\overline{\text{RMS } \Delta y_1}$ and $\overline{\text{RMS } \Delta P}$ and their standard deviations $\sigma(\text{RMS } \Delta y_1)$ and $\sigma(\text{RMS } \Delta P)$.

The nine objective functions of this study are indexed in Table 3. We also show in this table the relation of each residual to the errors in P and y_1. The development follows from the definition

$$r_i = \gamma_i^* - \gamma_i = \frac{y_i^*P^*}{x_iP_i^{\text{sat}}} - \frac{y_iP}{x_iP_i^{\text{sat}}} = \frac{(y_i^*P^* - y_iP)}{x_iP_i^{\text{sat}}}$$

The quantity in parentheses in the last term can be represented as the difference $\delta(y_iP)$. Since this difference

TABLE 3. CASES TREATED IN NUMERICAL STUDIES

Index	Objective function given by Equation	Residual	Notes
(a)	(11)†	$r_1 = \gamma_1^* - \gamma_1 = \gamma_1 \left(\frac{\delta P}{P} + \frac{\delta y_1}{y_1} \right)$	Case 1, Table 1
(b)	(11)†	$r_2 = \gamma_2^* - \gamma_2 = \gamma_2 \left(\frac{\delta P}{P} - \frac{\delta y_1}{y_2} \right)$	Case 1, Table 1
(c)	(11)†	$r = r_1 + r_2 = (\gamma_1 + \gamma_2) \frac{\delta P}{P} + \left(\gamma_1 - \frac{y_1}{y_2} \gamma_2 \right) \frac{\delta y_1}{y_1}$	Case 1, Table 2
(d)	(11)†	$r = r_1 - r_2 = (\gamma_1 - \gamma_2) \frac{\delta P}{P} + \left(\gamma_1 + \frac{y_1}{y_2} \gamma_2 \right) \frac{\delta y_1}{y_1}$	Case 1A, Table 2
(e)	(15)	$\left.\begin{array}{l} r_1 = \gamma_1^* - \gamma_1 \\ r_2 = \gamma_2^* - \gamma_2 \end{array}\right\}$ minimized simultaneously	(a) and (b) combined
(f)	(16)	$\left.\begin{array}{l} r_1 = \gamma_1^* - \gamma_1 \\ r_2 = \gamma_2^* - \gamma_2 \end{array}\right\}$ minimized simultaneously	Method of Fabries and Renon
(g)	(11)†	$r = \delta P$	Case 4, Table 2 Barker's method
(h)	(11)†	$r = \frac{\gamma_1^*}{\gamma_1} - \frac{\gamma_2^*}{\gamma_2} = \delta y_1 / y_1 y_2$	Case 7, Table 2
(i)	(18)	$\left.\begin{array}{l} r = \delta P \\ r' = \delta y_1 / y_1 y_2 \end{array}\right\}$ minimized simultaneously	(g) and (h) combined

† With $w = \sigma^2$.

is presumed small, it can be expanded to give as a good approximation

$$(y_i^* P^* - y_i P) = \delta(y_i P) = y_i \delta P + P \delta y_i$$

Therefore

$$r_i = \gamma_i^* - \gamma_i = \frac{y_i}{x_i P_i^{\text{sat}}} \delta P + \frac{P}{x_i P_i^{\text{sat}}} \delta y_i$$

Since

$$\gamma_i = \frac{y_i P}{x_i P_i^{\text{sat}}}$$

we get

$$r_i = \gamma_i^* - \gamma_i = \frac{\gamma_i}{P} \delta P + \frac{\gamma_i}{y_i} \delta y_i$$

or, finally

$$r_i = \gamma_i^* - \gamma_i = \gamma_i \left(\frac{\delta P}{P} + \frac{\delta y_i}{y_i} \right)$$

The other residuals of Table 3 follow from this and the fact that $\delta y_2 = -\delta y_1$. If the predicted values P^* and y_1^* approximate the true values, then δP and δy_1 approximate the errors in P and y_1. This should certainly be the case when the assumptions which underlie the principle of maximum likelihood are fulfilled.

Although only errors in y_1 and P were introduced into the data base, *this is not evident from the data themselves*. Therefore, in determining values for the variance of residuals, we assume that all terms of Equation (17) contribute, and we assign the following values to the estimates of the standard deviations in the measured variables:

$$\sigma_x = 0.003$$

$$\sigma_y = 0.003$$

$$\sigma_P = 0.3 \text{ mm Hg}$$

$$\sigma_T = 0.1°\text{K}$$

Values for $\pi_i \equiv d \ln P_i^{\text{sat}}/dT$ are constants evaluated from the Antoine equations for pyridine and tetrachloroethylene, and both are approximately $-0.04°\text{K}^{-1}$ at 60°C.

Results for data reductions based on ten data points are given in Table 4. With respect to y_1, each procedure leads to a fit well within the standard deviation of the data base. This is not true for each case with respect to P. Moreover, clearly superior results are provided by the objective functions indexed as (*e*), (*f*), (*g*), and (*i*). For these four, results are reported in Table 5 of a further study in which twenty randomly selected points form a data set. Again, twenty different data sets are reduced in each case. In addition, results of reduction of the entire 99-point data base are included in Table 5. All results for objective functions indexed (*e*), (*f*), (*g*), and (*i*) are displayed graphically in Figure 2, where the points represent the ΔP's and Δy's given in Tables 4 and 5, and the upper and lower limits show the corresponding values of $\pm\sigma$.

One wants small bias and small standard deviations of the fitting parameters from their true values. A compromise statistical parameter which incorporates both requirements is the mean-square error (Jenkins and Watts, 1968):

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^{n} \frac{[\sigma(B_j)]^2 + (\overline{B}_j - B_j^*)^2}{(B_j^*)^2} \qquad (20)$$

Values for this quantity are included in Tables 4 and 5.

DISCUSSION

The results clearly show that for the data base considered, parameters found by Barker's method are as successful in reproducing the perfect data set as are those found by any other maximum likelihood method. Indeed, for ten data points, representative of many real data sets, Barker's method is slightly superior to any other, as is evident from the MSE and RMS values given in Table 4. Since Barker's method makes use of just the x_1-P data subset, this conclusion tends to refute the often-stated opinion that measured values of y_1 provide valuable added information, which should be incorporated in the VLE data reduction scheme to im-

TABLE 4. RESULTS OF DATA REDUCTION WITH TEN DATA POINTS. THE IDENTIFYING INDEX REFERS TO TABLE 3. Δy_1 AND ΔP DETERMINED IN RELATION TO THE PERFECT DATA SET

Index	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
$\overline{A}_{21}$ (0.834)*	0.838	0.841	0.844	0.861	0.824	0.823	0.832	0.866	0.823
$\sigma(A_{21})$	0.019	0.025	0.026	0.031	0.020	0.013	0.013	0.026	0.013
$\overline{A}_{12}$ (0.874)*	0.885	0.868	0.888	0.898	0.860	0.863	0.870	0.903	0.856
$\sigma(A_{12})$	0.022	0.027	0.035	0.031	0.015	0.015	0.014	0.035	0.012
$\overline{D}$ (0.268)*	0.278	0.257	0.306	0.349	0.211	0.225	0.257	0.350	0.205
$\sigma(D)$	0.072	0.082	0.107	0.091	0.062	0.050	0.062	0.089	0.049
MSE	0.025	0.032	0.061	0.070	0.033	0.020	0.019	0.070	0.030
$\overline{\text{RMS }\Delta y_1}$	0.0012	0.0013	0.0014	0.0016	0.0011	0.0009	0.0008	0.0016	0.0010
$\sigma(\text{RMS }\Delta y_1)$	0.0004	0.0005	0.0007	0.0006	0.0006	0.0005	0.0004	0.0007	0.0004
$\overline{\text{RMS }\Delta P}$, mm Hg	0.20	0.24	0.20	0.32	0.13	0.10	0.09	0.38	0.11
$\sigma(\text{RMS }\Delta P)$, mm Hg	0.10	0.09	0.09	0.16	0.06	0.04	0.03	0.26	0.03

* True values.

TABLE 5. RESULTS OF DATA REDUCTION WITH TWENTY AND WITH 99 DATA POINTS. THE IDENTIFYING INDEX REFERS TO TABLE 3. Δy_1 AND ΔP DETERMINED IN RELATION TO THE PERFECT DATA SET

Index	(e)	(f)	(g)	(i)	
Twenty data points					
$\overline{A}_{21}$ (0.834)*	0.826	0.822	0.827	0.826	
$\sigma(A_{21})$	0.012	0.008	0.013	0.010	
$\overline{A}_{12}$ (0.874)*	0.861	0.857	0.866	0.863	
$\sigma(A_{12})$	0.014	0.011	0.012	0.013	
$\overline{D}$ (0.268)*	0.227	0.212	0.236	0.229	
$\sigma(D)$	0.051	0.039	0.053	0.045	
MSE	0.020	0.022	0.018	0.017	
$\overline{\text{RMS }\Delta y_1}$	0.0007	0.0007	0.0007	0.0006	
$\sigma(\text{RMS }\Delta y_1)$	0.0004	0.0004	0.0003	0.0004	
$\overline{\text{RMS }\Delta P}$, mm Hg	0.08	0.09	0.07	0.07	
$\sigma(\text{RMS }\Delta P)$, mm Hg	0.04	0.04	0.03	0.04	
99 data points	(e)	(f)	(g)	(i)	(h)
A_{21} (0.834*)	0.829	0.828	0.825	0.826	0.856
A_{12} (0.874)*	0.865	0.865	0.865	0.865	0.892
D (0.268)*	0.237	0.236	0.226	0.231	0.312
RMS Δy_1	0.0004	0.0004	0.0005	0.0004	0.0007
RMS ΔP, mm Hg	0.03	0.03	0.04	0.04	0.26

* True values.

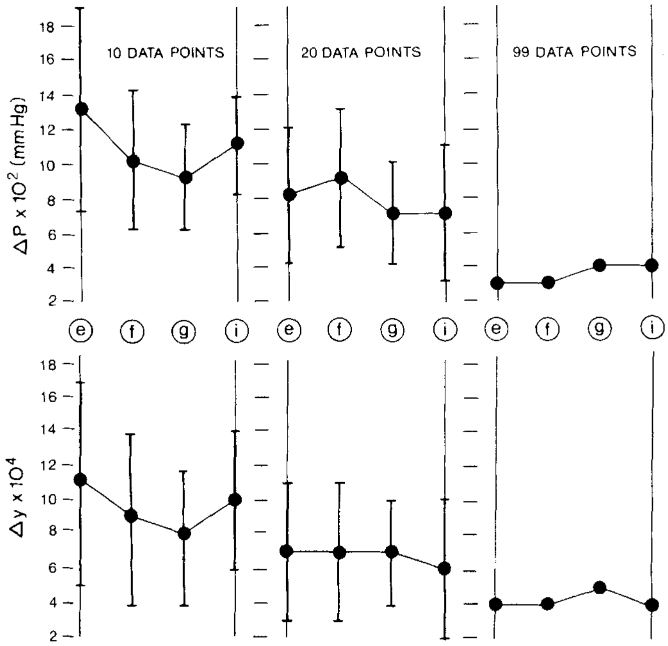

Fig. 2. Results of data reduction for methods *(e)*, *(f)*, *(g)*, and *(i)* (see Table 3). ΔP and Δy refer to the appropriate entries of RMS values in Tables 4 and 5. Upper and lower limits represent corresponding values of $\pm\sigma$.

prove the correlation. Actually, the only value of such information is for consistency testing. Indeed, consistency lies at the heart of the matter.

The 99-point data set of this study, constructed by the introduction of normally distributed random errors, is not consistent. None of the data reduction methods with all 99 points and the perfect P_i^{sat} values regenerates the original parameters from which the perfect data set is constructed, and no two methods lead to the same set of parameters. On the other hand, reduction of the perfect set of 99 points does regenerate the original parameters, regardless of the method. Thus, perturbation of a perfect data set with random errors such that the overall distribution of errors is Gaussian does not produce a consistent set of data. Consistency requires that the local distribution be Gaussian at every value of x_1. Such a data set would be consistent, and all maximum likelihood methods of data reduction would converge on the perfect values of the fitting parameters. Construction of such a data base is impractical; neither would it be representative of real data. All this is not to say that the data base as generated is badly inconsistent; on the contrary, the 99-point set would satisfy any reasonable consistency requirements that one might impose on real data. Moreover, all maximum likelihood methods do regenerate the perfect data set to within reasonable tolerances. Nevertheless, residual imperfections do taint the data base, making it less than perfectly consistent. No real set of data is likely to be better. These imperfections in the data underlie the diversity of results produced by different fitting methods.

When a set of x_1-y_1-P data is reduced, the effect of data imperfections enters in two ways, depending on whether the method is based on a single equation of equilibrium [Equation (1) with $i = 1$ or $i = 2$] or on both equations [$i = 1$ and $i = 2$]. In the first circumstance, represented by methods (a) and (b) of Table 3, data reduction is based on only half of the thermodynamic information available and is inherently deficient, leading to correlations of imperfect data that are not the best possible. The Gibbs-Duhem equation is not implicit in these methods. In the second circumstance, represented by methods (c), (d), (e), (f), and (i) of Table 3, the Gibbs-Duhem equation is implicit and imposes a constraint on the results. Neither δP nor δy_1 is minimized; the residual is some constrained function of these errors. With imperfect data, this constraint may act to distort the results by magnifying both δP and δy_1 in a compensating way, leading again to correlations that are not the best possible.

Finally, we have methods (g) and (h) of Table 3. Each is based on reduction of a data subset: x_1-P data for Barker's method (g) and x_1-y_1 data for method (h). Both methods make full use of applicable thermodynamic information, but neither is constrained by the Gibbs-Duhem equation. Barker's method (g) results in direct minimization of δP, thus yielding the best possible fit of the pressures in comparison with the actual data set, though not always quite the best in comparison with the perfect data set. Method (h) provides direct minimization of δy_1, yielding the best possible fit of the y_1's in comparison with the actual data set, but always the worst in comparison with the perfect data set.

Consistent with previous experience (Van Ness et al., 1973), this study demonstrates the superiority of fitting to P rather than to y_1. The nonlinear nature of the equations is such that imperfections in y_1 propagate into exaggerated errors in P, whereas imperfections in P propagate into minimal errors in y_1. This is seen in the results of Table 4. Whereas Barker's method (g) yields not only the smallest average value of ΔP but also of Δy_1 in comparison with the perfect data, method (h), based on x_1-y_1 data alone, gives the largest average RMS value of ΔP. Even with the complete set of 99 data points, method (h) still yields poor results, as is seen from Table 5. The net effect of imperfections in the y_1 values is severe distortion of the fitting equation and poor reproduction of pressure values. No method that combines the P data with the y_1 data succeeds in overcoming this negative influence of the y_1 imperfections to the point of improving significantly on Barker's method, wherein the y_1 data are not taken into account.

Parallel calculations by unweighted least-squares techniques ($w = 1$) were made corresponding to all maximum likelihood calculations reported here. For Barker's method (g), maximum likelihood and unweighted least squares produce identical results in every instance, because the maximum likelihood σ's show little variation from point to point, producing about equal weighting of all points. For all other methods, regardless of the number of data points, the maximum likelihood procedure produces superior results, decidedly so in all cases except for method (i), where the advantage is slight.

One cannot be sure that conclusions reached as a result of this study are completely general. They are, however, supported by the results of calculations done with a second 99 point set of artificial data, representative of the data of Kretschmer and Wiebe (1949) for ethanol-toluene at 55°C. For this highly nonideal system, Barker's method proved much superior to all others for both the maximum likelihood and unweighted least-squares procedures, and the two procedures gave the same results.

The primary conclusion reached as a result of this study is that Barker's method, either in the maximum likelihood or unweighted least-squares version, is at least as good as any other method for the reduction of VLE data taken in such a manner as to avoid systematic error. Even then, the data inevitably contain imperfections, the adverse influence of which cannot be overcome by any degree of statistical sophistication supposedly designed to improve on Barker's method through treatment of the complete x_1-y_1-P data set. Data which contain systematic error, detectable as unacceptable inconsistency, cannot be treated correctly by any procedure based on the assumption of random distribution of errors centered around correct values of the measured variables. In this situation, no method of data reduction which combines y_1 and P data is acceptable. If one cannot regard the P's as reasonably accurate, the data set should be disregarded. Consistency testing is best done by the comparison of y_1 values calculated by Barker's method with experimental values (Abbott and Van Ness, 1975). Data of acceptable consistency can be adequately reduced by methods other than Barker's, namely, by methods (e), (f), and (i) of Table 3, and here the maximum likelihood procedures should certainly be used. We cannot recommend other methods and specifically urge that routine methods based on minimization of residuals in G^E or G^E/x_1x_2 be abandoned.

ACKNOWLEDGMENT

The contribution to this work by H. C. Van Ness was largely made while he was a visiting professor at Instituttet for Kemiteknik, Danmarks tekniske Højskole. He is grateful for having had the opportunity to use and enjoy the splendid facilities of this university and for the help and many kindnesses of the staff.

NOTATION

A_{12}, A_{21}, D = parameters in Equation (19); represented in Equation (20) by B_j
$\overline{B}_j$ = mean (estimated) value of parameter B_j
B_j^* = true value of B_j
F = function; see Equation (12)
G^E = excess Gibbs function for liquid phase
P = pressure
P_i^{sat} = saturation vapor pressure of pure i
R = universal gas constant
S = sum of squares
T = absolute temperature
n = number of parameters B_j; here three
p_i = partial pressure of species $i (= y_i P)$
r = residual or difference between predicted and experimental values
w = weighting factor in Equation (11)
x_i = mole fraction of species i in liquid phase
y_i = mole fraction of species i in vapor phase
$\mathcal{G}$ = G^E/x_1x_2RT

Greek Letters

Γ = $1 + x_1(d \ln \gamma_1^*/dx_1)$
α, β = parameters, in general
γ_i = activity coefficient of species i in liquid phase
π_i = $d \ln P_i^{sat}/dT$
σ = a standard deviation
ΔP = difference between a computed pressure and the perfect value
Δy_1 = difference between a computed y_1 and the "perfect" value
δP = approximate error in experimental value of P
δy_1 = approximate error in experimental value of y_1

Superscripts

$*$ = predicted value

LITERATURE CITED

Abbott, M. M., and H. C. Van Ness, "Vapor-Liquid Equilibrium: Part III. Data Reduction with Precise Expressions for G^E," *AIChE J.*, 21, 62 (1975).

Barker, J. A., "Determination of Activity Coefficients from Total Pressure Measurements," *Austral. J. Chem.*, **6**, 207 (1953).

Fabries, J. F., and H. Renon, "Method of Evaluation and Reduction of Vapor-Liquid Equilibrium Data of Binary Mixtures," *AIChE J.*, **21**, 735 (1975).

Fried, V., P. Gallant, and G. B. Schneier, "Vapor-Liquid Equilibrium in the System Pyridine-Tetrachloroethylene," *J. Chem. Eng. Data*, **12**, 504 (1967).

Jenkins, G. M., and D. G. Watts, *Spectral Analysis and Its Applications*, pp. 98-99, Holden-Day, San Francisco, Calif. (1968).

Kretschmer, C. B., and R. Wiebe, "Liquid-Vapor Equilibrium of Ethanol-Toluene Solutions," *J. Am. Chem. Soc.*, **71**, 1793 (1949).

Péneloux, A., R. Deyrieux, E. Canals, and E. Neau, "The Maximum Likelihood Test and the Estimation of Experimental Inaccuracies. Application to Data Reduction for Liquid-Vapor Equilibrium," *J. Chim. Phys.*, **73**, 706 (1976).

Van Ness, H. C., "Precise Testing of Binary Vapour-Liquid Equilibrium Data by the Gibbs-Duhem Equation," *Chem. Eng. Sci.*, **11**, 118 (1959).

———, S. M. Byer, and R. E. Gibbs, "Vapor-Liquid Equilibrium: Part I. An Appraisal of Data-Reduction Methods," *AIChE J.*, **19**, 238 (1973).

Manuscript received October 17, 1977; revision received June 29, and accepted August 2, 1978.

Bubble Motion and Mass Transfer in Non-Newtonian Fluids:

S. M. BHAVARAJU
R. A. MASHELKAR
and
H. W. BLANCH

Department of Chemical Engineering
University of Delaware
Newark, Delaware 19711

Part I. Single Bubble in Power Law and Bingham Fluids

The Sherwood number and drag coefficient for a single gas bubble moving in a power law fluid and a Bingham plastic fluid are obtained using perturbation methods. The perturbation parameters for power law and Bingham plastic fluids are m $(= n - 1/2)$ and E $(= \tau_0 R/U\mu_0)$, respectively. It is found that in the case of power law fluid, mass transfer and drag increase with increasing pseudoplasticity. These theoretical results are found to be in good agreement with the available experimental data and the data obtained in the present study. In the case of Bingham plastic fluid, mass transfer and drag are found to increase with increase in the Bingham number N_B $(= 2\epsilon)$. Contours of plug flow regions, where local stresses are less than the yield stress, are obtained as a function of the Bingham number N_B. These results qualitatively predict the zero terminal velocity observed for bubble motion in liquids with very high yield stress. They are also in good agreement with the trends of the results obtained previously for solid sphere motion in Bingham plastic fluids.

SCOPE

There are numerous examples of bubble motion and mass transfer in non-Newtonian fluids in antibiotic fermentations, biological wastewater treatment, polymer processes, and food processing (Blanch and Bhavaraju, 1976; Astarita and Mashelkar, 1977). In order to understand these complex process situations, a first approach is presented in this study of single bubble motion and mass transfer in non-Newtonian fluids. The case of multiple bubbles is treated in a following paper. In general, polymer melts, fermentation broths, and some waste treatment processes exhibit pseudoplastic or Bingham plastic behavior. Hence, the effects of pseudoplasticity and yield stress are investigated in the present study.

R. A. Mashelkar is with the National Chemical Laboratory, Poona, India 411008.

H. W. Blanch is presently at The University of California, Berkeley, California 94720.

0001-1541-78-1734-1063-$00.95.