A GENERALIZED DATA REGRESSION SYSTEM
BASED ON THE MAXIMUM LIKELIHOOD ESTIMATION

by

SUPHAT WATANASIRI

S.B., MASSACHUSETTS INSTITUTE OF TECHNOLOGY
(1979)

Submitted in partial fulfillment
of the requirements for the
degree of
MASTER OF SCIENCE
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February, 1980

Signature redacted

Signature of Author

Department of Chemical Engineering

Signature redacted

Certified by

Thesis supervisor

Accepted by

Chairman, Department Committee

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 13 1980
LIBRARIES
A GENERALIZED DATA REGRESSION SYSTEM
BASED ON THE MAXIMUM LIKELIHOOD ESTIMATION

by

Suphat Watanasiri

Submitted to the Department of Chemical Engineering
on January 18, 1980, in partial fulfillment of the requirements
for the Degree of Master of Science

ABSTRACT

A generalized data regression system was developed for
the ASPEN Project (Advanced System for Precess EEngineers). The
system is an integral part of the ASPEN architecture in the
sense that it utilizes, to the greatest possible extent, the
input language, utility routines and physical property
calculation capability of ASPEN. The Britt-Luecke's
Generalized Least-Squares method based on the maximum
likelihood principle was chosen as the regression algorithm.
This method is a very general formulation for algebraic models
which are nonlinear implicit in the unknown parameters.
Furthermore, the system is Table-driven. Information that may
be modified in the future are stored in tables rather than
being hard-wired into the system codes. This feature
simplifies future modification of the system since major change
of system codes is not necessary. The data regression system
developed is very flexible, and can handle a wide variety of
problem specifications. For example, it is possible to specify
a problem that uses different types of experimental data.
Several problems can also be performed in a single run. Both
linear and nonlinear regression problems can be treated by the
present system. The output from the system includes the best
estimate of the parameters and statistical information that is
useful in the analysis of the models as well as the
experimental data. The capability of the Data Regression
System was demonstrated via sample problems.
January 18, 1980

Professor George C. Newton
Secretary of the Faculty
Massachusetts Institute of Technology
Cambridge, MA 02139

Dear Professor Newton:

In accordance with the rules of the faculty, I herewith submit a thesis, entitled "A generalized Data Regression System Based on the Maximum Likelihood Estimation", in partial fulfillment of the requirement for the degree of Master of Science in Chemical Engineering at the Massachusetts Institute of Technology.

Respectfully submitted

Signature redacted

Suphat Watanasiri
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to the following persons who have generously contributed to the completion of this thesis

Professor Lawrence B. Evans for advising the work undertaken in this thesis.

Dr. Joseph F. Boston and Dr. Herbert I. Britt of the ASPEN project for their supervision, valuable suggestions, stimulating discussions and encouragements throughout this work.

Mr. Fred Ziegler and Mr. Williams Kessel for their assistance in countless technical problems.

Mr. C. C. Chen for his advice and assistance during the initial stage of the system development.

Finally, my parents who have always been so loving and caring.
To My Parents
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>11</td>
</tr>
<tr>
<td>2.</td>
<td>Parameter Estimation Methods</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.1 Least Squares Techniques for Parameter Estimation</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.2 Development of Generalized Least Squares Criterion</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2.3 Estimated Parameters Analysis</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>2.3.1 Variance of the Observables of Unit Weight</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>2.3.2 Variance-Covariance Matrix</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2.3.3 Residual Analysis</td>
<td>33</td>
</tr>
<tr>
<td>3.</td>
<td>The Data Regression System</td>
<td>35</td>
</tr>
<tr>
<td>4.</td>
<td>Input Language</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>4.1 ASPEN Input Language</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>4.1.1 Title and Description</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>4.1.2 Component Specification</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>4.1.3 Physical Property Calculation Options</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>4.1.4 Data Input</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>4.1.5 Units Specification</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>4.2 DRS Input Language</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>4.2.1 Parameter Specification</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>4.2.2 Data-Group Specification</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>4.2.3 Case Specification</td>
<td>62</td>
</tr>
<tr>
<td>5.</td>
<td>Parameter Estimation: Case Studies</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>5.1 Phase Equilibrium Calculation</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>5.2 VLE Calculation Using Activity Coefficient Model</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>5.2.1 Discussion</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>5.2.2 Application to Parameter Evaluation</td>
<td>73</td>
</tr>
</tbody>
</table>
# Table of Contents (continued)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3 VLE Calculation Using Equation of State</td>
<td>76</td>
</tr>
<tr>
<td>5.3.1 Discussion</td>
<td>76</td>
</tr>
<tr>
<td>5.3.2 Application to Parameter Evaluation</td>
<td>77</td>
</tr>
<tr>
<td>5.3.3 Equation of State Pure Parameter Estimation</td>
<td>78</td>
</tr>
<tr>
<td>5.4 Vapor Pressure Parameters Estimation</td>
<td>79</td>
</tr>
<tr>
<td>5.5 Ideal Gas Heat Capacity Parameters Estimation</td>
<td>81</td>
</tr>
<tr>
<td>6. Results and Discussions</td>
<td>82</td>
</tr>
<tr>
<td>6.1 Evaluation of Parameters for Wilson Activity Coefficient Model</td>
<td>82</td>
</tr>
<tr>
<td>6.2 Equation of State Binary-Interaction Parameters</td>
<td>84</td>
</tr>
<tr>
<td>6.3 Equation of State Unary parameters determination</td>
<td>85</td>
</tr>
<tr>
<td>6.4 Vapor Pressure Parameters Estimation</td>
<td>86</td>
</tr>
<tr>
<td>6.5 Ideal Gas Heat Capacity Parameters Estimation</td>
<td>86</td>
</tr>
<tr>
<td>7. Conclusions and Recommendations</td>
<td>109</td>
</tr>
</tbody>
</table>

NOTATION 111
REFERENCE 114
APPENDIX A 116
APPENDIX B 123
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Calculated phase diagram by Wilson equation</td>
<td>87</td>
</tr>
<tr>
<td>6.2</td>
<td>Confidence Ellipses for Wilson parameters.</td>
<td>88</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Data Regression Section of the System Definition File for ASPEN</td>
<td>44</td>
</tr>
<tr>
<td>4.1</td>
<td>Example of Input Language</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Unit Options</td>
<td>68</td>
</tr>
<tr>
<td>4.3</td>
<td>Data Regression System Problem Specification Options</td>
<td>69</td>
</tr>
<tr>
<td>4.4</td>
<td>Data Regression System Input Language</td>
<td>70</td>
</tr>
<tr>
<td>6.1</td>
<td>Wilson Parameter Estimation Input Data File</td>
<td>89</td>
</tr>
<tr>
<td>6.2</td>
<td>Wilson Parameter Estimation Results</td>
<td>91</td>
</tr>
<tr>
<td>6.3</td>
<td>Wilson Parameter Estimation Results</td>
<td>92</td>
</tr>
<tr>
<td>6.4</td>
<td>Wilson Parameter Estimation with TPx data</td>
<td>93</td>
</tr>
<tr>
<td>6.5</td>
<td>Wilson Parameter Estimation with TPx data</td>
<td>93</td>
</tr>
<tr>
<td>6.6</td>
<td>Redlich-Kwong-Soave Equation of State Binary Parameter Estimation Input Data</td>
<td>94</td>
</tr>
<tr>
<td>6.7</td>
<td>Redlich-Kwong-Soave Equation of State Binary Parameter Estimation Results</td>
<td>95</td>
</tr>
<tr>
<td>6.8</td>
<td>Peng-Robinson Equation of State Binary Parameter Estimation Input Data</td>
<td>96</td>
</tr>
<tr>
<td>6.9</td>
<td>Peng-Robinson Equation of State Binary Parameter Estimation Results</td>
<td>97</td>
</tr>
<tr>
<td>6.10</td>
<td>Redlich-Kwong-Soave Equation of State Acentric Factor Estimation Input Data</td>
<td>98</td>
</tr>
<tr>
<td>Table</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>6.11</td>
<td>Redlich-Kwong-Soave Equation of State</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Acentric Factor Estimation Results</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>Extended Antoine Parameters Estimation Input Data</td>
<td>101</td>
</tr>
<tr>
<td>6.13</td>
<td>Extended Antoine Parameters Estimation Results</td>
<td>103</td>
</tr>
<tr>
<td>6.14</td>
<td>Ideal Gas Heat Capacity Parameter Estimation Input Data</td>
<td>107</td>
</tr>
<tr>
<td>6.15</td>
<td>Ideal Gas Heat Capacity Parameter Estimation Results</td>
<td>108</td>
</tr>
</tbody>
</table>
CHAPTER 1 INTRODUCTION

Good physical property models have become increasingly more important in view of the trend in the industry, for economic reasons, to perform fewer experiments, to eliminate or reduce the size of pilot-plants, and to design better equipment. A large number of empirical and semi-empirical models have been used quite successfully to predict various thermodynamic properties such as vapor pressure, molar volume, activity coefficient, etc. Notably the phase equilibrium behavior of mixtures can be predicted accurately by many models. This capability is particularly useful in aiding the design of separation equipment. These models not only provide accurate predictions of thermodynamic properties but also provide means of summarizing large amounts of data, permitting interpolation and extrapolation of data beyond regions in which measurements have been made. The importance of these models in process calculations is very clear. However, these models contain adjustable parameters which must be determined by fitting to experimental data. The best estimates of these parameters are desirable if the data were to be best represented, and the errors in the property predicted by the model using the parameters were to be minimized. Unfortunately, most of the models are nonlinear in their adjustable parameters. Consequently, the determination of these parameters is not straightforward. However, it is
generally recognized that the best estimates of these parameters can be obtained from a formalized method which correctly treats the statistical behavior of the errors associated with all the experimental data that are used in the determination of the parameters.

It is usually assumed that there are two types of measurement errors: systematic and random. The former are due to inherent bias in the measurement procedure resulting in a consistent deviation of the observable from its true value. Usually, only the skilled and experienced experimenters can detect and avoid these type of errors. Random or statistical errors are assumed to be the result of a large number of small disturbances. Such errors have been found to be distributed according to simple laws that make it possible to treat them with statistical methods. The major consequence of these random errors is the corresponding presence of errors in the estimated parameters. Because of these errors in the data and because of inaccuracies in the model, it is not possible for a model to represent the experimental data exactly. However, a method of parameter estimation which correctly utilizes all the pertinent information available will give a best fit of the model to the data and minimize the parameter uncertainty. This includes taking into account the nature and magnitude of the random errors in all the observables. Such a method will give not only the parameters but also provide estimates of their uncertainties.
Many parameter estimation methods had been proposed. The oldest and most widely used method was the Least-Squares method first developed by Gauss. The Least-Squares criterion was stated as: "The best estimates of the parameters are those which minimize the weighted sum of the squared deviations of the observed values from their calculated values." Application of this method is generally restricted to the model which is linear in the unknown parameters. However, it can also be applied to the model that is nonlinear in its parameters. The Least-Squares method distinguishes between independent variables and dependent or response variables. A basic assumption of this method is that the values of independent variables are known exactly. This assumption ignores the fact that independent as well as dependent variables are subject to statistical errors. Therefore, this method does not utilize all available information in estimating the parameters.

Deming in 1943 formulated a general criterion for estimating model parameters, taking into account the errors in all measured variables, called the Generalized Least-Squares Criterion (GLSC). The exact solution satisfying his criteria was almost impossible at the time, due to the unavailability of high-speed digital computers. Therefore, he developed an algorithm which gives fairly accurate estimates of the desired parameters. Exact solutions satisfying his generalized criteria have since been proposed by a number of authors. Some
of these methods were concerned only with fitting data by straight lines or polynomials.\(^{17,23,29}\) The other method proposed by Fabris and Renon\(^ {14}\) is not general, in fact, it is equivalent to the Deming's method. The completely general formulation was proposed by Britt and Luecke\(^ {6}\), Bard\(^ {4}\), and Celmins.\(^ {9}\) Their algorithms do not distinguish between dependent and independent variables, thus taking all measurement errors into account. The formulation was based on the maximum likelihood principle. Their algorithm not only provides the best estimates for the parameters, but yields information useful in analysis of both the model and the data. Part of this additional information is obtained in the form of variance-covariance matrix of the estimated parameters. This matrix gives an estimate of the uncertainties of the parameters, which may be used to determine the significance of the parameters or to estimate the uncertainties in the properties calculated using the parameters. Additional information is obtained in the form of estimates of the true values of all measured variables. The deviations of the measured variables from the estimates of the true values, called the residuals, are useful for comparison of different models and detection of obviously bad data points.

The generalized least-squares algorithm formulated by Britt and Luecke provides a powerful and versatile criterion for the estimation of model parameters from experimental data. From a theoretical standpoint it is superior to the
conventional ordinary least-squares method because it utilizes all the measurement errors in estimating the model parameters, and also provides information useful in the critical evaluation of both the models and the data. Moreover, the parameters estimated by this method when applied to process calculation, was found to give better property prediction than those parameters estimated by the ordinary least-squares method,$^{(28)}$ especially when highly non-ideal mixtures are being studied.

This generalized parameter estimation algorithm is very desirable to any process evaluation system where critical evaluation of models and data are frequently performed. Currently at MIT, an advanced computer-aid process simulation system called ASPEN (Advanced System for Process ENgineers) is being developed. This system, when completed, will be used in the evaluation of fossil-fuel conversion processes. Evidently, a rigorous and flexible data regression system is required to perform numerous data reductions and model evaluations. The Britt-Luecke's algorithm was chosen to be implemented for ASPEN. This thesis describes the ASPEN data regression system.

The formulation of Britt-Luecke's generalized least-squares algorithm is discussed in Chapter 2. The Data Regression System developed is described in Chapter 3. The user-oriented input language for the system is described in Chapter 4. The capabilities of the DRS are demonstrated through sample problems which are illustrated in Chapter 5 and Chapter 6.
2.1 Least Squares Techniques for Parameter Estimation

The method of least squares is the oldest and most widely used method of parameter estimation. It was first developed by Gauss\(^{(15)}\) in 1809. The least squares criterion is stated as: "the best estimates of the parameters are those which minimize the weighted sum of the squared deviations of the observed values from their calculated values." In the traditional application of this criterion, it is assumed that only dependent variables are subject to observation, or measurement error. This partial application of Gauss' principle is called the Ordinary Least Squares Criterion (OLSC). OLSC is well established and broadly applied for treating data when the model employed is linear in the unknown parameters\(^{(11,13)}\). The same criterion is also utilized, in principle, when the model is non-linear in its parameters\(^{(10)}\).

The mathematical representation of OLSC when there is only one response variable is

\[
\min_{\theta} \sum_{j=1}^{k} \left[ \frac{y_{Mj} - f\left(x_{Mj}, \theta\right)}{\sigma_{yj}} \right]^2
\]

\(2-1\)
where \( k \) is the number of measurements; \( y_M \) is the response or the dependent variable; \( f \) is the predictor function; \( x_M \) is the vector of the predictor variables or the independent variables; \( \sigma_y \) is the standard deviation of the error in \( y \) and \( \theta \) is the set of unknown parameters.

It is important to note that this formulation requires that a distinction be made between response and predictor variables, and that a function that predicts values of the response from values of the predictor variables and parameters be furnished. The distinction between response and predictor variables is not always inherent and is often arbitrary. Thus the analyst is forced to think in terms of a predictor function for the dependent variable which has to be expressed in terms of independent variables. Furthermore, as implied in the definition of OLSC, the independent variables are presumed to be known exactly, which is almost never absolutely true in reality. Evidently, the significance of this assumption should not be just of theoretical concern; it may cause significant error in the parameter estimates.

In 1943, Deming\(^{(12)}\) considered in detail the more general problem of fitting parameters to data while accounting for the error in all the measured variables. This full application of Gauss' principle is called the Generalized Least Squares Criterion (GLSC).

Mathematically, in a very general formulation, GLSC may be stated as follows:
\[
\begin{align*}
\min_{\theta, z} & \quad \sum_{j=1}^{k} \sum_{i=1}^{q} \frac{(z_{Mij} - z_{ij})^2}{\sigma_{zij}^2} \\
\text{subject to the constraints} & \\
& \quad g_j(z_j, \theta) = 0 \quad j = 1, \ldots, k \quad (2-3) \\
& \quad \text{and} \quad \text{lb}_i \leq \theta_i \leq \text{ub}_i \quad i = 1, \ldots, n \quad (2-4)
\end{align*}
\]

where \( q \) is the number of measured variables; \( \sigma_{zij}^2 \) is the variance of the error in \( z_{Mij} \); and \( g_j \) is the vector of the constraint functions. The value of \( z_{ij} \) is the estimate of the true value of the constraint variable whose observed value \( z_{Mij} \) is assumed to be subject to error. Eq. (2-4) represents inequality constraints for the parameters. Each parameter is bound with an upper bound \( \text{ub} \), and a lower bound \( \text{lb} \). Note that the major difference between this formulation and the OLSC is the use of constraints rather than response functions.

Deming developed an algorithm for this general formulation of the GLSC. He was primarily concerned with making a one-step adjustment to fairly accurate a priori estimates of the unknown parameters. An iterative solution to convergence, without a digital computer at that time, would have been totally out of the question for all but the simplest
of problems. However, even if this method is iterated until convergence is achieved, the result is not, in fact, the exact minimum point of GLSC. The term exact, as used here, means that the statistical criteria are satisfied mathematically within the convergence tolerance of the iterative procedures. Deming's method is not exact because no matter how tightly his algorithm is converged, the result will only approximately satisfy the generalized least-squares criterion. Nevertheless, Deming's method gives a very close approximation to the exact solution of the least squares problem under most circumstances.

A number of authors have proposed various methods for the exact solution of special cases of the generalized least squares problem as posed by Deming. These methods are concerned only with fitting data by straight lines or polynomials. An algorithm for the exact solution to Deming's formulation of the generalized least squares problem was given at about the same time by Britt and Luecke, Bard and Celmins. Their algorithm may be derived by using the maximum-likelihood principle with the assumption that the observation errors have zero mean and are normally distributed with known variances. The resulting algorithm is of the Gauss-Newton type and requires only first derivatives of the constraint functions. Their formulation, like that of Deming, is very general for algebraic models, and in fact, their algorithm can be thought of as a modification of Deming's method.
Powell and McDonald(19) have also developed an algorithm to find the exact least squares solution for a special case of the generalized least squares problem. Their models are limited to scalar explicit algebraic models with one independent variable subject to error. In addition to this somewhat less general formulation, their algorithm is of the Newton-Raphson type, and thus requires the calculation of the second derivatives of the model functions. Both the Powell-McDonald and the Britt-Luecke algorithms require good initial estimates of the parameters to ensure rapid convergence, and these estimates are usually obtained by using Deming's method as a starting algorithm for their algorithms. The Britt-Luecke algorithm actually automatically includes the Deming algorithm as an initialization procedure.

Southwell(23) has also developed an algorithm to obtain the exact solution to Powell and McDonald's formulation of the generalized least squares problem. However, his algorithm is not very competitive as compared to the others, since it requires higher order derivatives of the model functions. Anderson et al(1) proposed an algorithm similar to that of Britt-Luecke's but failed to utilize the full advantage of the concept of constraints formulation. Their method distinguishes between dependent and independent variables even though their independent variables are subject to error.

Another method for parameter estimation is the "Method of Effective Variance" (MEV) proposed by Hust and McCarty(16)
which has certain characteristics of both the OLSC and GLSC methods. MEV, like GLSC, accounts for errors in all measured variables. However, its mathematical formulation is similar to that of OLSC in that it involves explicit models for expressing dependent variables as functions of independent variables.

Mathematically, the MEV may be stated as follows:

$$\min_{\theta, \mathbf{x}} \sum_{j=1}^{k} \left[ \frac{y_{Mj} - f(x_{Mj}, \theta)}{\sigma_{\text{eff},yj}} \right]^2$$

(2-5)

where

$$\sigma_{\text{eff},yj}^2 = \sigma_{yj}^2 + \sum_{i=1}^{p} \left( \frac{\partial f}{\partial x_{ij}} \right)^2 \sigma_{xij}^2$$

(2-6)

where $p$ is the number of predictor variables, $\sigma_{\text{eff},yj}^2$ is the effective variance of the response variables $y_j$, $\sigma_{xij}^2$ is the variance of the predictor variable $x_{ij}$, and $\sigma_{yj}^2$ is the variance of the response variables $y_j$.

The most important point to note about MEV is that it uses the propagation of error formula to calculate an effective variance of the response variable that takes into account the variance of the response variable and all of the predictor variables.

Even though the classical MEV formulation would suggest that a distinction must be made between predictor and
response variables and that predictor functions must be chosen for the dependent variables, this is, in fact, not so. This point is further clarified by considering an alternative formulation in which the objective function is expressed in terms of the constraints of GLSC formulation:

$$\min_{\theta, z} \sum_{j=1}^{k} \frac{g(z_j, \theta)^2}{\sigma_{\text{eff}, gj}^2}$$

(2-7)

where

$$\sigma_{\text{eff}, gj}^2 = \sum_{i=1}^{q} \left( \frac{\partial g}{\partial z_{ij}} \right)^2 \sigma_{zij}^2$$

(2-8)

where $q$ is the number of the measured variables, $\sigma_{\text{eff}, gj}^2$ is the effective variance of the constraint $g$, and $\sigma_{zij}^2$ is the variance of the measured variable $z_{ij}$.

Using the MEV formulation and depending on how the objective function is treated, the solution will be either the exact GLSC solution or the Deming approximation to it. The exact solution is obtained if the objective function is minimized rigorously with respect to both $\theta$ and $z_j$. In this case, the partial derivatives in the propagation equations for $\sigma_{\text{eff}, y}^2$ or $\sigma_{\text{eff}, g}^2$ would be evaluated using adjusted values of the measured variables, obtained from the
minimizing algorithm. The Deming solution is obtained if an OLSC algorithm is applied (assuming constant effective variances), except that the variances are re-computed between iterations using the current $\theta$ and $z_j = 2M_j$ or $x_j = x_{Mj}$, depending on the formulation. Moreover, Deming's estimate for the true values of the measured variables can be obtained from Equation (A-16) of Appendix A by replacing $2_J$ with $2_M$. 
2.2 Development of Generalized Least Squares Criterion

Throughout this work the Britt-Luecke algorithm is used to apply the GLSC for estimating model parameters. In this section, a description of the algorithm is presented.

The Britt-Luecke formulation is a very general formulation for algebraic models. There may be any number of constraints involving measurements from any number of experiments. This general formulation can handle not only any number of constraints but also problems where a particular constraint involves data from more than one experiment. However, there are two limitations to this general formulation: first, it cannot handle constraints that involve only the unknown parameters,

\[ g_j(\theta) = 0 \quad j=1,\ldots,k \quad (2-9) \]

This situation arises very infrequently. Second, the formulation does not include inequality constraints for the measured variables,

\[ lb_{z_i} \leq z_i \leq ub_{z_i} \quad (2-10) \]

The Britt-Luecke algorithm described in this section is a special case of the Britt-Luecke general formulation. The most significant difference is that the algorithm cannot manage
those problems with constraints involving data from more than one experiment. However, this situation arises very infrequently, and not at all in the present work.

The original Britt-Luecke formulation is based on a vector of data representing all experiments, and a vector of models relating the true values of the measured variables to the true values of the unknown parameters exist. In the present work, it is assumed that $k$ experiments have been made; and for each experiment, one or more observable variables are subject to error. The true values of the measured variables are related to the true values of the parameters by one or more implicit models for each experiment.

Consider a system with an $n$-vector of unknown parameters whose true value is $\theta_0$ and on which $k$ experiments have been made. For each experiment a $q$-vector of observed variables with true values at the time of measurement of $z_{0j}$ ($j = 1,...,k$), is related to $\theta_0$ through an $m$-vector of constraint functions, $g_j(z_j, \theta)$, with the property that

$$g_j(z_{0j}, \theta_0) = 0 \quad j=1,...,k \quad (2-11)$$

In the above discussion, $n$ is the number of parameters, $q$ is the number of measured variables per experiment, $k$ is the number of experiments, $m$ is the number of constraints per experiment.
The measurements of \( z_{o j} \) contain random experimental errors, so that

\[
Z_{Mj} = Z_{o j} + e_j \quad j=1,\ldots,k \tag{2-12}
\]

where \( Z_{Mj} \) is the q-vector of measurements and \( e_j \) is the q-vector of measurement errors for the j-th experiment. Actually \( Z_{o j} \) is also a vector of unknown parameters since it represents the true, error free values of the observed variables. In order to find a maximum likelihood estimate of \( \theta_0 \), it will be necessary to also estimate, simultaneously, \( Z_{o j} \) (\( j=1,\ldots,k \)).

It is necessary that the functions \( g(z_j, \theta) \) be well behaved in the region of interest and it will be assumed that in a suitably large neighborhood of \( (z_{o1}, \ldots, z_{ok}, \theta_0) \):

(i) \( g(z_j, \theta) \) is a continuous function from \( \mathbb{R}^{(q+n)} \) to \( \mathbb{R}^m \), \( j=1,\ldots,k \);

(ii) the partial derivative of each component of \( g(z_j, \theta) \) with respect to \( z_j \) and \( \theta \) exists and is continuous, \( j=1,\ldots,k \);

(iii) the second partial derivative of each component of \( g(z_j, \theta) \) with respect to each pair of arguments exists and is bounded;

(iv) the \( (mxn) \) Jacobian matrices of \( g(z_j, \theta) \) with
The error vectors \( \mathbf{e}_j \) are assumed to be normally distributed random vectors having zero mean and a known positive definite covariance \((q \times q)\) matrix \( R_j \):

\[
\begin{align*}
E(\mathbf{e}_j) &= 0 \quad (2-13) \\
E(\mathbf{e}_j^T \mathbf{e}_j) &= R_j \quad (2-14) \\
E(\mathbf{e}_j^T \mathbf{e}_i) &= 0 \quad i \neq j \quad (2-15)
\end{align*}
\]

The last equation states that the errors in different experiments are independent of one another.

The joint probability density function for the measurement vector \( \mathbf{z}_{Mj} \) at each data point is given by

\[
d(\mathbf{z}_{Mj}) = (2\pi)^{-q/2} \left| R_j \right|^{-\nu/2} \exp \left\{ -\frac{1}{2} (\mathbf{z}_{Mj} - \mathbf{z}_{Oj})^T R_j^{-1} (\mathbf{z}_{Mj} - \mathbf{z}_{Oj}) \right\} \quad (2-16)
\]

The likelihood function is obtained from the joint probability density function for the measurement \( \mathbf{z}_{Mj} \), by considering the parameters \((\mathbf{z}_j, \mathbf{0})\) as the variables. Thus the likelihood function for each data point is

\[
\mathcal{L}_j(\mathbf{z}_j, \mathbf{0}) = (2\pi)^{-q/2} \left| R_j \right|^{-\nu/2} \exp \left\{ -\frac{1}{2} (\mathbf{z}_{Mj} - \mathbf{z}_j)^T R_j^{-1} (\mathbf{z}_{Mj} - \mathbf{z}_j) \right\} \quad (2-17)
\]
where \((z_j, \theta)\) is restrained to those points which satisfy

\[ g_j(z_j, \theta) = 0 \]

\[ j=1, \ldots, k \]  

(2-18)

Since experiments are assumed to be independent of one another the joint likelihood function is simply the sum over all experiments:

\[ L(z_1, \ldots, z_k, \theta) = \sum_{j=1}^{k} g_j(z_j, \theta) \]

(2-19)

The maximum likelihood estimate of \((z_{o1}, \ldots, z_{ok}, \theta_o)\) is found by maximizing eq. (2-19) subject to the constraints

\[ g_j(z_j, \theta) = 0 \]

\[ j=1, \ldots, k \]  

(2-20)

This is equivalent to minimizing

\[ I = \sum_{j=1}^{k} \sum_{Mj}^{k} (z_{Mj} - z_j)^T R_j^{-1} (z_{Mj} - z_j) \]

\[ (2-21) \]

with the same constraints.

The constrained minimization of Equation (2-21) with respect to \(z_j\) and \(\theta\) yields the maximum likelihood estimate of the true values of the unknown parameters \(\theta_o\) and of the measured variables \(z_o\). The detail derivation of the algorithm to achieve this minimization is shown in Appendix A. The resulting estimates of \(\theta_o\) and \(z_{oj}\) are designated by \(\hat{\theta}\), and \(\hat{z}_j\), for \(j=1, \ldots, k\).
2.3 Estimated Parameters Analysis

One of the important reasons for fitting model equations to experimental data is to obtain a representation that can be used confidently for predictions. The confidence in the predictions depends strongly on the confidence in the data and the model. The Britt-Luecke parameter estimation method based on the maximum likelihood principle provides information useful in the assessment of models, and in the critical evaluation of data. This information is readily available once the parameters have been determined, and may be divided into three categories. In this section, the analysis of the parameters estimated from the Britt-Luecke GLSC method is presented. However, it should be pointed out that the error analysis presented here is exact only for linear Ordinary Least-Squares problems, and is approximate for nonlinear Ordinary Least-Squares or Generalized Least Squares problems. Its application to nonlinear OLS and GLS problems should be carried out with this fact in mind. Nonetheless, the information it provides can serve as guidelines for models and data evaluations.

2.3.1 Variance of the Observables of Unit Weight

Estimate of the variance of each measured variable is required by the GLSC formulation. However, if the absolute values of the variances are not known, only the relative values
need be assigned. This flexibility is due to the nature of the maximum likelihood procedure. Therefore, \( R \) may be replaced by \( R \sigma^2 \) where \( \sigma^2 \) is a scalar unknown called the variance of the observables of unit weight. The estimated parameters are independent of \( \sigma^2 \), whereas the parameter variance-covariance matrix is not. As discussed by Deming,\(^{(12)}\) the sum of the weighted squared deviations \( I \), defined by equation (2-21), divided by \( \sigma^2 \) is a \( \chi^2 \) distribution with \( \nu = k - n \) degrees of freedom.

\[
\chi^2 = I/\sigma^2 \tag{2-22}
\]

Since the mean or expected value of \( \chi^2 \) is equal to the number of degree of freedom, the value of \( \chi^2 \) may be estimated from equation (2-22)

\[
s^2 = E(\sigma^2) = I/(k-n) \tag{2-23}
\]

The quantity \( s^2 \) can be used in selecting the best model to represent the data. For example, if two different models are used with a given data set, the one that results in a lower value of \( s^2 \) provides a better representation of the data. The statistical significance of the difference can be ascertained by calculating the ratio

\[
F = s^2_1/s^2_2 \tag{2-24}
\]
which will be distributed as the F distribution, assuming errors due to lack of fit are normally distributed. At a significance level \( \alpha \), the value \( F \) can be compared to a tabulated \( F_{\alpha/2}(v, \mu) \) where \( v \) and \( \mu \) are the degrees of freedom for the two models. If \( F \) is greater than \( F_{\alpha/2}(v, \mu) \), then the second model is indeed better than the first. Similarly, for the case of several sets of data and only a single model, it can be determined if one set of data is significantly better represented by the model than another set.

2.3.2 Variance-Covariance Matrix

The variance \( s^2 \), in equation (2-23) is used to estimate the errors in the calculated parameters in the form of variance-covariance matrix,

\[
\text{VAR}(\theta) = E\left\{ \left( \theta - \theta_0 \right) (\theta - \theta_0)^T \right\} = \left\{ G_j^T (G_j^T R_j G_j)^{-1} G_j \right\}^{-1}
\]

which is an \( nxn \) symmetric matrix. \( G \) are the Jacobian matrices defined in Appendix A.

The diagonal elements of this matrix represent the variance of the corresponding parameters. The square roots of these variances are estimates of the standard deviations of the parameters and, in effect, are measure of the uncertainties in the estimated parameters. The significance of the parameters (whether the parameter is statistically different from zero) can also be determined by the standard t test:
\[ t_i = \frac{\theta_i}{s_{\theta_i}} \quad (2-26) \]

Comparing \( t_i \) with the value \( t_{\alpha/2, v} \) determined from a tabulated \( t \) distribution, for \( v \) degrees of freedom and \( \alpha \) level of significance, if \( t_i \) exceeds \( t_{\alpha/2, v} \), the parameter \( \theta_i \) is therefore significantly different from zero.

Another by-product of the variance-covariance matrix is the correlation coefficient which is the measure of the degree of correlation between two parameters. This can be determined from the off-diagonal elements of the variance-covariance matrix

\[ \text{CORR}(\theta)_{ij} = \frac{\text{VAR}(\theta)_{ij}}{\sqrt{\text{VAR}(\theta)_{ii}} \sqrt{\text{VAR}(\theta)_{jj}}} \quad (2-27) \]

If parameters are completely independent, their correlation coefficient is zero. As parameters become more correlated, the correlation coefficient approaches a value of \( \pm 1 \). When the parameters are highly correlated, it becomes very difficult to determine them uniquely. For an \( nxn \) variance-covariance matrix resulting from the estimation of \( n \) parameters, there exists \( n \) eigen values \( \lambda_j \) associated with \( n \) eigen vectors \( V_j \), for \( j=1,\ldots,k \). It can be shown\(^{14}\) that the \( n \) linear combinations \( V'_j \) of the optimal parameters defined by

\[ V'_j = \sum_{k=1}^{n} V_{jk} \lambda_k \quad j=1,\ldots,n \quad (2-28) \]
are statistically independent, with variances $\lambda_j$. The smallest value of $\lambda_j$ is related to the combination of parameters which is known with most accuracy. The confidence region of the calculated parameters is the hyperellipsoid:

$$\_\theta^T \text{VAR}^{-1}(\theta) \_\theta = \beta^2$$  \hspace{1cm} (2-26)

where $\beta$ is related to a probability level. The confidence region is a region in which parameters can be varied without significant effect on the weighted sum of squares in Eq. (2-21).

The variance-covariance matrix can also be used to calculate the uncertainty in any property that is a function of the estimated parameters. If $P(\theta)$ is such a property, the variance of this property $\sigma_P^2$ is given by

$$\sigma_P^2 = P^T(\theta) \text{VAR}(\theta) P(\theta)$$  \hspace{1cm} (2-27)

where $P(\theta)$ is an n vector of the partial derivatives of $P(\theta)$ with respect to the parameters $\theta$. Equation (2-27) provides a direct means for determination of the expected errors in predictions made with the fitted model.

2.3.3 Residual Analysis

The generalized least-squares method used yields not only the best parameter estimates, but also the estimates of the true values of all measured variables. The difference
between these estimated true values and their corresponding experimentally determined values are called residuals. Statistical analysis of these residuals can indicate how well the model represents the data and how precise the data are. Ideally, for the case with numerous data points, the mean value of the residuals should be zero, the standard deviations equal to the assumed values, and the distribution of deviations gaussian. It is also generally useful to plot these residuals versus some system variables (eg. T or x or almost anything that could affect experiments). Useful information can be extracted from these plots. For example, if the points are distributed close to a curve which is different from the x-axis, it indicates either a lack of flexibility of the model, or systematic error in one or several measured variables, or incorrect data point. If all deviations are scattered around the x-axis with a normal distribution but with standard deviations very different from the assumed values, it means that the assumed standard deviations were incorrect.
CHAPTER 3  THE DATA REGRESSION SYSTEM

The ASPEN Data Regression System (DRS) determines model parameters by fitting to experimental data using the Britt-Luecke Generalized Least-Squares method. The system is one of several run classes for ASPEN, such as simulation run, economic analysis run, etc. The important features of the DRS can be conveniently divided into four categories:

1. The DRS is an integral part of ASPEN. It uses the same Input Translator (IT), the same System Definition File (SDF), the same Data Management System (DMS), the same executive system, etc., as ASPEN. The DRS uses the ASPEN Physical Property System (PPS) for all its thermophysical property calculations. The users must specify the methods and models to be used via ASPEN keywords described in Section 4.1. The dependency of DRS of PPS insures that the parameters estimated by the DRS are compatible with the PPS which used them in the subsequent calculations.

2. The DRS employs the Britt-Luecke's Generalized Least-Squares (GLSQ) method as its regression algorithm. This method is based on the maximum likelihood principle, and is a very general formulation for algebraic models as discussed in Chapter 2. The unknown model parameters are determined by minimizing an objective function I, subject to constraints, see eqs. (2-2), (2-3) and (2-4). All experimentally determined
quantities are used in the objective function, and the constraints are applied to every data point. For a particular data regression problem, the DRS automatically sets up appropriate objective function and necessary constraints depending on the users' problem specification. All the constraints are evaluated by a system-supplied constraint calculation subroutine which is capable of handling all the possible constraints in DRS.

Generally, there are three types of problem specifications:

a). Phase equilibrium type which involves fitting model parameters to phase equilibrium data. The constraints are the phase equilibrium conditions, while the measured variables include temperature T, pressure P, and mole fractions of all components in both equilibrium phases. In DRS, all phase equilibrium conditions are expressed in terms of fugacity coefficients:

\[ \phi^I_i(T,P,x^I)x^I = \phi^{II}_i(T,P,x^{II})x^{II} \quad i=1,\ldots,nc \]

(3-1)

where nc is the number of components in the mixture.

The constraints in Eq. (3-1) are applied to all data points. The fugacity coefficients can be calculated using any methods and models available on ASPEN. Many types of phase equilibrium problems can be performed, especially the DRS provides the capability of treating three-phase vapor-liquid-liquid
equilibrium problem. Details are given in Chapter 4. The objective function for this type of problem involves only the state variables,

\[ I = I \text{ (state variables)} \quad (3-2) \]

b). Property type problem which involves fitting model parameters to properties, such as density, vapor pressure, heat capacity, etc. More than one properties can be used for a problem. The objective function involves state variables, which usually include temperature, and all the properties specified.

\[ I = I \text{ (state variables, properties)} \quad (3-3) \]

There is one constraint function for every property,

\[ P_{M_i} = P_i \quad i=1, \ldots, p \quad (3-4) \]

where \( P_{M_i} \) is the \( i \)th measured property, \( P_i \) is the corresponding model-calculate property value, and \( p \) is the number of properties specified. Each constraint is applied to all experimental data points.

c). Problem involving the combination of the first two specifications. For example, it is possible to determine parameters using both phase equilibrium and enthalpy data. In
this case, the objective function involves the state variables, and the properties,

\[ I = I \text{ (state variables, properties)} \]  \hspace{1cm} (3-5)

The constraints are the phase equilibrium conditions in eq. (3-1) and the property constraints in eq. (3-4).

The constraints of problem specifications described above are evaluated by the general constraint calculation subroutine of DRS. However, it is possible for the users to supply their own subroutines. The specification procedure and the requirements of the routine are discussed in Section 4.2.3.

3. The DRS is Table-driven. This feature has been an important aspect of the design of ASPEN system, and hence of DRS. In this approach, important informations concerning the system are stored in Table forms. The System Definition File (SDF) is an example. Modifications of the system are done by modifying these tables. Any new physical property models added to the ASPEN will be done through these tables and will be completely transparent to DRS. The DRS section in the SDF Table is shown in Table 3.1. Possible future modification to the DRS includes the provision for additional phase equilibrium and property specifications.
4. The DRS generally treats regression problems using the Generalized Least-Squares Criterion (GLSC). However, it is also possible to apply the Ordinary Least-Squares Criterion (OLSC) to a problem. In DRS, the OLSC is a special case of GLSC. By assigning zero standard deviation values to variables to be treated as independent variables, they are automatically excluded from the GLSC objective function I, which the system sets up, and are therefore not adjusted by the algorithm. In this way, the GLSC reduces to the OLSC.

The input to the Data Regression System by the users may be conveniently divided into four groups.

1. **Parameter specification**
   
   This data defines all the unary and/or binary parameters to be determined in the DRS run. It includes the parameter names, the parameter ID's, the component ID's associated with the parameters, the element number, the initial estimate, the lower bound, the upper bound and the scale factor of the parameters. The first five data entries define a parameter, while the rest of the data are used to initialize, set bounds and properly scale the parameter.

2. **Data group specification**
   
   This data defines a group of experimental data to be used in a regression problem within a DRS run. It specifies component ID's associated with the data, the measured variables to be treated by the DRS as state variables, and those to be considered as properties. It also defines the type of phase
equilibrium constraints as well as the component ID's associated with it. The DRS sets up objective function and constraints according to these specifications. The standard deviation values as well as the measured data are entered together with the above specifications. Other experimental data with different data specifications can be described in another data group. Therefore, in a DRS input, there can be as many data group specification sections as necessary.

3. **Case specification**

This data defines a data regression case study. It specifies the unary and/or the binary parameters to be determined, and the data groups to be used in the problem. In a data regression case, any number of unary or binary parameters can be specified using any number and combination of experimental data groups defined above. Information about the regression algorithm such as the maximum number of iterations allowed, the tolerance, the adjustment stepsize, etc., are optional with default values being supplied. The default constraint calculation subroutine is DRCSTR. However, if the user provides his own subroutine, the name of the routine must be specified.

Each case specification completely defines a data regression problem. By defining additional cases, more than one regression problems can be performed within a single DRS run.
4. **Physical property specification**

All the property calculations in DRS are performed by the ASPEN Physical Property System (PPS). It is therefore necessary to define the methods and models to be used in calculating any properties that the DRS requires. This is done by specifying the physical property option-set which is a collection of information about the methods and models to be used for the calculation of physical properties. Details are discussed in Section 4.1. The components associated with the system as well as the sources of physical property data must also be defined.

The data regression problem specifications discussed above are accomplished by the use of system defined keywords, which are fully described in Chapter 4.

There are two files created by the DRS that are of interest to the users.

1. **History file**

This is an output generated by the main regression routine containing details about the progress of the regression algorithm at every iteration. The information includes the weighted sum of squares of data adjustments, the current parameter estimates, etc. Final results are given when convergence has been obtained. The results include total number of iterations taken to achieve convergence, final parameter estimates, the measured variables, estimates of the true values of the variables, data adjustments, parameter
correlation matrix, variance-covariance matrix, the execution time, etc. If the problem does not converge within the allowed number of iterations, a warning message is printed along with the parameters and other results estimated at the last iteration. The final results are always written on the history file, however, the information concerning the progress of the algorithm at each iteration is controllable through the input language by specifying the history file message level code. The history file is intended for debugging, checking algorithm performance, etc., by system programmers. It is of little use to ordinary users but may be useful to advanced users.

2. **Report file**

   This is the primary output file and contains the following information:

   1) The input data,

   2) The final parameter estimates,

   3) The number of iterations used,

   4) For each variable, the measured value, the estimate of the true value, and the data adjustment (residual), the root mean squares deviation, average absolute deviation and maximum deviation are given,

   5) The parameter correlation matrix and the parameter variance-covariance matrix.

   There is no output file generated by the system that is readable by an ASPEN simulation run. Parameters must be
transmitted by the user from the DRS Report file to an ASPEN Input file. A History file and a Report file generated by the DRS for the problem in which Ideal Gas Heat Capacity parameters are estimated, as discussed in Chapter 6, are shown in Appendix B.
PARAMETERS PRIORITY 3200

<table>
<thead>
<tr>
<th>PKW NAME</th>
<th>ALTERNATES</th>
<th>CODE</th>
<th>IDS</th>
<th>IDS</th>
<th>MISSING</th>
<th>END OF</th>
<th>SEMANTIC</th>
<th>--TKW LENGTH--</th>
<th>RUN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SKWS ROUTINE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3200</td>
<td>0</td>
<td>0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>2</td>
<td>48</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SKW NAME</th>
<th>ALTERNATES</th>
<th>ID</th>
<th>CODE= 0</th>
<th>MODE=2</th>
<th>LEFT</th>
<th>JUSTIFY(0)</th>
<th># OF TKWS=</th>
<th>COUNT TKWS(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.1 Data Regression System section of the System Definition File of ASPEN.
<table>
<thead>
<tr>
<th>DATA-GROUP</th>
<th>PKW NAME</th>
<th>ALTERNATES</th>
<th>ID CODE= 0</th>
<th>MODE=0</th>
<th>LEFT JUSTIFY(0)</th>
<th># OF TKWS= 2</th>
<th>COUNT TKWS(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHASE-EQ</td>
<td>PHASEEQ</td>
<td>PEQ</td>
<td>ID CODE= 0</td>
<td>MODE=1</td>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS= 2</td>
<td>COUNT TKWS(0)</td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STD-DEV</td>
<td>STDDEV</td>
<td>DEV</td>
<td>ID CODE= 0</td>
<td>MODE=2</td>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS= 2</td>
<td>COUNT TKWS(0)</td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATA</td>
<td>DATA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARAMETERS</td>
<td>PARAMS</td>
<td>PARM</td>
<td>ID CODE= 0</td>
<td>MODE=0</td>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS= 2</td>
<td>COUNT TKWS(0)</td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATA-GROUPS</td>
<td>DATAGROUPS</td>
<td>DGRP</td>
<td>ID CODE= 0</td>
<td>MODE=1</td>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS= 2</td>
<td>COUNT TKWS(0)</td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKW NAME</td>
<td>ALTERNATES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROPERTIES</td>
<td>PROPERTIES</td>
<td>PROP</td>
<td>ID CODE= 0</td>
<td>MODE=0</td>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS= 1</td>
<td>COUNT TKWS(0)</td>
</tr>
<tr>
<td>TKW NAME---------</td>
<td>DATATYPE</td>
<td>VECTOR</td>
<td>UNIT-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER</td>
<td>NUMBER</td>
<td>VALUE</td>
<td>VALUE------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### System Definition File Report

**CASE**

<table>
<thead>
<tr>
<th>SKW NAME</th>
<th>ALTERNATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALGORITHM</td>
<td>ALG-PARAM</td>
</tr>
<tr>
<td>ID CODE= 0</td>
<td>MODE=0</td>
</tr>
<tr>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS=</td>
</tr>
<tr>
<td>COUNT TKWS(0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-----------</th>
<th>DATATYPE</th>
<th>VECTOR UNIT</th>
<th>------- VAR. LIMIT ------</th>
<th>UPPER LIMIT</th>
<th>LOWER LIMIT</th>
<th>--- DEFAULT ---</th>
<th>EDIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL</td>
<td>MODEL</td>
<td>3 0 0 0 0 0 0 0 0 0 0 0</td>
<td>None</td>
<td>None</td>
<td>2</td>
<td>Model</td>
<td></td>
</tr>
<tr>
<td>MAXIT</td>
<td>MAXIT</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0</td>
<td>500</td>
<td>1 2</td>
<td>50</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>TOLERANCE</td>
<td>TOLERANCE</td>
<td>2 0 0 0 0 0 0 0 0 0 0 0</td>
<td>.1000D+75</td>
<td>-.1000D+75</td>
<td>2</td>
<td>.1000D-05</td>
<td>-1 0</td>
</tr>
<tr>
<td>DEDEM</td>
<td>IDEM</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0</td>
<td>2147483647 -2147483647</td>
<td>2 0 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IODS</td>
<td>IODS</td>
<td>1 0 0 0 0 0 0 0 0 0 0 0</td>
<td>2147483647 -2147483647</td>
<td>1 13 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STEP-SIZE</td>
<td>STEP- SIZE</td>
<td>5S 2 0 0 0 0 0 0 0 0 0 0</td>
<td>None</td>
<td>None</td>
<td>2</td>
<td>.1000D+75</td>
<td>-.1000D+75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PKW NAME</th>
<th>ALTERNATES</th>
<th>SORT CODE</th>
<th>IDS</th>
<th>IDS</th>
<th>OPT.</th>
<th>CALL IF</th>
<th>CALL AT SEMANTIC</th>
<th>SKW CODE</th>
<th>ROUTINE #</th>
<th>TensorFlow RUN</th>
<th>INTEGER</th>
<th>REAL</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF-SUBS-ATT</td>
<td>DEFSUBSATTR</td>
<td>SSAT</td>
<td>4000</td>
<td>2 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>102</td>
<td>11 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF-STREE-MAT</td>
<td>DEFSTREAMATT</td>
<td>SAT</td>
<td>4100</td>
<td>2 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>103</td>
<td>11 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF-SUBS-CLA</td>
<td>DEFSUBSCLASS</td>
<td>SSCL</td>
<td>4200</td>
<td>1 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>13</td>
<td>13 4 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PKW NAME</th>
<th>ALTERNATES</th>
<th>SORT CODE</th>
<th>IDS</th>
<th>IDS</th>
<th>OPT.</th>
<th>CALL IF</th>
<th>CALL AT SEMANTIC</th>
<th>SKW CODE</th>
<th>ROUTINE #</th>
<th>TensorFlow RUN</th>
<th>INTEGER</th>
<th>REAL</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF-SUBS-ATT</td>
<td>DEFSUBSATTR</td>
<td>SSAT</td>
<td>4000</td>
<td>2 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>102</td>
<td>11 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF-STREE-MAT</td>
<td>DEFSTREAMATT</td>
<td>SAT</td>
<td>4100</td>
<td>2 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>103</td>
<td>11 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF-SUBS-CLA</td>
<td>DEFSUBSCLASS</td>
<td>SSCL</td>
<td>4200</td>
<td>1 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>13</td>
<td>13 4 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ADF NAME**

<table>
<thead>
<tr>
<th>SKW NAME</th>
<th>ALTERNATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITION</td>
<td>DEF</td>
</tr>
<tr>
<td>ID CODE= 0</td>
<td>MODE=0</td>
</tr>
<tr>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS=</td>
</tr>
<tr>
<td>COUNT TKWS(0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-----------</th>
<th>DATATYPE</th>
<th>VECTOR UNIT</th>
<th>------- VAR. LIMIT ------</th>
<th>UPPER LIMIT</th>
<th>LOWER LIMIT</th>
<th>--- DEFAULT ---</th>
<th>EDIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE</td>
<td>TYPE</td>
<td>11 0 0 0 0 0 0 0 0 0 0 0</td>
<td>None</td>
<td>None</td>
<td>1</td>
<td>Missing</td>
<td></td>
</tr>
<tr>
<td>ATTRIBUTES</td>
<td>ATTRIBUTES</td>
<td>ATTR</td>
<td>11 0 0 1 0 0 0 0 0 0 0 0</td>
<td>None</td>
<td>None</td>
<td>1</td>
<td>Missing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PKW NAME</th>
<th>ALTERNATES</th>
<th>SORT CODE</th>
<th>IDS</th>
<th>IDS</th>
<th>OPT.</th>
<th>CALL IF</th>
<th>CALL AT SEMANTIC</th>
<th>SKW CODE</th>
<th>ROUTINE #</th>
<th>TensorFlow RUN</th>
<th>INTEGER</th>
<th>REAL</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF-SUBS-ATT</td>
<td>DEFSUBSATTR</td>
<td>SSAT</td>
<td>4000</td>
<td>2 0</td>
<td>YES(1)</td>
<td>PARAGRAPH(0)</td>
<td>14</td>
<td>14 3 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ADF NAME**

<table>
<thead>
<tr>
<th>SKW NAME</th>
<th>ALTERNATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITION</td>
<td>DEF</td>
</tr>
<tr>
<td>ID CODE= 0</td>
<td>MODE=1</td>
</tr>
<tr>
<td>LEFT JUSTIFY(0)</td>
<td># OF TKWS=</td>
</tr>
<tr>
<td>COUNT TKWS(0)</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4 INPUT LANGUAGE

The ASPEN input language is a free format, user-oriented language. The key terminologies and general concepts of ASPEN input language is discussed below. The DRS input data is described by ASPEN input language. The information entered in a DRS input file may be conveniently divided into two groups:

1. The specification of the components, the source of physical property constants and known parameters, the units of the input data and the output to be printed in the Report file, and the physical property methods and models to be used in the problem. These informations are entered by using the regular ASPEN input language as described in Section 4.1.

2. The information specific to the regression problem, for instance, the parameter specification, the data group specification, etc. These are entered by using input keywords that pertain only to the DRS. Detail description of these keywords are given in Section 4.2.

A DRS run is signaled by the word NEW starting in column one of the first input line, followed by the word DRS starting in column 6. The first input line establishes run class and is used to distinguish a data regression run from a normal ASPEN simulation run.

The ASPEN input language is made up hierarchically of paragraphs, sentences and individual data entries.
Corresponding to this hierarchy are three levels of keywords: primary, secondary and tertiary. Keywords are system defined words identifying the paragraph, sentence or individual data entry. An example of the input language is shown in Table 4.1.

A paragraph is a logical group of input data such as the data describing the parameters to be estimated. Paragraphs consist of one or more sentences but always begin with a primary keyword starting in column one. A paragraph is terminated when another primary keyword starting in column one is encountered or when the end of file is reached. In the example, COMPONENTS, PARAMETERS, CASE, etc. are primary keywords and the input following them are paragraphs. Primary keywords sometimes have one or two data entries associated with them directly rather than through sentences. These data entries are ID's which serve the purpose of completely identifying the paragraph to follow. In the first DATA-GROUP paragraph of the example, Gl is a data group ID which identifies the first data group being entered.

A sentence is a free format string of data entries on a single line beginning with a secondary keyword, with the restriction that the secondary keyword cannot begin in column one. Also, the first sentence of a paragraph may be on the same line as the primary keyword. Sample sentences are the PARAMETER and BIPARAMETER sentences in the example. A sentence may be continued on more than one line by placing an ampersand (&) at the end of the line to be continued. The PARAMETER
sentence is an example.

Sentence data may be entered positionally or by tertiary keyword. Data entered positionally must be entered in the system prescribed order, for example temperature followed by pressure, etc. A positional entry may be "skipped" by entering an asterisk (*) or by terminating the sentence before all values have been specified. ASPEN sentences have been structured, so far as possible, so that required entries precede entries that have defaults in order to facilitate positional entry. The SYSTEM-DEF sentence of data group G1 shows the component-list entered positionally.

The symbol / used at the end of a sentence indicates that another sentence with the same secondary keyword is following on the same line or the next line. In the COMPONENTS paragraph of the example there are actually three sentences, one for each component.

It is also possible to insert comments in the ASPEN input language. The line end special character (;), or the sentence continuation special character (&), logically terminates an input line. Anything following these symbols on the same line is treated as a comment and is ignored by the input processor. If a (;) is entered in column one then the entire line is treated as a comment. The example 4.1 has a comment following the PROPERTIES primary keyword.
4.1 *ASPEN Input Language*

The general ASPEN input language are described in the ASPEN User Manual. Only a selected set of the general ASPEN keywords is used in the Data Regression System. These keywords can be conveniently divided in to five groups:

4.1.1 **Title and description**

The user may supply descriptive information to be printed in the Report. The TITLE and DESCRIPTION primary keywords are used for this purpose. The TITLE is printed on each page of the report while the DESCRIPTION is printed at the beginning of the Report.

The recommended form of input for TITLE is:

```
TITLE 'upto sixty four characters enclosed in quotes'
```

If the character string contains no blanks or other ASPEN special characters the quotes are not required.

The recommended form of input for DESCRIPTION is:

```
DESCRIPTION
"any amount of text containing any symbols entered on any number of lines and enclosed in double quotes"
```

The first column of each line following DESCRIPTION must be left blank.

Example is shown in Table 4.1.

4.1.2 **Component specification**

There are four primary keywords that may be used to specify the components to be used in ASPEN simulation run along
with their types. Only the COMPONENTS primary keyword is of interest to the DRS. The COMPONENTS statements must be used to assign ID's to all components to be included in a DRS run. The order of the component list is unimportant. For those components for which some or all of the required properties are to be retrieved from one or more data banks, the components statement is also used to indicate the data bank component names. The recommended form of input for this keyword is as follows:

COMPONENTS cid cname /

where cid is a component ID. An ID must be entered in the COMPONENTS statement for every component to be included in a DRS run.

cname is a character string of upto 32 characters designating a unique component name or alias. For a component for which data is to be retrieved from a data bank, this character string must match the name or alias used for that component in the data bank. If the data is not to be retrieved, the component name is not required.

4.1.3 Physical property calculation option

The DRS depends on the physical property calculation capability of ASPEN. The calculation methods and models that are used in a particular DRS run must be specified. The
collection of information that defines both the methods and models to be used for the calculation of properties is referred to as option set. There are a variety of built-in option sets that are described in Appendix D of the ASPEN User Manual.\(^{(3)}\) If built-in option sets are not sufficient to satisfy the user's needs, the users may create additional option sets using the keywords described in Section 3.2.1.2 of the ASPEN User Manual.\(^{(3)}\) The primary keyword PROPERTIES is used to assign option set to the DRS. The recommended form of input which is relevant to DRS is as follows:

\[
\text{PROPERTIES opsetid seclist / . . .}
\]

where  \textit{opsetid} is a option set ID.

\textit{seclist} is an optional GLOBAL specification.

4.1.4 Data input

The physical property constants and known parameters required by the models for all the components included in the DRS run must be provided. There are basically two ways in which the data may be made available: (1) by retrieval from one or more data banks, and (2) by direct input using the PROP-DATA keyword. These two modes of input are described in the ASPEN User Manual.\(^{(3)}\)

4.1.5 Units specification

ASPEN uses the International System (SI) of units for all calculations and internal storage. The user may choose
from a wide variety of optional units for input and Report output. Two sets of units are available in addition to SI, English Engineering (ENG) and Metric Engineering (MET). In addition, the user may choose units for individual quantities that override the set. The unit options available and set constants are shown in Table 4.2.

With respect to the DRS, input units may be specified globally, for a paragraph or for a sentence. Output units may be specified globally or for a Report section. The History File is always written in the internal SI units.

The IN-UNITS keyword is used to specify the units in which the input data is entered. This keyword may be used as a primary keyword or as a secondary keyword anywhere within a paragraph. When used as a primary keyword, it specifies the global input units. When used as a secondary keyword within a paragraph, IN-UNITS indicates that the data following in that paragraph will be in the units specified.

The recommended form of input for an IN-UNITS paragraph or sentence is:

IN-UNITS [setname] [kwd=unitoption] . . .

where setname is the unit set name, SI, ENG, or MET. ENG is the default.

kwd may be any of the quantities listed in the first column of Table 4.2.

unitoption may be any of the unit options listed in row kwd of Table 4.2. This specification overrides
the set option for quantity kwd. When a unit option includes a (/), unit option must be included in quotes.

Example
1) IN-UNITS ENG TEMPERATURE=R
2) IN-UNITS TIME=HR DENSITY='LB/GAL'
3) DATA-GROUP Gl
   IN-UNITS SI TEMPERATURE=C

The OUT-UNITS keyword is used to specify the units in which the Report will be written. When used as a primary keyword it specifies the global Report units. When used as a secondary keyword it specifies the units for the Report of a case. The recommended form of input for OUT-UNITS is identical to that for IN-UNITS.
4.2 DRS Input Language

The topic discussed in this section can be conveniently divided into three main parts. The first part explains the input language used to specify unary and binary parameters to be determined in a DRS run. The second part discusses the specification of the data to be used in the determination of the parameters. The last part explains the keywords used to describe a regression problem by combining the parameters and the datasets that have been specified in the first two parts.

4.2.1 Parameter specification

The primary keyword PARAMETERS is used to specify all unary and binary parameters to be determined in a DRS run. The unary parameters are described by PARAMETER sentences while the binary parameters are defined by BIPARAMETERS sentences. Initial estimates of the parameters are not generated by the system and must be supplied by the user. The DRS assumes that unknown parameters are bounded, therefore it is necessary to specify their lower bound, and their upper bound. The recommended forms of input for these keywords are as follows:
PARAMETERS

PARAMETER pid pnam cid elem init lb ub [scale]
BIPARAMETER pid pnam cid1 cid2 elem init lb ub [scale]

where **pid** is an integer-value parameter ID.

**pnam** is a character string of upto 8 characters
designating a parameter name. This name is the same as
the labelled common name appears in the physical property
model which uses the parameter, for example the name for
Peng-Robinson equation of state binary parameter is PRKIJ.

**cid** is a component ID of the component associated with
the parameter. This ID is taken from the list if ID's in
the COMPONENTS sentence. For binary parameters, two
ID's, one for each component, are required.

**elem** is the element number of the parameter which has
multiple elements. For example, the Extended Antoine
parameter consists of 9 elements, all stored in the
labelled common called PLXANT.

**init** is the initial estimate of the parameter which the
user must supply. No default value is provided.

**lb** is the lower bound of the parameter

**ub** is the upper bound of the parameter.

**scale** is an optional keyword that is used to scale the
parameters to be determined in a particular problem so
that they all have the same order of magnitude. The
default value for this keyword is unity.
Example: 1) The first and third elements of the Extended Antoine parameters for component Cl.

```
PARAMETER 1 PLXANT Cl 1 x x x x
PARAMETER 2 PLXANT Cl 3 x x x x
```

2) The Peng-Robinson equation of state binary parameter for components Cl and C2.

```
BIPARAMETER 1 PRKIJ Cl C2 1 x x x x
```

4.2.2 Data group specification

The primary keyword DATA-GROUP is used to describe a set of experimental data that will be used in the regression problem. If more than one sets of data are used, each dataset is described in a DATA-GROUP paragraph to which a unique group ID has been assigned. The data entered in each group can be of any units that are allowed in ASPEN. The unit specification is done via the secondary keyword IN-UNITS which is described in Section 4.1.5. The experimental data of each group are defined using four secondary keywords. SYSTEM-DEF sentence defines the measured variables to be treated as state variables, and those to be treated as properties. The distinction between state variables and properties is made because the DRS treats them differently in the formulation of constraint functions for the regression algorithm. The property forms constraint functions which depends on the state variables and the unknown parameters. The possible state variables and property
specifications are listed in Table 4.3. This secondary keyword also specifies the components associated with the data group. PHASE-EQ secondary keyword is used to specify the type of phase equilibrium condition that the data is required to satisfy. With this specification, the DRS automatically sets up appropriate phase equilibrium constraints for the Generalized Least-Squares regression algorithm. The phase equilibrium options that are available in the DRS are listed in Table 4.3 together with the types of data that should accompany each option. In DRS, it is possible to determine parameters using three-phase equilibrium data such as vapor-liquid-liquid equilibrium data, by specifying appropriate phase equilibrium options and state variables as shown at the end of this section. The components in each equilibrium phase must also be defined. STD-DEV sentence defines the standard deviations of the measured variables for all data points. A set of standard deviation values may be assigned to a single data point, several data points or all data points in the data group. This is accomplished by assigning an index number to each set of the standard deviation. If only one set of values is given, it will be assigned to all data points. If several sets are provided, however, the difference between the indices of two adjacent sets indicates the number of data points to which the standard deviation values of the first set will be assigned. The number of standard deviation values in each set is the same as the number of measured variables. For GLSC,
quantities must be assigned standard deviations. However, the DRS is capable of handling problems using the OLSC as well. This is accomplished by simply assigning zero standard deviations to the variables which are treated as independent variables. The DRS automatically excludes them from the objective function $I$, given in eqs. (2-2), and (3-3). The GLSC method reduces to the Barker's method when the standard deviations of $T$ and $x$ are set to zero. DATA sentences are used to enter all the experimental data. All measured variables of a data point are entered in one sentence. Each data point is given an index which is used in conjunction with the index of the standard deviations to assign standard deviation set to the data. The number of measured variables to be entered in each sentence is determined by the state variable and property specifications. For example, with a TP state variable specification and a liquid density property specification, the DRS expects three measured variables, $T$, $P$ and liquid density, to be entered in that order. With state variable specification involving component mole fractions $x$ or $y$, for an $n$ component system, the mole fraction values of the first $n-1$ components must be supplied. The DRS determines the $n^{th}$ mole fraction by difference. The order of the components is defined by the cidlist of the SYSTEM-DEF sentence. The recommended forms of input for these keywords are as follows:
DATA-GROUP grid

SYSTEM-DEF statespec cidlist [PROP-LIST=proplist]
PHASE-EQ peqtype cidlist / . . .
STD-DEV num val-list / . . .
DATA num val-list / . . .

where grid is the data group ID.

statespec is a character string of upto 8 characters
designating state variables. The valid statespec options
are listed in Table 4.3.

cidlist is a list of component ID's which are associated
with the data. The order of the ID's in the list
specifies the order of the component mole fractions to be
entered in the DATA sentences.

proplist is a list of property names. Each name is a
character string of up to 8 characters. It defines the
variables to be treated as properties. A list of valid
property specifications is given in Table 4.3. This
information is optional and is omitted when all measured
variables are state variables.

peqtype is a character string of up to 8 characters
designating the phase equilibrium type, for example, VL
for vapor-liquid equilibrium. A list of valid peqtype is
given in Table 4.3.

cidlist is a list of component ID's associated with the
deqtype. Generally, this list is the same as the cidlist
of the secondary keyword SYSTEM-DEF because all
components in an equilibrium system should satisfy the phase equilibrium requirement. However, in certain cases, the user may wish to determine the parameters for which the phase equilibrium conditions are satisfied only for a selected number of components. In such cases, cidlist gives a list of those components.

num is the index number of the data or standard deviation sentence.

val-list is a vector containing the values of standard deviations or measured variables.

The symbol " / . . . " indicates that there can be more than one sentences of the same secondary keyword, with the keyword omitted from subsequent sentences.

Example 1) Vapor-liquid-liquid equilibrium data of components Cl and C2

DATA-GROUP G1
SYSTEM-DEF TPXXY C1 C2
PHASE-EQ VL C1 C2 / LL C1 C2
STD-DEV 1 x x x x x
DATA 1 x x x x x /
     .  
     . 
     .  
6 x x x x x
2) Density data of component Cl, with 10 data points and 2 sets of standard deviations. The first set is assigned to the first three data points while the second set is assigned to the remaining data.

```
DATA-GROUP G2

SYSTEM-DEF T Cl PROP-LIST=RHOL

STD-DEV 1 x x/
   4 x x

DATA 1 x x/
   2 x x
      .
      .
      .
   10 x x
```

4.2.3 Case specification

The primary keyword CASE is used to describe a parameter estimation problem. Several problems can be performed within a DRS run. Each case is uniquely identified by its case ID. There are three secondary keywords that can be used to describe a regression case. The PARAMETERS secondary keyword is used to specify the unary or the binary parameters to be determined in the problem. The DATA-GROUPS sentence defines the data groups to be used. Each data group is identified by its group ID. Since different set of
experimental data are not equally valuable, each data group may be weighted differently depending on its accuracy and its importance to the problem. However, the relative weight specification is optional, and is defaulted to unity. Another way of assigning relative importance to different data groups is via the standard deviation specification discussed in Section 4.2.2. ALGORITHM is an optional keyword that defines the control parameters for the GLSQ algorithm as well as specifying the name of the constraint routine to be used in the evaluation of the constraint functions. All the control parameters have default values. The DRS uses its own general constraint evaluation subroutine unless the users specify that their routines should be used instead. The physical property methods and models to be used in the problem is also defined via the keyword PROPERTIES as explained in section 4.1. The recommended forms of input for these keywords are as follows:

CASE  caseid

PARAMETERS [PURE=pidlist] [BINARY=pidlist]

DATA-GROUPS grid [relwt]/ . . .

ALGORITHM [modelname] [MAXIT=maxit] [TOL=tol]
[IDEM=idem] [IODS=iods] [SS=ss] [SSND=ssnd]

where caseid is the case ID.

pidlist is the list of parameter ID's. The unary parameter ID's are entered via the tertiary keyword PURE while the binary parameter ID's are entered through the tertiary keyword BINARY.
grid is the data group ID.
relwt is the relative weight of the data group.
modelname is the name of the subroutine which calculates the constraint functions given the current values of the estimated parameters and the adjusted variables. The system supplied subroutine can handle all the phase equilibrium and property constraints for any phase equilibrium and property options that are available on the DRS system (refer to Table 4.3). The routine uses a direct indexing approach in locating necessary data, and is therefore very efficient. It is also possible for the users to supply their own subroutines. In that case, the subroutine's name which is a character string of up to six characters must be specified via this keyword. The user-supplied subroutine must do the followings:

(1) Store current values of the parameters into appropriate labelled commons which have the same names as the parameters.

(2) Evaluate appropriate constraint functions, phase equilibrium type and/or property types, using the current values of the parameters and the adjusted variables. The constraint evaluations must be performed for all data points.

maxit is the maximum number of iterations allowed. The default value is 50.
**tol** is the convergence tolerance. Convergence is achieved if the root mean square fractional change in the parameters is less than the tolerance specified. The default value is $10^{-6}$.

**idem** is the Deming method parameters.

If $idem < maxit$, Deming's method is used until convergence is obtained or until idem iterations have been made. The program will then switch to the Britt-Luecke method.

If $idem = -1$, Deming's method is used for all iterations. The exact least squares solution is obtained in only certain special cases.

**iods** is the key for one dimensional search during the Deming iterations.

- $iods = 0$ means that there is no one dimensional search. The program takes $ss$ times the predicted Gauss-Newton step on each iteration.
- $iods = 1$ Search for the minimum in the Gauss-Newton direction on each iteration. On the first iteration, begin the search by taking $ss$ times the predicted Gauss-Newton step. On the subsequent iterations, begin the search with the optimum value of $ss$ from the previous iteration.
- $iods = 2$ Search for the minimum in the Gauss-Newton direction on each iteration. Begin the search on each iteration by taking $ss$ times the predicted Gauss-Newton step.
The default value for iods is 1.

ss is the stepsize parameter. The default value is 0.1.

ssnd is the stepsize to be used in calculating numerical derivatives of the constraints. The default value is $10^{-4}$.

**Example**

**CASE C1**

PARAMETERS PURE=1 2 3

DATA-GROUPS G1 1.5 / G2 1.0

**CASE C2**

PARAMETERS BINARY=4 5

DATA-GROUPS G3

ALGORITHM MYMODEL MAXIT=30 TOL=1d-5

All the keywords explained in this Section are summarized in Table 4.4.
Table 4.1  Example of Input Language.

NEW DRS
TITLE 'Wilson binary parameter estimation'
DESCRIPTION "This is an example of the DRS input language
intended to show its main features"
COMPONENTS  C1 ACETONE/  C2 BENZENE
PROPERTIES SYSOP8;  Specifies property option-set.
PROP-SOURCES
GLOBAL ASPEN1A COMPS=ALL
PARAMETERS

| PARAMETER 1 | PLXANT C2 1 X X & | X X |
| PARAMETER 2 | PLXANT C2 2 X X X X |
| PARAMETER 3 | PLXANT C2 3 X X X X |
| BIPARAMETER 4 | GMWSNA C1 C2 1 X X X X |
| BIPARAMETER 5 | GMWSNA C2 C1 1 X X X X |
| BIPARAMETER 6 | GMWSNB C1 C2 1 X X X X |
| BIPARAMETER 7 | GMWSNB C2 C1 1 X X X X |

DATA-GROUP G1
IN-UNITS TEMPERATURE=C PRESSSURE=TORR
SYSTEM-DEF TPXY C1 C2
PHASE-EQ VL C1 C2
STD-DEV 1 X X X X
DATA 1 X X X X/
    2 X X X X/
    .
    .
    12 X X X X

DATA-GROUP G2
IN-UNITS SI
SYSTEM-DEF TPX C1 C2
PHASE-EQ VL C1 C2
STD-DEV 1 X X X/
    7 X X X
DATA 1 X X X/
    2 X X X/
    .
    .
    15 X X X

CASE C1
PARAMETERS BINARY= 4 5 6 7
DATA-GROUPS G1 2.0/ G2 1.0
PROPERTIES SYSOP8
<table>
<thead>
<tr>
<th>QUANTITY</th>
<th>KEYWORD</th>
<th>SI SET</th>
<th>ENG SET</th>
<th>MET SET</th>
<th>OTHER</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREA</td>
<td>SET</td>
<td>SQM</td>
<td>SQFT</td>
<td>SQM</td>
<td>SQCM  SQIN  SQMILE</td>
</tr>
<tr>
<td>COMPOSITION</td>
<td>SET</td>
<td>MOL-FR</td>
<td>MOL-FR</td>
<td>MOL-FR</td>
<td>MASS-FR</td>
</tr>
<tr>
<td>DENSITY</td>
<td>SET</td>
<td>KG/CUM</td>
<td>LB/CUFT</td>
<td>GM/CC</td>
<td>LB/GAL</td>
</tr>
<tr>
<td>DIFFUSIVITY</td>
<td>SET</td>
<td>SQM/SEC</td>
<td>SQFT/HR</td>
<td>SPCM/SEC</td>
<td>KCAL</td>
</tr>
<tr>
<td>ENERGY</td>
<td>SET</td>
<td>J</td>
<td>BTU</td>
<td>CAL</td>
<td>KWHR</td>
</tr>
<tr>
<td>ENTHALPY</td>
<td>SET</td>
<td>J/KMOL</td>
<td>BTU/LBMOL</td>
<td>CAL/MOL</td>
<td>KWHR/SEC</td>
</tr>
<tr>
<td>ENTROPY</td>
<td>SET</td>
<td>J/KMOL-K</td>
<td>BTU/LBMOL-R</td>
<td>CAL/MOL-K</td>
<td>LB/SEC</td>
</tr>
<tr>
<td>FISCAL</td>
<td>SET</td>
<td>M-DOLL</td>
<td>M-DOLL</td>
<td>M-DOLL</td>
<td>MSG/HR</td>
</tr>
<tr>
<td>FLOW</td>
<td>SET</td>
<td>KG/SEC</td>
<td>LB/HR</td>
<td>KG/HR</td>
<td>L/SEC</td>
</tr>
<tr>
<td>MASS-FLOW</td>
<td>SET</td>
<td>KG/SEC</td>
<td>LB/HR</td>
<td>KG/HR</td>
<td>L/SEC</td>
</tr>
<tr>
<td>MOLE-FLOW</td>
<td>SET</td>
<td>KMOL/SEC</td>
<td>LBMOL/HR</td>
<td>KMOL/HR</td>
<td>L/SEC</td>
</tr>
<tr>
<td>VOLUME-FLOW</td>
<td>SET</td>
<td>CUM/SEC</td>
<td>CUFT/HR</td>
<td>CUM/SEC</td>
<td>MB/HR</td>
</tr>
<tr>
<td>ENERGY-FLOW</td>
<td>SET</td>
<td>WATT</td>
<td>BTU/HK</td>
<td>BTU/HK</td>
<td>MB/HR</td>
</tr>
<tr>
<td>FORCE</td>
<td>SET</td>
<td>NEWTON</td>
<td>LBF</td>
<td>LBF</td>
<td>MB/HR</td>
</tr>
<tr>
<td>HEAT-CAPACIT</td>
<td>SET</td>
<td>J/KMOL-K</td>
<td>BTU/LBMOL-R</td>
<td>LBF</td>
<td>MB/HR</td>
</tr>
<tr>
<td>HEAT-TRANS-COEFF</td>
<td>SET</td>
<td>WATT/SQM-K</td>
<td>BTU/HR-SQFT-R</td>
<td>CAL/MOL-K</td>
<td>L/SEC</td>
</tr>
<tr>
<td>LENGTH</td>
<td>SET</td>
<td>METER</td>
<td>BTU/HR-SQFT-R</td>
<td>MET/SQC-M-K</td>
<td>MF/SEC</td>
</tr>
<tr>
<td>MASS</td>
<td>SET</td>
<td>KG</td>
<td>LG</td>
<td>KG</td>
<td>MB/HR</td>
</tr>
<tr>
<td>POWER</td>
<td>SET</td>
<td>WATT</td>
<td>HP</td>
<td>HP</td>
<td>MB/HR</td>
</tr>
<tr>
<td>PRESSURE</td>
<td>SET</td>
<td>N/SQM</td>
<td>PSI</td>
<td>PSI</td>
<td>MB/HR</td>
</tr>
<tr>
<td>SURFACE-TENSION</td>
<td>SET</td>
<td>N/M</td>
<td>DYNE/CM</td>
<td>DYNE/CM</td>
<td>MB/HR</td>
</tr>
<tr>
<td>TEMPERATURE</td>
<td>SET</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>MB/HR</td>
</tr>
<tr>
<td>THERMAL-CONDUCTIVITY</td>
<td>SET</td>
<td>WATT/M-K</td>
<td>BTU-FT/HR-SQFT-R</td>
<td>KAC-M/HR-SQCM-K</td>
<td>MB/HR</td>
</tr>
<tr>
<td>TIME</td>
<td>SET</td>
<td>SEC</td>
<td>HR</td>
<td>HR</td>
<td>MB/HR</td>
</tr>
<tr>
<td>VELOCITY</td>
<td>SET</td>
<td>M/SEC</td>
<td>FT/SEC</td>
<td>M/SEC</td>
<td>MB/HR</td>
</tr>
<tr>
<td>VISCOSITY</td>
<td>SET</td>
<td>N-SEC/SQM</td>
<td>CUFT</td>
<td>CP</td>
<td>MB/HR</td>
</tr>
<tr>
<td>VOLUME</td>
<td>SET</td>
<td>CUM</td>
<td>RPM</td>
<td>L</td>
<td>MB/HR</td>
</tr>
<tr>
<td>FREQUENCY</td>
<td>SET</td>
<td>HZ</td>
<td>$/KG</td>
<td>$/KG</td>
<td>MB/HR</td>
</tr>
<tr>
<td>UNIT-PRICE</td>
<td>SET</td>
<td>$/LB</td>
<td>$/LB</td>
<td>$/LB</td>
<td>MB/HR</td>
</tr>
<tr>
<td>ENERGY-PRICE</td>
<td>SET</td>
<td>$/BTU</td>
<td>$/BTU</td>
<td>$/BTU</td>
<td>MB/HR</td>
</tr>
<tr>
<td>DELTA-T</td>
<td>SET</td>
<td>K</td>
<td>F</td>
<td>K</td>
<td>MB/HR</td>
</tr>
<tr>
<td>ANGLE</td>
<td>SET</td>
<td>RAD</td>
<td>DEG</td>
<td>DEG</td>
<td>MB/HR</td>
</tr>
<tr>
<td>HEAD</td>
<td>SET</td>
<td>J/KG</td>
<td>FT-LBF/HR</td>
<td>M-KG/KG</td>
<td>MB/HR</td>
</tr>
<tr>
<td>BOND-WORK-INDEX</td>
<td>SET</td>
<td>J-8.5**.5/KG</td>
<td>KWHR-MU**.5/TCN</td>
<td>KWHR-MU**.5/TON</td>
<td>MB/HR</td>
</tr>
</tbody>
</table>
Table 4.3 Data Regression System Problem Specification Options.

State Variable Specification

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>TP</th>
<th>Tx</th>
<th>TPx</th>
<th>TPy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Txy</td>
<td>TPxy</td>
<td>Txx</td>
<td>TPxx</td>
<td>Tpxxy</td>
</tr>
</tbody>
</table>

Property Specification Options

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>Liquid vapor pressure</td>
</tr>
<tr>
<td>PS</td>
<td>Solid vapor pressure</td>
</tr>
<tr>
<td>RHO_p</td>
<td>Density of pure component or mixture</td>
</tr>
<tr>
<td>H_p</td>
<td>Enthalpy of pure compound or mixture</td>
</tr>
<tr>
<td>C_P_p</td>
<td>Heat capacity of pure compound or mixture</td>
</tr>
<tr>
<td>GXS</td>
<td>Excess Gibbs free energy</td>
</tr>
<tr>
<td>HXS</td>
<td>Excess enthalpy</td>
</tr>
<tr>
<td>KVL</td>
<td>Vapor-liquid K value</td>
</tr>
<tr>
<td>GAMMA</td>
<td>Activity coefficients of components in the mixture</td>
</tr>
<tr>
<td>VOLV</td>
<td>Vapor molar volume</td>
</tr>
<tr>
<td>VOLL</td>
<td>Liquid molar volume</td>
</tr>
</tbody>
</table>

Phase Equilibrium Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Possible Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>VL</td>
<td>TPxy, TPx, Txy, Tx</td>
</tr>
<tr>
<td>LL</td>
<td>TPxxxy, TPxx, Txx, Txxxy</td>
</tr>
<tr>
<td>VS</td>
<td>TPy</td>
</tr>
<tr>
<td>LS</td>
<td>TPx, Tx</td>
</tr>
<tr>
<td>VL1</td>
<td>same as VL</td>
</tr>
<tr>
<td>VL2</td>
<td>same as VL1</td>
</tr>
<tr>
<td>LS</td>
<td>same as LS</td>
</tr>
<tr>
<td>L2S</td>
<td>same as LS</td>
</tr>
</tbody>
</table>

Subscript:

p = V for vapor
   = L for liquid
   = S for solid
Table 4.4 Data Regression System Input Language

PARAMETERS

PARAMETER pid pnam cid elem init lb ub [scale]

BI-PARAMETER pid pnam cidl cid2 elem init lb ub [scale]

DATA-GROUP grid

SYSTEM-DEF statespec cidlist [PROP-LIST=proplist]

PHASE-EQ peqtype cidlist / ...

STD-DEV num val-list / ...

DATA num val-list / ...

CASE caseid

PARAMETERS [PURE=pidlist] [BINARY=pidlist]

DATA-GROUPS grid [relwt] / ...

PROPERTIES opsetid

ALGORITHM [modelname] [MAXIT=maxit] [TOL=tol]

[IDEM=idem] [IODS=iods] [SS=ss] [SSND=ssnd]
CHAPTER 5 PARAMETER ESTIMATION: CASE STUDIES

The capability of the data regression system developed in this work is best illustrated by means of sample problems. The problems are divided into two categories: 1) the problems with phase equilibrium specification, and 2) the problems with property specification.

The most important application of any data regression algorithm is probably in the reduction of vapor-liquid equilibrium (VLE) data to determine the model parameters. This is because the prediction of vapor liquid phase equilibrium behaviour is a very important aspect of the design and successful operation of many separation processes. Such predictions often utilize semi-empirical models with several adjustable parameters to represent the thermodynamics of each equilibrium phase. Activity coefficient is such a model generally applied to the liquid phase. Equation of state model, however, can be used to represent both phases. These model parameters are determined by fitting to experimental binary VLE data. The resulting parameters can be used to accurately predict the multi-component VLE behaviour.\(^{(20)}\) This fact is of tremendous importance in chemical technology where multi-component mixtures are more common than binaries.

The VLE parameter estimations are discussed in Sections 5.2 and 5.3 with a brief review of the methods of phase equilibrium calculation presented in Section 5.1.
The determination of parameters for the Antoine and Ideal Gas Heat Capacity equations are chosen as examples for regression problems that involve property specifications. The Antoine parameters estimation is discussed in Section 5.4, while the Ideal Gas Heat Capacity parameters estimation is described in Section 5.5.

5.1 Phase Equilibrium Calculation

For a mixture at equilibrium, it is required that the partial fugacity of each component in both phases are equal. This requirement yields the thermodynamic relationship:

\[ f_i^I = f_i^{II} \]  \hspace{1cm} (5-1)

where I and II represent the two equilibrium phases.

Define \( \phi_i = f_i / (x_i P) \)  \hspace{1cm} (5-2)

Eq. (4-1) becomes,

\[ \phi_i(T, P, x_i^I) x_i^I = \phi_i(T, P, x_i^{II}) x_i^{II} \]  \hspace{1cm} (5-3)

There are two methods of calculating fugacity coefficients: the activity coefficient method and the equation of state method. The first method is generally used to calculate the fugacity coefficient of components in the
condensed phase while the second method can be applied to both the vapor and the condensed phases.

The first method is discussed in Section 5.2 and the second method is treated in Section 5.3.

5.2 Vapor Liquid Equilibrium Calculation using activity Coefficient Model

4.2.1 Discussion

For the activity coefficient method, Eq. (5-3) becomes,

\[ \phi_i^V(T, P, y, \theta) = \gamma_i(T, x, \theta) x_i \phi_i^{OL}(T, P) \]  (5-4)

The vapor fugacity coefficient \( \phi_i^V \) is a function of T, P, mole fraction y, and unknown parameters, but usually can be assumed to be unity at sufficiently low pressure. The pure component standard state fugacity coefficient \( \phi_i^{OL} \) is a function of T and P.

4.2.2 Application to parameter evaluation

Analysis of the isothermal data of Brown and Smith\(^7\) for the system Acetone(1)/Benzene(2) illustrates the application of the data regression system for analysis of vapor-liquid equilibrium data. The vapor fugacity coefficient is assumed to be unity since the pressure is sufficiently low. The standard-state fugacity is also assumed to be equal to pure component vapor pressure. The Wilson equation\(^{30}\) is used to
calculate the liquid-phase activity coefficients. The equation is expressed as a function of temperature and composition by

\[
\ln \gamma_1 = 1 - \ln(x_1 + A_{12}x_2) - \frac{x_1}{x_1 + A_{12}x_2} - \frac{A_{21}x_1}{A_{21}x_1 + x_2}
\]

\[
\ln \gamma_2 = 1 - \ln(A_{21}x_1 + x_2) - \frac{A_{12}x_1}{x_1 + A_{12}x_2} - \frac{x_2}{A_{21}x_1 + x_2}
\]

(5-5)

where \(A_{12}\) and \(A_{21}\) are the adjustable binary parameters.

Since these parameters are known to be temperature dependent, they are represented in the following form:

\[
A_{ij} = \exp(a_{ij} + b_{ij}/T)
\]

(5-6)

where \(T\) is absolute temperature. Therefore, for the Acetone(1)/Benzene(2) system, there are the maximum of four adjustable parameters, \(a_{12}, b_{12}, a_{21}, \) and \(b_{21}\). In practice, unless there are several sets of isothermal data, \(a_{12}\) and \(a_{21}\) are set equal to zero.

In general, the measured quantities for binary vapor-liquid equilibrium system are temperature \(T\), pressure \(P\), liquid phase mole fraction \(x\), and vapor phase mole fraction \(y\). The Generalized Least-Squares Criterion with the maximum likelihood estimates for VLE data reduction is therefore

\[
I = \sum_{i=1}^{k} \left( \frac{(T_i, M - T_i)^2}{\sigma_{T_i}^2} + \frac{(P_i, M - P_i)^2}{\sigma_{P_i}^2} + \frac{(x_{1i, M} - x_{1i})^2}{\sigma_{x_{1i}}^2} \right. \\
\left. + \frac{(y_{1i, M} - y_{1i})^2}{\sigma_{y_{1i}}^2} \right)
\]

(5-7)
The constraints for equation (5-7) are the phase equilibrium relationships, equation (5-4). The estimated parameters and their significance are discussed in Section 6.1

In the present example, the data used in the determination of the Wilson parameters are T-P-x-y data. Occasionally, however, the VLE data found in the literature are T-P-x or T-x data only. In such cases, the generalized least-squares criterion in eqs. (5-4) and (5-7) do not apply as stated, since the y information is required to calculate the left hand side of eq. (5-4). In the present version of DRS, estimates of the missing information, P or y, are being supplied to the Britt-Luecke's algorithm so that the generalized least-squares criterion as stated above, are still fully applicable. However, very large values of standard deviations are assigned to these estimated data. Assigning large deviation to a variable means the information provided by that variable is of lesser importance to the GLSC. In effect, that variable is excluded from the summation in eq. (5-4). In such cases, the resultant parameter estimates are expected to be less reliable or less accurate than those obtained if T-P-x-y data were used. It should be pointed out that the Barker's ordinary least-squares method utilizes only the T-P-x data in the parameter determination process. Therefore, the y information, even if available, is unimportant to the Barker's method. The Britt-Luecke's generalized least-squares method is equivalent to the Barker's method if the standard deviation of
T and x are set equal to zero. For the case with only T-P-x data, it is therefore interesting to compare the results obtained from using both methods.

The same set of isothermal data for Acetone(1)/Benzene(2) is used with the exception that only the T-P-x data are the input to the DRS. The Wilson binary parameters are determined using the Britt-Luecke's generalized least-squares method and the Barker's method. The results are discussed in section 6.1.

5.3 Vapor Liquid Equilibrium Calculation Using Equation of State

5.3.1 Discussion

This method of phase equilibrium calculation relies solely on an equation of state which is applicable to both the vapor and the condensed phases in calculating fugacity coefficients. From eqs. (5-4), the phase equilibrium relationship is:

$$y_i \phi_i^V(T,P,y) = x_i \phi_i^L(T,P,x)$$ (5-8)

This approach is particularly useful because the difficulty in specifying standard-state for the fugacity when there are supercritical component in the mixture can be avoided. However, its limitation lies in the fact that many of the equations of state currently available are not accurate for all densities, or when applied to polar compounds or large
molecules. Nevertheless, for many practical applications, this method has proved very successful.

Generally the one-fluid theory is assumed when an equation of state is applied to a mixture, that is the equation of state of the mixture is the same as that of a pure fluid except that the characteristic constants depend on composition according to some arbitrary mixing rules. Usually at least one of these mixing rules contain an adjustable binary parameter which must be determined by fitting to binary data. The calculated results are often highly sensitive to the mixing rules and the values of the adjustable parameters. Therefore, the need to determine the best adjustable parameters for the equation of state chosen to do the VLE calculation is apparent.

5.3.2 Equation of State Binary Parameter Estimation

The binary-interaction parameters for any equations of state that are available in ASPEN physical property system can be determined readily by the DRS using VLE data. For an illustrative purpose, the binary parameters $k_{ij}$ for the Redlich-Kwong-Soave (22) and the Peng-Robinson (18) equations of state are determined from $T$, $P$, $x$, and $y$ data of the Acetone(1)/Benzene(2) system. (7) The generalized least-squares criterion for this case is the same as that for the Wilson parameters estimation discussed in Section 5.2.2. However, the constraint equations are slightly different; the activity coefficients are being replaced by the fugacity
coefficients:

\[ Y_i \phi_i^V(T, P, Y_i, \theta) = x_i \phi_i^L(T, P, x_i, \theta) \]  

(5-9)

The results for both equations of state are discussed in Section 6.2.

5.3.3 Equation of State Pure Parameter Estimation

Generally, other parameters are also required by an equation of state, for example, critical temperature, critical pressure, acentric factor, orientation parameter etc. Some of these parameters can be determined experimentally, such as the critical temperature and critical pressure except for compounds with very high molecular weight. Other parameters must be determined by fitting to experimental data. Acentric factor can be determined from the definition, if the vapor pressure at the reduced temperature of 0.7 is available.

\[ \omega = -\ln\left(\frac{P'}{P_C}\right)_{T_r=0.7} - 1.0 \]  

(5-10)

However, better acentric factor values can be determined to fit a specific equation of state by regressing on vapor pressure data which cover a wide range of temperature. Characteristic temperature \( T^* \), characteristic volume \( v^* \), and the parameter \( c \) used in the Perturbed-Hard-Chain equation of state\(^{(21)} \) are determined by vapor pressure and density data reduction. Other
thermodynamic properties such as enthalpy departure were also used, together with vapor pressure and density, to determine the parameters required by the modified BWR equation of state.\(^{(24)}\) In this example, equation of state unary parameter estimation using several types of data is illustrated. It is intended to show a feature of the DRS that the data used in a regression problem can come from several data groups, each with different data-type. The determination of acentric factor for the Redlich-Kwong-Soave equation of state is chosen as an example. Vapor pressure and density data of Vargaflok\(^{(25)}\) for 1-Butene are used. The GLSC for this case is

\[
I = \sum_{j=1}^{k} \left( \frac{(T_j, M - T_j)^2}{\sigma^2_{T_j}} + \frac{(P'_j, M - P'_j)^2}{\sigma^2_{P'_j}} + \frac{(\rho_j, M - \rho_j)^2}{\sigma^2_{\rho_j}} \right) \quad (5-11)
\]

subject to the following constraints:

\[
\phi^V(T, P, \theta) y_i = \phi^L(T, P, \theta) x_i \quad (5-12)
\]

\[
\rho_j, M - \rho_j = 0 \quad (5-13)
\]

The results are presented in Section 5.3.

5.4 Vapor Pressure Parameters Estimation

The most widely used vapor pressure correlation is the Antoine equation.\(^{(24)}\)

\[
\ln P' = A + B/(T+C) \quad (5-14)
\]
Other modification of this equation includes the Rankine vapor pressure equation which has an additional term involving logarithm of temperature,

\[ \ln P' = A + B/T + C \ln(T) \]  \hspace{1cm} (5-15)

Further modification results in the Extended Antoine equation

\[ \ln P' = A + B/(T+C) + D \ln(T) + ET + FT^N \]  \hspace{1cm} (5-16)

which is proved to be extremely accurate in representing vapor pressure data for a wide range of compounds.

The extended Antoine equation reduces to the Antoine equation if the parameters D, E, and F are set equal to zero.

The parameters appeared in the above equations are determined by fitting to experimental vapor pressure data. Eq. (5-16) is chosen as an example. The generalized least-squares criterion for this case is

\[ I = \sum_{j=1}^{k} \left( \frac{T_j - M_j}{\sigma_{T_j}} \right)^2 + \left( \frac{P'_j - M_j}{\sigma_{P'_j}} \right)^2 \]  \hspace{1cm} (5-17)

and the constraint functions are

\[ P'_{j,M} - P'_{j} = 0 \quad j=1,\ldots,k. \]  \hspace{1cm} (5-18)

The results are discussed in Section 6.4.
5.5 **Ideal Gas Heat Capacity Parameter Estimation**

The most widely used ideal gas heat capacity correlation is probably the polynomial heat capacity equation

\[ C_{p}^{IG} = c_1 + c_2 T + c_3 T^2 + c_4 T^3 + c_5 T^4 + c_6 T^5 \]  \hspace{1cm} (5-19)

The parameters \(c_1\) through \(c_6\) can be determined by fitting to experimental ideal gas heat capacity data. The data of Vargaftik\(^{(25)}\) are used to determine these parameters for 1-Butene. The generalized least-squares criterion for this problem is:

\[
I = k \sum_{j=1}^{k} \left( \frac{(T_{j,M} - T_{j})^2}{\sigma_{T_j}^2} + \frac{(C_{p,j}^{IG} - C_{p,j}^{IG})^2}{\sigma_{C_{p,j}}^2} \right) \]  \hspace{1cm} (5-20)

and the constraint functions are:

\[ C_{p,j}^{IG, M} - C_{p,j}^{IG} = 0 \hspace{1cm} j=1, \ldots, k \]  \hspace{1cm} (5-21)

The results are presented in Section 6.5.
CHAPTER 6 RESULT AND DISCUSSION

6.1 Evaluation of Parameters for Wilson Activity Coefficient Model.

The vapor-liquid equilibrium data of Acetone(l)/Benzene(2)\(^{7}\) were used in the determination of binary parameters for the Wilson activity coefficient model as described in section 5.2.2. The assumed standard deviations in the measured variables were: \(\sigma_T = 0.3^\circ\text{C}\), \(\sigma_P = 1.0\ \text{mm Hg}\), \(\sigma_x = 0.001\), and \(\sigma_y = 0.01\). The parameters \(a_{12}\) and \(a_{21}\) were set equal to zero. The input data is shown in Table 6.1. The resultant parameters are shown in Table 6.2. Table 6.3 gives the measured data, the estimates of the true values, and deviations of the measured values from model predictions. The phase diagram of the estimates of the true values of the variables, together with the measured data are shown in Figure 6.1.

When two parameters are determined, it is useful to plot the confidence ellipses of the estimate parameters. These are obtained from the eigenvalues \(\lambda\), and eigenvectors \(V\) of the variance-covariance matrix of the parameters\(^{30}\) as discussed in Section 2.3. The eigenvalues are 18781.76 and 65.54. The major axis of the ellipse is therefore, given by the eigenvector \(b_{21} = -3.8572b_{12}\), and the minor axis is given by \(b_{21} = 0.2593b_{12}\). The confidence ellipse for this example are shown in Figure 6.2. The region shown in the
Figure represents the areas within which the parameters can be expected to lie at the confidence level associated with the ellipse. The small confidence region observed indicates that the model fits the data quite well and the random error associated with the data is also small.

As discussed in 5.2.2, when the experimental data to be used in the estimation of activity coefficient parameters consist only of T-P-x or T-x, it is necessary to supply the generalized least-squares algorithm with the estimates of the missing P or y data. Large standard deviation values are also assigned to the estimates. The DRS checks for missing VLE data (P or y) and automatically furnishes the regression algorithm with reasonable estimates of the data. A set of T-P-x data is taken from the T-P-x-y data for Acetone(1)/Benzene(2) used in the previous section. They are used to determine the Wilson binary parameters by the Generalized least-squares method as described in 5.2.2. The assumed standard deviations in T, P, and x are: $\sigma_T = 0.3^\circ$C, $\sigma_P = 1.0$ mm-Hg, and $\sigma_x = 0.001$. The input in Table 6.1 is also used for this problem. The results are summarized in Tables 6.4 and 6.5. These results are also compared with those obtained using the Barker's ordinary least-squares method. Evidently, the parameter values obtained using both methods are very close to one another. In fact, one lies in the confidence region of the other. Since the measured values of y are actually known, the difference between these values and the estimates of true y that both
methods generated, is a measure of how well the estimated parameters of each method can represent the VLE data. The measured $y$ and $P$ together with the estimates of the true values resulted from both methods are summarized in Table 6.5. Both methods can reproduce $y$ data equally well, though not nearly as well as the previous case where all $T$-$P$-$x$-$y$ data were used. For the prediction of $P$, however, the generalized least-squares method is somewhat better. This example shows that it is possible to determine reasonably good estimates of binary parameters even when the $y$ data is not available. However, it is always desirable to use all the available VLE information in the estimation of binary parameters because better parameters can be obtained.

6.2 Equations of State Binary-interaction Parameters

The same set of VLE data for Acetone(1)/Benzene(2)\(^{(7)}\) that was used in the determination of Wilson parameter was used to estimate the binary-interaction parameter of the Redlich-Kwong-Soave equation of state. The assumed standard deviations for the measured variables are $\sigma_T = 0.3^\circ C$, $\sigma_P = 1.0$ mm-Hg, $\sigma_x = 0.001$, and $\sigma_y = 0.01$, as before. The input data for this case is shown in Table 6.6. The resultant parameters $k_{12}$ is 0.03456 with a parameter standard deviation of 0.001466. The estimates of the true values of the measured variables are summarized in Table 6.7.
The binary-interaction parameters for any other equations of state that are available in the ASPEN physical property system can be determined in a similar manner. The binary-interaction parameters for the Peng-Robinson equation was determined by the reduction of the same set of VLE data. The same standard deviation values were also assumed. The input data is given in Table 6.8. The estimated parameter as well as the residuals of the observed variables are given in Table 6.9. In both cases, the equations of state can represent the VLE data very well. The standard errors in the estimated parameters are also very small, indicating that the estimated parameters are quite accurate.

6.3 Equation of State Pure Parameters Determination

The acentric factor for the Redlich-Kwong-Soave equation of state was determined by the reduction of vapor pressure and density data\(^{(25)}\) for 1-Butene. This example is intended to demonstrate the capability of the DRS in combining different data groups of different data types for a particular regression task. In this case, two data groups were used, one contains vapor pressure data, the other contains density data. The input data are shown in Table 6.10. The assumed standard deviations for the measured data are: \(\sigma_T = 0.5\) K, \(\sigma_P = 0.1\) mm Hg, and \(\sigma_P = 0.01\) gm/cc. The results are summarized in Table 6.11. The estimate standard deviation of the estimates of true values of density is 0.05 gm/cc. If the assumed
standard deviations had been based on analysis of replicated experiments, then this inconsistency could indicate either systematic error in the data or lack of fit of the model to the data. In this case, however, they were a priori estimates, and the results of the parameter estimation procedure serve merely to provide better estimate of the standard deviations.

6.4 **Vapor Pressure Parameters Estimation**

The parameters for the Antoine and Extended Antoine vapor pressure equations were determined by fitting to vapor pressure data of n-Butane. The parameters A, B, and D for the Extended antoine equation were determined. The parameter C was set equal to zero. The input data and results for this case are shown in Tables 6.12 and 6.13.

6.5 **Ideal Gas Heat Capacity Parameters Estimation**

The coefficients in the ideal gas heat capacity equation are determined using ideal gas heat capacity of 1-Betene. The equation is fitted only to the third power in temperature, that is the coefficients $c_5$ and $c_6$ are set equal to zero. The assumed standard deviation for temperature is $\sigma_T = 0.5K$ and for $Cp^{IG}$ is $\sigma_{Cp} = 50.0 \text{ J/kmole-K}$. The input data is shown in Table 6.14. The resultant parameters are shown in Table 6.15 together with the parameter variance-covariance matrix and the estimates of the true values of the measured variables.
Figure 6.1 Calculated phase diagram for Acetone(1)/Benzene(2) system.
Figure 6.2 Confidence ellipses for Wilson parameters. Acetone(1)/Benzene(2) system at 45°C.
NEW DRS
TITLE 'WILSON PARAMETERS ESTIMATION'
COMPONENTS C3H6O ACETONE / C6H6 BENZENE
PROPERTIES SYSOP8
PROP-SOURCES
  GLOBAL ASPEN1A COMPS=ALL
PROP-DATA
  COMP-LIST C3H6O C6H6
  BCVAL GMWSNA 1 1 C3H6O 0.0 0.0
  BCVAL GMWSNA 1 1 C6H6 0.0 0.0
PAR AMETERS
  BIPARAMETER 1 GMWSNB C3H6O C6H6 120. -40000. 40000. 1.0
  BIPARAMETER 2 GMWSNB C6H6 C3H6O 1 -417. -40000. 40000. 1.0
DATA-GROUP G1
  IN-UNITS TEMP=C PRES=TORR
  SYSTEM-DEF TPXY C3H6O C6H6
  PHASE-EQ VL C3H6O C6H6
  STD-DEV 1 0.3 1.0 0.001 0.01
  DATA
  1 45.0 250.73 0.0470 0.144/
  2 45.0 275.02 0.0963 0.2574/
  3 45.0 324.25 0.2207 0.4417/
  4 45.0 348.40 0.2936 0.5204/
  5 45.0 379.88 0.4011 0.6139/
  6 45.0 399.79 0.4759 0.6697/
  7 45.0 432.95 0.6125 0.7614/
  8 45.0 453.99 0.7045 0.8201/
  9 45.0 475.39 0.8081 0.8805/
 10 45.0 495.32 0.9084 0.9418/
 11 45.0 503.96 0.9529 0.9699
DATA-GROUP G2
  IN-UNITS TEMP=C PRES=TORR
  SYSTEM-DEF TPX C3H6O C6H6
  PHASE-EQ VL C3H6O C6H6
  STD-DEV 1 0.3 1.0 0.001
  DATA
  1 45.0 250.73 0.0470 /
  2 45.0 275.02 0.0963 /
  3 45.0 324.25 0.2207 /
  4 45.0 348.40 0.2936 /
  5 45.0 379.88 0.4011 /
  6 45.0 399.79 0.4759 /
  7 45.0 432.95 0.6125 /
  8 45.0 453.99 0.7045 /
  9 45.0 475.39 0.8081 /
 10 45.0 495.32 0.9084 /
 11 45.0 503.96 0.9529
DATA-GROUP G3
  IN-UNITS TEMP=C PRES=TORR
  SYSTEM-DEF TPX C3H6O C6H6
  PHASE-EQ VL C3H6O C6H6
  STD-DEV 1 0.0 1.0 0.0
  DATA
  1 45.0 250.73 0.0470 /
  2 45.0 275.02 0.0963 /
  3 45.0 324.25 0.2207 /
  4 45.0 348.40 0.2936 /
  5 45.0 379.88 0.4011 /
  6 45.0 399.79 0.4759 /

Table 6.1 Wilson parameter estimation input data
Table 6.1 (continue)

```
7  45.0  432.95  0.6125  /
8  45.0  453.99  0.7045  /
9  45.0  475.39  0.8081  /
10  45.0  495.32  0.9084  /
11  45.0  503.96  0.9529  /
```

CASE C1
PARAMETERS BINARY=1 2
DATA-GROUPS G1 1.0
PROPERTIES SYSOP8
ALGORITHM DRCSTR MAXIT=50 TOL=1D-6 IODEM=1 SS=1 SSND=1D-4

CASE C2
PARAMETERS BINARY=1 2
DATA-GROUPS G2 1.0
PROPERTIES SYSOP8
ALGORITHM DRCSTR MAXIT=50 TOL=1D-6 IODEM=1 SS=.7 SSND=1D-4

CASE C3
PARAMETERS BINARY=1 2
DATA-GROUPS G3 1.0
PROPERTIES SYSOP8
ALGORITHM DRCSTR MAXIT=50 TOL=1D-6 IODEM=1 SS=1 SSND=1D-4

OUT-UNITS TEMP=C PRES=TORR
HISTORY MSG-LEVEL SIM=4
Table 6.2  Estimated Wilson Binary Parameters from the Reduction of VLE Data for Acetone(1)/Benzene(2). (7)

Parameter estimates
\[ b_{12} = 188.2751 \]
\[ b_{21} = -556.2737 \]

Variance-covariance matrix
\[
\begin{pmatrix}
1244.3 & -4546.7 \\
-4546.7 & 17603.0
\end{pmatrix}
\]

Parameter correlation matrix
\[
\begin{pmatrix}
1.000 & -0.9715 \\
-0.9715 & 1.000
\end{pmatrix}
\]

Table 6.4  Comparison of Estimated Wilson Parameters from the Britt-Luecke's and the Barker's Methods, Using the T-P-x Data of Acetone(1)/Benzene(2). (7)

Britt-Luecke's method          Barker's method

Parameter estimates
\[ b_{12} = 210.3660 \]
\[ b_{21} = -637.7195 \]
\[ 216.9074 \]
\[ -665.3300 \]

Variance-covariance matrix
\[
\begin{pmatrix}
527.38 & -2349.6 \\
-2349.6 & 11215.0
\end{pmatrix}
\]
\[
\begin{pmatrix}
474.46 & -2135.0 \\
-2135.0 & 10376.0
\end{pmatrix}
\]

Parameter correlation matrix
\[
\begin{pmatrix}
1.000 & -0.96612 \\
-0.96612 & 1.000
\end{pmatrix}
\]
\[
\begin{pmatrix}
1.00000 & -0.96225 \\
-0.96225 & 1.000
\end{pmatrix}
\]
Table 6.3 Measured Variables and Estimates of Their True Values for Acetone(1)/Benzene(2) (7) System in Wilson Parameters Determination.

<table>
<thead>
<tr>
<th>Temperature, °C</th>
<th>Pressure, mm Hg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>1</td>
<td>45.0</td>
</tr>
<tr>
<td>2</td>
<td>45.0</td>
</tr>
<tr>
<td>3</td>
<td>45.0</td>
</tr>
<tr>
<td>4</td>
<td>45.0</td>
</tr>
<tr>
<td>5</td>
<td>45.0</td>
</tr>
<tr>
<td>6</td>
<td>45.0</td>
</tr>
<tr>
<td>7</td>
<td>45.0</td>
</tr>
<tr>
<td>8</td>
<td>45.0</td>
</tr>
<tr>
<td>9</td>
<td>45.0</td>
</tr>
<tr>
<td>10</td>
<td>45.0</td>
</tr>
<tr>
<td>11</td>
<td>45.0</td>
</tr>
</tbody>
</table>

Standard deviation 0.39

<table>
<thead>
<tr>
<th>Liquid composition, x_1</th>
<th>Vapor composition, y_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>1</td>
<td>0.0470</td>
</tr>
<tr>
<td>2</td>
<td>0.0963</td>
</tr>
<tr>
<td>3</td>
<td>0.2207</td>
</tr>
<tr>
<td>4</td>
<td>0.2936</td>
</tr>
<tr>
<td>5</td>
<td>0.4011</td>
</tr>
<tr>
<td>6</td>
<td>0.4759</td>
</tr>
<tr>
<td>7</td>
<td>0.6125</td>
</tr>
<tr>
<td>8</td>
<td>0.7045</td>
</tr>
<tr>
<td>9</td>
<td>0.8081</td>
</tr>
<tr>
<td>10</td>
<td>0.9084</td>
</tr>
<tr>
<td>11</td>
<td>0.9529</td>
</tr>
</tbody>
</table>

Standard deviation 0.00004

* Dev = calc - meas
Table 6.5  Measured Variables and Estimates of Their True Values for Acetone(1)/Benzene(2) System. Only T-P-x data were used.

<table>
<thead>
<tr>
<th>Temp, °C</th>
<th>Brit-Luecke's method</th>
<th>Barker's method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>1</td>
<td>45.0</td>
<td>250.73</td>
</tr>
<tr>
<td>2</td>
<td>45.0</td>
<td>275.02</td>
</tr>
<tr>
<td>3</td>
<td>45.0</td>
<td>324.25</td>
</tr>
<tr>
<td>4</td>
<td>45.0</td>
<td>348.40</td>
</tr>
<tr>
<td>5</td>
<td>45.0</td>
<td>379.88</td>
</tr>
<tr>
<td>6</td>
<td>45.0</td>
<td>399.79</td>
</tr>
<tr>
<td>7</td>
<td>45.0</td>
<td>432.95</td>
</tr>
<tr>
<td>8</td>
<td>45.0</td>
<td>453.99</td>
</tr>
<tr>
<td>9</td>
<td>45.0</td>
<td>475.39</td>
</tr>
<tr>
<td>10</td>
<td>45.0</td>
<td>495.32</td>
</tr>
<tr>
<td>11</td>
<td>45.0</td>
<td>503.96</td>
</tr>
</tbody>
</table>

standard deviation 1.02  6.32

<table>
<thead>
<tr>
<th>x1</th>
<th>Vapor composition, y1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Meas</td>
</tr>
<tr>
<td>1</td>
<td>0.0470</td>
</tr>
<tr>
<td>2</td>
<td>0.0963</td>
</tr>
<tr>
<td>3</td>
<td>0.2207</td>
</tr>
<tr>
<td>4</td>
<td>0.2936</td>
</tr>
<tr>
<td>5</td>
<td>0.4011</td>
</tr>
<tr>
<td>6</td>
<td>0.4759</td>
</tr>
<tr>
<td>7</td>
<td>0.6125</td>
</tr>
<tr>
<td>8</td>
<td>0.7045</td>
</tr>
<tr>
<td>9</td>
<td>0.8081</td>
</tr>
<tr>
<td>10</td>
<td>0.9084</td>
</tr>
<tr>
<td>11</td>
<td>0.9529</td>
</tr>
</tbody>
</table>

standard deviation 0.0265  0.0252

* Dev = Calc - Meas

-93-
NEW DRS
TITLE 'REDLICK-KWONG-SOAVE EOS BINARY-PARAMETER ESTIMATION '
COMPONENTS C3H60 ACETONE /C6H6 BENZENE
PROPERTIES SYSOP3
PROP-SOURCES GLOBAL ASPENIA COMPS=ALL
PROP-DATA COMP-LIST C3H60 C6H6
    BCVAL RKSKIJ 1 1 C3H60 .0 .1
    BCVAL RKSKIJ 1 1 C6H6 .1 .0
PARAMETERS
BIPARAMETER 1 RKSKIJ C6H6 C3H60 1 .1 -1.0 1.0 1.0
DATA-GROUP GI
    INUNITS TEMP=C PRES=TORR
    SYSTEM-DEF TPXY C3H60 C6H6
    PHASE-EQ VL C3H60 C6H6
    STD-DEV 1 0.3 1.0 0.001 0.01
    DATA 1 45.0 250.73 0.0470 0.1444/
      2 45.0 275.02 0.0963 0.2574/
      3 45.0 324.25 0.2207 0.4417/
      4 45.0 348.40 0.2336 0.5204/
      5 45.0 379.88 0.4011 0.6139/
      6 45.0 399.79 0.4759 0.6697/
      7 45.0 432.95 0.6125 0.7614/
      8 45.0 453.99 0.7045 0.8201/
      9 45.0 475.39 0.8081 0.8805/
     10 45.0 495.32 0.9064 0.9418/
     11 45.0 503.96 0.9529 0.9699
CASE C1
    PARAMETERS BINARY=1
    DATAGROUPS GI 1.0
    PROPERTIES SYSOP3
    ALGORITHM DRCSTR MAXIT=50 TOL=1D-6 IDEM=10 IODS=0 SS=1 SSND=1D-4
    OUT-UNITS TEMP=C PRES=TORR
    HISTORY MSG-LEVEL SIM=4

Table 6.6 Redlich-Kwong-Soave equation of state binary parameter estimation input data.
Table 6.7  Measured Variables and Estimates of Their True Values for Acetone(1)/Benzene(2) System. Binary parameter estimation for the Redlich-Kwong-Soave Equation of State

<table>
<thead>
<tr>
<th>Temperature, °C</th>
<th>Pressure, mm Hg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>45.0</td>
<td>45.28</td>
</tr>
<tr>
<td>45.0</td>
<td>45.24</td>
</tr>
<tr>
<td>45.0</td>
<td>44.98</td>
</tr>
<tr>
<td>45.0</td>
<td>44.90</td>
</tr>
<tr>
<td>45.0</td>
<td>44.87</td>
</tr>
<tr>
<td>45.0</td>
<td>44.91</td>
</tr>
<tr>
<td>45.0</td>
<td>45.06</td>
</tr>
<tr>
<td>45.0</td>
<td>45.23</td>
</tr>
<tr>
<td>45.0</td>
<td>45.38</td>
</tr>
<tr>
<td>45.0</td>
<td>45.58</td>
</tr>
<tr>
<td>45.0</td>
<td>45.69</td>
</tr>
</tbody>
</table>

std dev: 0.35 | 0.28

<table>
<thead>
<tr>
<th>Liquid composition, x_1</th>
<th>Vapor composition, y_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>0.0470</td>
<td>0.0472</td>
</tr>
<tr>
<td>0.0963</td>
<td>0.0964</td>
</tr>
<tr>
<td>0.2207</td>
<td>0.2206</td>
</tr>
<tr>
<td>0.2936</td>
<td>0.2935</td>
</tr>
<tr>
<td>0.4011</td>
<td>0.4010</td>
</tr>
<tr>
<td>0.4759</td>
<td>0.4758</td>
</tr>
<tr>
<td>0.6125</td>
<td>0.6125</td>
</tr>
<tr>
<td>0.7045</td>
<td>0.7046</td>
</tr>
<tr>
<td>0.8081</td>
<td>0.8082</td>
</tr>
<tr>
<td>0.9084</td>
<td>0.9085</td>
</tr>
<tr>
<td>0.9529</td>
<td>0.9530</td>
</tr>
</tbody>
</table>

std dev: 0.0001 | 0.0071

Parameter estimates:  k_{12} = 0.0346

parameter variance = .21496d-05

* Dev = calc - meas
Table 6.8 Input data for the estimation of Peng-Robinson equation of state binary parameters.
Table 6.9  Measured Variables and Estimates of Their True Values for Acetone(1)/Benzene(2) (7) System. Binary parameter estimation for the Peng-Robinson Equation of State

<table>
<thead>
<tr>
<th>Temperature, °C</th>
<th>Pressure, mm Hg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meas</strong></td>
<td><strong>Calc</strong></td>
</tr>
<tr>
<td><strong>Meas</strong></td>
<td><strong>Calc</strong></td>
</tr>
<tr>
<td>45.0</td>
<td>44.80</td>
</tr>
<tr>
<td>45.0</td>
<td>44.92</td>
</tr>
<tr>
<td>45.0</td>
<td>44.95</td>
</tr>
<tr>
<td>45.0</td>
<td>44.93</td>
</tr>
<tr>
<td>45.0</td>
<td>44.92</td>
</tr>
<tr>
<td>45.0</td>
<td>44.94</td>
</tr>
<tr>
<td>45.0</td>
<td>45.01</td>
</tr>
<tr>
<td>45.0</td>
<td>45.10</td>
</tr>
<tr>
<td>45.0</td>
<td>45.17</td>
</tr>
<tr>
<td>45.0</td>
<td>45.28</td>
</tr>
<tr>
<td>45.0</td>
<td>45.33</td>
</tr>
</tbody>
</table>

| **Meas**       | **Calc**       | **Dev** |
|----------------|----------------|
| 250.73         | 251.50         | 0.77    |
| 275.02         | 275.27         | 0.25    |
| 324.25         | 324.39         | 0.14    |
| 348.40         | 348.59         | 0.18    |
| 379.88         | 380.09         | 0.21    |
| 399.79         | 399.94         | 0.15    |
| 432.95         | 432.93         | -0.02   |
| 453.99         | 453.77         | -0.22   |
| 475.39         | 475.03         | -0.36   |
| 495.32         | 494.78         | -0.54   |
| 503.96         | 503.32         | -0.64   |

std dev 0.17  0.41

<table>
<thead>
<tr>
<th>Liquid composition, $x_1$</th>
<th>Vapor composition, $y_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meas</strong></td>
<td><strong>Calc</strong></td>
</tr>
<tr>
<td><strong>Meas</strong></td>
<td><strong>Calc</strong></td>
</tr>
<tr>
<td>0.0470</td>
<td>0.0470</td>
</tr>
<tr>
<td>0.0963</td>
<td>0.0965</td>
</tr>
<tr>
<td>0.2207</td>
<td>0.2207</td>
</tr>
<tr>
<td>0.2936</td>
<td>0.2936</td>
</tr>
<tr>
<td>0.4011</td>
<td>0.4011</td>
</tr>
<tr>
<td>0.4759</td>
<td>0.4759</td>
</tr>
<tr>
<td>0.6125</td>
<td>0.6125</td>
</tr>
<tr>
<td>0.7045</td>
<td>0.7046</td>
</tr>
<tr>
<td>0.8081</td>
<td>0.8082</td>
</tr>
<tr>
<td>0.9084</td>
<td>0.9085</td>
</tr>
<tr>
<td>0.9529</td>
<td>0.9530</td>
</tr>
</tbody>
</table>

std dev 0.0001  0.0072

parameter estimate: $k_{12} = 0.02759$

parameter variance = 0.77576d-06

* Dev = calc - meas
NEW DRS
TITLE 'REDLICH-KWONG-SOAVE EOS ACENTRIC FACTOR ESTIMATION'
COMPONENTS C4H10 BUTANE / C4H8 BUTENE
PROPERTIES SYSOP3
PROP-SOURCES
GLOBAL ASPEN1A COMPS=ALL
PARAMETERS
PARAMETER 1 OMGRKS C4H10 1 0.30 0.10 1.0 1.0
PARAMETER 2 OMGRKS C4H8 1 0.20 0.10 1.0 1.0
DATAGROUP G1
INUNITS TEMP=R PRES=PSI
SYSTEM-DEF TP C4H8
PHASE-EQ VL C4H8
STD-DEV 1 0.5 0.1
DATA
1 257.6698 0.0006
2 275.6698 0.0029
3 293.6698 0.0108
4 311.6698 0.0342
5 329.6698 0.0928
6 347.6698 0.2230
7 365.6698 0.4844
8 383.6698 0.9678
9 401.6698 1.7991
10 419.6698 3.1480
11 437.6698 5.2277
12 455.6698 8.2968
13 473.6698 12.6579
14 491.6698 18.6374
15 509.6698 26.6579
16 527.6698 37.0862
17 536.6698 43.3807
18 545.6698 50.3570
19 563.6698 67.0073
20 581.6698 87.3127
21 599.6698 110.6638
22 617.6698 139.0912
23 635.6698 173.0301
24 653.6698 212.9155

DATAGROUP G2
INUNITS TEMP=R PRES=PSI DENS='LB/CUFT'
SYSTEM-DEF TP C4H8 PROP-LIST=RHOL
STD-DEV 1 0.5 0.1 0.01
DATA
1 347.6698 0.2230 0.7798
2 355.6698 0.4844 0.7698
3 383.6698 0.9678 0.7581
4 401.6698 1.7991 0.7467
5 419.6698 3.1480 0.7351
6 437.6698 5.2277 0.7231
7 455.6698 8.2968 0.7109
8 473.6698 12.6579 0.7009
9 491.6698 18.6374 0.6885
10 509.6698 26.6579 0.6727
11 527.6698 37.0862 0.6594

Table 6.10 Input data for Acentric factor estimation
DATAGROUPS G1 1.0 /G2 1.0
PROPERTIES SYSOP3
   ALGORITHM DRCSTR MAXIT=50 TOL=1D-6 IDEM=10 IODS=0 SS=1 SSND=1D-4
OUT-UNITS TEMP=R PRES=PSI DENS='LB/CUFT'
HISTORY MSG-LEVEL SIM=4

Table 6.10 (continue)
Table 6.11  Redlich-Kwong-Soave Equation of State Acentric Factor Estimation for 1-Butene.

<table>
<thead>
<tr>
<th>Temperature R</th>
<th>Vapor pressure psi</th>
<th>Density lb/ft$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
<td>Dev</td>
</tr>
<tr>
<td>257.67</td>
<td>257.67</td>
<td>0.00</td>
</tr>
<tr>
<td>275.67</td>
<td>275.67</td>
<td>0.00</td>
</tr>
<tr>
<td>293.67</td>
<td>293.67</td>
<td>0.00</td>
</tr>
<tr>
<td>311.67</td>
<td>311.67</td>
<td>0.00</td>
</tr>
<tr>
<td>329.67</td>
<td>329.67</td>
<td>0.00</td>
</tr>
<tr>
<td>347.67</td>
<td>347.67</td>
<td>0.00</td>
</tr>
<tr>
<td>365.67</td>
<td>365.68</td>
<td>0.01</td>
</tr>
<tr>
<td>383.67</td>
<td>383.69</td>
<td>0.02</td>
</tr>
<tr>
<td>401.67</td>
<td>401.72</td>
<td>0.05</td>
</tr>
<tr>
<td>419.67</td>
<td>419.78</td>
<td>0.11</td>
</tr>
<tr>
<td>437.67</td>
<td>437.86</td>
<td>0.19</td>
</tr>
<tr>
<td>455.67</td>
<td>455.94</td>
<td>0.27</td>
</tr>
<tr>
<td>473.67</td>
<td>474.01</td>
<td>0.34</td>
</tr>
<tr>
<td>491.67</td>
<td>492.02</td>
<td>0.35</td>
</tr>
<tr>
<td>509.67</td>
<td>510.06</td>
<td>0.39</td>
</tr>
<tr>
<td>527.67</td>
<td>528.07</td>
<td>0.40</td>
</tr>
<tr>
<td>536.67</td>
<td>537.11</td>
<td>0.44</td>
</tr>
<tr>
<td>545.67</td>
<td>546.04</td>
<td>0.37</td>
</tr>
<tr>
<td>563.67</td>
<td>564.08</td>
<td>0.41</td>
</tr>
<tr>
<td>581.67</td>
<td>581.98</td>
<td>0.31</td>
</tr>
<tr>
<td>599.67</td>
<td>599.08</td>
<td>-0.59</td>
</tr>
<tr>
<td>617.67</td>
<td>616.59</td>
<td>-1.08</td>
</tr>
<tr>
<td>635.67</td>
<td>634.34</td>
<td>-1.33</td>
</tr>
<tr>
<td>653.67</td>
<td>652.19</td>
<td>-1.48</td>
</tr>
</tbody>
</table>

std dev  0.46  0.0286  0.0510

Parameter estimate = 0.1911

Parameter variance = 0.1241d-04
NEW DRS
TITLE 'EXTENDED ANTOINE PARAMETERS ESTIMATION'
COMPONENTS C4H10 BUTANE / C4H8 BUTENE
PROPERTIES SYSOP0
PROP-SOURCES
GLOBAL ASPEN1A COMPS=ALL

DATA
PROP-LIST PLXANT 1
PV AL C4H10 0.661499D2 -.430138D4 0.00 0. &
0.0 0.0 0.0 224.321D0 425.200

PARAMETERS
PARAMETER 1 PLXANT C4H10 1 .66D2 .1D2 .9D2 0.015
PARAMETER 2 PLXANT C4H10 2 -.4304 -.9D4 0.1D2 0.000233
PARAMETER 3 PLXANT C4H10 3 0.005 -1.0 1.0 1.0
PARAMETER 4 PLXANT C4H10 4 0.005 -1.0 1.0 1.0

DATAGROUP G1
INUNITS TEMP=K PRES='N/SQM'
SYSTEM-DEF TP C4H10
PHASE-EQ VL C4H10
STD-DEV 1 0.3 5.0/
2 0.3 40.0/
10 0.3 100.0/
20 0.3 1000.0/
30 0.3 10000.0/
35 0.3 30000.0

DATA 1 170.08 116.06/
2 175.08 203.22/
3 180.08 342.125/
4 185.08 556.0233/
5 190.08 875.43/
6 195.08 1339.43/
7 200.08 1996.93/
8 205.08 2907.981/
9 210.08 4144.98/
10 215.08 61141.80/
11 220.08 75169.0/
12 225.08 91634.99/
13 230.08 110823.1/
14 235.08 133030.4/
15 240.08 158566.9/
16 245.08 187753.8/
17 250.08 220922.5/
18 255.08 258413.6/
19 260.08 300575.4/
20 265.08 347763.1/
21 270.08 400337.9/
22 275.08 458666.7/
23 280.08 530337.9/
24 285.08 61141.80/
25 290.08 75169.0/
26 295.08 91634.99/
27 300.08 110823.1/
28 305.08 133030.4/
29 310.08 158566.9/
30 315.08 187753.8/
31 320.08 220922.5/
32 325.08 258413.6/
33 330.08 300575.4/
34 335.08 347763.1/
35 340.08 400337.9/
36 345.08 458666.7/
37 350.08 530337.9/
38 355.08 61141.80/
39 360.08 75169.0/
40 365.08 91634.99/
41 370.08 110823.1/
42 375.08 133030.4/
43 380.08 158566.9/
44 385.08 187753.8/
45 390.08 220922.5/
46 395.08 258413.6/
47 400.08 300575.4/
48 405.08 347763.1/
49 410.08 400337.9/
50 415.08 458666.7/
51 420.08 530337.9/
52 425.08 61141.80/
53 430.08 75169.0/
54 435.08 91634.99/
55 440.08 110823.1/
56 445.08 133030.4/
57 450.08 158566.9/
58 455.08 187753.8/
59 460.08 220922.5/
60 465.08 258413.6/
61 470.08 300575.4/
62 475.08 347763.1/
63 480.08 400337.9/
64 485.08 458666.7/
65 490.08 530337.9/
66 495.08 61141.80/
67 500.08 75169.0/
68 505.08 91634.99/
69 510.08 110823.1/
70 515.08 133030.4/
71 520.08 158566.9/
72 525.08 187753.8/
73 530.08 220922.5/
74 535.08 258413.6/
75 540.08 300575.4/
76 545.08 347763.1/
77 550.08 400337.9/
78 555.08 458666.7/
79 560.08 530337.9/
80 565.08 61141.80/
81 570.08 75169.0/
82 575.08 91634.99/
83 580.08 110823.1/
84 585.08 133030.4/
85 590.08 158566.9/
86 595.08 187753.8/
87 600.08 220922.5/
88 605.08 258413.6/
89 610.08 300575.4/
90 615.08 347763.1/
91 620.08 400337.9/
92 625.08 458666.7/
93 630.08 530337.9/
94 635.08 61141.80/
95 640.08 75169.0/
96 645.08 91634.99/
97 650.08 110823.1/
98 655.08 133030.4/
99 660.08 158566.9/
100 665.08 187753.8/
101 670.08 220922.5/
102 675.08 258413.6/
103 680.08 300575.4/
104 685.08 347763.1/
105 690.08 400337.9/
106 695.08 458666.7/

Table 6.12 Input data for Extended Antoine parameter estimation.
Table 6.12 (continue)
Table 6.13  Extended Antoine parameter estimation results.

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>24.91784</th>
<th>-3229.868</th>
<th>-5499968D-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Iterations</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
### Table 6.13 Extended Antoine Parameter estimation results. (continue)

<table>
<thead>
<tr>
<th>MEASURED</th>
<th>ESTIMATED</th>
<th>ADJUSTMENTS</th>
<th>% ADJUSTMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>-103.0700</td>
<td>-103.4558</td>
<td>-0.3858356</td>
<td>0.374341</td>
</tr>
<tr>
<td>-98.0700</td>
<td>-98.08493</td>
<td>-0.1493317</td>
<td>0.1227055</td>
</tr>
<tr>
<td>-93.0700</td>
<td>-93.09624</td>
<td>-0.2624201</td>
<td>0.2189600</td>
</tr>
<tr>
<td>-88.0700</td>
<td>-88.11111</td>
<td>-0.4110867</td>
<td>0.3667727</td>
</tr>
<tr>
<td>-83.0700</td>
<td>-83.12556</td>
<td>-0.5555806</td>
<td>0.5068810</td>
</tr>
<tr>
<td>-78.0700</td>
<td>-78.13023</td>
<td>-0.6023450</td>
<td>0.5571447</td>
</tr>
<tr>
<td>-73.0700</td>
<td>-73.11197</td>
<td>-0.4197337</td>
<td>0.3742668</td>
</tr>
<tr>
<td>-68.0700</td>
<td>-68.06223</td>
<td>0.7770460</td>
<td>-0.1141540</td>
</tr>
<tr>
<td>-63.0700</td>
<td>-62.98787</td>
<td>0.8212591</td>
<td>-0.1302139</td>
</tr>
<tr>
<td>-58.0700</td>
<td>-56.00772</td>
<td>0.6227516</td>
<td>-0.1072415</td>
</tr>
<tr>
<td>-53.0700</td>
<td>-52.95347</td>
<td>0.1065342</td>
<td>0.2007428</td>
</tr>
<tr>
<td>-43.0700</td>
<td>-47.52236</td>
<td>0.1436362</td>
<td>0.2988064</td>
</tr>
<tr>
<td>-38.0700</td>
<td>-42.90428</td>
<td>0.1657196</td>
<td>0.3847880</td>
</tr>
<tr>
<td>-33.0700</td>
<td>-37.09908</td>
<td>0.1709181</td>
<td>0.4489575</td>
</tr>
<tr>
<td>-26.0700</td>
<td>-32.91109</td>
<td>0.1589073</td>
<td>0.4805180</td>
</tr>
<tr>
<td>-23.0700</td>
<td>-27.93745</td>
<td>0.1325488</td>
<td>0.4722081</td>
</tr>
<tr>
<td>-18.0700</td>
<td>-22.97454</td>
<td>0.9546038</td>
<td>-0.4137858</td>
</tr>
<tr>
<td>-16.0700</td>
<td>-18.01813</td>
<td>0.5187436</td>
<td>-0.5170462</td>
</tr>
<tr>
<td>-13.0700</td>
<td>-13.06483</td>
<td>0.5170462</td>
<td>-0.5170462</td>
</tr>
<tr>
<td>-6.0700</td>
<td>-8.078978</td>
<td>0.8978486</td>
<td>-0.112576</td>
</tr>
<tr>
<td>-3.0700</td>
<td>-3.033324</td>
<td>0.2332352</td>
<td>0.7597237</td>
</tr>
<tr>
<td>1.930000</td>
<td>1.888247</td>
<td>0.4175315</td>
<td>-0.2163376</td>
</tr>
<tr>
<td>6.930000</td>
<td>6.867095</td>
<td>0.6290461</td>
<td>-0.9077144</td>
</tr>
<tr>
<td>11.93000</td>
<td>11.84554</td>
<td>0.8445827</td>
<td>-0.7079486</td>
</tr>
<tr>
<td>16.93000</td>
<td>16.82662</td>
<td>1.033754</td>
<td>-0.610649</td>
</tr>
<tr>
<td>21.93000</td>
<td>21.81392</td>
<td>1.161766</td>
<td>-0.5297610</td>
</tr>
<tr>
<td>26.93000</td>
<td>26.81076</td>
<td>1.192231</td>
<td>-0.4427147</td>
</tr>
<tr>
<td>31.93000</td>
<td>31.82103</td>
<td>1.108971</td>
<td>-0.3412828</td>
</tr>
<tr>
<td>36.93000</td>
<td>36.84793</td>
<td>0.8207449</td>
<td>-0.222434</td>
</tr>
<tr>
<td>41.93000</td>
<td>41.92852</td>
<td>0.1475370</td>
<td>-0.3523180</td>
</tr>
<tr>
<td>46.93000</td>
<td>46.93164</td>
<td>0.1636680</td>
<td>-0.3491770</td>
</tr>
<tr>
<td>51.93000</td>
<td>51.93716</td>
<td>0.7158970</td>
<td>-0.1378510</td>
</tr>
<tr>
<td>56.93000</td>
<td>56.94608</td>
<td>0.1607991</td>
<td>-0.2824506</td>
</tr>
<tr>
<td>61.93000</td>
<td>61.95965</td>
<td>0.2965446</td>
<td>-0.4788338</td>
</tr>
<tr>
<td>66.93000</td>
<td>66.93587</td>
<td>0.5869873</td>
<td>-0.8770167</td>
</tr>
<tr>
<td>71.93000</td>
<td>71.93927</td>
<td>0.9267979</td>
<td>-1.2884722</td>
</tr>
<tr>
<td>76.93000</td>
<td>76.94399</td>
<td>1.399193</td>
<td>-1.618788</td>
</tr>
<tr>
<td>81.93000</td>
<td>81.95042</td>
<td>2.041528</td>
<td>-1.2491795</td>
</tr>
<tr>
<td>86.93000</td>
<td>86.95899</td>
<td>2.898698</td>
<td>-1.3334519</td>
</tr>
<tr>
<td>91.93000</td>
<td>91.97024</td>
<td>4.023931</td>
<td>-1.4377168</td>
</tr>
</tbody>
</table>

ROOT MEAN SQUARES DEVIATION = .9858910D -01
AVERAGE ABSOLUTE DEVIATION = .6837118D -01
MAXIMUM DEVIATION = -.3858355
<table>
<thead>
<tr>
<th>MEASURED</th>
<th>ESTIMATED</th>
<th>ADJUSTMENTS</th>
<th>% ADJUSTMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1145423D-02</td>
<td>.1381751D-02</td>
<td>.2363276D-03</td>
<td>20.63234</td>
</tr>
<tr>
<td>.2005625D-02</td>
<td>.2369606D-02</td>
<td>.3634373D-03</td>
<td>18.12090</td>
</tr>
<tr>
<td>.3376521D-02</td>
<td>.3797935D-02</td>
<td>.4214139D-03</td>
<td>12.48071</td>
</tr>
<tr>
<td>.5487523D-02</td>
<td>.5933788D-02</td>
<td>.4462644D-03</td>
<td>8.132346</td>
</tr>
<tr>
<td>.6598222D-02</td>
<td>.9056573D-02</td>
<td>.4167502D-03</td>
<td>4.823597</td>
</tr>
<tr>
<td>.1321915D-01</td>
<td>.1353752D-01</td>
<td>.318373D-03</td>
<td>2.408456</td>
</tr>
<tr>
<td>.1970817D-01</td>
<td>.1986722D-01</td>
<td>.1590536D-03</td>
<td>.8070441</td>
</tr>
<tr>
<td>.2869954D-01</td>
<td>.2867810D-01</td>
<td>.144178D-04</td>
<td>-.7471124D-01</td>
</tr>
<tr>
<td>.4090777D-01</td>
<td>.4074026D-01</td>
<td>.1675153D-03</td>
<td>.408451</td>
</tr>
</tbody>
</table>

<p>| SEQUENCE 1 REGRESSION ANTOINE CONSTANTS DATA REGRESSIONSECTION (CONTINUED) |</p>
<table>
<thead>
<tr>
<th>ADJUSTMENTS</th>
<th>% ADJUSTMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.2363276D-03</td>
<td>20.63234</td>
</tr>
<tr>
<td>-.3634373D-03</td>
<td>18.12090</td>
</tr>
<tr>
<td>-.4214139D-03</td>
<td>12.48071</td>
</tr>
<tr>
<td>-.4462644D-03</td>
<td>8.132346</td>
</tr>
<tr>
<td>-.4167502D-03</td>
<td>4.823597</td>
</tr>
<tr>
<td>-.318373D-03</td>
<td>2.408456</td>
</tr>
<tr>
<td>-.1590536D-03</td>
<td>.8070441</td>
</tr>
<tr>
<td>-.144178D-04</td>
<td>-.7471124D-01</td>
</tr>
<tr>
<td>-.1675153D-03</td>
<td>.408451</td>
</tr>
</tbody>
</table>

Table 6.13 (continue)

ROOT MEAN SQUARES DEVIATION = .1292676
AVERAGE ABSOLUTE DEVIATION = .4913472D-01
MAXIMUM DEVIATION = -.5453543
PARAMETER CORRELATION MATRIX

1.0000000  
-0.9970090 1.0000000  
-0.9973785 0.9889814 1.0000000

PARAMETER VARIANCE COVARIANCE MATRIX

-1.2463D-04 
-0.25429D-04 0.52195D-04  
-0.15493D-05 0.31438D-05 0.19360D-06

Table 6.13 (continue)
Table 6.14 Input data for Ideal Gas Heat Capacity Parameters Estimation.
Table 6.15 Measured Variables and Estimates of the True Values of the Variables for 1-Butene in the Determination of Ideal Gas Heat Capacity Parameters.

<table>
<thead>
<tr>
<th>Temperature, °C</th>
<th>Ideal gas heat capacity (J/kmole-K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meas</td>
<td>Calc</td>
</tr>
<tr>
<td>1 298.16</td>
<td>298.05</td>
</tr>
<tr>
<td>2 300.00</td>
<td>299.64</td>
</tr>
<tr>
<td>3 400.00</td>
<td>399.33</td>
</tr>
<tr>
<td>4 500.00</td>
<td>500.43</td>
</tr>
<tr>
<td>5 600.00</td>
<td>600.72</td>
</tr>
<tr>
<td>6 700.00</td>
<td>700.14</td>
</tr>
<tr>
<td>7 800.00</td>
<td>799.73</td>
</tr>
<tr>
<td>8 900.00</td>
<td>899.57</td>
</tr>
<tr>
<td>9 1000.00</td>
<td>999.74</td>
</tr>
<tr>
<td>10 1100.00</td>
<td>1100.07</td>
</tr>
<tr>
<td>11 1200.00</td>
<td>1200.26</td>
</tr>
<tr>
<td>12 1300.00</td>
<td>1300.22</td>
</tr>
<tr>
<td>13 1400.00</td>
<td>1399.79</td>
</tr>
</tbody>
</table>

Estimated parameters

\[
\begin{align*}
C_1 &= -1721.81 \\
C_2 &= 347.01 \\
C_3 &= -0.1904 \\
C_4 &= 0.41301d-4
\end{align*}
\]

Parameter variance-covariance matrix

\[
\begin{bmatrix}
0.1976d+7 & -7214.2 & 26.924 \\
8.0914 & -0.0307 & 0.3546d-4 \\
-0.2841d-2 & 0.1091d-4 & -0.1274d-7 & 0.4625d-11
\end{bmatrix}
\]
A generalized data regression system was developed. The algorithm implemented was the Britt-Luecke's Generalized least-squares method based on the maximum likelihood principle. This method presents a departure from the conventional Ordinary Least-Squares data regression formulation which associates the parameter estimation with the minimization of a user-construct objective function. Predictor functions are used in the objective function and have to be identified. Distinction also have to be made between response variables, and independent variables. Instead, the present formulation associates the parameter determination with the minimization of the weighted sum of the squared deviations of all measured variables from the estimates of the true values of the variables, subject to constraints. The constraint functions are based on physical grounds eg, in VLE data reduction, the constraints are simply the phase equilibrium condition. All measured variables are subject to errors therefore distinction need not be made between response variables and independent variables. From the users' standpoint, this formulation is easier to use. From the theoretical standpoint, it rests on a better basis because all experimental uncertainties are incorporated into the parameter estimation procedure.

The system uses free format, easy to use input language with keywords which are self explanatory. The system is general and the algorithm employed is very robust. The
system can treat a wide variety of regression problems with many problem specifications, and can be used for both linear and non-linear regression studies. Output from the system includes the best estimates of the model parameters as well as information useful in the analysis of the models and the experimental data.

The limitations of the Britt-Luecke's algorithm implemented in the System, as discussed in Section 2.2, includes the followings:

1. The present algorithm does not handle inequality constraints of the measured variables, and
2. The algorithm cannot handle equality constraints that only involve the unknown parameters.

Modification of the present algorithm to properly handle these constraints would definitely contribute to the capability of the DRS system and is recommended.
NOTATION

Unless otherwise stated, number in parentheses after description refer to equations. Single underlined symbols are vectors, while double underlined symbols are matrices. Symbols that appear infrequently may not be listed.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{ij}$</td>
<td>Wilson Equation parameter</td>
</tr>
<tr>
<td>$C_{p_{IG}}$</td>
<td>Ideal gas heat capacity</td>
</tr>
<tr>
<td>CORR</td>
<td>parameter correlation matrix</td>
</tr>
<tr>
<td>$d$</td>
<td>joint probability density function</td>
</tr>
<tr>
<td>$e$</td>
<td>measurement error</td>
</tr>
<tr>
<td>$f$</td>
<td>predictor function</td>
</tr>
<tr>
<td>$f_i$</td>
<td>partial fugacity of component $i$ in a mixture</td>
</tr>
<tr>
<td>$f^o_i$</td>
<td>standard-state liquid fugacity of pure component $i$</td>
</tr>
<tr>
<td>$g$</td>
<td>constraint functions</td>
</tr>
<tr>
<td>$G_{zj}$</td>
<td>Jacobian matrix of $g(z_j, \theta)$ with respect to $z_j$</td>
</tr>
<tr>
<td>$G_{\theta j}$</td>
<td>Jacobian matrix of $g(z_j, \theta)$ with respect to $\theta$</td>
</tr>
<tr>
<td>$I$</td>
<td>objective function</td>
</tr>
<tr>
<td>$k$</td>
<td>number of measurements</td>
</tr>
<tr>
<td>$k_{ij}$</td>
<td>binary-interaction parameter for equations of state</td>
</tr>
<tr>
<td>$L$</td>
<td>Joint likelihood function</td>
</tr>
<tr>
<td>$m$</td>
<td>number of constraints</td>
</tr>
</tbody>
</table>
\( n \) \hspace{1cm} \text{number of parameters}
\( n_c \) \hspace{1cm} \text{number of components in a mixture}
\( P_i' \) \hspace{1cm} \text{vapor pressure of pure component } i
\( P \) \hspace{1cm} \text{Total pressure}
\( q \) \hspace{1cm} \text{number of measured variables}
\( R \) \hspace{1cm} \text{covariance matrix}
\( T \) \hspace{1cm} \text{temperature}
\( T_r \) \hspace{1cm} \text{reduced temperature}
\( \text{VAR} \) \hspace{1cm} \text{parameter variance-covariance matrix}
\( x_i \) \hspace{1cm} \text{liquid mole fraction of component } i
\( y_i \) \hspace{1cm} \text{vapor mole fraction of component } i
\( z \) \hspace{1cm} \text{measured variables}
\( z_M \) \hspace{1cm} \text{estimates of the true values of the measured variables}

**Greek Symbols**

\( \sigma^2 \) \hspace{1cm} \text{variance of a quantity}
\( \phi \) \hspace{1cm} \text{vapor fugacity coefficient of component } i
\( \phi_{iL}^0 \) \hspace{1cm} \text{standrad-state pure component fugacity coefficient}
\( \mathcal{L} \) \hspace{1cm} \text{likelihood function}
\( \gamma_i \) \hspace{1cm} \text{activity coefficient of component } i
\( \theta \) \hspace{1cm} \text{model parameters}
Superscripts

J at the J-th iteration
T transpose of a matrix
V vapor phase
L Liquid phase
-1 inverse of a matrix

Subscripts

j at the j-th measurement
M measured value of the quantity
REFERENCE


In this appendix, the algorithm for minimizing Equation (2-18) with respect to \( z_j \) and \( \theta \) is derived.

The constraint equation (2-17) may be combined with equation (2-18) by the use of Lagrange multiplier m-vectors \( \lambda_j \), \( j=1,...,k \), to form a new unconstrained objective function

\[
Q(z_1, \ldots, z_k, \theta, \lambda_1, \ldots, \lambda_k) = \frac{1}{2} (z_M - z_j)^T R_j^{-1} (z_M - z_j) + \lambda_j^T g(z_j, \theta) \quad \text{(A-1)}
\]

Conditions (ii) and (iv) in section (2-2) are sufficient to guarantee that the unconstrained problem, Equation (A-1), has the same stationary points as the constrained problem, equations (2-17) and (2-18). Necessary conditions for a stationary point of Equation (A-1) are

\[
\frac{\partial Q}{\partial z_j} = 0 \quad j=1,\ldots,k \quad \text{(A-2)}
\]

\[
\frac{\partial Q}{\partial \theta} = 0 \quad \text{(A-3)}
\]

\[
\frac{\partial Q}{\partial \lambda_j} = 0 \quad j=1,\ldots,k \quad \text{(A-4)}
\]
which give:

\[-R_j^{-1} (z_{Mj} - z_j) + G_j^T (z_j, \theta) \lambda_j = 0 \quad j = 1, \ldots, k \quad (A-5)\]

\[G_{\theta j}^T (z_j, \theta) \lambda_j = 0 \quad (A-6)\]

\[g(z_j, \theta) = 0 \quad (A-7)\]

These equations are analogous to the normal equations for the standard parameter estimation problem. Any algorithm designed to solve simultaneous, non-linear equations can be applied to them if the number of equations \((q.k + n + m.k)\) is not too large. However, in most applications the number of experiments, \(k\), is large enough so that it is impractical, if not impossible, to solve the system equation \((A-5)\) through \((A-7)\).

The key to the present approach is to linearize only the constraint equation \((2-17)\). Expanding Equation \((2-17)\) in a Taylor's series about the point \((z_1^J, \ldots, z_k^J, \theta^J)\) where \((z_1^J, \ldots, z_k^J, \theta^J)\) represents the latest estimate of \((z_{01}, \ldots, z_{0k}, \theta_0)\) in an iterative process, and dropping higher order terms, an objective function with linear terms is obtained.\(^1\)
\begin{align*}
\frac{1}{2} (z_{Mj} - z_j)^T R_j^{-1} (z_{Mj} - z_j) + \lambda_j^T \varphi (z_j^J, \theta_j^J) + \\
G_{\theta_j} \left( \theta - \theta_j^J \right) + G_{z_j} \left( z_j - z_j^J \right)
\end{align*}

where \( G_{\theta_j} \) and \( G_{z_j} \) are the values of the Jacobian matrices of \( \varphi (z_j, \theta) \) with respect to \( \theta \) and \( z_j \), respectively, at the points \( (z_j^J, \theta_j^J) \). Necessary conditions for a stationary point of equation (A-8) are

\begin{align*}
-R_j^{-1} (z_{Mj} - z_j) + G_{z_j} \lambda_j &= 0 \quad j=1, \ldots, k \\
G_{\theta_j} \lambda_j &= 0 \quad j=1, \ldots, k
\end{align*}

\begin{align*}
\varphi (z_j^J, \theta_j^J) + G_{\theta_j} \left( \theta - \theta_j^J \right) + G_{z_j} \left( z_j - z_j^J \right) &= 0
\end{align*}

Multiplying each Equation (A-9) by \( G_{z_j} R_j \), replacing \( (z_{Mj} - z_j) \) by \( (z_{Mj} - z_j^J) + (z_j^J - z_j) \), and substituting the result into equation (A-11) gives

\begin{align*}
\varphi (z_j^J, \theta_j^J) + G_{\theta_j} \left( \theta - \theta_j^J \right) + G_{z_j} \left( z_j - z_j^J \right) - G_{z_j} R_j G_{z_j}^T \lambda_j &= 0 \\
&j=1, \ldots, k
\end{align*}
Multiplying each Equation (A-12) by $G^T_{\theta_j}W_j$, where $W_j = (G_{z_j}R_jG_{z_j})^{-1}$, and summing over, yields

$$
\sum_{j=1}^{\infty} W_j \cdot z_j G(z_j, \theta_j^J) + \sum_{j=1}^{\infty} (z_j - z_j) \cdot \lambda_j = 0
$$

(A-13)

Using Equation (A-10), Equation (A-13) becomes

$$
\theta - \theta^J = - \sum_{j=1}^{\infty} W_j \cdot z_j G(z_j, \theta_j^J) + \sum_{j=1}^{\infty} (z_j - z_j) \cdot \lambda_j = 0
$$

(A-14)

From Equation (A-12)

$$
\lambda_j = \frac{W_j}{z_j} \cdot z_j G(z_j, \theta_j^J) + \sum_{j=1}^{\infty} (z_j - z_j) \cdot \lambda_j = 0
$$

(A-15)

and from Equation (A-9),

$$
G_{z_j} \cdot z_j = R_j \cdot G^T_{z_j} \cdot \lambda_j
$$

(A-16)

Equation (A-15) becomes
\[
\begin{align*}
    z_j &= z_{Mj} - R_j G_j^T W_j \varphi(z_j^J, \theta^J) + G_{\theta_j} (\theta - \theta^J) + G_{z_j} (z_{Mj} - z_j) \\
    j &= 1, \ldots, k \\
\end{align*}
\] (A-17)

The \((z_1', \ldots, z_k', \theta)\) computed from Equations (A-14) and (A-17) is a stationary point of Equation (A-8). Since Equation (A-8) is only an approximation to the true objective function, Equation (A-1); \((z_1^{J+1}, \ldots, z_k^{J+1}, \theta^{J+1})\) is set equal to \((z_1', \ldots, z_k', \theta)\), the constraint relinearized, and a new estimate \((z_1^{J+2}, \ldots, z_k^{J+2}, \theta^{J+2})\) is computed. The iteration are continued until \(\theta^{J+1} - \theta^J\) and \(z_j^{J+1} - z_j^J, j = 1, \ldots, k\) satisfy a convergence criterion.

It is worth noting that the Deming approximate algorithm has a similar derivation. It may be obtained by expanding the same constraint equation (2-17) about the point \((z_{Mj}', \theta^J)\), rather than \((z_j^J, \theta^J)\), for each iteration. The resulting linearized objective function is

\[
\begin{align*}
    l/2 (z_{Mj} - z_j)^T R_j^{-1} (z_{Mj} - z_j) + \\
    \lambda_j^T \varphi(z_{Mj} - \theta^J) + G_{\theta_j} (\theta - \theta^J) + G_{z_j} (z_j - z_{Mj}) \\
\end{align*}
\] (A-18)
where the Jacobian matrices are evaluated at the points $(z_{Mj}, \theta)$. 
APPENDIX B
Table B-1  Input data for Ideal Gas Heat Capacity Parameter Estimation.

-124-
Table B-2  DRS Report file
SUBSECTION: CASE

INPUT MEASURED VARIABLES

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>298.160</td>
<td>85726.9</td>
</tr>
<tr>
<td>300.000</td>
<td>86119.6</td>
</tr>
<tr>
<td>400.000</td>
<td>109010</td>
</tr>
<tr>
<td>500.000</td>
<td>129488</td>
</tr>
<tr>
<td>600.000</td>
<td>147105</td>
</tr>
<tr>
<td>700.000</td>
<td>162084</td>
</tr>
<tr>
<td>800.000</td>
<td>175044</td>
</tr>
<tr>
<td>900.000</td>
<td>186260</td>
</tr>
<tr>
<td>1000.00</td>
<td>196030</td>
</tr>
<tr>
<td>1100.00</td>
<td>204560</td>
</tr>
<tr>
<td>1200.00</td>
<td>211961</td>
</tr>
<tr>
<td>1300.00</td>
<td>218413</td>
</tr>
<tr>
<td>1400.00</td>
<td>224023</td>
</tr>
</tbody>
</table>

Table B-2 Report file (continue)

-126-
### Parameter Names

| CPIG | CPIG | CPIG |

### Parameter Estimates

-1721.814 347.0078 -1.1904403 .4130051D-04

### Number of Iterations

7

### Variable Number 1

<table>
<thead>
<tr>
<th>Measured</th>
<th>Estimated</th>
<th>Adjustments</th>
<th>% Adjustments</th>
</tr>
</thead>
<tbody>
<tr>
<td>298.1600</td>
<td>298.0488</td>
<td>-1.111518</td>
<td>-3727925D-01</td>
</tr>
<tr>
<td>300.0000</td>
<td>299.6458</td>
<td>-0.3541698</td>
<td>-1180565</td>
</tr>
<tr>
<td>400.0000</td>
<td>399.3322</td>
<td>-0.6678449</td>
<td>-1669612</td>
</tr>
<tr>
<td>500.0000</td>
<td>500.4257</td>
<td>-0.4257154</td>
<td>0.8514308D-01</td>
</tr>
<tr>
<td>600.0000</td>
<td>500.7168</td>
<td>0.7158359</td>
<td>0.1194728</td>
</tr>
<tr>
<td>700.0000</td>
<td>700.1359</td>
<td>0.1359355</td>
<td>0.1941936D-01</td>
</tr>
<tr>
<td>800.0000</td>
<td>799.7319</td>
<td>-0.2681227</td>
<td>-3351533D-01</td>
</tr>
<tr>
<td>900.0000</td>
<td>899.5735</td>
<td>-0.4264808</td>
<td>-4738676D-01</td>
</tr>
<tr>
<td>1000.00</td>
<td>998.7416</td>
<td>-0.2584123</td>
<td>-2584123D-01</td>
</tr>
<tr>
<td>1100.00</td>
<td>1100.071</td>
<td>0.7084026D-01</td>
<td>0.6440024D-02</td>
</tr>
<tr>
<td>1200.00</td>
<td>1200.258</td>
<td>0.2583667</td>
<td>0.2153055D-01</td>
</tr>
<tr>
<td>1300.00</td>
<td>1300.223</td>
<td>0.2227560</td>
<td>0.1713308D-01</td>
</tr>
<tr>
<td>1400.00</td>
<td>1399.790</td>
<td>-0.2095900</td>
<td>-1.497072D-01</td>
</tr>
</tbody>
</table>

- Root mean squares deviation = .370
- Average absolute deviation = .3174018
- Maximum deviation = .7169369

Table B-2  Report file (continue)
### Table B-2 Report file (continue)

#### Variable Number 2

<table>
<thead>
<tr>
<th>Measured</th>
<th>Estimated</th>
<th>Adjustments</th>
<th>% Adjustments</th>
</tr>
</thead>
<tbody>
<tr>
<td>85726.90</td>
<td>85740.53</td>
<td>13.63376</td>
<td>.1590371D-01</td>
</tr>
<tr>
<td>86119.60</td>
<td>86162.93</td>
<td>43.32613</td>
<td>.5030926D-01</td>
</tr>
<tr>
<td>109010.0</td>
<td>109110.8</td>
<td>100.8009</td>
<td>.9246941D-01</td>
</tr>
<tr>
<td>129488.0</td>
<td>129414.4</td>
<td>-73.59221</td>
<td>-.5683323D-01</td>
</tr>
<tr>
<td>147104.7</td>
<td>146962.1</td>
<td>-142.5659</td>
<td>-.9691460D-01</td>
</tr>
<tr>
<td>162084.4</td>
<td>152053.2</td>
<td>-31.22124</td>
<td>-.1926234D-01</td>
</tr>
<tr>
<td>175044.5</td>
<td>175115.9</td>
<td>71.41590</td>
<td>.4079871D-01</td>
</tr>
<tr>
<td>186260.0</td>
<td>186392.1</td>
<td>132.0579</td>
<td>.7089974D-01</td>
</tr>
<tr>
<td>196030.0</td>
<td>196123.0</td>
<td>92.97005</td>
<td>.4742644D-01</td>
</tr>
<tr>
<td>204560.0</td>
<td>204530.6</td>
<td>-29.44618</td>
<td>-.1439489D-01</td>
</tr>
<tr>
<td>219961.0</td>
<td>21838.5</td>
<td>-122.4657</td>
<td>-.5778737D-01</td>
</tr>
<tr>
<td>218413.0</td>
<td>218295.1</td>
<td>-117.8530</td>
<td>-.5395879D-01</td>
</tr>
<tr>
<td>224023.0</td>
<td>224142.9</td>
<td>119.9205</td>
<td>.5353046D-01</td>
</tr>
</tbody>
</table>

**Root Mean Squares Deviation** = 93.75  
**Average Absolute Deviation** = 83.94542  
**Maximum Deviation** = -142.5659  

**Parameter Correlation Matrix**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.9890919</td>
<td>1.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9665461</td>
<td>-0.9931911</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>-0.9399250</td>
<td>0.9775313</td>
<td>-0.9952378</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

**Parameter Variance Covariance Matrix**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.19759D+07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-7214.2</td>
<td>26.924</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.0914</td>
<td>.30689D-01</td>
<td>.35461D-04</td>
<td></td>
</tr>
<tr>
<td>-.2841D-02</td>
<td>.10908D-04</td>
<td>-.12745D-07</td>
<td>.46249D-11</td>
</tr>
</tbody>
</table>
INPUT TRANSLATION SUCCESSFULLY COMPLETED
SIMULATION PROGRAM MAY BE EXECUTED

************
* ASPEN INPUT PROCESSOR ENDS EXECUTION *
************

REPORT WRITER ENTERED  TIME = 0.02/ 0.04

BEGIN CASE 1  TIME = 0.04/ 0.06

MINIMUM WORK VECTOR LENGTHS,REAL,INTEGER 220 16

********
***** ITERATION NO. 1

THIS ITERATION BY DEMINGS METHOD

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .21886555
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 67.479749

ESTIMATES FOR PARAMETERS ARE
-1712.4861  346.98246  -.19042057  .41296140D-04

********
***** ITERATION NO. 2

THIS ITERATION BY DEMINGS METHOD

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .91237874D-03
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 66.458722

ESTIMATES FOR PARAMETERS ARE
-1709.4215  346.96317  -.19039064  .41283130D-04

********
***** ITERATION NO. 3

THIS ITERATION BY DEMINGS METHOD

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .43924386D-05
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 66.458958

ESTIMATES FOR PARAMETERS ARE
-1709.4066  346.96309  -.19039052  .41283082D-04

********
***** ITERATION NO. 4

Table B-3 DRS History file
THIS ITERATION BY DEMINGS METHOD

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .52454257D-08
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 66.458959

ESTIMATES FOR PARAMETERS ARE

-1709.4066  346.96309  -0.19039052  .41283082D-04

**********

****** ITERATION NO.  5

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .365919270-02
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 68.815142

ESTIMATES FOR PARAMETERS ARE

-1721.8856  347.00807  -0.19044054  .41300596D-04

**********

****** ITERATION NO.  6

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .21156518D-04
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 68.815987

ESTIMATES FOR PARAMETERS ARE

-1721.8129  347.00782  -0.19044027  .41300508D-04

**********

****** ITERATION NO.  7

ROOT MEAN SQUARE OF RELATIVE DELP ADJUSTMENTS IS .30973423D-06
WEIGHTED SUM OF SQUARES OF DATA ADJUSTMENTS IS 68.815987

ESTIMATES FOR PARAMETERS ARE

-1721.8140  347.00782  -0.19044028  .41300509D-04

Table B-3  History file (continue)
### GENERALIZED LEAST SQUARES PROGRAM

#### BRITT-LUECKE ALGORITHM

#### FINAL RESULTS

**NUMBER OF ITERATIONS**: 7
**NUMBER OF CALLS TO MODEL**: 49

**PARAMETER ESTIMATES**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1721.814</td>
<td>2</td>
<td>347.0078</td>
</tr>
</tbody>
</table>

**FINAL FUNCTION VALUES**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.1455192D-10</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-.1637090D-07</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>-.9150528D-07</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>.1263108D-07</td>
<td>11</td>
</tr>
</tbody>
</table>

**OBSERVED VARIABLES**

<table>
<thead>
<tr>
<th>K</th>
<th>M</th>
<th>MEASURED</th>
<th>ESTIMATED</th>
<th>ADJUSTMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>298.1600</td>
<td>298.0488</td>
<td>.1111518</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>85726.90</td>
<td>85740.53</td>
<td>-.1363376</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>300.0000</td>
<td>299.6458</td>
<td>.3541696</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>86119.50</td>
<td>86162.93</td>
<td>-.4332613</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>400.0000</td>
<td>399.3322</td>
<td>.6678445</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>109010.0</td>
<td>109110.8</td>
<td>-100.8009</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>500.0000</td>
<td>500.4257</td>
<td>-.4257154</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>129488.0</td>
<td>125414.4</td>
<td>.7359222</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>600.0000</td>
<td>600.7168</td>
<td>-.7168368</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>147104.7</td>
<td>146962.1</td>
<td>142.5659</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>700.0000</td>
<td>700.1359</td>
<td>-.1359354</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>162054.4</td>
<td>162053.2</td>
<td>31.22121</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>800.0000</td>
<td>799.7319</td>
<td>.2681228</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>175044.5</td>
<td>175115.9</td>
<td>-.7141692</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>900.0000</td>
<td>899.5735</td>
<td>.4264809</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>186260.0</td>
<td>186392.1</td>
<td>-.132.5079</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1000.0000</td>
<td>999.7416</td>
<td>.2581423</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>196030.0</td>
<td>196123.0</td>
<td>-.9297004</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1100.0000</td>
<td>1100.071</td>
<td>-.7084033D-01</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>204560.0</td>
<td>204530.6</td>
<td>29.44621</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1200.0000</td>
<td>1200.258</td>
<td>-.2563667</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>211961.0</td>
<td>211838.5</td>
<td>122.4867</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1300.0000</td>
<td>1300.223</td>
<td>-.2227560</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>218413.0</td>
<td>218295.1</td>
<td>117.8530</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1400.0000</td>
<td>1399.790</td>
<td>.1399900</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>224023.0</td>
<td>224142.9</td>
<td>-119.9206</td>
</tr>
</tbody>
</table>

Table B-3 History file (continue)
PARAMETER CORRELATION MATRIX

1.000000
-0.9690919 1.000000
0.9666461 0.9931911 1.0000000
-0.9399250 0.9775313 -0.9952378 1.0000000

WEIGHTED SUM OF SQUARES
SIGMA SQUARED ESTIMATE 0.6861599D+02
SIGMA ESTIMATE 0.7646221D+01

PARAMETER VARIANCE-COVARIANCE MATRIX

-19759D+07
-7214.2 26.924
8.0914 -.30689D-01 .35461D-04
-.28414D-02 .10908D-04 -.12745D-07 .46249D-11

REPORT GENERATED TIME = 2.67/ 2.99

Table B-3 History file (continue)