CHAPTER 3

FLOWSHEET MODELING

3.1 Introduction

Measured process data obtained directly from instrument are commonly contains inherent information in nature. When these data are utilized in constructing the model, the result usually misrepresents the process leading to misunderstanding the process as a whole. It is noted that when the data contains a significant amount of error, the model made from such data will contain even higher orders of error. Thus, to encounter these measurement errors one should learn how these errors occurred, what caused them to occur and how to eliminate them. The technique commonly used to track down errors and is called "Data Reconciliation". In order to obtain more reliable information for process operation, the data must be repeatedly and systematically collected from the plant.

3.2 Flowsheet Modeling

The usage and advancement of computer-aided simulation over the past few years have been gradually increasing. In the past, it was rare for the graduate chemical engineer to have any experience in using a chemical process simulator. Most material and energy balances were still done by hand by teams of engineers. The rigorous simulation of multistage separation equipment and complicated reactors was generally unheard of. A combination of simplified analyses, short-cut methods, and years of experience achieved design of such equipment. At present, however, companies now expect their engineers to be conversant with a wide variety of computer programs, especially a process simulator

To some extent, the knowledge base required to simulate successfully a chemical process will depend on the simulator used. Currently there are several process

simulators in the market for example, CHEMCAD, ASPEN PLUS, HYSIM, PRO/II. The availability of such powerful software is a great asset to the experience process engineer.

3.2.1 The Structure of a Process Simulator

The six main features of all process simulators are illustrated in the left-hand column of Figure 3.1. These elements are:

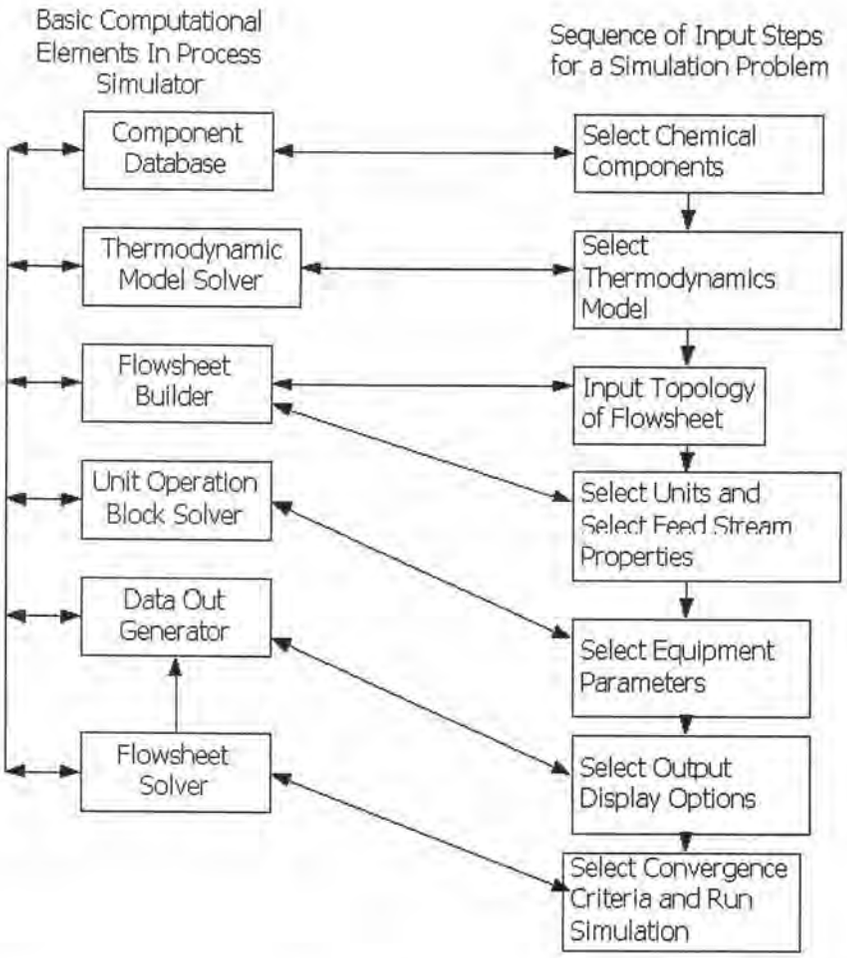

Figure 3.1 Relationship between Basic Computational Elements and Required Input to Solve a Process Simulation Problem

1. Component Database – This contains the constants required to calculate the physical properties from the thermodynamic models.

2. Thermodynamic Model Solver – A variety of options for vapor-liquid (VLE) and liquid-liquid (LLE) equilibrium, enthalpy calculations, and other thermodynamics property estimation are available.

3. Flowsheet Builder – This part of the simulator keeps track of the flow of streams and equipment in the process begin simulated. Often this information can be both input and displayed graphically.

4. Unit Operation Block Solver – Computational blocks or modules are available that allow energy and material balances and some design calculations to be performed for a wide variety of process equipment.

5. Data Output Generator – This part of the program serves to customize the results of the simulation in terms of an output report. Often, graphical displays of tower profiles, heating curves, and a variety of other useful process data can be produced.

6. Flowsheet Solver – This portion of this simulator controls the sequence of the calculations and the overall convergence of the simulation.

There are several other elements commonly found in process simulators that are not shown in Figure 3.1. For example, there are file control options, the option to use different engineering units, and possibly some additional features associated with regressing data for thermodynamic model, and so on. The availability of these other options are dependent on the simulator used and will not be discussed further.

Also shown on the right-hand side of the diagram in Figure 3.1, are the seven general steps to setting up a process simulation problem. The general sequence of events that a user must follow in order to set up a problem on a simulation are as follows:

1. Select all of the chemical components that are required in the process from the component database.

2. Select the thermodynamics models required for the simulation. These may be different for the different pieces of equipment. For example, to simulate correctly a liquid-liquid extractor, it is necessary to use a thermodynamic model that can predict liquid-phase activity coefficients and the existence of two liquid phases. However, for a pump in the same process, a less sophisticated model could be used.

3. Select the topology of the flowsheet to be simulated by specifying the input and output streams for each piece of equipment.

4. Select the properties (temperature, pressure, flow rate, vapor fraction, and composition) of the feed streams to the process.

5. Select the equipment specifications (parameters) for each piece of equipment in the process.

6. Select the way in which the result are to be displayed

7. Select the convergence method and run the simulation.

The interaction between the elements and steps and the general flow of information is shown by the lines on the diagram. Of the seven input steps given above, steps 2, 5 and 7 are the cause of most problems associated with running process simulations.

There basically three types of solution algorithm for process simulators: sequential modular, equation solving (simultaneous non-modular), and simultaneous modular.

In the sequential modular approach, the equations describing the performance of equipment units are grouped together and solved in modules, that is, the process is solved equipment piece by equipment piece. In the equation solving, or simultaneous non-modular, technique, all the relationships for the process are written out together and then the resulting matrix of nonlinear simultaneous equations is solved to yield the solution. This technique is very efficient in terms of computation time, but requires a lot of time to set up and is unwieldy. The final technique is simultaneous modular approach, which combines the modularizing of the equation relating to specific equipment with the efficient solution algorithms for the simultaneous equations solving technique.

Of these three types, the sequential modular algorithm is by far the most widely used. In the sequential modular method, each piece of equipment is solved in sequence, starting with the first then followed by the second, and so on. It is assumed that all the input information required to solved each piece of equipment has been provided. Therefore, the output from a given piece of equipment, along with specific information on the equipment, becomes the input of the next piece of equipment in the process. Clearly, for a process without recycle streams, this method requires only one flowsheet

iteration to produce a converged solution. The term flowsheet iteration means each piece of equipment is only solved once. However, there may be many iterations for any one given piece of equipment. This concept is illustrated in Figure 3.2

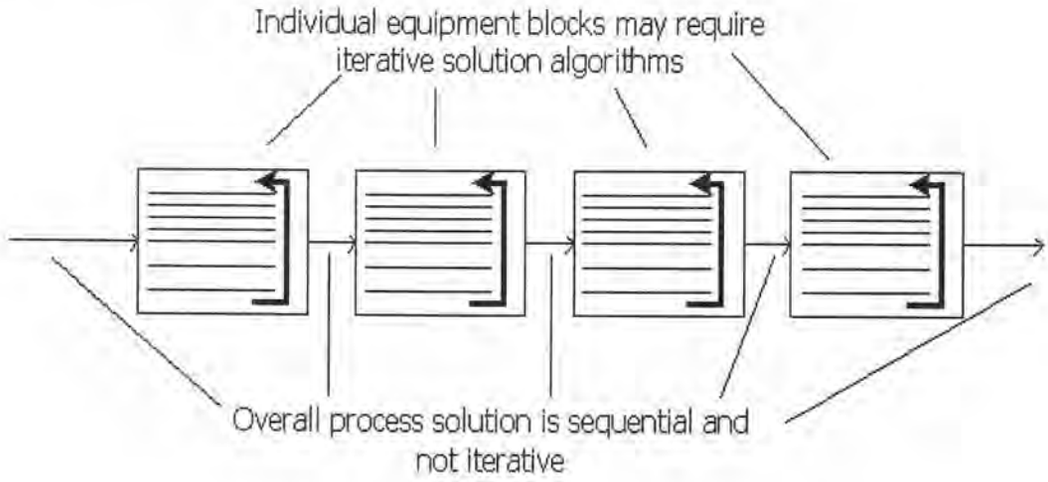

Figure 3.2 Solutions sequence using sequential modular simulator

In order to insert an equipment or process unit into a flowsheet, an understanding of a process model should be considered as this would help ease clarify the type of model and errors within in.

3.2.2 Types of a Process Model

A process model is a mathematical model purposely built for simulating the process. For that purpose, it is thus very crucial to clearly understand the process model before constructing the objective function with provided constrains of the process. A process model will be change according as to the manipulated variables of the process change at a set point in the control component. A process model is often built at steady state and mostly has a non-linear relationship (Darby and White, 1988).

A model can be divided into three types as follows:

1. Theoretical or white block model

This type of model employed knowledges in physics, chemistry (Newton's Law, mass and energy balance, thermodynamics, reaction rate equations) for constructing the model. The basic laws are used to describe the behavior of a phenomenon or a process.

2. Empirical or black box model

This type of model is generally developed from the relationship among empirical data, operation data and plant data but doesn't include any theoretical concern knowledge. The empirical process model obtained could be then related to those theories. Although the empirical process model might be easy to be constructed, but it described and represented a limited validity of the process. The model can represent the process upon the operating condition at a certain working point, at which conditions and parameters of the process changes slightly. However, this type of model could be too difficult to be constructed since it requires a lot of data while some data are not measured and containing many disturbances.

3. Semi-empirical or gray block model

In process of a model construction, it has been found that theoretical model approach is sometime not a reasonable model. Also, it was previously mention that theoretical model includes unmeasured parameters and variables. Although, theoretical model is constructed from the basic law of physic and chemistry but necessary that one have to apply empirical data to estimate those parameters and variables.

Thus, a model that built by using theoretical concerns and utilizing empirical data is called a semi-empirical model. This type of model could be used in a wider range than empirical model and can also be used with unmeasured variables, or process variables that change when operating objective changes. More over, the model is also classified into different types (Edgar and Himmelblau, 1989) as follows:

- **Linear and nonlinear**

Linear model exhibits the important properties of superposition while the nonlinear ones do not poses that characteristic. Equation and models are linear if the dependent variables or their derivatives appear only to the first power otherwise they are defined as a nonlinear model. In practice, use a linear model for a process is of great significance since ability of manipulation and solution of linear models required during optimization calculations is in an order of magnitude easier than nonlinear ones.

- **Steady and unsteady state**

A steady-state model is referred to a process in which the values of the dependant variable remains uncontacted with respect to time. An unsteady state model

or dynamic model represents the situation in which the process-dependent variables change with time.

- **Distributed and lumped parameters**

A lumped parameter representation means spatial variations are ignored and that the various properties and the state of the system can be considered homogeneous throughout the entire volume. Distributed parameter representation takes into account detailed variations in behavior from point to point throughout the system. A good rule of thump is that if the response of the process is essentially the same at all points then the process can be lumped as a single unit.

- **Continuous and discrete variables**

Continuous means that a variable can assume any value within an interval. Discrete means a variable can take only one distinct value in the interval.

The process simulation package could be used to simulate the process model. Process simulation, in general, is heat arid mass balancing either steady state or dynamic of a chemical process, usually performed by a computer. Each physical unit could be represented by a set of equations: the heat and material balances, equilibrium equations, physical property equations and other internal unit equations such as kinetic rate equations. The variables in these equations include those in the inlet and outlet streams, equipment parameters, internal unit variables and intermediate physical property variables. The goal of process simulation is to solve equations for all units in the process, as well as to connect equations related to the outlet streams of a unit to the inlet streams of other units (Gallier, 1987).

3.3 Data Reconciliation

Plant process data inherently contains random and gross errors that lead to inaccurate mass and energy balances. One of the most used methods to detect and reduce those errors is data reconciliation. Data reconciliation is defined as the estimation of measured process data variables to reduce measurement errors through the use of temporal and spatial or functional redundancies. In other words, data reconciliation is the optimal solution to a constrained least squares or maximum likelihood objective function. First, one must understand the type of errors and the nature of it.

3.3.1 Types of Error

As one understands, errors are very important factors in constructing a model for effective plant operation. Therefore, it is very valuable to be able to identify and measure types of error appearing mostly in the process. In general, types of error are classified into three categories as follows:

1. **Small random error**

Small random error is typically assumed to be zero-mean and normally distributed. Small random error is usually attributed to the irreproducibility of the measurement device (Mah, 1990)

2. **Systematic bias**

This type of error occurs when measurement devices provide consistent erroneous values, either high or low. Thus the expected value of the measurement error is not zero. Bias may arise from sources search as incorrect installation or calibration of the measurement device.

3. **Gross error**

This type of error is usually caused by nonrandom events. The measurement values bear little or no relation to the true value of the desired property. Gross errors could be subdivided into measurement related errors such as malfunctioning sensors, process-related errors such as process leaks.

All measurements are subjected to errors. Random errors are assumed to be normally distributed and have zero mean. Gross errors are caused by non-random events such as instrument biases, malfunctioning measuring devices, inaccurate or incomplete process models and process leaks (Tjao and Beigler, 1991; Mah, 1990; Liebman, Edgar and Lasdon, 1992). Figure 3.3 illustrates the probabilities involved for random and gross errors.

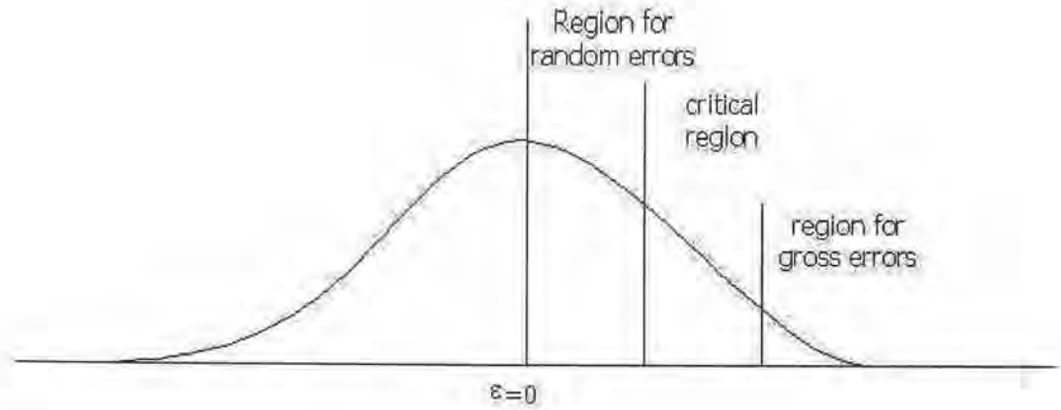

Figure 3.3 Error probability distribution for process measurement

Normally, not all process variables are measured for the reasons of cost, inconvenience or technical unfeasibility. Therefore, other related measurement, spatial (functional) redundancies is employed in this research.

Redundant measurement means obtaining the same process information or measurement from two or more device sources. These spatial redundancies arise from the fact that same measurement data is known and related to another unknown or not measured value.

There are two types of spatial redundancies, which could be categorized by availability of data.

1. Spatial redundant: measurements are excessive in information to completely define the process model at any instant time.
2. Temporal redundant: measurements that follow these are measurements that if part of them are available and could be used for estimation purpose.

On the other hand, the one must estimate data if possible, this process was referred as coaptation by Mah et. al., (1976). Figure 3.4 illustrates the general steps for processing measurement data.

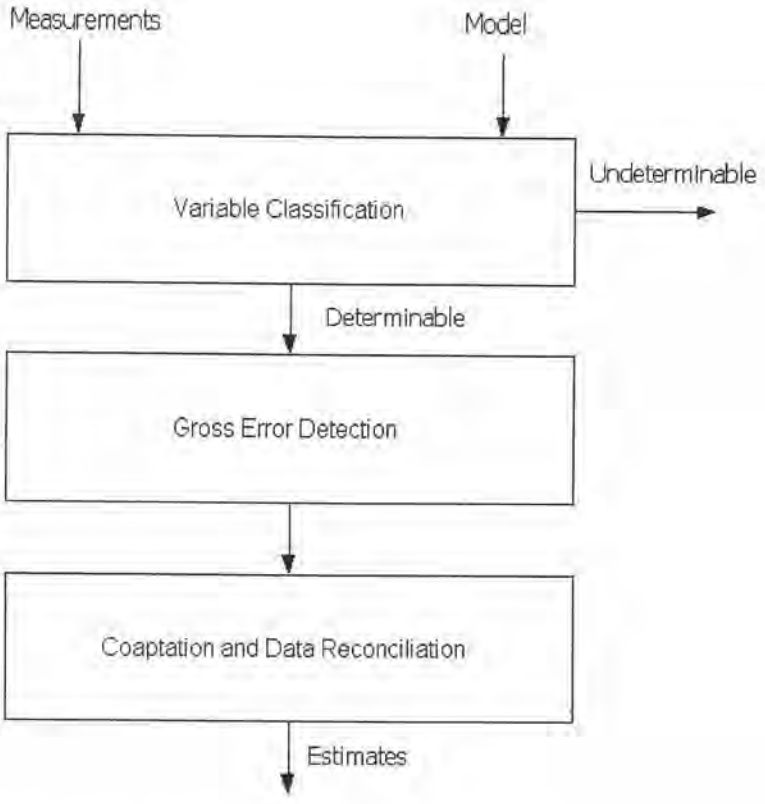

Figure 3.4 General steps for processing measurement data

Description of the general steps for processing measurement data is as follows:

1. Variable classification (Stanley and Mah, 1981; Crowe, 1986)

To determine type of variable one should be able to identify observable or unobservable and redundant or undetermined process data. Those undeterminable process variables are not useful for improvement.

2. Gross error detection (Mah, 1987; Mah, 1990)

In this step, the gross errors are identified and removed. Gross error detection strategies are typically based on statistical hypothesis testing. Gross error detection must be performed prior to (or simultaneously with) data reconciliation since a key assumption made during the reconciliation step is that the errors are normally distributed.

At this stage, consistent estimates are obtained and sequentially satisfy all of the specified model equations while staying as close as possible to the actual measurements.

Coaptation, the estimation of all unmeasured but observable variables could be treated simultaneously with measured data reconciliation. There were few researches reported comparing the difference of data reconciliation and gross error detection methods. Mah, (1987) and Mah, (1990) provided probably the best review in this area to date. However, development of better nonlinear and dynamic data reconciliation and gross error detection methods are still a common on-going area of research. Practically, data reconciliation is mathematically calculated first then followed by gross error detection.

In real operation, the plant variables are the actual values of the measurement. The model variables are the predicted values of the measurement. The offset variables are the difference between the plant and the model variables.

3.3.2 Linear, Steady-State Data Reconciliation

For data reconciliation, the simplest case is linear steady-state data type. Mah (1990) and Crowe et al. (1983) revealed very extensively on this type of data reconciliation. The following equation relates these three quantities in sequential modular form:

$$\gamma = \zeta + \varepsilon \tag{3.1}$$

Where y = (s x 1) vector of measurement variables

ς = (s x 1) vector of true variable values

ε = (s x 1) vector of random measurement error

It is usually assumed that:

1. the expected value of ε, $\mathbf{E}(\varepsilon) = 0$

2. the successive vectors of measurements are independent, i.e., $E(\varepsilon_i \varepsilon_i^T) = 0$, for $i \neq j$

3. the covariance matrix is known and positive definite, i.e., $cov(\varepsilon) = E(\varepsilon_i \varepsilon_i^T) = Q$, and Q is positive definite and known

The linear constraints are formed by stoichiometric constraint, energy, mass and other balances. In this case, they are linear and homogenous in nature, and could be given by:

$$Ax = 0 \tag{3.2}$$

where:

A = (n x s) coefficient matrix representing the linear process model (A is of full row rank)

If one assumes that the measurement errors are normally distributed, the data reconciliation problem may be formulated as the following constrained least square estimation problem:

$$\min_{x} [\ (y-x)^T Q^{-1} (y-x)\ ] \tag{3.3}$$

subject to $Ax = 0$

The solution is given by:

$$\hat{X} = y - QA^T (AQA^T)^{-1} Ay \tag{3.4}$$

Crowe et al. (1983) used a projection matrix to obtain a reduced set of balance equations, and also to include a statistical test for the removal of gross errors. Mah et al. (1990) also disclosed the decomposition and solution of these equations in detail.

The linear constraints given by stoichiometric constraints, energy, mass and other balances are generalized into the form:

$$A_1 x + A_2 u = c \tag{3.5}$$

where A_1 = (n x p) matrix of coefficient of known constants

A_2 = (n x m) matrix of coefficients of known constants

u = (m x 1) vector of parameters not directly related to the measurements through the basic model equation

c = (n x 1) vector of known constants

3.3.3 Nonlinear Steady-State Data Reconciliation

Nonlinear steady-state data reconciliation is the extension of the previously mentioned method for including nonlinear constraints. These occur when variables are measured indirectly by other physical properties, e.g., concentration by density, pH or thermal conductivity. Bilinear constraints occur when variables appear in two balances, e.g., temperatures measured along with concentrations will appear in both energy and component material balances. Bilinear constraints also occur when a stream is split into two or more streams of the same temperature and composition. No general analytical solution is available to bilinear and nonlinear data reconciliation problems. Nonlinear constraints are more difficult to determine whether unmeasured variables may be estimated, and how to decompose stress the problem with missing measurements.

A simple method of solving the nonlinear data reconciliation problem is to extend the linear case to handle the nonlinear constraints:

$$\min_{x} [(y - x)^{T} Q^{-1} (y - x)] \tag{3.6}$$

Subject to $h(x) = 0$

$g(x) \geq 0$

where $h(x)$ = set of equality constraints representing the process model

$g(x)$ = any inequality constraints present in the process model

There is no longer any analytical solution to the problem given in equation (3.5). The problem must be solved by using an optimization algorithm. For practical application of this method, weights are added to the least square objective function to place emphasis on key measured variables. A large weighing factor for a given variable will force the optimization to reduce the degree of adjustment for that variable. These

weights are usually chosen based on operational experience. The objective function for the nonlinear data reconciliation then becomes:

$$\min_{x}\left[\sum_{i=1}^{n} w_i\left(\frac{y_i - x_i}{s_i}\right)^2\right] \tag{3.7}$$

Subject to $h(x) = 0$

where w = weighting factor associated with each measured variable

y = measured variable

x = true variable value

s = the standard deviation of the measured value

The construction of the projection matrices is direct made and each requires the inversion of a matrix needed for subsequent steps in the solution. The method was introduced by Crowe (1983), Crowe (1986). It could be applied by using two successive projection matrices to the equation.