CHAPTER II
THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1 Errors in Measurement

Measured process data certainly contain some inaccurate information by errors during the measurement, processing and transmission of the measured signal. Errors in measured data can lead to significant deterioration in plant operation. The total error in a measurement which is the difference between the measured value and the true value of a variable can be classified into two types, random and gross errors.

2.1.1 Random Errors

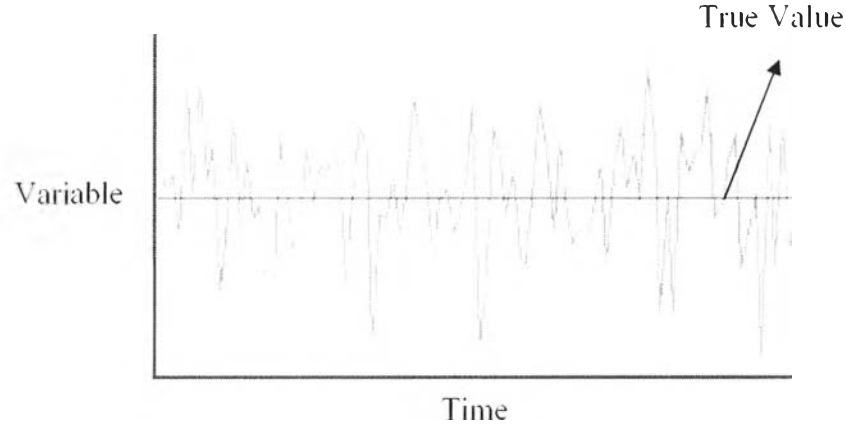

Figure 2.1 Example of random errors

Random errors are errors in measurement that lead to measurable values being inconsistent when repeated measures of a constant attribute or quantity are taken. Random errors cannot be completely eliminated and are always present in any measurement. Although the measurement is repeated with the same instrument under identical process conditions, random errors may be appeared.

Random errors are caused by unpredictable fluctuations in the readings of a measurement apparatus, or in the experimenter's interpretation of the instrumental reading; these fluctuations may be in part due to interference of the environment with

the measurement process. And random errors occur high frequently but their magnitudes are typically less than the other errors such as gross errors.

2.1.2 Gross Errors

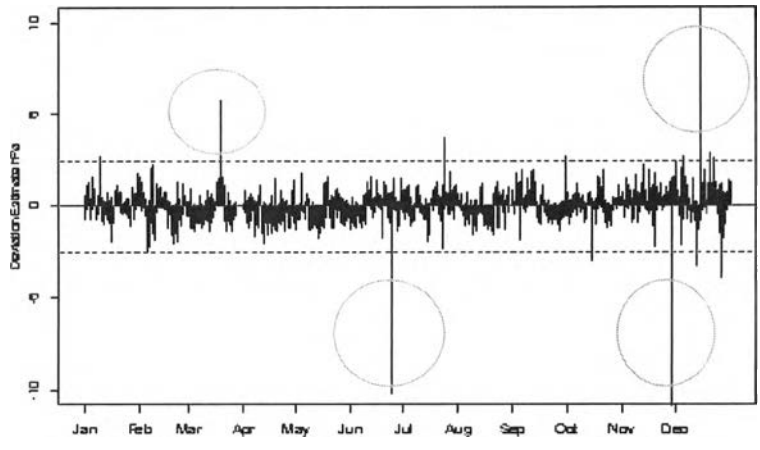

Figure 2.2 Example of gross errors.
(http://www.univie.ac.at/IMG-Wien/daquamap/Interpretation.html)

Gross errors are caused by nonrandom events such as instrument malfunctioning, miscalibration, wear or corrosion of sensor, and solid deposits. If the measurement is repeated with the same instrument under identical conditions, the contribution of a systematic gross error to the measured value will be the same. These errors occur less frequently but their magnitudes are typically larger than random errors. However, gross errors can be prevented by good installation and maintenance procedures.

2.2 Data Improvement Technique

A simplified view of measurement data improvement techniques can be divided into three basic steps as shown in Figure 2.3. The first step, variable classification, involves determining which variables are observable or unobservable and which ones are redundant or underdetermined. Several authors have published algo-

rithms for variable classification such as Crowe (1986), Stanley and Mah (1981), Mah (1990).

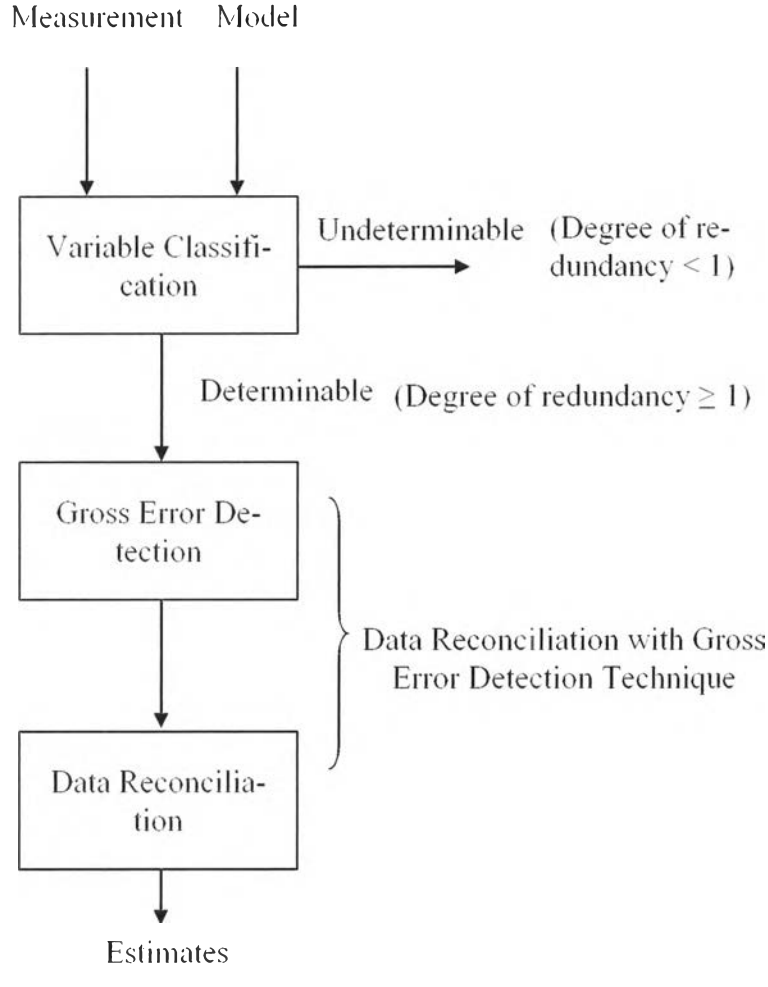

Figure 2.3 Steps for data improvement.
(Thomas F. Edgar, "Optimization of chemical processes", Second Edition, 2011)

If the system is undeterminable, it is not available for improvement but if the system is determinable, the next step will be continued: gross errors are all detected and eliminated in this step. Several methods proposed for gross error detection have been evaluated by Mah (1990), Rollins *et al.* (1996) and Tong and Crowe (1997). After that the data reconciliation will be purposed in the final step to remove the remaining small, random measurement errors from data.

2.2.1 Process Variable Classification

It is also important to clarify some concepts in data reconciliation techniques. Measured variables are classified as redundant and non-redundant, whereas unmeasured variables are classified as observable and non-observable. The classification of process variables is shown in Figure 2.4.

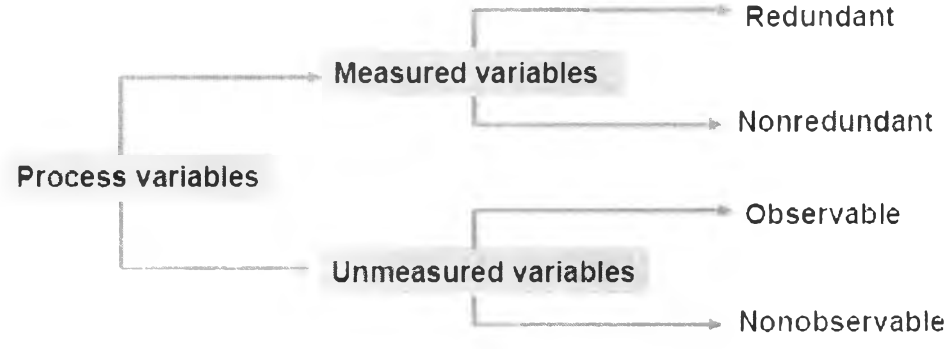

Figure 2.4 Classification of process variables.
(University of Ottawa & North Carolina State University, "Introduction to Data Reconciliation", 2003)

- A redundant variable is a measured variable that can be estimated by other measured variables via process models, in addition to its measurement.
- A non-redundant variable is a measured variable that cannot be estimated other than by its own measurement.
- An observable variable is an unmeasured variable that can be estimated from measured variables through physical models.
- A non-observable variable is a variable for which no information is available.

2.2.1.1 Redundancy

Sensor and topological redundancy

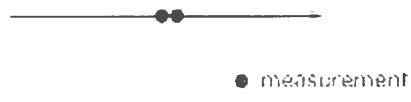

Figure 2.5 Sensor redundancy.
(http://en.wikipedia.org/wiki/Data_Validation_and_Reconciliation)

Sensor redundancy arising from multiple sensors of the same quantity at the same time at the same place.

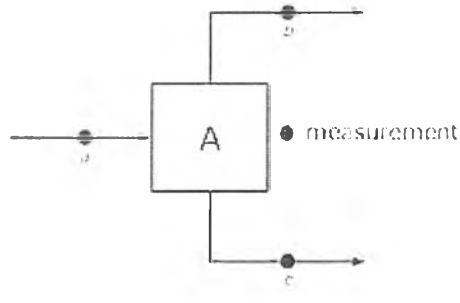

Figure 2.6 Topological redundancy.
(http://en.wikipedia.org/wiki/Data_Validation_and_Reconciliation)

Topological redundancy arising from model information, using the mass conservation constraint $a = b + c$, for example one can calculate c, when a and b are known.

Data reconciliation and gross error detection both achieve error reduction only by exploiting the redundancy property of measurements. Redundancy is a source of information that is used to correct the measurements in order to satisfy the process constraints. Redundancy can be due to sensor redundancy, where sensors

are duplicated in order to have more than one measurement of the same quantity. Redundancy can also arise from topological redundancy, where a single variable can be estimated in several independent ways, from separate sets of measurements. Topological redundancy is intimately linked with the degree of freedom (DOF) of a mathematical system, i.e. the minimum number of pieces of information that are required in order to calculate all of the system variables. From the example above, the flow conservation requires that $a = b + c$, and it is clear that one needs to know the value of two of the 3 variables in order to calculate the third one. Therefore, DOF in that case is equal to 2.

About topological redundancy, we denote x is the unmeasured variables and y is the measured variables. Then the system of the process constraints becomes $F(x,y) = 0$, which is a nonlinear system in y and x. If the system $F(x,y) = 0$ is calculable with the n measurements given, then the level of topological redundancy is defined as $redundancy = n - DOF$, the number of additional measurements which are required in order to just calculate the system. Another way of viewing the level of redundancy is to use the definition of DOF, which is the difference between the number of variables (measured and unmeasured) and the number of equations. Then one gets $redundancy = n - DOF = n - (n + m - p) = p - m$, the difference between the number of equations, p, and the number of unmeasured variables, m. The level of total redundancy is the sum of sensor redundancy and topological redundancy.

In any process, the variables are related to each other through physical constraints such as material or energy conservation laws. Given a set of such system constraints, a minimum number of measurements are required in order to calculate all of the system parameters and variables. If there are more measurements than this minimum, then redundancy exists in the measurements that can be exploited. This type of redundancy is called spatial redundancy and the system of equations is said to be over-determined.

Data reconciliation cannot be performed without spatial redundancy. With no extra measured information, the system is just determined and no correction to erroneous measurements is possible. Further, if fewer variables than necessary to determine the system is measured, the system is underdetermined and the values

of some variables can be estimated only through other means or if additional measurements are provided.

A second type of redundancy is temporal redundancy. This arises due to the fact that measurements of process variables are made continually in time at a high sampling rate, producing more data than necessary to determine a steady-state process. If the process is assumed to be in a steady-state condition, then temporal redundancy can be exploited by simply averaging the measurements, and applying steady-state data reconciliation to the averaged values. If the process state is dynamic, however, the evolution of the process state is described by differential equations corresponding to mass and energy balances, which provide both spatial and temporal redundancy of measured variables.

2.2.1.2 Error Reduction Methods

Digital filters have been used to reduce random errors (high-frequency noise) in process values. These filters are very helpful tools for data conditioning before data reconciliation. Various classical digital filters have been designed. Each type has its own advantages, as well as related defects. Some reduce significantly the errors, but with large delay. Others have less delay, but overshoot undershoot after a true step process change. In general, a compromise between the amount of noise attenuation and the time delay in the filtered results is required in order to achieve the best performance for any types of filter. This can be accomplished by tuning the filter parameters which, unfortunately, is not easy to tune.

Data filtering is different from data smoothing which deals with the past data. Data filtering estimates the current value based on the current and past measurements and it is of primary concern in process control. Data smoothing estimates the value of the central point from past and recent measurements (values from both side of the central point) and it is mainly used for fault diagnostic and steady-state process optimization.

2.2.2 Data Reconciliation and Gross Error Detection Technique

In modern petroleum and petrochemical plant consist of a large number of process units such as reaction vessels, distillation columns, storage tanks, etc., which are interconnected together by a complicated network of streams. Measurements such as flow rates, temperatures, pressures, levels, concentrations of components and automatically recorded are routinely made for the purpose of process control, online optimization, or process performance evaluation. The use of computers to collect data not only allows data to be obtained a greater frequency, but has also remove errors present in manual recording. Hence, the accuracy and validity of process data are greatly improved by using the technique called "Data Reconciliation and Gross Error Detection". These two technique have receive considerable attention in chemical engineering literature consequently such as Almasy and Szatno (1975), Crowe, Campos and Hrymak (1983), Knepper and Gorman (1980), Kuehn and Davidson (1960), Madron, Veverka and Venecek (1977), Mah, Stanley and Downing (1976), Mah and Tamhane (1982), Murthy (1973), Nogita (1972), Reilly and Carpani (1963), Ripps (1965), Romagnoli and Stephanopolous (1981) (Tamhane and Mah, 1985).

2.2.2.1 Data Reconciliation

Data reconciliation is a technique that has been developed to improve the accuracy of measurements by reducing the effect of random errors in the data. A solution to the data reconciliation problem was first proposed by Kuehn and Davison (1961) who used Lagrange multipliers for the case where all component flow rates were measured. Also Crowe et al. (1983) and Crowe (1986) proposed an approach which computes the optimal solution using matrix projection and so on (Kim *et al.*, 1997). The major difference between data reconciliation and other filtering techniques is data reconciliation use the process model constraints and obtains estimates of process variables by adjusting process measurements so that the estimates satisfy the constraints and the reconciled estimates would be more accurate than the measurements and also consistent with the known relationships between process variables as defined by the constraints. The word, "More Accurate", is usually defined as the optimal solution to a constrained least-square or maximum likelih-

ood objective function. It is important to understand what is wrong with the value obtained by measurement and why they must be adjusted (Romagnoli and Sanchez, 1999).

Data reconciliation can make the process data more useful for making decision and control by smoothing, eliminating outliers and adjusting for bias and drift, so leading to better quality control, detection of faulty instrumentation and increased the profits.

Data reconciliation techniques for error reduction can be applied to industrial processes as part of integrated strategy referred to as data conditioning or data rectification. Figure 2.7 shows the various operations and the position of data reconciliation in data conditioning for online industrial applications.

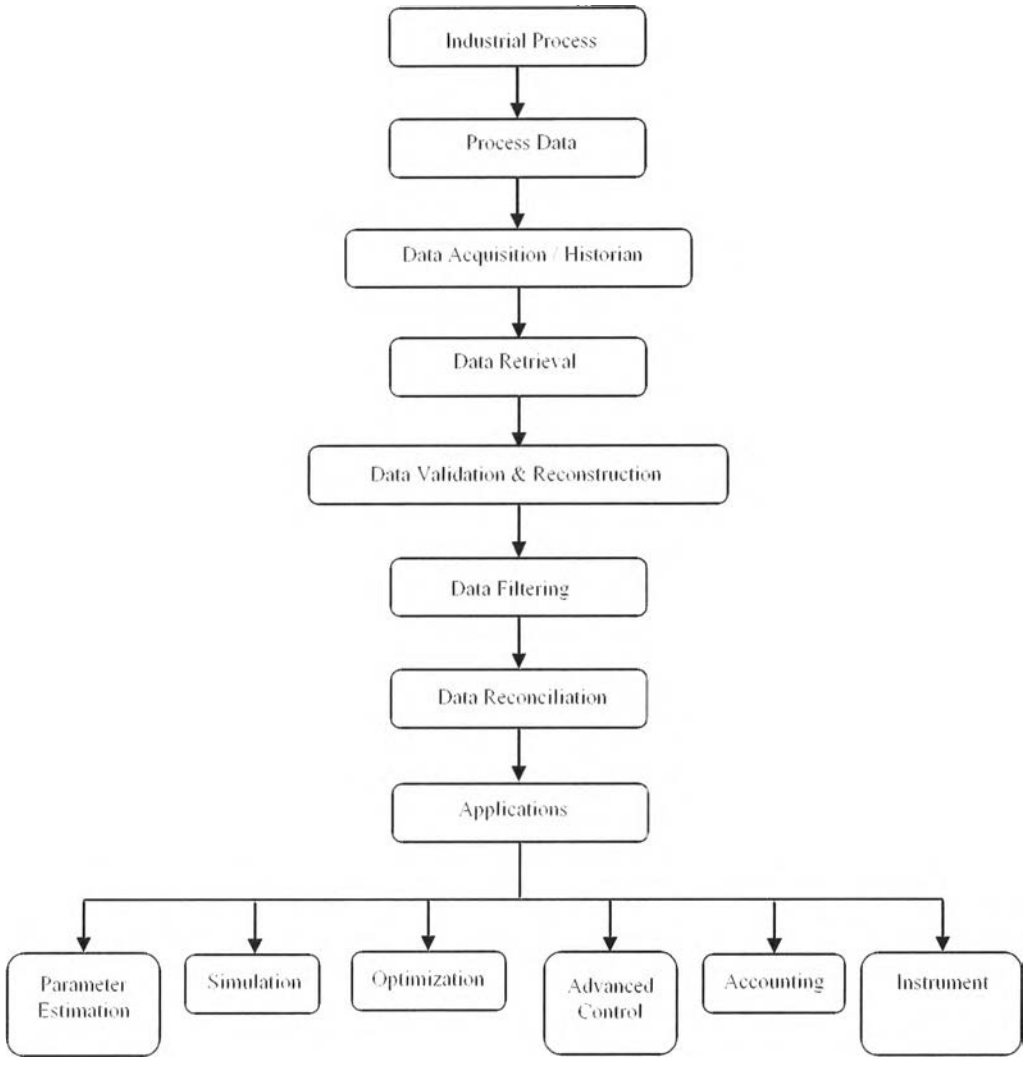

Figure 2.7 Online data collection and conditioning system.
(Narasimhan and Jordache, "Data Reconciliation & Gross Error Detection", 2000.)

In doing the data reconciliation, it is supposed that the relationship between a measurement of variable and its true value can be represented by

$$y_i = x_i + \varepsilon_i \tag{2.1}$$

From equation 2.1, let y_i represent the measured value, x_i represent the true value (Estimated reconcile value) and ε_i represent the random measurement error (assuming no gross error existing in the system).

For steady-state data and processes, Kuehn and Davidson (1961) presented the seminal paper describing the data reconciliation problem based on least-square optimization. For dynamic data and processes, Kalman filtering (Gelb, 1974) has been successfully used to recursively smooth measurement data and estimate parameter. Both techniques were developed for linear systems and weighted least-square objective function.

The amount of adjustment made to the measurements is minimized since the random errors in the measurements are expected to be small. In general, data reconciliation can be formulated by the following constrained weighted least-square optimization problem (generalized least-square objective function).

$$Min\ (y-x)^T \sum^{-1} (y-x) \qquad \text{(a)}$$

From equation a, if we consider the matrix $\sum$ to be diagonal, equation a becomes

$$Min\ \sum_{i=1}^{n} w_i \left(\frac{y_i - x_i}{\sigma_i}\right)^2 \qquad \text{(b)}$$

Subject to $g_k(x_i, u_j) = 0 \qquad k = 1, \ldots, m \qquad (2.2)$

Where w_i are the weights

y_i is the measured value

x_i is the reconciled estimate for variable, i

u_j are the estimates of unmeasured variables, j

The objective function, equation b, defines the total weighted sum square of adjustments made to measurements. Equation 2.2 defines the set of model constraints. The weights, w_i, are chosen depending on the accuracy of different measurements.

The model constraints are generally material and energy balances, but could include inequality relations imposed by feasibility of process operations. The laws of conservation of mass or energy are typically used as constraints for data

reconciliation because they are usually known. Empirical or other types of equations involving many unmeasured parameters are not recommended, since they are well-known only approximately. Forcing the measured variables to obey inexact relations can cause inaccurate data reconciliation solution and incorrect gross error diagnosis.

Any mass or energy conservation law can be expressed in the following general form:

$$Input - Output + Generation - Consumption - Accumulation = 0 \qquad (2.3)$$

The quantity for which the above equation is written could be the overall material flow, the flow of individual components, or the flow of energy. If there is no accumulation of any of these quantities, then these constraints are algebraic in character and define as steady-state operation. For a dynamic process, however, the accumulation terms cannot be neglected and the constraints are differential equations. For most process units, there is no generation or depletion of material. In the case of reactors, the generation or depletion of individual components due to reaction should be taken into account.

The types of constraints that are imposed in reconciliation depend on the scope of the reconciliation problem and the type of process units. Furthermore, the complexity of the solution techniques used depends strongly on the constraints imposed. For example, if we are interested in reconciling only the flow rates of all streams, then the material balances constraints are linear in the flow variables and linear data reconciliation problem results. On the other hand, if we wish to reconcile composition, temperature or pressure measurements along with flows, then a nonlinear data reconciliation problem occurs.

Another important question is whether to perform reconciliation using a steady-state or a dynamic model of the process. Practically, a process is never truly at a steady-state condition. However, a plant is normally operated for several hours or days in a region around a nominal steady-state operating point. For applications such as online optimization where reconciliation is performed once every few hours, it is appropriate to employ steady-state reconciliation on measurements averaged over the time period of interest.

During transient conditions (such as during a changeover to a new crude type in a refinery) when the departure from steady state is significant, steady-state reconciliation should not be applied because it will result in large adjustments to measured values. Measurements taken during such transient periods can be reconciled, if necessary, using a dynamic model of the process. Similarly for process control applications where reconciliation needs to be performed every few minutes, dynamic data reconciliation is appropriate.

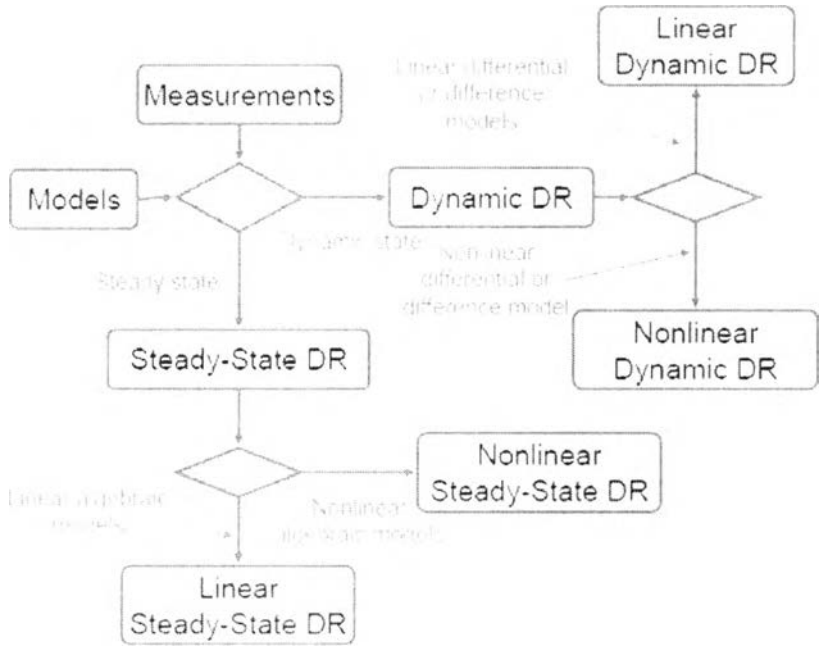

Figure 2.8 Sub-problems in data reconciliation.
(University of Ottawa & North Carolina State University, "Introduction to Data Reconciliation.", 2003)

Dynamic data reconciliation models were first presented by Stanley and Mah (1977), who adapted Kalman filtering in a quasi steady-state condition

Extended Kalman filtering has been a popular method used to solve the dynamic data reconciliation problem (Muske and Edgar, 1998). As an alternative, the nonlinear dynamic data reconciliation problem with a weighted least-

square objective function can be expressed as a moving horizon problem (Liebman *et al.*, 1992).

The nonlinear objective function is

$$Min\ f(y(t), x(t)) \tag{2.4}$$

Subject to the dynamic model

$$h\left(\frac{dx(t)}{dt}, x(t)\right) = 0 \tag{2.5}$$

And inequality constraints

$$g(x(t)) \geq 0 \tag{2.6}$$

This problem can be solved using a combined optimization and constraint model solution strategy (Muske and Edgar, 1998) by converting the differential equations to algebraic constraints using orthogonal collocation or some other model discretization approach.

Weiss *et al.*, (1996) successfully used data reconciliation to an industrial pyrolysis reactor in linear and nonlinear method. Both methods were used to solve the data reconciliation problem. The first method, linear, which included successive linearization, yielded results similar to yield from nonlinear method. The overall heat transfer coefficient, one of the operating parameters of the pyrolysis reactor, calculated by using reconciled data showed a trend consistent with plant experience and could be used to determine the better regeneration cycle time of the reactor.

Bagajewicz *et al.*, (2000) presented the comparison of the performance of integral approach to dynamic data reconciliation and steady-state data reconciliation. Dynamic data, subject to averaging and steady-state reconciliation will be compared with the average of the results of dynamic data reconciliation and the

averages of the true values. It is shown that in the absence of biases and leaks, the performance of both approaches is similar. Moreover, it is proven that once the appropriate variance is chosen, both methods are identical in the absence of accumulation terms. Finally, an analysis is made on how large the discrepancies are when there are accumulation terms.

Lid *et al.*, (2008) studied data reconciliation and optimal operation of a catalytic naphtha reformer. The reformer also has an important function as the producer of hydrogen to the refinery hydrotreaters. A process model based on a unit model structure is used for estimation of the process condition using data reconciliation. The process model is fitted to 21 data sets from the naphtha reformer that collected in a two year period and includes feed and product analysis and process measurements. Measurements are classified as redundant or non-redundant and the model variables are classified as observable, barely observable or unobservable. The computed uncertainty of the measured and unmeasured variables shows that even if a variable is observable it may have a very large uncertainty and may thereby be practically unobservable.

The general assumptions of data reconciliation are only random errors are presented in the measurements which follow a normal (Gaussian) distribution, with known variances, this means the mean of measurement errors is assumed to be zero (Kim *et al.*, 1997). And a system is operated at steady-state condition. If a gross error due to a measurement bias is present in some measurement or if a significant process leak is present which has not been accounted for in the model constraints, then the reconciled data may be very inaccurate. It is therefore necessary to identify and remove such gross errors. Therefore no gross errors either in the measurements or in the process model constraints make the data reconciliation effectively

2.2.2.2 Gross Error Detection

Gross error detection is a companion technique to data reconciliation that has been developed to identify and eliminate gross errors. Thus, data reconciliation and gross error detection are applied together to improve accuracy of measured data.

From steps for data improvement, as shown in Figure 2.3, show that the gross error will be detected and eliminated in the second step but actually in practice, gross error detection and elimination are usually performed iteratively along with the final step of data reconciliation.

The gross error detection problem has been studied since 1960s. There are several review papers on this subject such as Mah (1987), Crowe (1994), and Yang (1992). Many methods have been created to solve the problem so they can be categorized as follows (Yang *et al.*, 1995):

- Based on the assumption of data normal distribution:
  - Global Test (1965)
  - Maximum Power Test (1975)
  - Constraint or Nodal Test (1976)
  - Measurement Test (1962)
  - Akaike Information Criterion Method (1986)
  - Generalized Likelihood Ratios (1989)
  - Principle Component Analysis (1994)
- Based on data abnormal distribution:
  - Bivariate Likelihood Distribution (1991)
  - Non-central Probability Density Function (1993)
- Neural Network Methods:

There are many main methods to do the gross error detection based on *linear steady-state processes* such as Global Test (GT), Measurement Test (MT) and Nodal Test (NT), we will discuss each of them below.

Global Test (GT)

Figure 2.9 Chi-squared distributions. (Montgomery and Runger, 2003)

For detection of gross error, many authors, Almasy and Szatno (1975), Madron *et al.* (1977), Rip (1965), have suggested the use of a global Chi-squared statistic constructed from the observed discrepancies in the constraints.

The test is based on the vector of balance residuals, r, which is given by

$$r = Ay - c \tag{2.7}$$

Where r is the vector of residual

A is the vector of balance matrix

y is the measured value

c is the true value

Under the null hypothesis, H_0, that there are no gross errors present, the vector r is a multivariate normal distribution with zero mean value and variance-covariance matrix, V, given by

$$V = cov(r) = A\sum A^T \tag{2.8}$$

Where $\sum$ is the measurement variance-covariance matrix

And Global test uses the test statistic given by

$$\gamma = r^T V^{-1} r \tag{2.9}$$

Under H_0, the above statistic follows a Chi-squared distribution with degrees of freedom, μ, where μ is the rank of matrix A. If the test criterion is chosen as $X^2_{1-\alpha,\mu}$, where $X^2_{1-\alpha,\mu}$ is the critical value of Chi-squared distribution at the chosen level of significant, α, which generally taken as α = 10%, 5% and 1% (with 5% level of significant, it means that 95% is trusted the gross error are absent), then H_0 is rejected and a gross error is detected, if $\gamma > X^2_{1-\alpha,\mu}$.

Basic global test statistic (the conventional gross error detection) can be applied by using the objective function of linear or nonlinear data reconciliation because the objective function of data reconciliation by least-square method is also a Chi-squared probability distribution, with degree of freedom, μ, which is equal to degree of redundancy, where degree of redundancy (DOR) is equal to the number of equation discarded by the number of unmeasured variable.

$$Min \sum_{i=1}^{n} w_i \left(\frac{y_i - x_i}{\sigma_i}\right)^2 \ll X^2_{1-\alpha,\mu} \qquad (2.10)$$

If the objective function value is less than $X^2_{1-\alpha,\mu}$ then gross errors are not expected in the system.

Nodal Test (NT)

Reilly and Capani (1963) and Mah *et al.* (1976) independently proposed performing a separate test on each nodal imbalance. The vector r can also be used to derive test statistics, one for each constraint i, given by

$$Z_{r,i} = \frac{|r_i|}{\sqrt{V_{ii}}} \qquad i = 1, 2, \ldots, m \qquad (2.11)$$

Or, written in vector form

$$Z_r = [diag(V)]^{-\frac{1}{2}} r \qquad (2.12)$$

Where $diag(V)$ is a diagonal matrix whose diagonal elements are V_{ii}

The nodal test uses the test statistics, $Z_{r,i}$ for gross error detection. It can be proved that $Z_{r,i}$ follows a standard normal distribution. If any of the test statistics $Z_{r,i}$ exceed the test criterion $Z_{1-\alpha/2}$ where $Z_{1-\alpha/2}$ is the critical value of the standard normal distribution for chosen level of significance, α, a gross error is detected.

And the test criterion can be performed in another form by using Sidak (1967) inequality to decrease the probability of Type I error, in other words, the probability of Type I error may be more than the specified value of α. If we wish to control the Type I error probability, the following modified level of significance, β, can be used.

For any specified value of α, the modified value β can be compute by using equation 2.13

$$\beta = 1 - (1 - \alpha)^{1/m} \tag{2.13}$$

Measurement Test (MT)

Mah and Tamhane (1982) proposed the measurement test. The test based on the vector of measurement adjustment as below.

$$a = y - \hat{x} \tag{2.14}$$

Where reconciled estimates, $\hat{x}$, are obtained by using data reconciliation technique. Using this solution, the measurement adjustments can also be showed as

$$a = \Sigma A^T V^{-1} r \tag{2.15}$$

$$d = \Sigma^{-1} a \tag{2.16}$$

The result of a and d will be the same value if the variance matrix, Σ, is diagonal matrix and under H_0, d is also generally distributed with zero mean and a covariance matrix

$$W = cov(d) = A^T \Gamma^{-1} A \tag{2.17}$$

And the following test statistic is

$$Z_{d,j} = \frac{|d_j|}{\sqrt{W_{jj}}} \qquad j = 1, 2, \ldots, n \tag{2.18}$$

Similar to the nodal test, the measurement test also involves multiple univariate tests. Thus the criterion statistic test can be used the same as nodal test and if we choose to use Sidak (1967) inequality to increase the performance of criterion test, we can use equation 2.13 by replacing m to n, the number of univariate measurement test (the number of measured variable) and the statistical criterion test

Or the basic measurement test (the conventional gross error detection and elimination) can detect gross errors by using the confidence interval ($\pm Z\sigma$, depending on level of significant). The confidence interval can be calculated by using standard deviation of measurement adjustment, σ, (measurement adjustment is equal to measured value discarded by reconciled value), if the measurement adjustment value, a, is larger than the confidence interval, then gross error is detected.

And the estimator value, Z, can be calculated by using equation 2.19

$$\%confidence\ level = erf\left(\frac{|Z|}{\sqrt{2}}\right) \tag{2.19}$$

Where Z is the estimator value

Erf is the error function

There are many methods to identify the multiple gross error detection based on *linear steady-state processes* by categorizing as simultaneous strategy, serial strategy and combinatorial strategy below.

Simultaneous strategies

- Identification using single gross error test statistics
- Identification using combinatorial hypotheses
- Identification using simultaneous of gross error magnitudes

Serial strategies

- Serial Elimination
  - Iterative Measurement Test
  - Modified Iterative Measurement Test
- Serial Compensation
  - Simple Serial Compensation Strategy
  - Modified Serial Compensation Strategy
  - Bounded GLR

Combinatorial strategies

- The linear combination technique

Kim *et al.*, (1997) studied The Modified Iterative Measurement Test (MIMT) gross error detection algorithm using nonlinear programming (NLP) techniques to improve its robustness and performance. The algorithm has been tested on a continuous stirred tank reactor (CSTR) example. They found that the performance of data reconciliation using nonlinear programming in the presence of gross errors was better than the conventional data reconciliation algorithm using a linearization technique. The conventional gross error detection algorithm such as MIMT using linearization techniques was not able to compute the optimal gross error estimates as the number of gross errors in the measurement was increased. Thus the performance of this algorithm rapidly deteriorated as the number of gross errors increased. But the enhanced gross error detection algorithm using NLP not only showed robustness to the number of gross errors but also computed reliable reconciled estimates for the gross errors. Therefore, this enhanced algorithm should be able to reconcile the measured data effectively for highly nonlinear chemical processes.

The Modified Iterative Measurement Test (MIMT) (Kim *et al.*, 1997)

The root of Iterative Measurement Test (IMT) is purposed by Serth and Heenan, (1986), and Kim *et al.*, (1997) modified it by using NLP, namely the Modified Iterative Measurement Test and its algorithm is followed below and its flowsheet is shown in Figure 2.10.

Let set S be the set of original measured variables. Let set C be the set of measurements which are identified as containing gross errors. Let set T be the set of measured variables in reduced reconciliation problem

Step 1: Solve the initial reconciliation problem. Compute the vectors $\hat{x}$ (reconciled estimate value), a (equation 15), and d (equation 16)

Step 2: Compute the measurement statistics, $Z_{d,j}$, (equation 18) for each measured variable

Step 3: Find Z_{max} the maximum absolute value among all $Z_{d,j}$ from step 2 and compare it with the test criterion $Z_c = Z_{1-\beta/2}$ If $Z_{max} \leq Z_c$, proceed to step 6. Otherwise, select the measurement corresponding to Z_{max} and delete it from the set of measured variables (set S) and add it to the set of unmeasured variable (set C) If two or more measurement test statistics are equal to Z_{max}, select the measurement with the lowest index j to add to set C

Step 4: Remove the measurements contained in C from set S. Solve the data reconciliation problem treating the variables corresponding to set C also as unmeasured. Obtain T, the set of measurements in the reduced data reconciliation problem, and the vectors a and d corresponding to these measurements

Step 5: Determine if the reconciled value for all variables in set T and set C are within their prescribed lower and upper bounds. If all reconciled values are within the bounds, store the current solution and return to step 2. Otherwise, delete the last entry in C, replace it by measured variable corresponding to the next largest value of $|Z_{d,j}| > Z_c$ and return to step 4. If $|Z_{d,j}| \leq Z_c$ for all remaining variables, delete the last entry in set C and go to step 6.

Step 6: The measurements containing in set C are suspected of containing gross errors. The reconciled estimates after removal of these measurements are those obtained in step 4 of last iteration.

And all of data reconciliation steps use the nonlinear programming technique (Kim *et al.*, 1997).

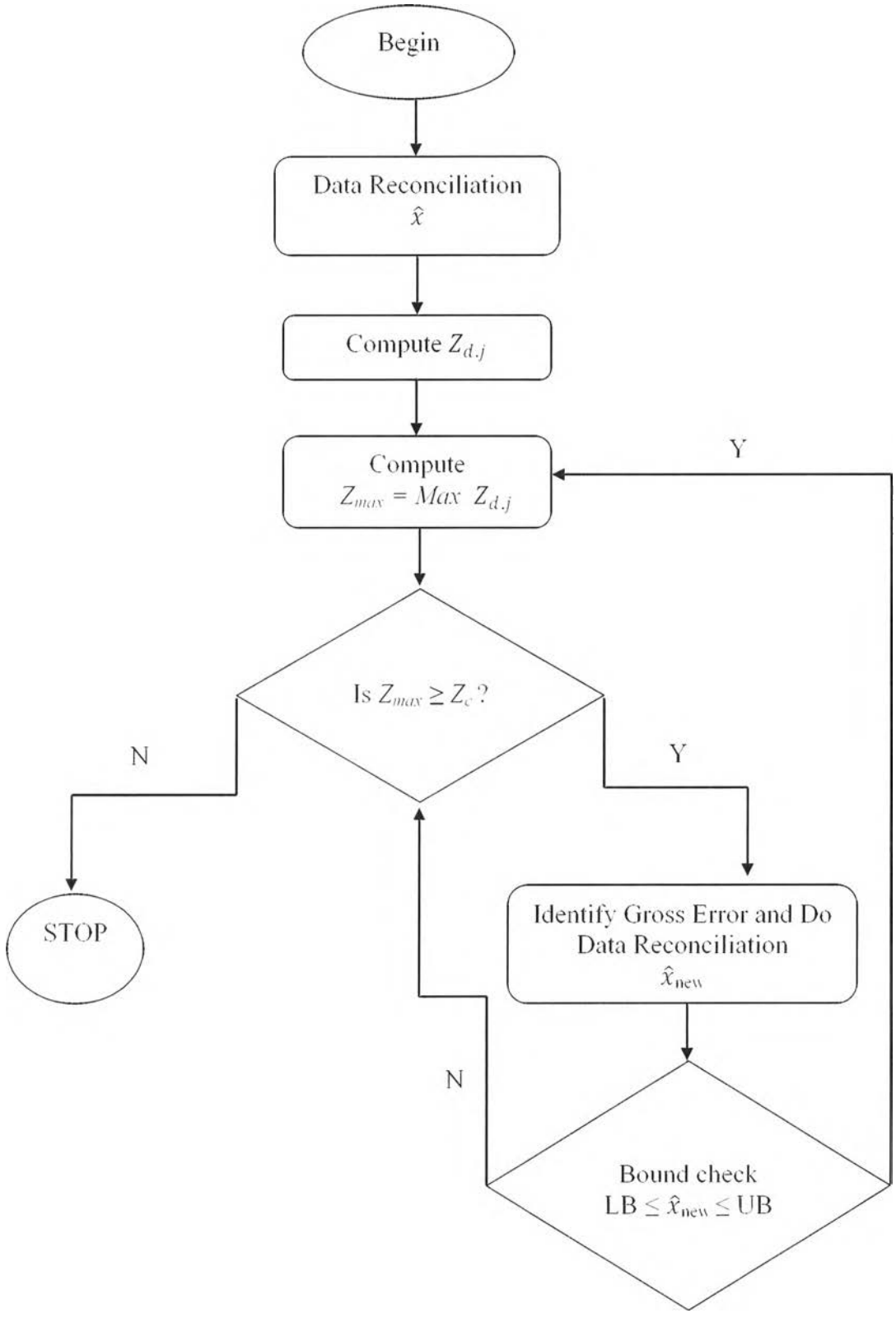

Figure 2.10 Flowsheet of MIMT using NLP algorithm. (Kim *et al.*, 1997)

Bagajewicz *et al.*, (1998) proposed a method to identify and estimate gross errors in plant linear dynamic data reconciliation. Three kinds of gross errors in process plants i.e. flow rate biases, holdup biases and tank leaks are considered and included in the integral model. The strategy to identify and estimate the gross error performs with efficiency even in the presence of a high number of errors. Gross errors are identified without the need for measurement elimination. The strategy is capable of effectively identifying a large number of gross errors by using the serial method.

Sanchez *et al.*, (1996) studied the method or strategy that allows the identification of gross errors for pyrolysis reactor measurement. A reactor model is performed in terms of mass and heat balance and input-output mapping based on available measurements. Gross error detection performance is evaluated by using Generalized Likelihood Ratio, GLR (Narasimhan *et al.*, 1987) and Simultaneous Estimation of Gross Errors, SEGE (Sanchez *et al.*, 1995) method to detect, identify and estimate the magnitude of simulated gross errors in the process variables and use the expected fraction of correct identification (OPF) as a performance measure (Rollins and Davis, 1992). Both GLR and SEGE behave satisfactorily but SEGE gives greater OPF than GLR around 90% of results.

Wang *et al.*, (2004) found that there are some defects and the decrease of the coefficient matrix rank is the most serious one. So, they presented an improved MT–NT combined method for solving this problem. The proposed method overcomes them by using a strategy of detecting and reconciling via successive iteration. The case study shows the improved performance compared to the existing MT–NT method. The improved MT–NT method can avoid this defect without increasing the operation since it uses the estimates to replace the measurements with gross errors so that the equations are not changed. An example shows that this method is efficient and reliable in practical use.