INTRODUCTION TO DATA RECONCILIATION AND GROSS ERROR DIAGNOSIS

Process Data Conditioning Methods

In any modern chemical plant, petrochemical process or refinery, hundreds or even thousands of variables such as flow rates, temperatures, pressures, levels, compositions, etc. are routinely measured and automatically recorded for the purpose of process control, online optimization or process economic evaluation. Modern computers and data acquisition systems facilitate collection and processing of a great volume of data, often sampled with a frequency of the order of minutes or even seconds. The use of computers not only allows data to be obtained at a greater frequency, but has also resulted in the elimination of errors present in manual recording. This in itself has greatly improved the accuracy and validity of process data. However, the increased amount of information can be exploited for further improving the accuracy and consistency of process data through a systematic data checking and treatment.

Process measurements are inevitably corrupted by errors during the measurement, processing and transmission of the measured signal. The total error in a measurement, which is the difference between the measured value and the true value of a variable can be conveniently represented as the sum of the contributions from two types of errors - *random errors* and *gross errors*. The term random error implies that neither the magnitude nor the sign of the error can be predicted with certainty. In other words, if the measurement is repeated with the same instrument under identical process conditions, a different value may be obtained depending on the outcome of the random error. The only possible way these errors can be characterized is by use of probability distributions. These errors can be caused by a number of different sources such as power supply fluctuations, network transmission and signal conversion noise, analog input filtering, changes in ambient conditions, etc. Since, these errors can arise from different sources some of which may be beyond the control of the design engineer, they cannot be completely eliminated and are always present in any measurement. They usually correspond to the high frequency components of a measured signal, and are usually small in magnitude except for some occasional spikes.

On the other hand, gross errors are caused by nonrandom events such as instrument malfunctioning (due to improper installation of measuring devices), miscalibration, wear or corrosion of sensors, solid deposits, etc. The non-random nature of these errors implies that at any given time they have a certain magnitude and sign, which may be unknown. Thus, if the measurement is repeated with the same instrument under identical conditions, the contribution of a gross error to the measured value will be the same. By following good installation and maintenance procedures, it is possible to ensure that gross errors are not present in the measurements at least for some time. Gross errors caused by sensor miscalibration may occur suddenly at a particular time and thereafter remain at a constant level or magnitude. Other gross error causes such as wear or fouling of sensors can occur gradually over a period of time and thus the magnitude of the gross error increases slowly over a relatively long time period. Thus, gross errors occur less frequently but their magnitudes are typically larger than those of random errors.

Errors in measured data can lead to significant deterioration in plant performance. Small random and gross errors can lead to deterioration in the performance of control systems, whereas larger gross errors can nullify gains achievable through process optimization. In some cases, erroneous data can also drive the process into a uneconomic, or even worse, an unsafe operating regime. It is therefore important to reduce, if not completely eliminate, the effect of both random and gross errors. Several data processing techniques can be used together to achieve this objective. In this text we describe methods which can play an important role as part of an integrated data processing strategy to reduce errors in measurements made in continuous process industries.

Research and development in the area of signal conditioning have led to the design of analog and digital filters which can be used to attenuate the effect of high frequency noise in the measurements. Large gross errors can be initially detected by using various data validation checks. These include checking whether the measured data and the rate at which it is changing is within predefined operational limits. Nowadays smart sensors are available which can perform diagnostic checks to determine whether there is any hardware problem in measurement and whether the measured data is acceptable. More sophisticated techniques include statistical quality control tests (SQC) which can be used to detect significant errors (outliers) in process data.. These techniques are usually applied to each measured variable separately. Thus, although these methods improve the accuracy of the measurements, they do not make use of a process model and hence do not ensure consistency of the data with respect to the inter-relationships between different process variables. Nevertheless, these techniques must be used as a first step to reduce the effect of random errors in the data and to eliminate obvious gross errors. It is possible to further reduce the effect of random error and also eliminate systematic gross errors in the data by exploiting the relationships that are known to exist between different variables of a process. The techniques of *data* reconciliation and *gross error detection* that have been developed in the field of chemical engineering during the past 35 years for this purpose is the principal focus of this book.

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. The principal difference between data reconciliation and other filtering techniques is that data reconciliation explicitly makes use of process model constraints and obtains estimates of process variables by adjusting process measurements so that the estimates satisfy the constraints. The reconciled estimates are expected to be more accurate than the measurements and, more importantly, are also consistent with the known relationships between process variables as defined by the constraints. In order for data reconciliation to be effective, there should be no gross errors either in the measurements or in the process model constraints. 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.

Data reconciliation and gross error detection both achieve error reduction only by exploiting the redundancy property of measurements. Typically, 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 error free measurements is 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 which can be exploited. This type of redundancy is usually called *spatial redundancy* and the system of equations is said to be *overdetermined*. 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 are 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 that exists in measurements 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, then temporal redundancy can be exploited by simply averaging the measurements, and applying steady state data reconciliation to the averaged values. However, if the process state is dynamic then the evolution of the process state is described by differential equations corresponding to mass and energy balances, which inherently capture both the temporal and spatial redundancy of measured variables. For such a process, dynamic data reconciliation and gross error detection techniques have been developed to obtain accurate estimates consistent with the differential model equations of the process.

Signal processing and data reconciliation techniques for error reduction can be applied to industrial processes as part of an integrated strategy referred to as *data conditioning* or *data rectification.* Figure 1-1 illustrates the various operations and the position occupied by data reconciliation in data conditioning for online industrial applications

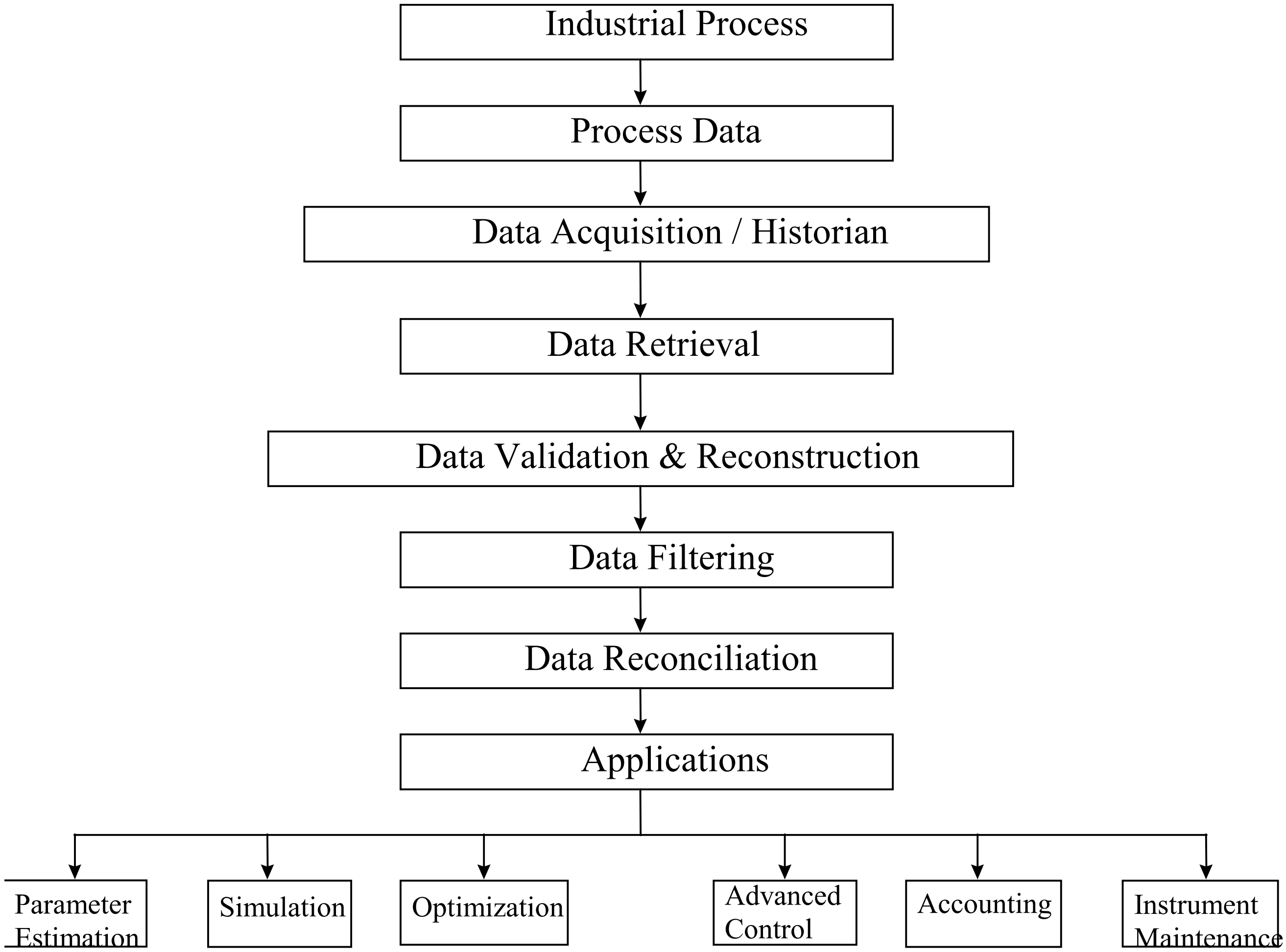

Figure 1.1 Online Data Collection and Conditioning System

Industrial Examples of Steady State Data Reconciliation

We briefly describe two examples of industrial applications of steady state data reconciliation drawn from our experience in order to illustrate the need for such a technique and the benefits that can be derived from it.

Crude Split Optimization in Preheat Train of a Refinery

In any refinery, the crude oil is initially heated by passing it through a interconnected set of heat exchangers called as the crude preheat train, before being fractionated. In a crude preheat train, typically the crude is split into one or more parallel streams each of which is heated, by passing it through a train of heat exchangers before being merged and sent to a furnace for further heating. The process streams that are used for heating the crude are the various product and pump-around streams from a downstream atmospheric or vacuum crude distillation column. In order to maximize energy recovery from these process streams the optimal flows of the crude splits through the different parallel heat exchanger trains should be determined online, say every few hours. For determining the optimal flows, the total inlet flow of crude and all hot process streams along with their inlet temperatures have to be specified. Moreover, details of all heat exchangers such as heat exchanger areas, overall heat transfer coefficients, etc. also have to be specified.

Generally, in a crude preheat train all the stream flows as well as all intermediate temperatures are measured. Thus there are more measurements than those required for performing the optimization. It is possible to ignore all measurements and use only the measurements of inlet flows and inlet temperatures of all streams for determining the optimal crude split flows. However, since all measurements contain errors any optimization exercise carried out using such measurements will not necessarily result in the predicted gains. In order to overcome this, steady state reconciliation and gross error detection is applied to measured data to eliminate measurements containing gross errors and obtain reconciled estimates of all stream flows and temperatures which satisfy the flow and enthalpy balances of the crude preheat train. As part of the reconciliation, the overall heat transfer coefficients of all exchangers are also estimated. These estimated heat transfer coefficients will more correctly reflect the actual current performance of the heat exchangers than their original design values. The reconciled estimates of inlet flows and temperatures of all streams, and the estimated overall heat transfer coefficients of all exchangers are used to determine the optimal values of the crude splits which are then implemented in the plant. Use of reconciled estimates in the optimization is likely to result in actual energy recovery from the process being close to the predicted optimal values.

It should be noted that the time periods selected for reconciliation and optimization are selected based on the time constants of the system. Since a change in the crude split flows has an effect on the downstream crude distillation column and hence affects all the distillate streams which are used for pre-heating the crude, it takes two hours for all the stream flows and temperatures to reach a new steady state after a new set of optimal crude splits values are implemented. The process is operated at this steady state for an additional two hours after which the optimization of the crude split flows is repeated. The measurements made during the preceding two hours of steady state operation are averaged and used as data for the reconciliation problem. This example is described in greater detail in the concluding chapter on industrial case studies.

Minimizing Water Consumption in Mineral Beneficiation Circuit

In a mineral beneficiation circuit, crushed ore is washed with water along with other additives in an interconnected network of classifiers or flotation cells in order to liberate the particles containing the minerals from the gangue material. In order to minimize the water consumption for a desired concentration of the beneficiated ore, the performance of the flotation cells have to be simulated for different flow conditions. The simulation model in turn requires data on the feed characteristics as well as parameters such as pulp densities. Generally, the flows of the feed stream and pure water streams are measured. Using samples drawn from different streams, the concentrations of different minerals in each stream and their pulp densities are also measured in the laboratory. These measurements contain errors and are also not consistent with the flow and component balances of the mineral beneficiation circuit. Steady state reconciliation and gross error detected can be applied to the measurements in order to obtain reconciled estimates of all stream flows, pulp densities and mineral concentrations such that they satisfy the material balances. The reconciled estimates are used in the detailed simulation of the flotation cells and to determine the minimal amount of water to be added. In one such exercise, it was possible to reduce water consumption by 5%.

Data Reconciliation Problem Formulation

As stated in the preceding sections, data reconciliation improves the accuracy of process data by adjusting the measured values so that they satisfy the process constraints. The amount of adjustment made to the measurements is minimized since the random errors in the measurements are expected to be small. In the general case, not all variables of the process are measured due to economic or technical limitations. The estimates of unmeasured variables as well as model parameters are also obtained as part of the reconciliation problem. The estimation of unmeasured values based on the reconciled measured values is also known as *data coaptation*. In general, data reconciliation can be formulated by the following constrained weighted *least squares optimization* problem.

$$\underset{x_i, u_j}{\text{Min}} \sum_{i=1}^{n} u_j (y_i - x_i)^2 \qquad (1\text{-}1)$$

$$\text{subject to } g_k(x_i, u_j) = 0 \qquad k = 1, \ldots, m \qquad (1\text{-}2)$$

The objective function (1-1) defines the total weighted sum square of adjustments made to measurements, where w_i are the weights, y_i is the measurement and x_i is the reconciled estimate for variable *i*, and u_j are the estimates of unmeasured variables. Equation (1-2) 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 deterministic natural 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 at best 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 [1]:

$$\text{input - output + generation - consumption - accumulation} = 0 \qquad (1\text{-}3)$$

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 a steady state operation. However for a dynamic process, the accumulation terms cannot be neglected and the constraints are differential equations. For most process units, there is no generation or depletion of material.

However, in the case of reactors the generation or depletion of individual components due to reaction should be taken into account.

For some simple units such as splitters, there is no change either in the composition or temperature of streams. For such units, the component and energy balances reduce to a simple form such as

$$x_i = x_j \qquad (1\text{-}4)$$

where the variable x_i represents either the temperature or composition of stream *i*. The above equation is also useful when two or more sensors are used to measure the same variable, say flow rate or temperature of a stream.

The type 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 depend 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 *a 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.

An issue to be addressed is the kind of constraints that we can legitimately impose in a data reconciliation application. Since data reconciliation forces the estimates of all variables to satisfy the imposed constraints, this issue assumes great importance. Usually, material and energy balance constraints are included because they are valid physical laws. It should be noted, however, these equations are generally written assuming that there is no loss of material or energy from the process unit to the environment. While this may be valid for material flow, significant losses in energy may occur for example from improperly insulated heat exchangers. In such cases, it is better not to impose the energy balances or alternatively include an unknown loss term in the balance equation which can be estimated as part of the reconciliation.

Other than material and energy conservation constraints, a model of a process unit can contain equations involving the unit parameters. For example, a heat exchanger model can include a rating equation relating the heat duty to the overall heat transfer coefficient, the exchanger area available for heat transfer, and the stream flows and temperatures. Equation (1-5) below describes this relationship.

$$Q - UA\Delta T_{\mathbf{ln}} = 0 \qquad (1\text{-}5)$$

where: Q is the heat duty, U is the overall heat transfer coefficient , A is the exchanger area , and ΔT_{ln} is the logarithmic mean temperature difference.

Should this equation be included as a constraint when applying data reconciliation to processes involving heat exchangers? Generally, since the overall heat transfer coefficient is unknown and has to be estimated from the measured data, this equation may be included and U estimated as part of the reconciliation problem. However, if there is no prior information about U and no feasibility restrictions on it, then inclusion of this constraint does not provide any additional information and estimates of all other variables will be the same regardless of whether this constraint is included or not. Thus, the data reconciliation problem can as well be solved without this constraint and U can subsequently be estimated by the above equation using the reconciled values of flows and temperatures. On the other hand, if U has to be within specified bounds or if there is a good estimate for U from a previous reconciliation exercise (as in the crude preheat train example discussed in the previous section, where the estimates of U from the reconciliation solution of the most recent time period can be used as good *a priori* estimates), then the constraint should be included along with the additional information about U as part of the reconciliation problem. The overall heat transfer coefficient can also be related to the physical properties of the streams, their flows, temperatures and the heat exchanger characteristics using correlations. It is not advisable to use such an equation in the reconciliation model since the correlations themselves can be quite erroneous and forcing the flows and temperatures to fit this equation may increase the inaccuracy of the estimates.

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. 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 (as in the case of crude split optimization example) 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.

Data reconciliation is based on the assumption that only random errors are present in the measurements which follow a normal (gaussian) distribution. 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. This is known as the *gross error detection problem*. Gross errors can be detected based on the extent to which the measurements violate the constraints or on the magnitude of the adjustments made to measurements in a preliminary data reconciliation. Although gross error detection techniques were developed primarily to improve the accuracy of reconciled estimates, they are also useful in identifying measuring instruments which need to be replaced or re-calibrated.

Examples of Simple Reconciliation Problems

In order to obtain a good understanding of the issues and underlying assumptions in data reconciliation, some of the simplest possible cases are introduced here. We assume a process operating at a steady state, constrained by a set of linear equations.

Systems with all measured variables. Let us first consider the simplest data reconciliation problem: the reconciliation of the stream flows of a process. Initially, all flow rates are assumed to be directly measured. The flow measurements contain unknown random errors. For that reason, the material input and output of every process unit and of the overall process do not balance. The aim of reconciliation is to make minor adjustments to the measurements in order to make them consistent with the material balances. The adjusted measurements, which are referred to as estimates, are expected to be more accurate than the measurements. Although the problem considered here is simple, it does have important industrial applications in accurate accounting for the material flows as, for example, in a lube blending plant, steam and water distribution subsystem of a plant, or in a complete refinery.

Example 1-1 : Flow Reconciliation Example

For example, let us consider a simple process of a heat exchanger with bypass as shown in Fig. 1-2. Let us also ignore the energy flows of this process and focus only on the mass flows. It is assumed that the flows of all six streams of this process are measured and that these measurements contain random errors. If we denote the true value of the flow of stream i by the variable x_i and the corresponding measured value by y_i, then we can relate them by the following equations

$$y_i = x_i + \varepsilon_i \qquad i = 1 \ldots 6 \tag{1-6}$$

where ε_i is the random error in measurement y_i.

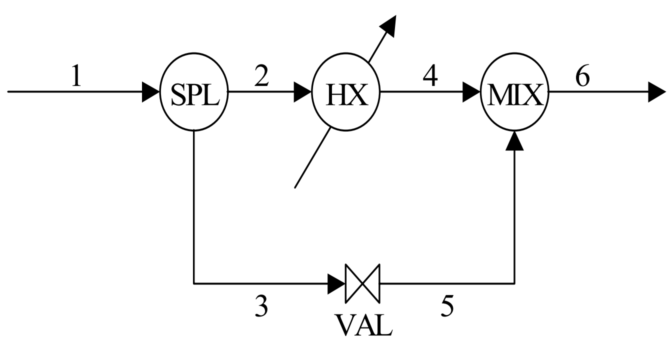

Figure 1-2 Heat Exchanger system with bypass.

The flow balances around the splitter, exchanger, valve and mixer can be written as

$$x_1 - x_2 - x_3 = 0 \tag{1-7a}$$
$$x_2 - x_4 = 0 \tag{1-7b}$$
$$x_3 - x_5 = 0 \tag{1-7c}$$
$$x_4 + x_5 - x_6 = 0 \tag{1-7d}$$

The measured values (given in Table 1-1) do not satisfy the above equations, since they contain random errors. It is desired to derive estimates of the flows that satisfy the above flow balances. Intuitively, we can impose the condition that the differences between the measured and estimated flows, also referred to as *adjustments*, should be as small as possible. As a first choice, we can represent this objective as

$$\underset{x_i}{Min} \sum_{i=1}^{6} (y_i - x_i)^2 \tag{1-8}$$

The above function is the familiar least squares criterion used in regression. Since it is immaterial whether the adjustments are positive or negative, the square of the adjustment is minimized. Although other types of criteria may be used such as minimizing the sum of absolute adjustment, they do not have a statistical basis and also make the solution of the problem more difficult.

The least squares criterion is acceptable, if all measurements are equally accurate. The adjustment made to one measurement is given the same importance as any other. However, in practice, it is likely that some measurements are more accurate than others depending on the instrument being used and the process environment under which it operates. In order to account for this, we can use a weighted least squares objective as a more general criterion, given by

$$\underset{x_i}{\mathrm{Min}} \sum_{i=1}^{6} w_i (y_i - x_i)^2 \tag{1-9}$$

where the weights w_i are chosen to reflect the accuracy of the respective measurements. More accurate measurements are given larger weights in order to force their adjustments to be as small as possible. Generally it is assumed that the error variances for all the measurements are known and the weights are chosen to be the inverse of these variances.

The reconciliation problem is thus a constrained optimization problem with the objective function given by Eq. (1-9) and the constraints given by Eq. (1-7). The solution of this optimization problem can be obtained analytically for flow reconciliation. Table 1-1 shows the true, measured and reconciled flows for the process of Fig 1-2. The reconciled flows shown in column 4 of this table are obtained by assuming that all measurements are equally accurate (weights are all equal). It can be easily verified that while the measured values do not satisfy the flow balances, Eq. (1-7), the reconciled flows satisfy them.

Table 1.1: Flow reconciliation for a completely measured process

Stream Number	True Flow Values	Measured Flow Values	Reconciled Flow Values
1	100	101.91	100.22
2	64	64.45	64.50
3	36	34.65	35.72
4	64	64.20	64.50
5	36	36.44	35.72
6	100	98.88	100.22

Systems with unmeasured variables. In the previous example, we have assumed that all variables are measured. However, usually only a subset of the variables are measured. The presence of unmeasured variables not only complicates the problem solution, but also introduces new questions such as whether an unmeasured variable can be estimated, or whether a measured variable can be reconciled as illustrated by the following example.

Example 1-2. Flow Reconciliation Example Revisited

Let us consider the flow reconciliation problem of the simple process shown in Fig. 1-2. However, we will not assume that all the flows are measured as before. Instead, we will assume that only selective flows are measured and in each case discuss the issues and problems involved in partially measured systems.

Case 1- Flows of streams 1, 2, 5, and 6 are measured, while the other two stream flows are unmeasured.

The objective in this case is to not only reconcile the measured flows, but also estimate all the unmeasured flows as part of the reconciliation problem. As in Eq. (1-6) we relate the measured and true stream flows,

$$y_i = x_i + e_i \qquad i = 1, 2, 5, 6 \qquad (1\text{-}10)$$

The constraints are still given by Eq. (1-7). It should be noted that the constraints involve both measured and unmeasured flow variables. The objective function is the weighted sum of squares of adjustments made to measured variables, and is given by

$$\underset{x_1,x_2,x_5,x_6}{\text{Min}} \quad w_1(y_1 - x_1)^2 + w_2(y_2 - x_2)^2 + w_5(y_5 - x_5)^2 + w_6(y_6 - x_6)^2 \qquad \text{(1-11)}$$

Since the unmeasured variables are present only in the constraint set, the simplest strategy for solving the problem is to eliminate them from the constraints. This will not affect the objective function since it does not involve unmeasured variables. Variable x_3 can be eliminated by combining Eqs.(1-7a) and (1-7c), while variable x_4 can be eliminated by combining Eqs.(1-7b) and (1-7d). Thus, we obtain a reduced set of constraints which involves only measured variables.

$$x_1 - x_2 - x_5 = 0 \qquad \text{(1-12a)}$$
$$x_1 - x_6 = 0 \qquad \text{(1-12b)}$$

The reduced data reconciliation problem is now to minimize (1-11) subject to the constraints of Eq. (1-12). It can be observed that this reduced problem involving the variables x_1, x_2, x_5, and x_6 is similar to the completely measured case, and an analytical solution can be used to obtain the reconciled values of the measured variables. Using the same measured values for x_1, x_2, x_5, and x_6 as given in Table 1-1, and assuming all measurements to be equally accurate, the reconciled values which are obtained are shown in Table 1-2 in the column under *Case 1*- Once the reconciled values for the measured variables are obtained, the estimates of the unmeasured variables can be calculated using the original constraints. Thus the estimate of x_4 is equal to that of x_2, and the estimate of x_3 is equal to that of x_5. These values are also indicated in Table 1-2. By comparing with the results of Table 1-1, it can be observed that since there are fewer measured variables in this case, the estimates of all variables are less accurate than those derived for the completely measured system. The central idea that is gained from this case, is that the reconciliation problem can be split or decomposed into sub-problems, the first being a reduced reconciliation problem involving only measured variables, followed by an estimation or coaptation problem for calculating the estimates of unmeasured variables.

Table 2.2: Flow reconciliation of partially measured process

Stream	Reconciled Flow Values		
	Case 1 - Streams 3 and 4 unmeasured	Case 2 – Streams 2 and 3 unmeasured	Case 2 - Streams 2,3,4,5 unmeasured
1	100.49	101.91	100.39
2	64.25	64.45	-
3	36.24	37.46	-
4	64.25	64.45	-

5	36.24	37.46	-
6	100.49	101.91	100.39

Case 2. Only flows of streams 1 and 2 are measured.

In this case, only Eqs. (1-7a) and (1-7b) contain measured variables and are useful in the reconciliation problem. The objective function is set up as before to minimize the adjustment made to measured variables and is given by

$$\min_{x_1, x_2} \; w_1(y_1 - x_1)^2 + w_2(y_2 - x_2)^2 \tag{1-13}$$

As in Case 1, we try to eliminate the unmeasured variables from the constraints (1-7a) and (1-7b). Our attempt to produce an equation involving only measured variables by suitably combining the original constraints ends in failure. Thus, the reconciliation problem we obtain is to minimize (1-13) without any constraints. It is immediately obvious that the best estimates of x_1 and x_2 are given by their respective measured values which results in the least adjustment of zero for (1-13). The estimates of the unmeasured variables can now be calculated using the constraints. The estimate of x_6 is equal to x_1, estimate of x_4 is equal to x_2, and the estimates of x_3 and x_5 are both equal to the difference between x_1 and x_2. These values are all given in Table 1-2 under *Case 2*.

Two important observations can be made in this case. First, no adjustment is made to the two measured variables x_1, and x_2. This is due to the fact that there is no additional information in the form of constraints that relate only the measured variables which can be exploited for adjusting their measurements. Such measured variables are also known as *nonredundant* variables. Second, a unique estimate for every unmeasured variable is obtained using the constraints and estimates of measured variables. These unmeasured variables are also known as *observable*. A formal definition of the concepts of observability and redundancy is given in chapter 3. It is sufficient at present to note that while the partially measured process in Case 1 gives a redundant and observable system, Case 2 gives rise to a nonredundant, observable system.

Case 3. Only flows of streams 1 and 6 are measured.

The reduced reconciliation problem we obtain for this case is

$$\min_{x_1, x_6} \; w_1(y_1 - x_1)^2 + w_6(y_6 - x_6)^2 \tag{1-14}$$

such that:

$$x_1 - x_6 = 0 \tag{1-15}$$

Equation (1-17) is obtained by adding all the constraints (1-9a) through (1-9d). Assuming that the measurements of x_1 and x_6 are equally accurate, their reconciled values obtained are given in Table

1-2 under *Case 3*. We now attempt to calculate the estimates of the remaining four variables. However, we will not be successful in obtaining unique estimates for these variables. In other words there are many solutions, in fact an infinite number, which can satisfy the constraints. For example, one possible solution is to take the estimates of x_3 and x_5 to be both equal to that of x_1, and the estimates of x_2 and x_4 to be equal to zero. Alternatively, we can choose the estimates of x_2 and x_4 to be equal to that of x_1, while the estimates of x_3 and x_5 are chosen to be zero. Without additional information there is no way of determining which of these myriad possible solutions is more accurate. The variables x_2, x_3, x_4, and x_5 are denoted as *unobservable* in this case. An interesting feature of this case is that though there are some unmeasured variables which cannot be uniquely estimated, reconciliation of the variables x_1 and x_6 can still be performed utilizing the available measurements. Therefore, Case 3 is a redundant, unobservable system.

System Containing Gross Errors

In all the cases considered in Example 1, the measurements did not contain any systematic error or bias. In such cases, data reconciliation does reduce the error in measurements. We will now examine the case when one of the measurements contains a systematic bias or gross error and demonstrate the need to perform gross error detection along with data reconciliation.

<u>Example 1-2 Reconciliation of Flow Measurements Containing a Gross Error</u>

We reconsider the flow process shown in Figure 1-2 for which the true stream flows are as given in Table 1-1- We will assume that all flows are measured with measurements as given in Table 1-1, except that the measurement of stream 2 contains a positive bias of 4 units, so that its measured value is 68.45 instead of 64.45. As before, we reconcile these measurements and obtain estimates which are shown in Table 1-3. in column 2 when all the measurements are used. A comparison of these estimates with those listed in Table 1-1, clearly shows that the accuracy of the estimates has decreased due to the presence of the gross error. Furthermore, although only the flow measurement of stream 2 contains a gross error, the accuracy of all the flow estimates has decreased. This is known as a *smearing effect* and occurs due to reconciliation which exploits the spatial constraint relations between different variables. In order for data reconciliation to be effective, it is therefore necessary to identify those measurements containing gross errors and either eliminate them or make appropriate compensation. The last column of Table 1-3 shows the reconciled estimates obtained when the flow measurement of stream 2 is discarded and not used in the reconciliation process. Clearly, the accuracy of the reconciled estimates has improved considerably, even though by discarding the measurement, the redundancy has decreased.

Table 1-3: Flow reconciliation when stream 2 measurements contains a gross error

Stream	Reconciled Flow Values	
	All measurements used	Stream 2 measurement eliminated
1	100.89	100.23
2	65.83	64.53
3	35.05	35.71
4	65.83	64.53
5	35.05	35.71
6	100.89	100.23

Thus far we have not considered the important question of how to identify the measurement containing a gross error based only on the knowledge of the measured values and constraint relations between variables. There are several ways of tackling this problem and in this example we illustrate one approach. Given a set of measurements, we can initially reconcile them assuming that there are no gross errors in the data. In the flow process example considered here, the reconciled estimates obtained under this assumption have already been shown in the second column of Table 1-3. From these reconciled estimates we can compute the differences between the measured and reconciled values (measurement adjustments) for all measured variables, and these are shown in Table 1-4.

Table 1-4: Measurement adjustments for flow process

Stream	Measurement adjustments
1	1.02
2	2.62
3	-0.40
4	-1.63
5	1.39
6	-2.01

If the constraints are linear as in this example, the expected variance of the adjustments can be analytically derived which will be a function of the constraint matrix and the measurement error variances. For the flow process example considered here, it can be shown that the standard deviation of measurement adjustments for every variable is 0.8165. A simple statistical test can be applied to determine if the computed measurement adjustments fall within a confidence interval, say within a $\pm 2\sigma$ interval. In this example the 2σ interval (95% confidence interval) is [-1-6 1-6]. From Table 1-4, we can observe that the measurement adjustments for the flows of streams 2, 4, and

6 fall outside this interval and as a first cut the measurements of these streams can be suspected of containing a gross error. Among these the measurement adjustment of stream 2 has the largest magnitude and can be identified to contain a gross error. After discarding the measurement of stream 2, we can again reconcile the data and compute the measurement adjustments to examine if any more gross errors are present.

The procedure used above is a sequential procedure for gross error detection and makes use of the statistical test known as the *measurement test*. A variety of statistical tests and methods for identifying one or more gross errors have been developed which are described in Chapters 7 and 8. Although, in this example we have only considered a gross error in measurements, it is possible for a gross error to be present in the constraints due to an unaccounted leak or loss of material. Some of the methods described in chapters 7 and 8 can also be used to identify such gross errors. The example also clearly demonstrates that data reconciliation and gross error detection have to be applied together for obtaining accurate estimates.

Additional References

1. Reklaitis, G.V. *Introduction to Material and Energy Balances*, New York: John Wiley & Sons, 1983
2. Kuehn, D.R. and Davidson,H. "Computer Control. II. Mathematics of Control", *Chem. Eng. Progress,* Vol.57, 44-47 (1961)
3. Vaclavek, V. “Studies on System Engineering, I. On the Application of the Calculus of Observations in Calculations of Chemical Engineering Balances”, *Coll. Czech. Chem. Commun.*, Vol. 34, 3653, (1968).
4. Vaclavek, V and Loucka, M. “Selection of Measurements Necessary to Achieve Multicomponent Mass Balances in Chemical Plant”, *Chem. Eng. Sci.,* Vol. 31, 1199-1205 (1976).
5. Mah, R.S.H., Stanley, G.M. and Downing,D.W. "Reconciliation and Rectification of Process Flow and Inventory Data", *Ind. & EC Proc. Des. Dev.,* Vol.15, 175-183 (1976).
6. Crowe, C.M., Campos Y.A.G. and Hrymak, A. "Reconciliation of Process Flow Rates by Matrix Projection. I: Linear Case", *AIChE Journal.*, Vol.29, 881-888 (1983).
7. Swartz, C.L.E, "Data Reconciliation for Generalized Flowsheet Applications", American Chemical Society National Meeting, Dallas, TX (1989).
8. Sanchez, M. and Romagnoli, J. "Use of Orthogonal Transformations in Data Classification-Reconciliation.” Computers Chem. Engng. Vol 20, 483-493 (1996).
9. Knepper, J.C., and Gorman, J.W. “Statistical Analysis of Constrained Data Sets”, *AIChE J.*, Vol. 26, 260-264 (1980).
10. Britt, H.I., and Luecke, R.H. “The Estimation of Parameters in Nonlinear Implicit Models”, *Technometrics*, Vol. 15, 233-247 (1973).
11. Pai, C.C.D. and Fisher, G.D. “Application of Broyden’s Method to Reconciliation of Nonlinearly Constrained Data”, *AIChE J.*, Vol. 34, 873-876 (1988).
12. Toja, I.B., and Biegler, L.T. “Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems”, *Comput. Chem. Eng.*, Vol. 15, 679-690 (1991).

13. Ravikumar, V., Singh, S.R., Garg, M.O., and Narasimhan, S. "RAGE - A Software Tool for Data Reconciliation and Gross Error Detection", in Foundations of Computer Aided Process Operations, CACHE/Elsevier, Netherlands (1993).
14. Wiegel, R.I. "Advances in Mineral Processing Material Balances", *Canad. Metall. Q.*, Vol. 11, 413-424 (1972).
15. Hodouin, D. and Everell, M.D. "A Hierarchical Procedure for Adjustment and Material Balancing of Mineral Processes Data", *Int. J. Miner. Proc.*, Vol. 7, 91-116 (1980).
16. Simpson, D.E., Voller, V.R., and Everett, M.G. "An Efficient Algorithm for Mineral Processing Data Adjustment", *Int. J. Miner. Proc.* Vol. 31, 73-96 (1991).
17. Heraud, N., Maquin, D., and Ragot, J. "Multilinear Balance Equilibration: Application to a Complex Metallurgical Process", *Min. Metall. Proc.*, Vol. 11, 197-204 (1991).
18. Reid K.J, Smith, K.A., Voller, V.R., and Cross, M. "A Survey of Material Balance Computer Packages in the Mineral Industry", in *17th Applications of Computers and Operations Research in the Mineral Industry* (Eds. T.B. Johnson, and R.J. Barnes), AIME, New York (1982).
19. Stanley, G.M., and Mah, R.S.H. "Estimation of Flows and Temperatures in Process Networks", *AIChE J.,* Vol.23, 642-650 (1977).
20. Almasy, G.A. "Principles of Dynamic Balancing", *AIChE J.*, Vol. 36, 1321-1330 (1991).
21. Liebman, M.J., Edgar, T.F. and Lasdon L.S., "Efficient Data Reconciliation and Estimation for Dynamic Processes Using Nonlinear Programming Techniques", *Comput. Chem. Eng.,* Vol.16, 963-986 (1992).
22. Ripps, D.L. "Adjustment of Experimental Data", *Chem. Eng. Progr. Symp. Ser. No. 55*, Vol 61, 8-13 (1965).
23. Reilly, P.M., and Carpani, R.E. "Application of Statistical Theory to Adjustment of Material Balances", 13th Can. Chem. Eng. Conf., Montreal, Quebec (1963).
24. Almasy, G.A, and Sztano, T. "Checking and Correction of Measurements on the Basis of Linear System Model", *Prob. Control Inform. Theory*, Vol. 4, 57-69 (1975).
25. Narasimhan, S. and Mah, R.S.H. "Generalized Likelihood Ratio Method for Gross Error Identification", *AIChE J.*, Vol. 33, 1514-1521 (1987).
26. Tamhane. A.C., Iordache, C. and Mah R.S.H. "A Bayesian Approach to Gross Error Detection in Chemical Process Data. Part I: Model Development", *Chemometrics and Intel. Lab. Sys.*, Vol.4, 33 (1988).
27. Tong, H. and Crowe C. M. "Detection of Gross Errors in Data Reconciliation by Principal Component Analysis" *AIChE Journal*, Vol. 41, 1712-1722 (1995).
28. Romagnoli, J.A., and Stephanopoulos, G. "Rectification of Process Measurement Data in the Presence of Gross Errors", *Chem. Eng. Sci.*, Vol. 36, 1849-1863 (1981).
29. Serth, R.W., and Heenan, W.A. "Gross Error Detecting and Data Reconciliation in Steam-Metering Systems", *AIChE J.*, Vol. 30, 243-247 (1986).
30. Rosenberg, J., Mah, R.S.H., and Iordache, C. "Evaluation of Schemes for Detecting and Identification of Gross Errors in Process Data", *I&EC Proc. Des. Dev.*, Vol. 26, 555-564 (1987).
31. Iordache, C., Mah, R.S.H., and Tamhane, A.C. "Performance Studies of the Measurement Test for Detecting of Gross Error in Process Data", *AIChE J.*, Vol. 31, 1187-1201 (1985).
32. Rollins, D.K., and Davis, J.F. "Unbiased Estimation Technique for Identification of Gross Errors", *AIChE J.*, Vol. 38, 563-571 (1992).
33. Murthy, A.K.S. "Material Balance around a Chemical Reactor, II", *Ind. Eng. Chem. Process. Des. Dev.*, Vol. 13, 347 (1974).

34. Madron, F., Veverka, V., and Vanacek, V. "Statistical Analysis of Material Balance of a Chemical Reactor", *AIChE J.*, Vol. 23, 482-486 (1977).
35. Wang and Stephanopoulos, G., "Application of Macroscopic Balances to the Identification of Gross Measurement Errors", *Biotechnol. Bioeng.*, Vol. 25, 2177-2208 (1983).
36. Crowe, C.M. "Reconciliation of Process Flow Rates by Matrix Projection, II. The Nonlinear Case", *AIChE J.*, Vol. 32, 616-623 (1986).
37. Sheel, J.P., and Crowe, C.M. *Can. J. Chem. Eng.*, Vol. 47, 183 (1969).
38. MacDonald, R.J, and Howat, C.S. "Data Reconciliation and Parameter Estimation in Plant Performance Analysis", *AIChE J.*, Vol. 34, 1-8 (1988).
39. Hlavacek, V. "Analysis of a Complex Plant - Steady State and Transient Behavior I - Plant Data Estimation and Adjustment", *Comput. Chem. Eng.*, Vol. 1, 75-81 (1977).
40. Mah, R.S.H., "Design and Analysis of Performance Monitoring Systems", in Chemical Process Control II (Eds. D.E. Seborg, and T.F. Edgar), Engineering Foundation, New York (1982).
41. Tamhane, A.C. and Mah R.S.H. "Data Reconciliation and Gross Error Detection in Chemical Process Networks", *Technometrics,* Vol. 27, 409-422 (1985).
42. Mah, R.S.H., "Data Screening", in Foundations of Computer-Aided Process Operations (Eds. G.V. Reklaitis and H.D. Spriggs), CACHE/Elsevier, Amsterdam, Netherlands (1987).
43. Crowe, C.M., "Data Reconciliation - Progress and Challenges", *J. Proc. Cont.*, Vol. 6, 89-98 (1996).
44. Mah, R.S.H. *Chemical Process Structures and Information Flows*. Boston: Butterworths, 1990.
45. Bodington, C.E. *Planning, Scheduling and Control Integration in Process Industries,* New York: McGraw-Hill, 1995.
46. Madron, F. *Process Plant Performance: measurement and data processing for optimization and retrofits*. Chichester, West Sussex, England: Ellis Horwood Limited Co., 1992
47. Veverka, V.V. and Madron, F. *Material and Energy Balancing in Process Industries: From Microscopic Balances to Large Plants*, Elsevier, Amsterdam, The Netherlands, 1997.