Chapter 8

ON-LINE OPTIMIZATION

Introduction

On-line optimization adjusts the operation of a plant based on product scheduling and production control to maximize the plant's profit. It provides the means for continuously driving a process toward its optimum operating point. In most industrial processes, the optimal operating point constantly moves in response to changing market demands for products, fluctuating costs of raw materials, products and utilities, variations in feed quality and availability and changing equipment efficiencies. The time frame over which these changes can occur ranges from minutes to months. Competitive economic environments require timely response to these changing factors. This means that the optimization must be done on-line to have the plant operate continually under the best conditions. Key benefits for on-line optimization have been a 10% improvement in plant profits, a reduction in energy use and waste generation and an increased understanding of plant operations. The terminology used in on-line optimization is given in a section at the end of the chapter.

The structure of on-line optimization is shown in Figure 8-1 along with the components that work together to maximize the profit from the operation of the plant. The key components of on-line optimization include the plant and economic models, gross error detection, data reconciliation and parameter estimation. Also, an efficient optimization algorithm is used to solve the three nonlinear optimization problems shown in Figure 8-1. Referring to Figure 8-1, plant data is sampled from the distributed control system, and gross errors are removed from the data. Then the data is reconciled to be consistent with material and energy balances of the process. This data is then used to update the parameters in the plant model to ensure the plant model predicts the operation of the plant. The updated plant model is used with an updated profit function (economic model) to generate optimal set points for the distributed control system for the best operating conditions for the plant. These optimal set points are sent to the plant distributed control system as new set points for the controllers. A coordinator program uses algorithms to identify when the plant is at steady state and sends steady-state data to the on-line optimization loop. Also, the controller program accepts the optimal set points and only uses them for new set points for the controllers only if the plant is at steady state.

The rapid development in computer hardware and software as well automation technology in the last ten years has made it possible to consider on-line optimization of chemical plants. On-line optimization improves the economic and environmental performance of chemical plants and refinery processes without requiring substantial capital investment, and it is a growth area for modeling technology and advanced control companies.

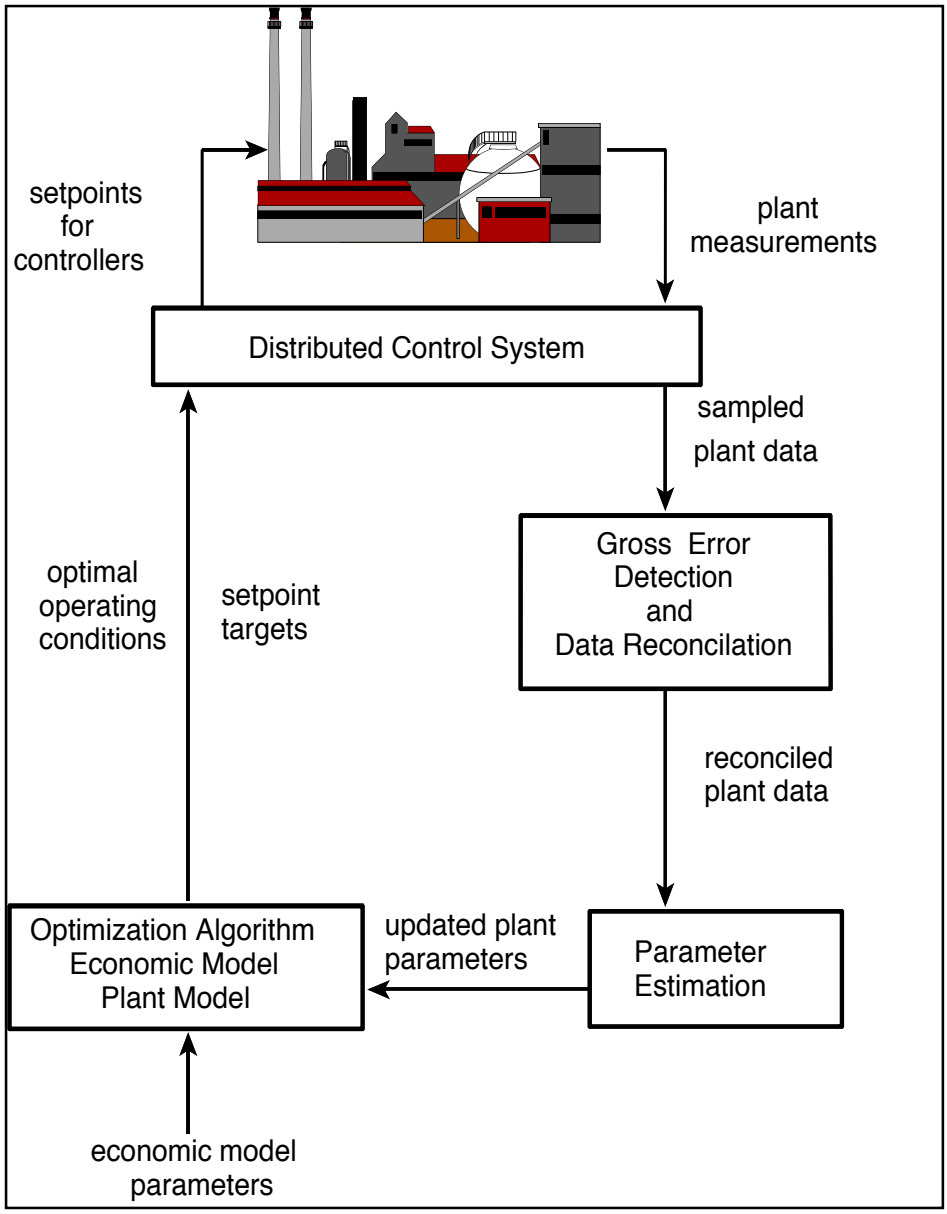


Figure 8-1 Structure of On-Line Optimization

With the availability of distributed control systems (DCS) for process control and data acquisition as well as the application of multivariable controllers, large-scale application of on-line optimization became feasible. DCS provides current plant operating data (plant measurements) for updating the parameters in plant models to avoid the plant-model mismatch. Multivariable controllers ensure the control ability to quickly and accurately response to new optimal set points. Moreover, the decline in cost of computer hardware and software and the increase in the cost of energy and pollution prevention have stimulated manufacturers to improve and optimize their processes, which have boosted the development of on-line optimization.

A chemical plant or refinery's distributed control system will run the control algorithm three times a second. A tag for a control point in the distributed control system contains about 20 values, e.g., set point, limits, alarm setting, temperature, pressure, mass flow rate, etc. A refinery and large chemical plant will have more than 10,00 tags. The control system's data historian stores instantaneous values for each tag every five seconds or as specified. The historian includes a relational data base for laboratory and other measurements not from the control system. Values in the historian are stored for one year usually and require a very large amount of storage. Data in the historian is available over a LAN in various forms, e.g. averages, Excel files.

Operating a newly installed on-line optimizer on a chemical plant's distributed control system requires steps in operator training according to Dow Chemical Company's seminars. First, operators are presented the result from the optimizer to evaluate the proposed control set-point moves being acceptable and will not cause a plant upset. Operators tend to be concerned that the optimizer has the potential of replacing them, and some of the recommended set-points are counter-intuitive based on their experience. The second step is to have operators change set-points specified by the optimizer, if they are acceptable to the operators. Having gained operator acceptance at this step, the third step has the optimizer change the control system set-points with operators monitoring these changes. Seeing that the optimizer can operate with supervision, the fourth step is continuing without close operator monitoring. In case of a plant upset the optimizer is turned off with procedures established to turn the optimizes back on after the plant returns to steady-state. In case of equipment or process changes that are new and not included in the optimizer's process simulation, the process model must be revised to reflect these process changes to accurately represent plant performance required by the optimizer.

Industrial Applications of On-Line Optimization

On-line optimization or real-time optimization, these names are used interchangeably, is a mature technology that is routinely used with digital control systems in refineries and chemical plants. This methodology went from a research topic to an established industrial application in the past 20 years. On-line optimization programs are available from engineering design and control system companies such as Aspen Technology, Honeywell, and others as listed in Figure 8-2. These companies can provide both the plant's distributed control system and the online optimizer or provide the on-line optimizer for the existing control system. Just providing the on-line optimizer is a multi-million-dollar project and exceed budgets available for on-line optimization in small to moderately sized plants.

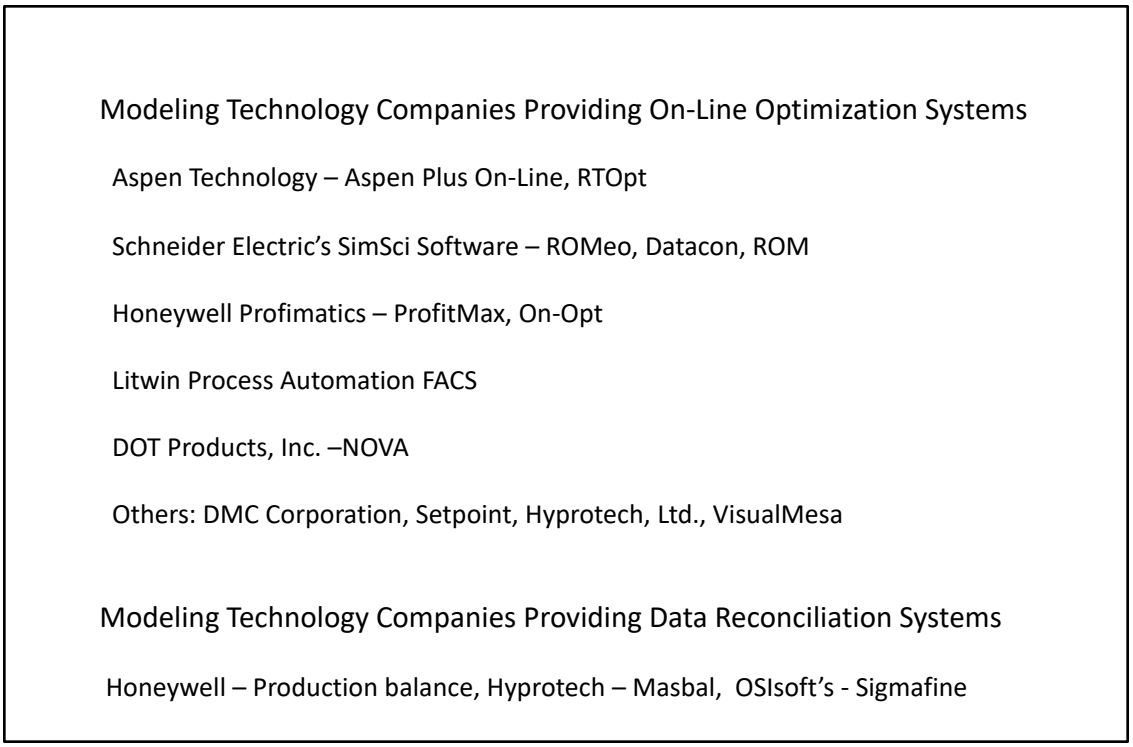


Figure 8-2 Modeling Technology Companies

Some companies shown in Figure 8-2 provide off-line data reconciliation and gross error detection programs without on-line optimization capability that use data from the distributed control system for economic evaluations and to close material and energy balances on a whole refinery and chemical plant. Also, these tools are used to reconcile flow, temperature and composition measurements to satisfy material and energy balances around each unit in a process plant, as well as to estimate parameters for the units.

Companies in the U. S. and Europe that gave reported applying on-line optimization are shown in Figure 8-3. Industrial applications have been mainly for crude units and ethylene plants where small improvements in capacity can mean very large increases in profits. Reported improvements in plant operations and economics in a range of 3% to 20%. However, details of methodology used are sketchy because proprietary processes are being used as described below.

Lauks, et al., (1992) reviewed the industrial applications of on-line optimization reported in the literature from 1983 to 1991 and cited nine applications for five ethylene plants, a refinery, a gas plant, a crude unit and a power station. These results showed a profitability increase of 3% or $4M/year. Also, intangible profits from a better understanding of the plant behavior were significant. In addition, they gave results for the OMV Deutschland GmbH complex including a refinery unit, an ethylene plant and downstream treating units in Burghausen, Germany. An equation oriented flowsheeting program was used for the process model having more than 5,000 linear and nonlinear equations that led to an optimization problem with 106 constraints and 37

Some Companies Reported Using On Line Optimization

United States	Europe
Dow	Dow Benelux
Cheveron	Shell
Amoco	OMV Deutchland
Conoco	Penex
Texaco	DSM Hydrocarbons
Sunoco	OEMV
Lyondel	Borealis
Phillips	
Marathon	
British Petroleum	
NOVA Chemicals (Canada)	

Applications: Mainly crude units in refineries and ethylene plants

Figure 8-3 Some Companies Reporting Applying On-Line Optimization

decision variables. Data reconciliation involved 450 points, and there were about 300 tuning parameters. The program was run on a DG-AVIION 4200 Unix system with a total computation time of 60 minutes. Optimization results were summarized in a setpoint report and manually implemented by plant operators on a TDC 2000 system. The improvement in profitability has been between 1-3% depending on price structure, and it has provided better insight to operation of the plant.

Scott, et al., (1995 and 1994) reported that Texaco Refining and Marketing Inc. (TRMI) has implemented ROM from Simulation Sciences Inc. on a four-unit complex. This on-line optimization package provides integrated modeling of reaction units, optimization across multiple units, validation of laboratory and plant data, higher quality control, and a large amount of operating information. It was expected that the benefits from this project would exceed $1.0 million annually. Also, this can be used as a versatile tool for troubleshooting, planning, and training of the processes.

Krist, et al., (1994) described the development and implementation of a generic system for on-line optimization (SOLO) in a benzene plant of Dow Benelux N.V. SOLO contains generic modules and plant specific modules. The generic modules are used for data-retrieved, data analysis, data reconciliation and decision mechanism; and the plant specific modules are used for

parameter estimation and final optimization. This optimization increased the plant's margin by an average of 4%.

Fatora, et al., (1992) reported that the use of closed-loop real-time optimization and dynamic matrix control technology has achieved significant economic benefits in an olefin plant. The payback period for the total project was less than one year. In addition, benefits of this on-line optimization system were that it pushed the unit to the most profitable constraints based on current economics and operating objectives. This increased the plant capacity, reduced energy requirement, and improved product qualities.

Van Wijk and Pope (1992) described on-line optimization of the catalytic cracking complex at Shell's Stanlow refinery in the UK. The on-line optimization system received process and economic data from the refinery supervisory control system and performed optimizations on a three-hour cycle providing targets to the process controllers. The process and economic models were nonlinear, and a reduced gradient algorithm was used for the optimization. Data reconciliation was performed on several hundred points, and rotating equipment efficiencies and heat transfer coefficients were two of the parameters updated in the process model. Benefits of on-line optimization were a 10% increase in feed rate, a 9% increase in catalyst circulation rate that resulted in a 9% increase in gasoline production.

OEMV, an Austrian company, had successfully installed an on-line control and optimization system in the fluid catalytic cracking units (FCCU) in 1987 (Rhemann, et al., 1989). The advanced control and optimization project schedule were included in an overall project providing a new digital instrument control system (DCS) for FCCU, gas plant and treating units, consolidated in one common control area. The new DCS was installed and commissioned without a plant shutdown during normal plant operations. The improved control from advanced control and on-line optimization translated into a large reduction in the standard deviation of control variables. The advanced control and on-line optimization gave a 4.3% increase in the maximum operating feed rate for the FCCU.

Sourander, et al., (1984) described the on-line optimization of an ethylene plant using refinery heavy feedstocks. The plant produced 200,000 tpa of ethylene using nine cracking furnaces which had a computer control system with set point supervisory controls of analog controllers. Gas chromatographs using dedicated microcomputers sampled feed and product streams, and analyses were sent to the main process computer. Seven different feedstocks and three different recycle streams were sent to the nine heaters at varying rates to meet production demand for seven products. The economic model was based on gross margin, and linear programming was used to maximize gross margin subject to market demand, feed availability and the plant constraints (material and energy balances and process unit capacities). The on-line optimization cycle was executed every four hours. Error detection was very important, especially for the heater effluent, and a bad analysis not detected and included in the model updating caused errors to be carried through to the control system. The results of using on-line optimization were reported to be increased furnace run times of 30%, efficiencies of 3%, capacities of 4% and increased ethylene yields of 2%.

Saha, et al., (1990) of Amoco Production Company reported results for the on-line optimization of a 240 MMscfd gas-processing plant in Evanston, Wyoming using the ChemShare ProCAM system which has data reconciliation and a proprietary process modeling system using a simultaneous solution technique. More than 550 data points were taken from the plant's distributed control system (DCS) and reconciled for optimization using a plant model with 170 pieces of equipment and detailed economic model. The optimization analysis determined the best operating conditions for 40 process variables that were reported to the plant operator for implementing via the DCS. Preliminary estimates were approximately $9,000 per day for an increased pretax profit and 50% higher than this for a high ethane recovery mode.

Moore and Corripio (1991) reported on the on-line optimization of distillation columns in series that used dynamic programming with steepest descent and a simple model for product recovery for two and three distillation columns in series. Applied to a two and three column train at Dow Chemical Company's Louisiana Division, the control system performed successfully to reduce operating costs beyond what was anticipated.

Bailey, et al., (1993) reported on the on-line optimization of a hydrocracker fractionation plant using MINOS as optimizer. The full plant model contains 2891 variables with 10 degree of freedom. Detailed methodologies including modeling and numerical techniques were outlined. They showed that the important factors for implementing the model-based optimizer were scaling, starting points, sparsity patterns and thermodynamic approximations. The on-line optimization system gave an 3% increase in profit.

Gott, Roubidoux and Heersink (1991) described an on-line optimization system for the Conoco's Billings refinery fluid catalytic cracking (FCC) units using Profimatics Inc. FCC-SIMOPT package. The on-line optimizer generates both optimal control targets as well as the optimal operating strategy for the advanced FCC constraint control. The on-line optimization was divided into five phases: 1) process data monitoring, 2) program scheduling, 3) data reconciliation, 4) model update, 5) optimization. The results are sent to the advanced control system. They concluded that this system increased the profit and provided better insight into the operation of the FCC units.

Ozyurt, et al., 2003, have described the optimal application of on-line optimization to the alkylation process in refinery in the lower Mississippi River corridor along with pinch analysis and pollution assessment using an advances process analysis system. A significant increase in profit and energy savings were projected through reduced steam use and a small decrease in sulfuric acid catalyst consumption.

In summary, on-line optimization significantly improved profitability, plant operation, and emission reduction; and it provided better understanding of processes. Typically, profitability was increased by 5 to 10% with comparable improvements in plant operations. Also, it was reported that a more thorough understanding of the plant performance was very valuable but is difficult to quantify economically.

Methodology

On-line optimization for chemical processes includes three important steps: combined gross error detection and data reconciliation, simultaneous data reconciliation and parameter estimation, and plant economic optimization. In combined gross error detection and data reconciliation, a set of accurate plant measurements are generated from data extracted from the plants distributed control system (DCS). This set of data is used for estimating the parameters in plant model, and parameter estimation is necessary to have the plant model match the current performance of the plant. Then, economic optimization is conducted to optimize the economic model using this current plant model as constraints. This optimization provides set points for the DCS to move the plant to the new optimal operating conditions.

Each optimization problem in on-line optimization has a similar mathematical statement as following:

Optimize: **Objective function**
Subject to: **Constraints from plant model**

where the objective function is obtained from a joint distribution function for data reconciliation and parameter estimation and a profit function (economic model) for plant economic optimization. The constraint equations describe the relationship among variables and parameters in the process, and they are material and energy balances, rate equations, and thermodynamic equilibrium relations.

On-line optimization takes advantage of the fact that chemical plants operate at steady state with transient periods that are short compared to steady state operations. Consequently, steady-state process models are used to describe the plant. These plant models are complicated and highly nonlinear. The general mathematical statement for on-line optimization is:

$$\begin{aligned} &\textit{Optimize}: && P(\mathbf{y}, \mathbf{x}) \\ &\textit{Subject to}: && \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ & && \mathbf{g}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \leq 0 \\ & && \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U,\ \mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U \end{aligned} \qquad 8\text{-}1$$

where the objective function P is subject to a process model that includes the equality constraints **f**, inequality constraints **g**, and bounds on the variables. In Equation 8-1, the vector **y** represents a set of measurements sampled from distributed control system for measured variables, and vector **x** denotes the true values (that satisfy material and energy balances) of the measured variables **y**. The vector **z** represents a set of unmeasured process variables that include all process variables except the measured ones in plant model, and $\boldsymbol{\theta}$ is the vector of process parameters. The equality constraints **f** represents material and energy balances, rate equations and equilibrium relations. The inequality constraints **g** represents the demand for products, the availability of raw materials, the limitation on the capacity of equipment, the allowable operating conditions, and the restrictions on waste and pollutant emission. In addition, $\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$ and $\mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U$ give upper and lower bounds on process variables. The following equation gives the relation between **y** and **x:**

$$\mathbf{y} = \mathbf{x} + \mathbf{e} \qquad 8\text{-}2$$

where the vectors **e** represents the measurement errors that could be random or gross errors.

Key Elements of On-Line Optimization

The objective of on-line optimization is to determine optimal process set points based on plant's current operating and economic conditions. As shown in Figure 8-1, the key elements of on-line optimization are:

- Gross Error Detection
- Data Reconciliation
- Parameter Estimation
- Economic Model (Profit Function)
- Plant Model (Process Simulation)
- Optimization Algorithm

The relationship between these key elements is outlined in Figure 8-1 where both plant model and optimization algorithm are required in the three steps of on-line optimization. On-line optimization involves solving three nonlinear optimization problems: economic optimization, parameter estimation, and data reconciliation. The plant model serves as the constraint equations in these three nonlinear optimization problems, and the optimization algorithm is used to solve the nonlinear optimization problems. For economic optimization, the plant model is used with the economic model to maximize the plant profit and to provide the optimal set points for the distributed control system to operate the plant. For parameter estimation, parameters in the plant model are estimated by optimizing an objective function, such as minimizing the sum of squares of measurement errors, subject to the constraints in the plant model. For data reconciliation, the errors in plant measurements are rectified by optimizing a function based on the joint probability distribution function for the plant measurements subject to plant model, and a test statistic is used to detect gross errors in the measurements.

Data Reconciliation

Data reconciliation has been the subject of several texts, (Madron, 1992, Narasimhan and Jordache, 2000 and Romagnoli and Sanchez, 2000, Veverka and Madron, 1997) which also describe gross error detection. Ozyurt and Pike, 2004, have described the theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. Results of research on data reconciliation and gross error detection were reviewed and evaluated in detail through 1988 by Mah (1990) for steady state processes. Evaluations of methods and research results have been reported through 2015 by Zhang and Chen, 2015.

Generally, raw process data is subject to two types of errors, random and gross errors. Random errors come from the randomness in measurements and are commonly assumed to be independently and normally distributed with zero mean. Gross errors are caused by non-random event such as process leaks, biases in instrument measuring or malfunction of instrument measuring, etc., and they are not considered to be random events.

Data reconciliation is a procedure to adjust or reconcile process data and to obtain more accurate values for the measurements by requiring the reconciled data to be consistent with material and energy balances. The data reconciliation problem can be formulated as a constrained optimization problem, e.g., least squares estimation problem if the measurements contain only random errors. A set of reconciled data $\mathbf{x} = \mathbf{y} + \mathbf{a}$ is determined where **a** is called the vector of measurement adjustments.

Using Equation 8-2, the vector of measurement errors **e** is defined as:

$$\mathbf{e} = \mathbf{y} - \mathbf{x} \tag{8-3}$$

where vector **y** represents the measured process variables (sampled values) and vector **x** denotes the true values of the measured variables (satisfy material and energy balances).

If measurement *x* is subject to only random errors with known normal distributions, the Gaussian or normal probability distribution function for the individual measurement error is:

$$P(x:\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2\right] \tag{8-4}$$

where σ is the standard deviation of a measurement error, e, and μ is the mean. The colon after *x* indicates that μ and σ are parameters that determine the shape of the normal probability distribution function. In general, probability is the relative frequency of an event.

The distribution function given by Equation 8-4 is for one variable, *x*. For n variables that has data extracted from the distributed control system, each has a probability distribution function associated with it. If the measurement error in each of the n variables are independent, then the joint probability distribution for all measurement errors (or likelihood function) is the product of distributions functions for individual measurement errors given by Equation 5, i.e.,

$$L = \prod_{i=1}^{n}\left[\frac{1}{(2\pi\tau_i^2)^{1/2}}e^{-(x_i-\mu_i)^2/2\tau_i^2}\right] \tag{8-5}$$

The above equation can be written as:

$$L = \frac{1}{(2\pi)^{n/2}\prod_{i=1}^{n}\sigma_i}e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\tau_i}\right)^2} \tag{8-6}$$

Taking the ln of both sides of Equation 6 gives:

$$\ln L = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 - \ln\left[\left(2\pi\right)^{n/2}\prod_{i=1}^{n}\sigma_i\right] \qquad 8\text{-}7$$

The mathematical argument is that maximizing the likelihood function, Equation 8-5, will maximize the likelihood of the probability of the estimation, Barlow, 1989. For data reconciliation, maximizing the likelihood function will give the best reconciliation of the measurements. Then to maximize the likelihood function, L, Equation 8-5, the negative of Equation 8-5 can be minimized. Equation 8-7 can be put in the following form where the standard deviation, σ_i, is a known parameter, and this gives the function to be minimized as the least squares function given below.

$$\textit{Minimize}: \quad \sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \qquad 8\text{-}8$$

Examining Equation 8-8 as applied to data reconciliation, the mean, μ_i is considered to be the true value of y_i as determined by the material and energy balance equations and is designated x_i. The measurements x_i are designated as y_i, the values measured from the plant's distributed control system. Also, σ_i is the standard deviation of measurement error and is considered a known parameter determined from the standard deviation of data taken previously from the distributed control system.

Using the form of Equation 8-1, the data reconciliation optimization problem is:

$$\begin{aligned} &\textit{Minimize}: && \sum_{i=1}^{n}[(y_i - x_i)/\sigma_i]^2 \\ &\mathbf{x} \\ &\textit{Subject to:} && \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ & && \mathbf{g}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \le 0 \\ & && \mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U,\ \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U \end{aligned} \qquad 8\text{-}9$$

Defining **Q** as the known variance matrix of measurement errors **e**, $\mathbf{Q} = \text{diag}\{\sigma^2_{ij}\}$ (Ozyurt and Pike, 2004), Equation 9 can be written as:

$$\begin{aligned} &\textit{Minimize}: && \mathbf{e}^T\mathbf{Q}^{-1}\mathbf{e} = (\mathbf{y} - \mathbf{x})^T\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{x}) = \sum_{i=1}^{n}[(y_i - x_i)/\sigma_i]^2 \\ &\mathbf{x} \\ &\textit{Subject to}: && \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ & && \mathbf{g}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \le 0 \\ & && \mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U,\ \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U \end{aligned} \qquad 8\text{-}10$$

Equation 8-10 is the nonlinear optimization problem for data reconciliation. Solving Equation 8-10 gives the reconciled values of process variables **x,** and the measurement adjustments **a** = **x** - **y** can be computed from these values.

If the constraints are linear, i.e., only material balances, these can be written in matrix form as:

$$\mathbf{Ax} = 0 \qquad 8\text{-}11$$

then, the optimization problem of Equation 8-10 has an analytical solution using Lagrange multipliers (Mah and Tamhane, 1982 and Knopf, 2012), which is:

$$\mathbf{x} = \mathbf{y} - \mathbf{QA^T}(\mathbf{AQA}^T)^{-1}\mathbf{Ay} \qquad 8\text{-}12$$

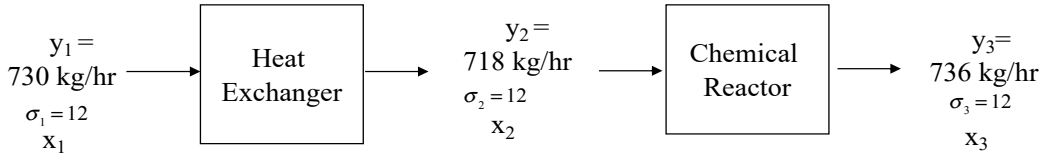
and the vector of measurement adjustments, **a**, is:

$$\mathbf{a} = \mathbf{x} - \mathbf{y} = -\mathbf{QA^T}(\mathbf{AQA}^T)^{-1}\mathbf{Ay} \qquad 8\text{-}13$$

In this case, Equation 8-12, all of the process variables are measured. More details about measured, unmeasured variables, parameters and constraints is given in a following section.

The following simple example illustrates the use of Equation 8-12:

Example 1 Application of Data Reconciliation to a Simple Process



Steady State Material Balances $x_1 = x_2$ $x_1 - x_2 = 0$ $x_2 = x_3$ $x_2 - x_3 = 0$

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$y = \begin{bmatrix} 730 \\ 718 \\ 736 \end{bmatrix} \qquad Q = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 12 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

$$x = y - QA^T(AQA^T)^{-1}Ay$$

Substituting in Equation 12 gives: $\mathbf{x} = [\ 728\ 728\ 728\ ]^T$

Mah (1990) gives extensions to this analytical solution for unmeasured and precisely measured variables, in addition to measured variables. The material balance equation, Equation 11, for measured, **x**, unmeasured, **u**, and precisely measured, **c',** variables is written as:

$$\mathbf{A_1 x} + \mathbf{A_2 u} + \mathbf{A_3 c'} = 0 \qquad 8\text{-}14$$

Equation 8-14 is multiplied by matrix **P** to have $\mathbf{PA_2 = 0}$ and Equation 14 can be written as $\mathbf{PA_1 x} = \mathbf{-PA_3 c'}$. Then defining $\mathbf{B\,x = C}$, where $\mathbf{B = PA_1}$ and $\mathbf{C = -PA_3 c'}$, the optimization problem becomes:

$$\textit{Minimize}: (\mathbf{y} - \mathbf{D\,x})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{D\,x}) \qquad 8\text{-}15$$

Subject to: $\mathbf{B\,x = C}$

The Lagrange multiplier solution is:

$$\mathbf{x = x_0 + (D^T Q^{-1} D)^{-1} B^T [B (D^T Q^{-1} D)^{-1} B^T]^{-1} (c - B x_0)} \qquad 8\text{-}16$$

$$\mathbf{x_0 = (D^T Q^{-1} D)^{-1} y} \qquad \mathbf{u = A_2^{-1} (C - A_1 x)}$$

There are extensions to include component material balances (species continuity equations with chemical reaction), energy flow treated as additional components, stoichiometric constraints and elemental balances (Mah, 1990). In component material balances, there are products of composition and total flow rate in the constraint equations, and these balance equations are bilinear. In the energy equation, species enthalpies are usually expressed as a nonlinear function of the measured variables (temperature and species mass flow rate). Hence, the energy balance equations are nonlinear. When constraints are nonlinear, the optimization problem must be solved by nonlinear programming techniques.

Example 8-2 Data Reconciliation with Nonlinear Constraint Equations

The process flow diagram shows a heat exchanger used to heat the feed to an isothermal continuously stirred chemical reactor (CSTR)

Material Balances	Heat Exchanger	Stirred Reactor
	$F_1 = F_2$	$F_2 = F_5$
	$F_3 = F_4$	$C_{AO} F_2 - C_A F_5 + k C_A V = 0$

Energy Balances for Heat Exchanger

$q_c = F_2 C_p (T_2 - T_0) - F_1 C_p (T_1 - T_0)$

$q_h = F_4 C_p (T_4 - T_0) - F_3 C_p (T_3 - T_0)$

$q_c - q_h = 0$

$q_c = U\,A\,[(T_3 - T_2) - (T_4 - T_1)]/\ln [(T_3 - T_2)/(T_4 - T_1)]$

Cost of Operation (cost of feed, heating feed and mixing): $C = C_f F_1 + C_h F_3 + C_m V$
where C_f – unit cost of feed, C_h – unit cost of heating feed, C_m - mixing cost per volume of reactor.

Measured variables: F_1, F_3, F_5, T_3, C_A Parameters: U, k, C_f, C_h, C_m Constants: C_p, V, C_{A0}, T_0

The data reconciliation optimization problem is:

Minimize: $\{[(728 - F_1)/8]^2 + [(878 - F_3)/7]^2 + [(736 - F_5)/9]^2 +$
$[(157 - T_1)/14]^2 + [(157 - T_3)/18]^2 + [(0.032 - C_A)/0.0020]^2\}$

Subject to: $F_1 - F_2 = 0$
$F_3 - F_4 = 0$
$F_2 - F_5 = 0$
$C_{AO} F_2 - C_A F_5 + kC_A V = 0$
$q_c - F_2 C_p (T_2 - T_0) + F_1 C_p (T_1 - T_0) = 0$
$q_h - F_4 C_p (T_4 - T_0) + F_3 C_p (T_3 - T_0) = 0$
$q_c - q_h = 0$
$q_c - U A [(T_3 - T_2) - (T_4 - T_1)]/\ln [(T_3 - T_2)/ (T_4 - T_1)] = 0$

An Excel Solver solution of the optimization problem gave the following results for the reconciled variables.

Variables	Measured Values	Standard Deviations	Reconciled Values	Standard Error
F_1	728 kg/hr	8	732 kg/hr	$\|728 - 732\|/8 = 1.85$
F_3	878 kg/hr	7	867 kg/hr	$\|878 - 867\|/7 = 1.57$
F_5	736 kg/hr	9	732 kg/hr	$\|736 - 732\|/9 = 0.44$
T_1	157°C	14	168°C	$\|157 - 168\|/14 = 0.79$
T_3	204°C	18	221°C	$\|204 - 221\|/18 = 0.94$
C_A	0.032 mol/m3	0.002	0.026 mol/m3	$\|0.032 - 0.026\|/0.0020 = 3.0$ gross error?

Note: The standard error, $|y_i - x_i|/\sigma_i$ was computed for the measured variables as shown in the last column. The standard error of 3.0 for C_A indicates that the measured value of C_A may be a gross error. Details about the standard error and its significance as a test statistic is given in the following section on the measurement test.

Number of variables: $F_1, F_2, F_3, F_4, F_5, T_1, T_2, T_3, T_4, q_c, q_h, C_A$: 12 variables
Number of constraint equations: 8, Degrees of freedom: 12 -8 = 4

Minimum number of measured variables for redundancy: 5, one more than the degrees of freedom.

Note: Specifying the four temperatures is not sufficient for a solution to the constraint equations.

In summary, the constrained least squares method was widely used to reconcile process data by assuming that the measurement errors are normally distributed and there are no gross errors. Data reconciliation is a nonlinear optimization problem that can be solved nonlinear programming techniques. Data reconciliation usually is conducted with gross error detection and/or parameter estimation as described below.

Gross Error Detection

Gross errors in some of the data extracted from the plant's distributed control can have gross errors caused by instrument errors, such as bias, drifting, precision degradation, instrument failure and process leaks. An example of time series of data illustrating these errors are shown in Figure 8-4.

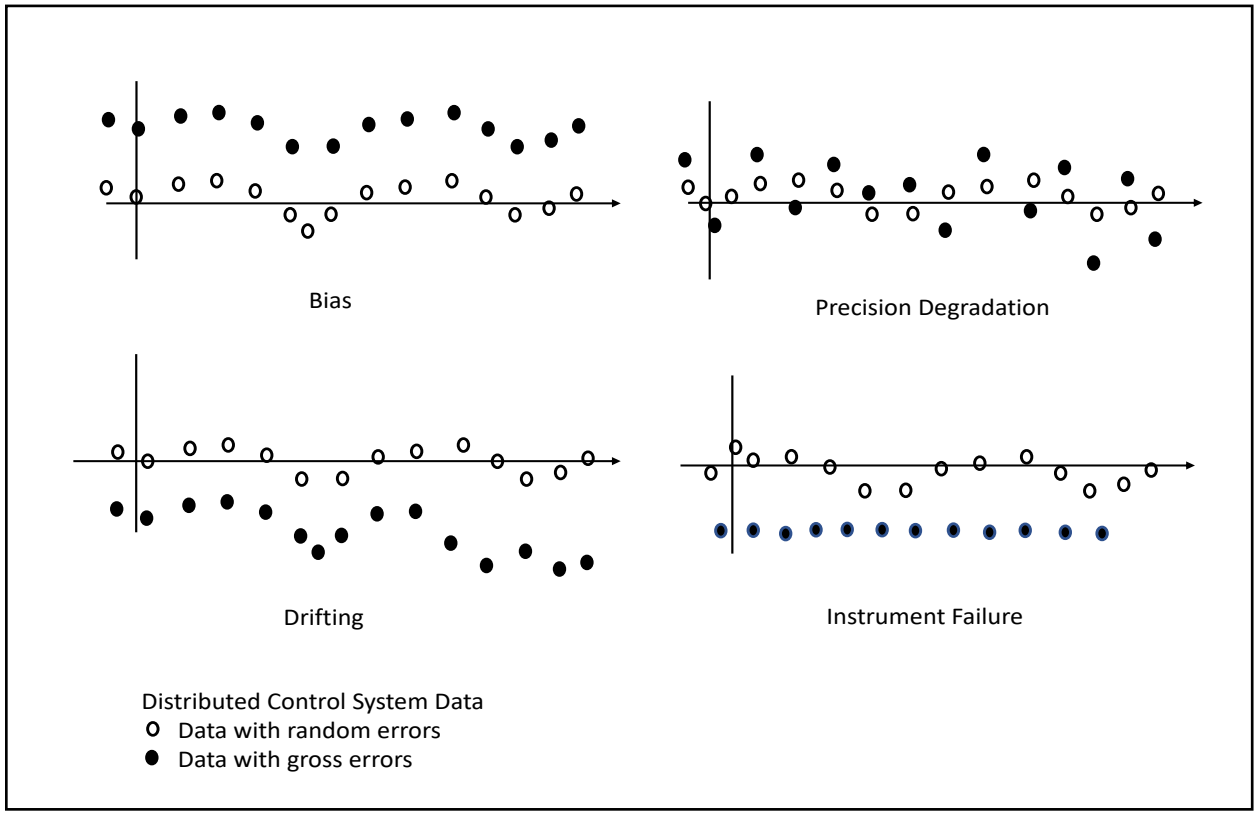


Figure 8-4 Illustrations of Some Types of Gross Errors in Data from a Distributed Control System

Gross errors are caused by non-random event including process leaks, biases in instrument measurements, malfunction of instruments, inadequate accounting of departures from steady state operations and/or inaccurate process models. The results for gross error detection have been reviewed and evaluated in detail through 1988 by Mah (1990), through 1993 by Crowe (1994) and through 2014 by Xu, et al. (2015).

There are numerous statistical methods for gross error detection, and the most successful method in industrial applications is called the robust function method and is based on robust statistics, Ozyurt and Pike, 2004. These methods require a detail plant model to relate the individual measurement and detect gross errors. They have been found to be very effective for detecting gross errors and usually require solving a nonlinear optimization problem. They use statistical hypothesis testing to determine if a gross error is present, and this requires selecting a

statistical test. A gross error is declared if the measurement error exceeds the value specified by the statistical test (the alternative hypothesis H_1 is accepted.) If the measurement error does not exceed this value, the measurement is said to not contain a gross error with a certain probability (the null hypothesis H_0 is accepted).

Significant reduction in product variability can be made through advanced control. However, there is a limitation of understanding instrumentation errors. Sanders (1995) reported that nearly two-thirds of the process upsets, which were severe enough to result in the restriction and downgrading of the product, could be traced to instrument faults. On-line gross error detection is the method for identifying instruments that produce abnormal information.

Several approaches, such as time series screening, statistical methods, or neural network method, have been practiced or proposed for gross error detection. Time series screening has been used in industrial applications. Vertical time screening is used to filter out gross errors in data sampled from the DCS, and horizontal time screening is used to test for steady state. These methods are straightforward but have to be performed manually. Also, they cannot detect persistent gross errors that include process leaks.

Methods for statistical hypothesis testing include global test, nodal or constraint test, measurement test, generalized likelihood ratio (GLR) method, Akaike's Information criterion (AIC) method, and unbiased estimation technique (UBET), and they have been described by a number of authors (Almasy and Sztano, 1975; Mah, et al., 1976; Willsky and Jones, 1974; Narasimhan and Mah, 1987 and 1988; Yamamura and coworkers, 1988; Rollins and Davis, 1992; Mah and Tamhane, 1982). If the covariance matrices of constraint residuals or measurement adjustments are not diagonal, the assumption that measurement errors are independent of each other is not satisfied, and this affects the power of the statistical tests. The methods of maximum power (MP) test (Tamhane, 1982) and principal component analysis (PCA) (Tong and Crowe, 1994 and 1995) were developed to overcome this weakness.

The traditional approach to describe gross error detection is to resent a test for a gross error in the data (Global Test) followed by a test for a gross error at a node in the data (Nodal Test). These tests typically determine if there is a gross error in the data but do not identify the specific variable with a gross error. Measurement tests attempt to do simultaneous data reconciliation and gross error detection, but this test can be confounded by gross errors and declare that a data point has a gross error when none is present.

There are two typical approaches for detecting gross errors using statistical methods. One is based on the distribution of constraint residuals, and the other is based on the distribution of measurement adjustments. The constraint residual **r** is given by (Mah, 1990)

$$\mathbf{r} = \mathbf{A}\mathbf{y} \qquad 8\text{-}14$$

where **A** is the coefficient matrix of constraint equations in Equation 8-11.

Methods based on the constraint residual **r** that do not require data reconciliation include the global test and the nodal test. These methods require linear constraints and all variables be measured (or the unmeasured variables be removed the projection matrix method). The global test only determines if there is a gross error in the plant data but not the measurement that contains the gross error. The nodal test only determines if there is a gross error in the measurements associated with a process unit but not the measurement that contains the error. These methods are not applicable to on-line optimization for complicated and highly nonlinear chemical processes., but they are worth a brief description as background for other methods that apply to on-line optimization. See Mah, 1990.

Methods based on the vector of measurement adjustments, $\mathbf{a} = \mathbf{x} - \mathbf{y}$, include the measurement test method, Tjoa and Biegler's contaminated Gaussian distribution method and the robust function method. They are combined gross error detection and data reconciliation methods. These methods reconcile the process data first, and then they examine the reconciled data to determine if a measurement contains a gross error. These methods can be applied to nonlinear constraints, and there can be unmeasured variables in the plant model.

Combined Gross Error Detection and Data Reconciliation

The process data from a distributed control system is subject to two types of errors: random and gross errors. Gross errors must be detected and removed, and the data reconciled before it is used to estimate plant parameters. Only combined gross error detection and data reconciliation algorithms can be used to detect and rectify the gross errors in measurements for on-line optimization. These algorithms are the measurement test method using a normal distribution, Tjoa-Biegler's method using a contaminated Gaussian distribution, and a robust statistical method using the Lorentzian robust function. These algorithms are described in the following section.

Measurement Test Method: This method assumes all measurements are subject to only random errors with known normal distributions under null hypothesis, and the measurement errors are independent of each other. Then the distribution probability function for measurement error i under null hypothesis is given by Equation 8-4, and the joint probability distribution for all measurement errors is the product of the distributions for individual measurement error given in Equation 8-5.

The measurement errors are estimated by maximizing the joint probability density function P, Equation 8-5, or minimizing the sum squares of standardized measurement errors, $\mathbf{e}^T\mathbf{Q}^{-1}\mathbf{e}$, subject to a set of constraints, Equation 8-6, which represent the material and energy balances, etc. This is the well-known least squares method, Equation 8-10, and it is expressed as:

$$\begin{aligned} &\textit{Minimize}: \quad \underset{\mathbf{x},\,\mathbf{z}}{} \quad \mathbf{e}^T\mathbf{Q}^{-1}\mathbf{e} = (\mathbf{y}-\mathbf{x})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{x}) \\ &\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = \mathbf{0} \\ &\qquad\qquad \mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U,\ \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U. \end{aligned} \tag{8-15}$$

where **x**, **y**, **z**, and **θ** have the same meaning as described in Equation 8-1 previously. In Equation 8-15, **x** and **z** are variables to be determined by the optimization. **θ** is a constant vector of parameters and **y** is a constant vector of measurements. Solving Equation 8-15 will estimate the values for the measured variables **x** and unmeasured variables **z**. Then, the measurement errors can be determined by **a** = **y** - **x**.

After data reconciliation, each measurement error is examined to see if it contains a gross error by a test statistic. The test statistic of measurement test method is:

$$| \varepsilon_i | = | a_i | / \sigma_i \sim N(0, 1) \qquad 8\text{-}16$$

Equation 8-16 means that the standardized measurement error, $\varepsilon_i = a_i / \sigma_i$, follows a standard normal distribution N(0, 1) if the measurement does not contain gross error.

If the value of test statistic, $| a_i | / \sigma_i$, exceeds the critical value C, then this measurement contains a gross error. Otherwise, there is no gross error in this measurement. The critical value C is selected from the table of the standard normal distribution function at the significant level β for individual measurement. If the overall significant level is specified as 0.05 (e.g., 95% confidential interval), α = 0.05, and 43 measurements are used, then the significant level for individual measurement is given by Narasimhan and Jordache, 2000:

$$\beta = 1 - (1-\alpha)^{1/m} = 1-(1-0.05)^{1/43} = 0.0012 \qquad 8\text{-}17$$

At the β/2=0.006 point, the critical value C is determined from the standard normal distribution with accumulated probability at 0.994, and the value is 3.2, i.e., C = 3.2. In Example 8-2, the standard error for 6 measurements, |0.032 - 0.026|/0.0020 = 3.0, and this point may contain a gross error. Equation 8-17 comes from hypothesis testing where β is the probability of a Type 2 error declaring there is no gross error when one is present, α i is the probability of a Type 1 error declaring a gross error present when there is none and m is the number of measured variables.

The measurement test method in Equation 8-15 and 8-16 is an option for data reconciliation in the interactive on-line optimization program. The algorithm for the measurement test in the program is as follows:

1. Conduct data reconciliation and evaluate **a** and **x**.

2. Compute test statistic $| \varepsilon_i | = | a_i | / \sigma_i$ and compare with criterion value of the test statistic, C.

If $| \varepsilon_i | \geq C$, then designate this value as a suspected gross error.

If $| \varepsilon_i | \leq C$, then designate this value as having a random error.

3. Remove variables that have suspected gross errors by combining process units and eliminating them as variables.

4. Conduct data reconciliation on the new system from step 3 and preform the evaluations in Step 2 to detect gross errors.

5 Repeat Steps 3 and 4 until all of the remaining variables contain random errors.

Serth and Heenan,1986 described this algorithm and modifications that were tested with a model of a process steam system for a methanol synthesis unit. They reported the various modifications each had advantages and disadvantages. The disadvantage of these measurement tests is that a gross error in one variable can cause other variables to seem to contain gross errors. Variables that have gross errors have to be eliminated by combining units and are not evaluated. In Example 1, if y_2 has a gross error, the heat exchanger and reactor must be combined and only y_1 and y_3 can be reconciled.

Contaminated Gaussian Distribution Function Method: The measurement test assumes all of the measurements are randomly distributed, but gross errors are not. In an attempt to replace the normal distribution with a distribution function that is insensitive to gross errors, Biegler, et al., (Tjoa and Biegler, 1991; Albuquerque and Biegler, 1995) proposed a contaminated Gaussian distribution function to describe random and gross measurement errors.

If a measurement is subject to either random or gross error, there are the two possible outcomes: G = {Gross error occurred} with prior probability η and R = {Random error occurred} with prior probability 1-η. Therefore, the distribution of a measurement error is:

$$P(y_i \mid x_i) = (1-\eta)P(y_i \mid x_i, R) + \eta\, P(y_i \mid x_i, G) \qquad 8\text{-}18$$

where $P(y_i \mid x_i, R)$ is the probability distribution of a random error and $P(y_i \mid x_i, G)$ is the probability distribution of a gross error.

It is assumed that the random errors are normally distributed with a zero mean and a known variance σi^2. The distribution function for a random error is:

$$P(y_i / x_{i,} R) = \frac{1}{\sqrt{2\pi}\sigma i} e^{\frac{(y-x)^2}{2\sigma^2}} \qquad 8\text{-}19$$

Also, it is assumed that the gross errors are subject to a contaminated normal distribution which has a zero mean and larger variance $(b\sigma)^2$, $(b >> 1)$. Therefore, the distribution function for a gross error is:

$$P(y_i / x_{i,} G) = \frac{1}{\sqrt{2\pi}\sigma i} e^{\frac{(y-x)^2}{2b^2\sigma^2}} \qquad 8\text{-}20$$

If the measurement errors are independent of each other, then the likelihood function for all measurements is the product of the distributions for individual measurement, i.e.,

$$P(x/y) = \prod_i P(y_i/x_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} \left\{ (1-\eta) e^{\frac{-(y-x)^2}{2\sigma^2}} + \frac{n}{b} e^{\frac{-(y-x)^2}{2b^2\sigma^2}} \right\} \quad 8\text{-}21$$

The measurement errors are estimated by maximizing the joint probability density function (likelihood function) in Equation 21 or minimizing the negative logarithm of Equation 8-21. The optimization problem for combined gross error detection and data reconciliation using the contaminated Gaussian distribution can be stated as:

$$\text{Minimize:} \quad \rho = -\sum_i \left\{ \ln\left[(1-n) e^{\frac{-(y-x)^2}{2\sigma^2}} + \frac{n}{b} e^{\frac{-(y-x)^2}{2b^2\sigma^2}} \right] - \ln\left[\sqrt{2\pi\sigma_i}\right] \right\} \quad 8\text{-}22$$

$\mathbf{x}, \mathbf{z}$

Subject to: $\mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0$

$$\mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U, \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U$$

This optimization problem is comparable to Equation 8-15 for the least squares (measurement test) method. Solving Equation 22 determines the values of measured and unmeasured variables (**x** and **z**). Then, the measurement errors are determined by **e** = **y** - **x**.

After data reconciliation, each measurement is examined with a test statistic to see if it contains a gross error. The test statistic for gross error detection is:

$$|\varepsilon_i| = \left| \frac{y_i - x_i}{\sigma_i} \right| \ge \sqrt{\frac{2b^2}{b^2 - 1} \ln\left[\frac{b(l-\eta)}{\eta} \right]} \quad 8\text{-}23$$

If $|\varepsilon_i|$ greater than the right-hand side of Equation 8-23, then measurement i contains gross error. Otherwise, no gross error is present in this measurement. Recommended values for the two parameters in Equation 8-23 are: $\eta = 0.05$ and $b = 10$, and the value of the right-hand side of Equation 8-23 is 2.157, i.e., if $|\varepsilon_i| > 3.26$, then measurement i contains a gross error.

In summary, the contaminated Gaussian distribution method is composed of the distribution functions for random and gross errors. The reconciled data from contaminated Gaussian distribution method is not sensitive to the presence of gross errors, and this method gives an unbiased estimation for the reconciled data. This can be seen by weight coefficients of measurements in the linearized joint distribution. The objective function in Equation 8-21 can be approximated as a linear function using a first order Taylor expansion, i.e., $\rho = \sum w_i [(y_i - x_i) - (y_i - x_i)^0] = \sum w_i (\varepsilon_i - \varepsilon_i^0)$, where w_i is the weight coefficient of measurement y_i on the joint distribution function (objective function in Equation 8-21) evaluated at the last feasible point x_i^0 or ε_i^0, and it is the partial derivatives of the joint contaminated Gaussian distribution function with respect to the variable x_i as given below:

$$W_1 = \frac{\frac{(y_i - x_i)}{\sigma^2}\left\{(1-\eta)e^{\frac{-(-y-x)^2}{2\sigma^2}\left(1-\frac{1}{b^3}\right)} + \frac{\eta}{b^3}\right\}}{(1-\eta)e^{\frac{-(y-x)^2}{2\sigma^2}\left(1-\frac{1}{b^2}\right)} + \frac{\eta}{b}} \qquad 8\text{-}24$$

$$= \frac{\frac{\varepsilon_i}{\sigma}\left\{(1-\eta)_e^{\frac{-\varepsilon^2}{2}\left(1-\frac{1}{b^2}\right)} + \frac{\eta}{b^3}\right\}}{(1-\eta)e^{\frac{-\varepsilon^2}{2}\left(1-\frac{1}{b^2}\right)} + \frac{\eta}{b}} / \varepsilon - \varepsilon^0$$

For smaller error, e.g., $\varepsilon_i < 2$, the exponential term in the Equation 8-24 is much larger than the second term η/b^3 (or η/b), The weight function can be simplified as $w_i \propto (y_i\text{-}x_i)/\sigma i^2 = \varepsilon_i / \sigma_i$. For larger error, e.g., $\varepsilon_i > 4$, the exponential term in the equation is much smaller than the second term η/b^3 (or η/b). The weight function can be simplified as wi $\propto (y_i\text{-}x_i)/(b\sigma_i)^2 = \varepsilon_i/(\sigma_i\, b2)$. Therefore, Equation 8-24 can be approximated as given in Equation 8-25:

$$w_i = \{\varepsilon_i / \sigma_i for \varepsilon_i < 2 \quad \{\varepsilon_i / (\sigma_i b^2) for \varepsilon_i > 4 \qquad 8\text{-}25$$

From the weight coefficient function in Equation 8-25 and the linearized objective function, it is seen that the measurement with a smaller error has a large weight coefficient (i.e., wi = ε_i/σi) in the linearized objective function than the measurement with a larger error (i.e., $w_i = \varepsilon_i/(\sigma_i\, b^2)$, where b>>1). This means the measurement with a larger error has a smaller effect on the objective function, and its value is determined mainly by the measurements with small errors.

The procedure to conduct contaminated Gaussian distribution method is:

1. Solve optimization problem, Equation 8-22, with a set of measurements, **y**, from the DCS to determine the reconciled values for measured variables, **x** and unmeasured variables **z**, and then the measurement adjustments, **a** = **x** - **y**, are determined.

2. Examine the standardized measurement adjustment ε_i, $\varepsilon_i = a_i / \sigma_i$, using the criterion given Equation 8-25 to determine if a measurement contains a gross error. If a measurement contains a gross error, then its value is replaced with the reconciled data. Then a new set of measurements is constructed using the reconciled data to replace the measurements containing gross errors along with the original measurements that contain only random errors. This new set of measurements contains only random errors, and it is used in simultaneous data reconciliation and parameter estimation to update plant parameters for on-line optimization.

Robust Statistical Methods: The motivation for robust function methods is reducing the impact of gross errors in estimating statistical parameters, e.g. variance. See Huber, 1981. The

basic idea of robust estimation is to build a robust distribution function ρ which is asymptotic to the normal distribution or any pre-assumed rigorous distribution function that describes the distribution pattern of measurement errors under some ideal assumptions. The estimator (mean or variance) determined by the robust distribution is insensitive to extreme observations and yet maintains a high efficiency (lower dispersion).

Two robust functions have been proposed in literature for mean estimation, and they are applicable for data reconciliation and gross error detection of on-line optimization. These robust functions are Lorentzian distribution proposed by Johnston and Kramer (1995), which was originally presented by Huber (1981), and Fair function proposed by Albuquerque and Biegler (1995). Chen (1998) has shown that the Lorentzian function gives an effective method for data reconciliation, and the Fair function is only slightly better than the measurement test.

Lorentzian distribution function of a measurement error is given as:

$$\rho(\varepsilon_i) = \frac{1}{1 + \frac{1}{2}\varepsilon_i^2} \qquad 8\text{-}26$$

where ε_i is the standardized measurement error, i.e., $\varepsilon_i = e_i / \sigma_i = (y_i - x_i)/\sigma_i$. The robust function of measurement errors using Lorentzian distribution is the sum of the individual distribution, i.e,

$$\rho(\varepsilon) = \sum_i \rho(\varepsilon_i) = \sum_i \frac{1}{1 + \frac{1}{2}\varepsilon_i^2} \qquad 8\text{-}27$$

The optimization problem for the combined gross error detection and data reconciliation using the Lorentzian distribution function is expressed as:

Maximize: $\rho(\varepsilon) = \sum_i \frac{1}{1 + \frac{1}{2}\varepsilon_i^2}$

$\mathbf{x}, \mathbf{z}$ 8-28

Subject to: $\mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0$

$\mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U, \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U$

The procedure to conduct gross error detection and data reconciliation with robust method is the same as the one for contaminated Gaussian distribution method. After solving the optimization problem in Equation 8-28, the reconciled data for measured variables is determined, and the measurement adjustments can be determined by $\mathbf{a} = \mathbf{y} - \sim\mathbf{x}$. Then, each measurement adjustment is examined to see if it contains a gross error by a test statistic.

The test statistic for robust method is established using a statistical hypothesis test procedure as measurement test method. If the standardized measurement adjustment, $|\varepsilon_i| = |a_i|/\sigma_i$, does not exceed the critical value C, then measurement i does not contain a gross error. Otherwise,

the measurement contains a gross error. The critical value C is determined by the robust function at the specified confidential interval or significant level β. For example, if 95% of confidential level is used, then the overall significant level α is 0.05 and the significant level for individual measurements β is calculated by Equation 8-17 from the given overall significant level α and the number of measurements m. Then, the critical value C is the error size that has an accumulated probability value as (1-β/2).

Comparison of Methods: The theoretical performance of four distribution functions: normal distribution of measurement test method, contaminated Gaussian distribution of Tjoa-Biegler's method, Lorentzian distribution and Fair function of robust method, were evaluated based on the influence function and relative efficiency of the distributions. In summary, the evaluation of influence functions of distributions showed that normal distribution causes significant biased estimation if measurements with gross errors were used to reconcile data and the degree of bias increased unboundedly with the increase of errors. Therefore, an iterative elimination strategy was required to avoid the bias whenever a gross error was detected. The comparisons of influence function and relative efficiency showed that both contaminated Gaussian and Lorentzian distributions had a better combination of influence function (gross error sensitivity) and relative efficiency (estimation accuracy). Therefore, they would have a better performance when reconciling data with both random and gross errors. The method using the contaminated Gaussian distribution would have the best performance for measurements with moderate size of gross errors among four distribution, and the method using the Lorentzian would be more effective for extremely large gross errors.

In a detailed study of robust statistics functions for data reconciliation and gross error detection, Ozurt and Pike, 2004, six different methods were derived from robust statistics and were investigated for along with weighted least squares and a modified version of measurement test for nonlinear models to compare the data reconciliation and gross error detection performance. This involve the use of the influence function (Hampel et al. 1986), which is defined for a sample x, an estimator T over an assumed distribution function F and a perturbed distribution function F_t as follows:

$$IF(x,T,F) = \lim_{x\to 0} \frac{T(F_t) - F(T)}{t} = \frac{\partial [T(F_t)]|_{t=0}}{\partial t} \qquad 8\text{-}29$$

The heuristic interpretation of this influence function is that "it describes the effect of an infinitesimal contamination at the point x on the estimate" (Hampel et al. 1986). The influence function is proportional to the derivative of the maximum likelihood function, and the weight given to any gross errors in the measurements while calculating the estimates can be seen in Figure 5.

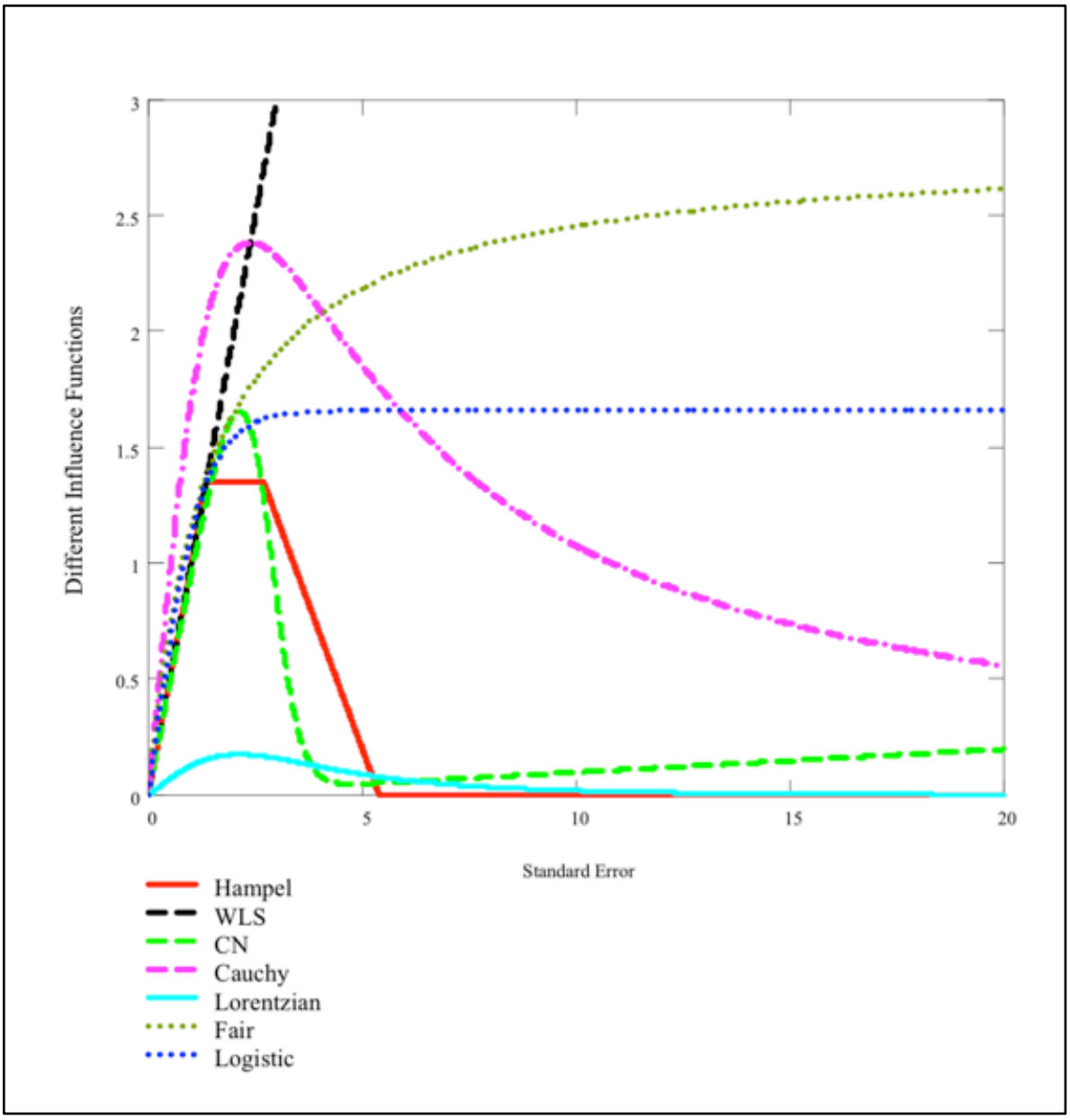


Figure 8-5 Comparison of Influence Functions for Robust Estimators, from Ozurt and Pike, 2004

Referring to Figure 8-5 the influence function for the WLS (measurement test) is proportional to the measurement error justifying the low breakdown point and unbounded effect of large errors. The effect of larger errors is reduced for the r function of the Cauchy distribution, "Lorentzian" function and Hampel's redescending M-estimator and is shown by gradually decreasing influence functions in the region of greater than 3.0 of the standard error. Therefore, these three r functions are called redescending r functions. Fair function and the r function of the Logistic distribution have a bounded influence by the large errors since their influence function increases slowly with respect to the measurement errors approaching a constant value for large errors. The influence of small measurement errors on the r function of the Contaminated Normal distribution is the same as on the WLS function. However, the influence decreases for larger errors and becomes proportional to very large errors after passing through a minimum (at standard error 4.7 in Figure 8-5).

The evaluation of the performance of a total of eight methods is undertaken using five small- scale examples from the literature and two cases involving industrial plants with real process data. The Monte Carlo study shows that the robust approaches for the simultaneous data reconciliation and gross error detection of chemical processes can provide similar or better results compared to a sequential method, with a single (two for Hampel's redescending M-estimator) solution of the NLP.

$$\begin{aligned} &\textit{Minimize}: \quad \mathbf{e^T Q^{-1} e} = (\mathbf{y} - \mathbf{x})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{x}) \\ &\quad \boldsymbol{\theta} \\ &\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ &\qquad\qquad \mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U, \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U, \boldsymbol{\theta}^L \le \boldsymbol{\theta} \le \boldsymbol{\theta}^U \end{aligned} \tag{8-30}$$

In simultaneous data reconciliation and parameter estimation the parameters in plant model are considered as variables along with the measured process variables. Both measured variables and parameters are estimated simultaneously when solving the nonlinear programming problem. The general mathematical statement for simultaneous data reconciliation and parameter estimation is written as:

$$\begin{aligned} &\textit{Minimize}: \quad \mathbf{e^T Q^{-1} e} = (\mathbf{y} - \mathbf{x})^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{x}) \\ &\quad \mathbf{x}, \mathbf{z}, \boldsymbol{\theta} \\ &\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ &\qquad\qquad \mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U, \mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U, \boldsymbol{\theta}^L \le \boldsymbol{\theta} \le \boldsymbol{\theta}^U \end{aligned} \tag{8-31}$$

where the equality constraints $\mathbf{f}$ denote the plant model, $\mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U$, $\mathbf{z}^L \le \mathbf{z} \le \mathbf{z}^U$, and $\boldsymbol{\theta}^L \le \boldsymbol{\theta} \le \boldsymbol{\theta}^U$ represent the bounds on process variables ($\mathbf{x}$ and $\mathbf{z}$) and parameters $\boldsymbol{\theta}$. Least squares are shown in Equation 31 for the objective function, but the objective function for the contaminated Gaussian distribution, and Lorentzian distribution or other robust distribution function could be used. Also, more precise values of the parameters may be obtained using multiple measurements of y_i and the objective function in Equation 8-31 would be replaced with:

$$\min_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{y_{ij} - x_j}{\sigma_j} \right) \qquad 8\text{-}32$$

Simultaneous Data Reconciliation and Parameter Estimation

In data reconciliation and gross error detection random and gross errors have been removed from the data from the distributed control system, and the data satisfies the material and energy balances, the plant model. These evaluations were performed using values of the parameters that were previously determined. In the case where parameters have slowly varying values such as heat transfer coefficients from fouling and catalyst deactivating, the values of these parameters need to be up-dated for the current performance of the process equipment using the reconciled plant data. The following optimization problem could be evaluated where the best values of the model parameters, **θ,** are determined**.**

Procedure for Data Reconciliation, Gross Error Detection and Parameter Estimation

As shown in Figure 8-6, combined gross error detection and data reconciliation is conducted by solving the nonlinear programming problem given by Equations 8-15. Then simultaneous data reconciliation and parameter estimation is conducted by solving the nonlinear programming problem given by Equation 8-31. In combined gross error detection and data reconciliation, data reconciliation is required to reconciled process data and to estimate the measurement errors for gross error identification. In simultaneous data reconciliation and parameter estimation, data reconciliation is required to estimate process parameters and process variables. These two data reconciliation optimization problems use the same plant model, and the only difference is that the process parameters are constants in combined gross error detection and data reconciliation and are variables in simultaneous data reconciliation and parameter estimation.

Data reconciliation requires current values of the parameters in the plant model. However, only the parameter values from the previous optimization cycle are available. Therefore, a strategy to avoid this dilemma is to detect and reconcile the measurements containing gross errors using the plant model with the parameter values from previous on-line optimization cycle in gross error detection and data reconciliation. Then a new set of measurements is constructed using the reconciled data to replace the measurements containing gross errors along with the original measurements that contain only random errors. This new set of measurements is supposed only containing random errors, and it can be used to conduct simultaneous data reconciliation and parameter estimation using least squares method with error-in-variables formulation.

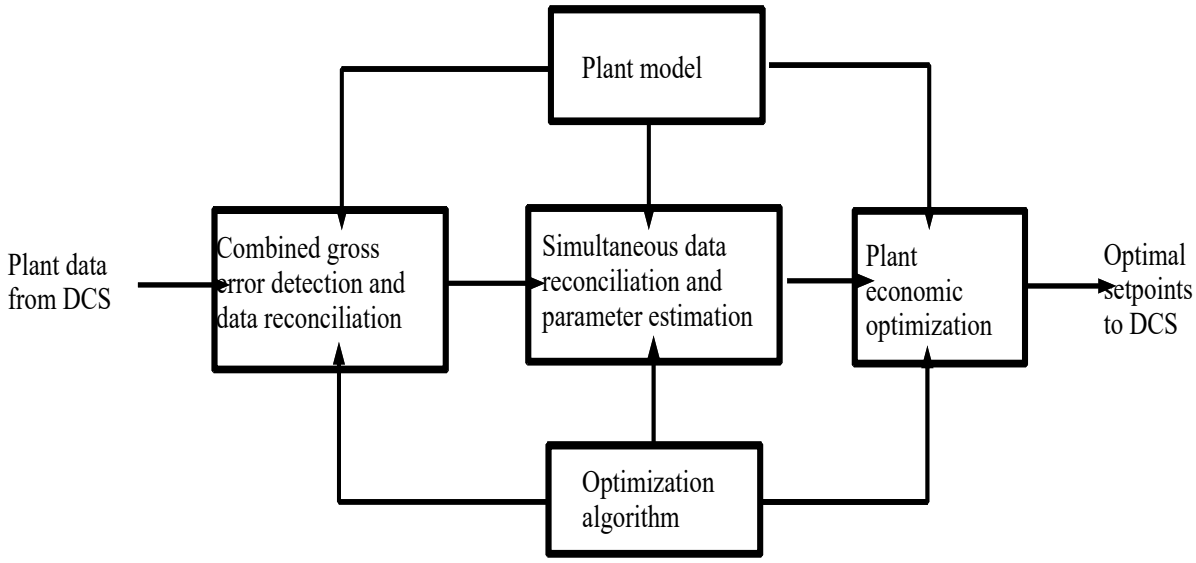


Figure 8-6 Relationship between Key Elements of On-Line Optimization

Plant Economic Optimization

The objective of plant economic optimization is to generate a set of optimal setpoints for the distributed control system. These setpoints will maximize the profit and minimize waste generation and energy use. The nonlinear programming problem for economic optimization is:

$$\begin{aligned} &\textit{Maximize}: && P(\mathbf{x}) \\ &\mathbf{x}, \mathbf{z} \\ &\textit{Subject to}: && \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \\ & && \mathbf{g}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \leq 0 \\ & && \mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U, \mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U \end{aligned} \qquad 8\text{-}33$$

where P(**x**) represents the economic model (e.g., profit function). The equality constraints **f** are the same as those in data reconciliation. The inequality constraints **g** represents the additional restrictions for the economic optimization, such as the demand for the main products and by products, availability of raw materials, maximum and minimum capacities of the process equipment, and restriction on the waste/pollutant emission. The bounds $\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$ and $\mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U$ represent the allowable minimum and maximum operating conditions for the process variables.

Economic Models: The economic model in Equation 8-33 can be an equation to maximize plant profit, minimize cost of operations, energy use, production of undesired by-products, waste/pollutant emission, or a combination of these objectives. Details about these methods are given in the companion volume Economic Decision Analysis for Chemical Engineering, Pike, 2015.

There are two types of industrial optimization problems: design optimization (new plant, plant expansion, debottlenecking, adding new technology) and operations optimization (process, plant, multi-plant) as shown in Figure 8-7. Both require economic models that describe the profit to be maximized or the cost to be minimized, and a plant simulation (process model) that is used to predict the performance of the plant. As shown in Figure 8-7, the economic model for optimal plant design is net present value, while the economic model for optimal plant operations is net profit. Net present value is the annual cash flows discounted to the present value after all the capital and operating expenses have been paid. The net profit is the difference between the funds received from selling the product and the manufacturing costs. The manufacturing costs include operating and raw material costs, taxes, administration, and other costs.

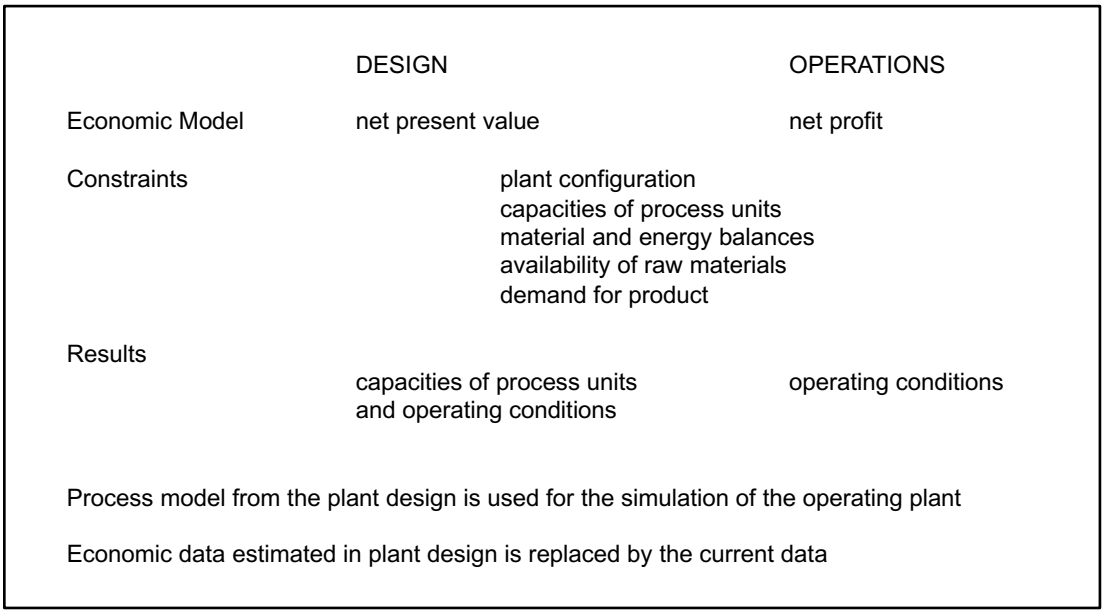


Figure 8-7 Comparison of Design and Operations Optimization

A simpler version of the net profit is the "value added" economic model that is the difference between the sales and raw material where operating costs and all other costs are assumed constant. It is expected that the net profit after taxes will be comparable to the estimates of the cash flows made for this time period when the plant was designed and that the plant will generate the anticipated return on the investment.

Unlike in the design optimization problem, the plant configuration is specified, thus making the operation optimization problem somewhat easier. However, there are multiple levels of optimization that must be considered as shown in Figure 8. One level is the optimal scheduling problem of corporate headquarters to distribute raw materials among the company's plants to maximize profits in producing, transporting, and marketing products to consumers worldwide. Also included is the optimal scheduling problem of the individual plant to set operating conditions to produce required products from allocated raw materials for a maximum net profit or minimum cost of operations. The best schedule is determined for steady-state daily or weekly average flow rates for the plant. Finally, there is on-line optimization of process operations to determine the set-points for the distributed control system of the individual process units in the plant which give the best operating conditions while producing the specified quality and quantity of products as shown in Figure 8. Also, on-line optimization keeps track of such things as catalyst deactivation and scaling in heat exchangers by parameter adjustments in the process models of the units from sampling plant data.

To summarize, optimization for design and plant operations are different in several ways. The economic model for design is net present value and for operations is net profit. The process model for operations includes the plant configuration, material and energy balances, availability of raw materials, and demand for products. The process model for design plant does not have a plant configuration, and it has to be determined, along with the capacities of process units. Finally, design optimization determines the capacities of individual units and plant operating conditions. The process model from plant design can be transferred to the simulation of the operation of the plant. Economic data estimated in plant design are replaced by actual data.

Classification of Variables and Determination of the Parameters

After the constraints for the plant model have been constructed, the variables in the model are divided into two groups: measured and unmeasured variables. There should be as many measured variables as possible. In general, more measurements will give a more accurate estimation of the reconciled data.

Measured variables are available from the distributed controlled system (DCS) and the plant control laboratory. The remaining variables in the process model are unmeasured variables. Some additional measurements may be required after evaluating observability and redundancy which will be discussed in the following section.

There are two types of parameters in the process model as shown in Figure 8-9. One type are constants, such as reaction activity energy, stoichiometry of chemical reactions and distillation column diameter and height. The other types are time-varying parameters, such as heat exchanger fouling factors and catalyst deactivation parameters. These parameters vary slowly with time, e.g., 10% per month. The values of these parameters are determined by the characteristics of the equipment and physical properties of materials but are not strongly relate to operating condition. Their values provide the information about equipment performance.

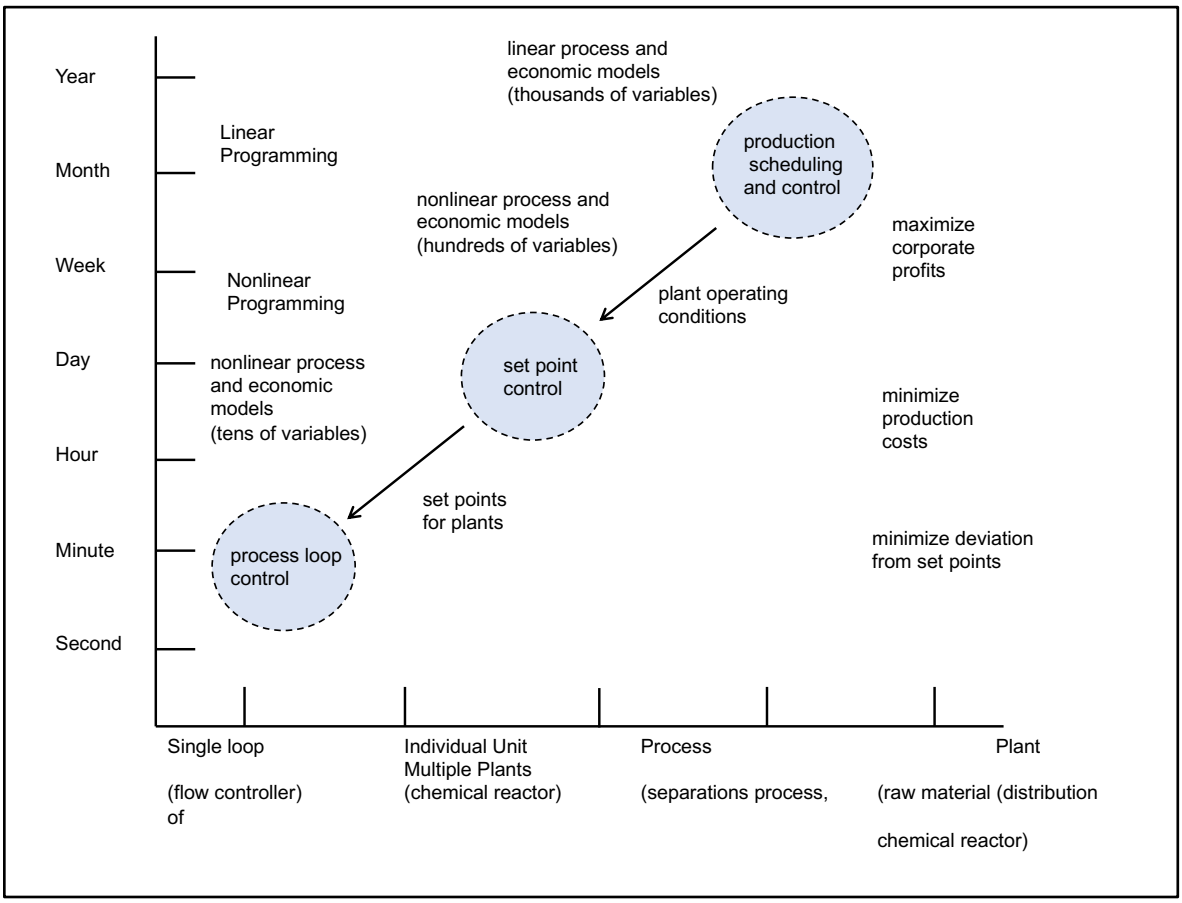


Figure 8-8 Plant and Time Scales in Process Optimization, after Koninckx, et al., 1988

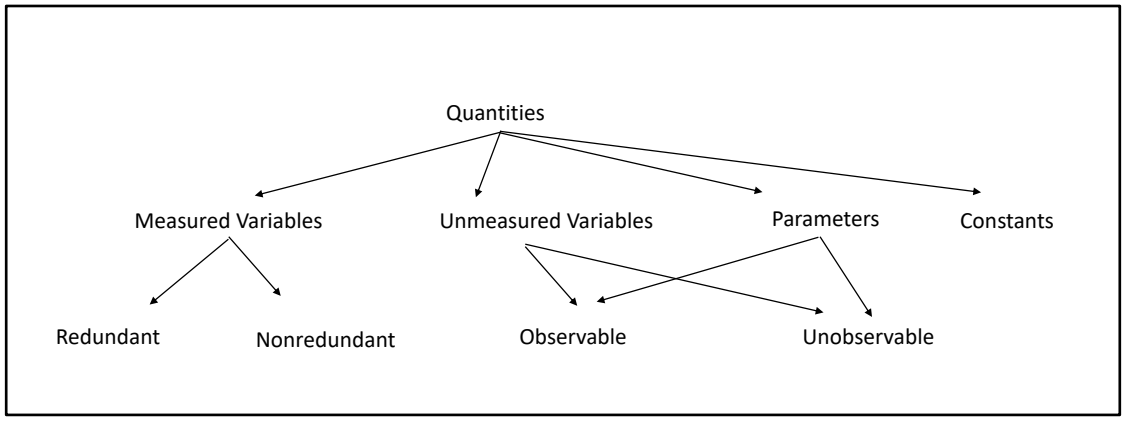


Figure 8-9 Classification of Quantities in Plant Model

As shown in Figure 8-9 variables can be measured or unmeasured. Measured variables can be redundant or nonredundant depending on the number of measured variables. Unmeasured variables can be observable or unobservable depending on the constraint equations.

A requirement for data reconciliation is that redundant measurements must be available. Redundant measurements occur when the reconciled values for the measured variables and any fixed variables complete a material or energy balance (Knopf, 2010). For a process variable to be reconciled, it must be measured, and its adjusted or reconciled value must appear in a useable material or energy balance constraint in the data reconciliation problem. Measuring a variable does not ensure that it can be used for data reconciliation (Knopf, 2010). To conduct data reconciliation, redundant measurements are required to reconcile errors in measurements. This leads to the definitions of observability and redundancy given by Bagajewicz, 2010:

Observability: "A non-measured variable is observable if it can be calculated in at least one way from the measurements."

Redundancy: "A measurement is redundant if it can be calculated in at least one way from the remaining measurements. That is, the measurement can be deleted, and the rest of the measurements can be used to calculate the value of that variable."

Redundancy is a desirable property of a system because if an instrument fails its variable can be estimated using the balance equations. Details about the qualifications of observability and redundancy are reported by Bagajewicz, 2010.

The following method for observability and redundancy is based on an analysis of the degree of freedom of the constraint equations, Chen, 1998. Consider a set of m equality constraint equations with p parameters and n variables, in which n_1 variables are measured. The unmeasured variables and parameters are observable if the number of measured variables n_1 is greater than or equal to the number of degrees of freedom. The number of degrees of freedom for a set of equations is the difference between the number of variables and of equations, i.e., $d = n - m$. For data reconciliation and gross error detection, the parameters are constants and not considered unmeasured variables. For combined data reconciliation and parameter estimation, the degrees of freedom are the number of variables and parameters subtracted from the number of equations, i.e., $n + p - m$.

Determination of observability and redundancy is conducted for each unit or each balance node or for entire process (multiple units). If it is conducted for each unit, then the examination result is called local observability and redundancy. If it is conducted for entire process, then the examination result is called global observability and redundancy.

For a set of constraint equations for a unit, the unmeasured variables and parameters are local observable, if the number of measured variables is greater than or equal to the degree of freedom of this set of equations. The degrees of freedom is the number of variables (measured and unmeasured) and parameters subtracted by the number of equations. For local observability and redundancy, the classification of measured variables and unmeasured variables is slightly

different from the definition given above. A class of dummy measured variables is introduced in local examination to represent the unmeasured flow rate variables that can be directly determined by available measured variables at the up or down stream. The number of measured variables equals the sum of the numbers of measured variables and dummy measured variables in the equations, and the number of unmeasured variables equals the number of unmeasured variables subtracted by number of dummy measured variables. For example, measured variable could be an instrument reading that gives a stream flow rate.

For a set of constraint equations for a unit, it is said that the measured variables have local redundancy if the number of measured variables is larger than the degree of freedom of this set of equations, and the number of local redundancy of measurements equals the number of measured variables subtracted by the number of degrees of freedom. For individual measured variables, it is said that a measured variable is redundant if all unmeasured variables and parameters are observable after the measured variable is changed to an unmeasured variable. Otherwise, the measured variable is not redundant.

In Figure 10 a process flow diagram there are three units, two heat exchangers and an adiabatic flash unit. In streams S1, S2, S3, S4, and S5, there are two components A and B. Consider having measured variables: F_{1A}, F_{1B}, T_1, P_1, F_{5A}, F_{5B}, T_5, and P_5 with the others being unmeasured variables. Also, consider having unmeasured variables: F_{2A}, F_{2B}, F_{4A}, and F_{4B} be dummy measured variables associated with the adiabatic flash unit because F_{2A}, F_{2B}, F_{4A}, and F_{4B} can be directly determined by measured variables F_{1A}, F_{1B}, T_1, P_1, F_{5A}, F_{5B} respectively through the component mass balances. However, temperatures T_2 and T_4 are not dummy measured variables because they cannot be directly determined by available measured variables.

For heat exchangers shown in Figure 10, each unit has nine equations which involved 13 variables e.g. for heat exchanger 1: F_1, F_2, F_3, F_4, T_1, T_2, T_3, T_4, H_1, H_2, H_3, H_4, and ΔT_m and two parameters, U and Q_{loss} if both cold and hot streams have single components. The degrees of freedom for this set of equations and variables are six. Therefore, six variables must be measured variables or dummy measured variables to satisfy the observability, and more than six variables must be measured or dummy measured variables to provided redundancy for error rectification.

After the unit by unit examination of observability and redundancy, the global observability and redundancy are examined for entire process based on the number of measured variables and degree of freedom for the entire process. In global observability and redundancy examination, all dummy measured variables belong to unmeasured variables.

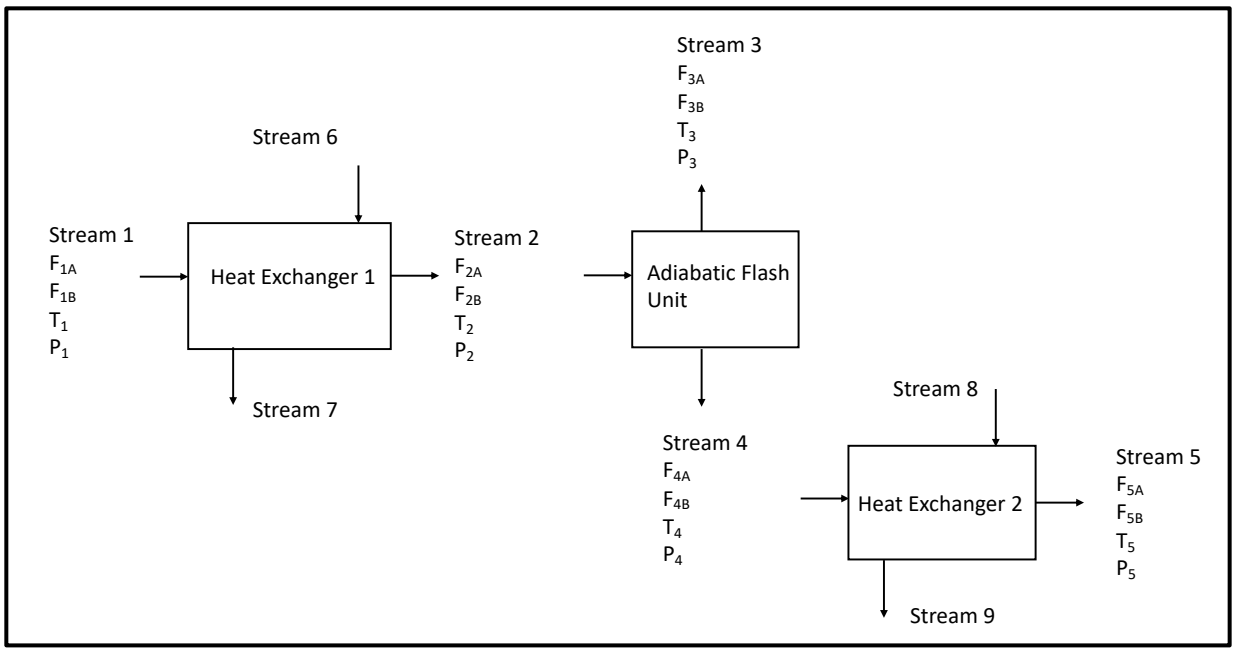


Figure 8-10 Simple Process Flow Diagram

Plant Model Formulation and Validation

After the plant model is completely formulated and the process variables are correctly classified into measured variables (**x**), unmeasured variables (**z**), and parameters (**θ**), the accuracy of the plant model must be examined. To assess precision of the plant model, the simulation results predicted by the plant model must be compared with accurate data from plant. Consistent and complete plant design data can be used to help ensure that the constraint equations are correctly describing the processes. This can be done by designating some of plant design values as measured data. Then this data is used to estimate the values of the unmeasured variables and the plant parameters, and the estimated parameters and process variables are compared with the plant design data. If the predicted results are very close to the design data with a less than 1% relative deviation, then it is said that the plant model precisely simulates the plant.

The paragraph above is the brief discussion on the development and examination of plant model. The following gives a general procedure and the steps necessary for formulating an effective and precise plant model for on-line optimization.

1. Develop the process constraints according to the conservation laws (material and energy balances) rate equations and equilibrium relations.

2. Select plant parameters, **θ**, to be updated by on-line optimization. Classify the variables in plant model into measured variables, **x,** and unmeasured variables, **z,** according to the

measurability and/or available measurements for variables. Incorporate as much measurement information as possible.

3. Evaluate the observability of unmeasured variables, **z,** and parameters, **θ,** and the redundancy of measured variables **x**. All unmeasured variables and parameters must be observable; and usually, the higher the redundancy the better the reconciliation.

4. Evaluate the precision of the process model by comparing the plant model with accurate information, such as the plant design data.

Execution Frequency for On-Line Optimization

The execution frequency of optimization is the time between conducting optimizations of the process, and it has to be determined for each of the units in the process. It depends on the settling time, i.e., the time required for the units in the process to move from one set of steady-state operating condition to another. The settling time can be estimated from the time constant determined by process step testing. The time period between two on-line optimization execution must be longer than the settling time to ensure that the units have returned to steady state operations before the optimization is conducted again. This is illustrated in Figure 8-10, after Darby and White (1988). The figure shows an execution frequency for optimization that was satisfactory for one process may be too rapid for another process which has a longer settling time. In Figure 8-11a, the process has returned to steady-state operations and held that position until the next optimization. However, in Figure 8-011b, the process did not have enough time to return to steady-state operations before the optimization altered the operating conditions; the process would not return to steady state operations if such optimization continued. The settling time for an ethylene plant is four hours according to Darby and White (1988), and this time for sulfuric acid contact process is twelve hours according Hertwig (1997).

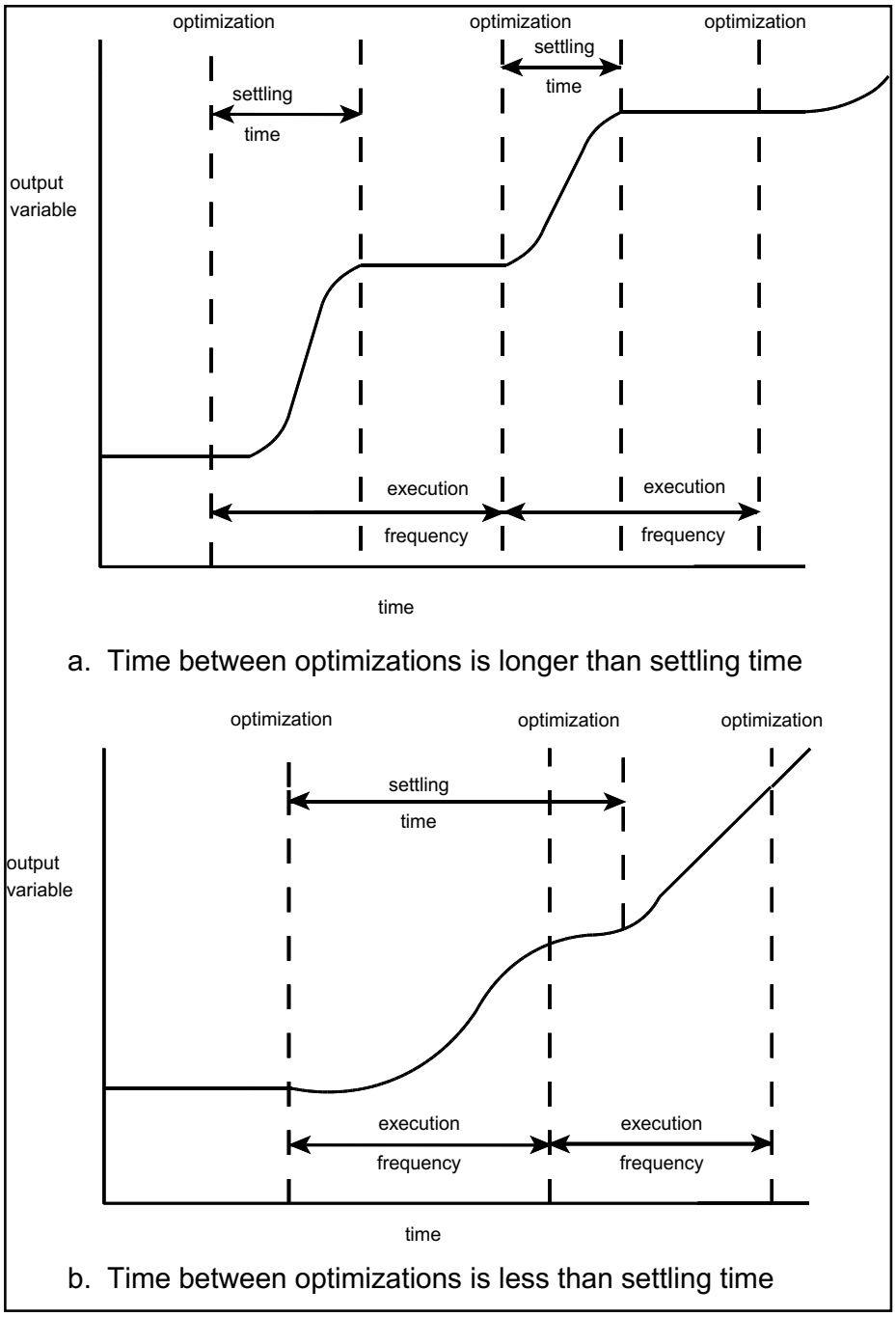


Figure 8-11 Comparison of Time between Optimization and Process Settling Time, after Darby and White (1988)

Steady-State Detection

The process and economic models are based on the plant being at steady-state, and detection of the plant operating conditions is required before on-line optimization can proceed. Several methods have been described that examine the time series of variables from the distributed control system, including statistical process control, analysis of variance, box plots, identifying sample means taken at two consecutive time periods and others. Some of these are described by Brown and Rinehart, 2000

For a steady state plant model, Figure 8-12 describes the implementation procedure for on-line optimization system modified from Kelly, et al., (1996). First, the selected key measurements are examined to test if the process is at steady state. If not, testing of the process is continuing until the process reaches steady state. When the process is at steady state, the plant measurements are extracted from distributed control system and are processed through the data validation step to remove or rectify the gross errors in the measurements. Then the reconciled plant data can be used to estimate the parameters in the plant model. These parameters are usually unmeasurable and time-varying constants, such as catalyst activity, heat exchanger fouling factors, and tray efficiencies of distillation columns. They reflect the equipment conditions that change with time and are relative independent of plant operation conditions. Estimating these parameters on-line has the plant simulation model match the plant operation at the current operating conditions.

The parameters in the economic model include sale prices and demand for products, costs and availability of raw materials, utility cost, etc., which are determined by conditions that are separate from process operations. These parameters have to be adjusted to have an accurate description of the profit. Economic optimization uses the current economic model incorporated with the updated and precise plant model to determine the best operating conditions (e.g. temperatures, pressures, and flow rates) for distributed control system to operate the plant.

After the optimal set points are obtained from economic optimization, the operating state must be examined again to ensure the process is at same steady state. If it is then the setpoints are sent to the distributed control system.

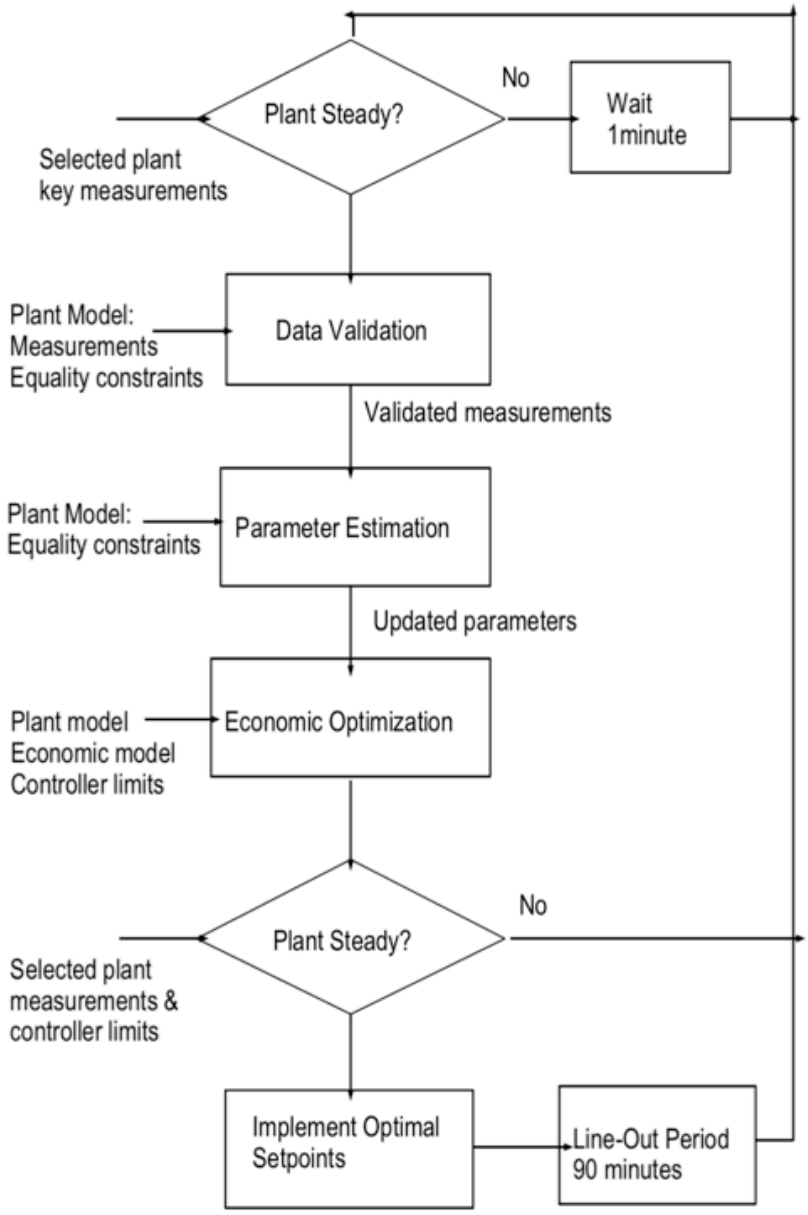


Figure 8-12 Implementation procedures after Kelly, et al., 1996

Summary

On-line optimization adjusts the operation of a plant based on product scheduling and production control to maximize the plant's profit. It provides the means for continuously driving a process toward its optimum operating point. Key benefits for on-line optimization have been a 10% improvement in plant profits, a reduction in energy use and waste generation and an increased understanding of plant operations. The structure of on-line optimization uses components that work together to maximize the profit from the operation of the plant. The components include the plant and economic models, gross error detection, data reconciliation and parameter estimation. Also, an efficient optimization algorithm is used to solve the three nonlinear optimization problems.

On-line optimization programs are available from engineering design and control system companies such as Aspen Technology, Honeywell, and others These companies can provide both the plant's distributed control system and the online optimizer or provide the on-line optimizer for the existing control system. Just providing the on-line optimizer is a multi-million-dollar project and exceed budgets available for on-line optimization in small to moderately sized plants.

Industrial applications have been mainly for crude units and ethylene plants where small improvements in capacity can mean very large increases in profits. Reported improvements in plant operations and economics in a range of 3% to 20%.

On-line optimization takes advantage of the fact that chemical plants operate at steady state with transient periods that are short compared to steady state operations. Consequently, steady-state process models are used to describe the plant. These plant models are complicated and highly nonlinear.

On-line optimization involves solving three nonlinear optimization problems: economic optimization, parameter estimation, and data reconciliation. The plant model serves as the constraint equations in these three nonlinear optimization problems, and the optimization algorithm is used to solve the nonlinear optimization problems. For economic optimization, the plant model is used with the economic model to maximize the plant profit and to provide the optimal set points for the distributed control system to operate the plant. For parameter estimation, parameters in the plant model are estimated by optimizing an objective function, such as minimizing the sum of squares of measurement errors, subject to the constraints in the plant model. For data recon ciliation, the errors in plant measurements are rectified by optimizing a function based on the joint probability distribution function for the plant measurements subject to plant model, and a test statistic is used to detect gross errors in the measurements.

Data reconciliation is a procedure to adjust or reconcile process data and to obtain more accurate values for the measurements by requiring the reconciled data to be consistent with material and energy balances. The data reconciliation problem can be formulated as a constrained optimization problem, e.g., least squares estimation problem if the measurements contain only random errors.

Gross errors in some of the data extracted from the plant's distributed control can have gross errors caused by instrument errors, such as bias, drifting, precision degradation, instrument failure and process leaks.

There are numerous statistical methods for gross error detection, and the most successful method in industrial applications is called the robust function method and is based on robust statistics. These methods require a detail plant model to relate the individual measurement and detect gross errors. They have been found to be very effective for detecting gross errors and usually require solving a nonlinear optimization problem. They use statistical hypothesis testing to determine if a gross error is present, and this requires selecting a statistical test. A gross error is declared if the measurement error exceeds the value specified by the statistical test (the alternative hypothesis H_1 is accepted.) If the measurement error does not exceed this value, the measurement is said to not contain a gross error with a certain probability (the null hypothesis H_0 is accepted).

Combined gross error detection and data reconciliation algorithms can be used to detect and rectify the gross errors in measurements for on-line optimization. These algorithms are the measurement test method using a normal distribution, Tjoa-Biegler's method using a contaminated Gaussian distribution, and a robust statistical method using the Lorentzian robust function

The theoretical performance of four distribution functions: normal distribution of measurement test method, contaminated Gaussian distribution of Tjoa-Biegler's method, Lorentzian distribution and Fair function of robust method, were evaluated based on the influence function and relative efficiency of the distributions. In summary, the evaluation of influence functions of distributions showed that normal distribution causes significant biased estimation if measurements with gross errors were used to reconcile data and the degree of bias increased unboundedly with the increase of errors. Therefore, an iterative elimination strategy was required to avoid the bias whenever a gross error was detected. The comparisons of influence function and relative efficiency showed that both contaminated Gaussian and Lorentzian distributions had a better combination of influence function (gross error sensitivity) and relative efficiency (estimation accuracy). Therefore, they would have a better performance when reconciling data with both random and gross errors.

The results for the influence function for the WLS (measurement test) is proportional to the measurement error justifying the low breakdown point and unbounded effect of large errors. The effect of larger errors is reduced for the r function of the Cauchy distribution, "Lorentzian" function and Hampel's redescending M-estimator and is shown by gradually decreasing influence functions in the region of greater than 3.0 of the standard error. Therefore, these three r functions are called redescending r functions. Fair function and the r function of the Logistic distribution have a bounded influence by the large errors since their influence function increases slowly with respect to the measurement errors approaching a constant value for large errors. The influence of small measurement errors on the r function of the Contaminated Normal distribution is the same as on the WLS function. However, the influence decreases for larger errors and becomes proportional to very large errors after passing through a minimum (at standard error 4.7 in Figure 8-5).

The evaluation of the performance of a total of eight methods is undertaken using five small- scale examples from the literature and two cases involving industrial plants with real process data. The Monte Carlo study shows that the robust approaches for the simultaneous data reconciliation and gross error detection of chemical processes can provide similar or better results compared to a sequential method, with a single (two for Hampel's redescending M-estimator) solution of the NLP.

After data reconciliation and gross error detection random and gross errors have been removed from the data from the distributed control system, and the data satisfies the material and energy balances, the plant model. These evaluations were performed using values of the parameters that were previously determined. In the case where parameters have slowly varying values such as heat transfer coefficients from fouling and catalyst deactivating, the values of these parameters need to be up-dated for the current performance of the process equipment using the reconciled plant data.

In simultaneous data reconciliation and parameter estimation the parameters in plant model are considered as variables along with the measured process variables. Both measured variables and parameters are estimated simultaneously when solving the nonlinear programming problem.

Data reconciliation requires current values of the parameters in the plant model. However, only the parameter values from the previous optimization cycle are available. Therefore, a strategy to avoid this dilemma is to detect and reconcile the measurements containing gross errors using the plant model with the parameter values from previous on-line optimization cycle in gross error detection and data reconciliation. Then a new set of measurements is constructed using the reconciled data to replace the measurements containing gross errors along with the original measurements that contain only random errors. This new set of measurements is supposed only containing random errors, and it can be used to conduct simultaneous data reconciliation and parameter estimation using least squares method with error-in-variables formulation.

In economic optimization, optimization for design and plant operations are different in several ways. The economic model for design is net present value and for operations it is net profit. The process model for operations includes the plant configuration, material and energy balances, availability of raw materials, and demand for products. The process model for design plant does not have a plant configuration, and it has to be determined, along with the capacities of process units. Finally, design optimization determines the capacities of individual units and plant operating conditions. The process model from plant design can be transferred to the simulation of the operation of the plant. Economic data estimated in plant design are replaced by actual data.

In on-line optimization, variables can be measured or unmeasured. Measured variables, data from the distributed control system, can be redundant or nonredundant depending on the number of measured variables. Unmeasured variables can be observable or unobservable depending on the constraint equations.

For a process variable to be reconciled, it must be measured, and its adjusted or reconciled value must appear in a useable material or energy balance constraint in the data reconciliation

problem. Measuring a variable does not ensure that it can be used for data reconciliation (Knopf, 2010). To conduct data reconciliation, redundant measurements are required to reconcile errors in measurements.

After the plant model is completely formulated and the process variables are correctly classified into measured variables (**x**), unmeasured variables (**z**), and parameters (**θ**), the accuracy of the plant model must be examined. To assess precision of the plant model, the simulation results predicted by the plant model must be compared with accurate data from plant. This can be done by designating some of plant design values as measured data, and this data is used to estimate the values of the unmeasured variables and the plant parameters. Then and the estimated parameters and process variables are compared with the plant design data. If the predicted results are very close to the design data with a less than 1% relative deviation, then it is said that the plant model precisely simulates the plant.

The execution frequency of optimization is the time between conducting optimizations of the process, and it has to be determined for each of the units in the process. It depends on the settling time, i.e., the time required for the units in the process to move from one set of steady-state operating condition to another. The settling time can be estimated from the time constant determined by process step testing. The time period between two on-line optimization execution must be longer than the settling time to ensure that the units have returned to steady state operations before the optimization is conducted again.

The process and economic models are based on the plant being at steady-state, and detection of the plant operating conditions is required before on-line optimization can proceed. Several methods have been described that examine the time series of variables from the distributed control system. After the optimal set points are obtained from economic optimization, the operating state must be examined again to ensure the process is at the same steady state. If it is then the setpoints are sent to the distributed control system.

Nomenclature

A a matrix whose elements are the coefficients of linear constraints in Equation 11 of the process model

a a vector of measurement adjustments in Equation 13 that are the differences between the measurements and the reconciled values for measured variables

B a matrix whose elements are the coefficients of the linear constraints for unmeasured variables, Equation 15

b a parameter in contaminated Gaussian distribution function in Equation 20 that represents the ratio of standard deviation of a gross error to one of a random error

C the critical value for a test statistic

c a constant vector that represents the constants in linear constraints of a process model

c a vector in the profit function (Equation 33) in which the elements with respect to the variables of raw materials are the costs of the corresponding raw materials, and other elements are zero

d a vector of measurement errors that is transferred from the vector of measurement errors **e**, i.e., $\mathbf{d} = \mathbf{\Sigma}^{-1}\,\mathbf{e}$ for maximum power test method

e a vector of measurement errors in Equation 2 that are the differences between the measurements and the true values for measured variables

f equality constraints in on-line optimization problems that describe the relation of variables and parameters in a chemical process, such as mass and energy balances

g inequality constraints (Equation 1) in on-line optimization problems that represent the demand of products, the availability of raw materials, the limitation on the capacity of equipment, the allowable operating conditions, and the restrictions on waste and pollutant emission

M number of constraints

N number of measurements

P probability distribution function in Equation 4 for all measurements

p_i a probability distribution function for measurement I in Equation 21

Q variance and covariance matrix of measurement errors in Equation 10

r a vector of constraint residuals in Equation 14

r_j constraint residual for constraint j in Equation 14

S denotes the set of the suspected measurements that contain gross errors for IMT and MIMT methods

w_i weight coefficient of measurement i in the joint probability distribution function in Equation 2

x a vector in Equation 1 that denotes the true values of the measured variables

y a vector of measurements in Equation 1 for measured variables

z a vector in Equation 1 that denotes the unmeasured variables in the process model

Greek

α the overall significant level for all measurements in Equation 17

β the significant level for individual measurement in Equation 17

ε_i standardized measurement error for measurement i, $\varepsilon_i = e_i / \sigma_i$ in Equation 16

η the prior probability of a gross error in contaminated Gaussian distribution in Equation 21

$\boldsymbol{\theta}$ a vector of parameters in a process model in Equation 1

ρ denotes a robust function or algorithm of a probability function p, i.e., $\rho = \ln p$, in Equation 22

$\mathbf{\Sigma}$ summation notation

σ_i^2 variance of measurement error i

σ_{ij}^2 covariance of measurement error i and measurement error j

σ_i standard deviation of measurement error i in Equation 4

Subscripts

i a index representing a measurement in Equation 1

j a index representing a constraint in Equation 1

k a index representing the repeated data in Equation 32

Terminology

Bounds - define the allowable range of process variables. The low and up bounds represent the allowable minimum and maximum operating conditions of the process variables and the raw material availability and product quality requirements.
Closed form sequent modular plant model - follows the traditional design rules, using the information for the input streams of a unit to determine the values of the output variables. Changes of variables in input streams can affect variables in output streams, but the changes of variables in output streams can not affect the determination of process variables in the input streams.
Control variables - are the variables whose values must be satisfied by adjusting the manipulated variables.
Data reconciliation - Data reconciliation is a procedure to adjust or reconcile process data obtained from distributed control system and obtain more accurate values by adjusting the data to be consistent with material and energy balances.
Distribution function - is used to describe the behavior pattern of measurement errors.
Economic model - is the objective function for economic optimization. It is a function that is used to maximize the plant profit; minimize the operation cost, emission or energy consumption; for example.
Economic optimization - is to determine the plant operation conditions that will optimize the economic objective (model) and satisfy the constraints of the plant model.
Equality constraint equations - are mass and energy balances, heat transfer equations, reaction rate equations (kinetic model), thermodynamic equilibrium equations, physical property functions, and others.
GAMS, General Algebraic Modeling System - was developed at the World Bank to solve large and complex mathematical programming models by using a programming language that makes concise algebraic statements of the models and was easily read by both the modeler and the computer (Brook et al., 1988).
Gross error detection - is a statistical procedure to detect and rectify gross errors in plant sample data sampled from distributed control system.
Gross error detection rate - is the ratio of number of gross errors that are correctly detected by the algorithm to the actual number of gross errors in measurements.
Inequality constraint equations - provide additional restrictions for the economic optimization. The inequality constraint equations for a chemical process are the demand for main and by products, availability of raw materials, maximum capacities of the equipment, restriction on the waste/pollutant emission, and others.
Influent function - is proportional to the derivative of the distribution function. It reflects the influence of contaminated measurements on the estimation.
Initial point - the starting values of variables in a optimization problem for the optimization algorithm to search for optimal solution. The default initial point of GAMS is zero or the bound whichever is closer to zero if the bounds are specified to be different from default values.
Key measured variables - are the variables that are directly related to the determination of plant parameters
Measurable variables - are the variables that can be measured by instruments, such as flow rate, temperature, pressure, composition, or other.

Manipulated variables - are the variables that are adjusted to satisfy the requirement on control variables.
Open form equation-based plant model - is written as a set of algebraic and/or differential equations in the form $\mathbf{f}(\mathbf{x}) = 0$. The equations are solved simultaneously for the values of variables, rather than sequentially.
Observability - An unmeasured variable in steady state model is observable if and only if it can be uniquely determined from a set of values for the measured variables, which are consistent with all of the given constraints. Any unmeasured variable which is not so determinable is unobservable (Crowe, 1989).
Optimization algorithm - is a mathematical method to solve an optimization problem, such as simplex method for linear optimization problems and successive linear programming, successive quadratic programming and the generalized reduced gradient method for nonlinear optimization problems.
Parameter estimation - is a statistical procedure to update the values of parameters in the plant model using the plant data reconstructed from the combined gross error detection and data reconciliation.
Plant (simulation) model - is consist of a set of equations that represent the relationship among process variables and describe the process behavior. These include the equality equations (material and energy balances, etc.) and inequality equations (availability of raw materials, demand of products, capacity of equipment, etc.).
Plant parameters - are parameters in plant model that are immeasurable and whose values change slowly with time and are not affected by the changes of operation conditions., e.g., heat exchanger fouling factors, catalyst effectiveness factors, or tray efficiency. These parameters usually describe the condition of process equipment.
Redundancy - A measured quantity is redundant if and only if it would be observable if that quantity was not measured. Otherwise, the measured quantity is non-redundant (Crowe, 1989).
Relative efficiency - represents the asymptotic efficiency of a distribution to normality. It indicates the estimation accuracy for normal measurements.
Relative error reduction - is the ratio of the remaining error after data reconciliation to the original measurement error.
Set points - are the operating points of the controllers in the distributed control system that are adjusted by n-line optimization.
Type I error - is the event that the algorithm has incorrectly identified a normal measurement (no gross error) as an abnormal measurement (measurement containing gross error).
Type II error - is the event that the algorithm has incorrectly identified an abnormal measurement (measurement containing gross error) as normal measurement.
Unmeasured variables - are the variables that are not sampled from plant distributed control system. Their values will be determined by the measured variables through constraint equations.

References

Albuquerque, J.S. and L.T. Biegler, (1995), "Gross Error Detection and Variables Classification in Dynamic Systems," *AIChE Annual Meeting*, Miami Beach.

Almasy, G. A. and R. S. H. Mah, (1984), "Estimation of Measurement Error Variances from Process Data," *Ind. Eng. Chem. Process Des. Dev.*, Vol. 23, No. 4, pp. 779-784.

Almasy, G. A. and T. Sztano, (1975), *Problems of Control and Information Theory*, 4, (1), 57-69.

Anonymous, 1982, *Oil and Gas Journal*, 394, (May, 1982).

Bailey, J. K., A. N. Hrymak, S. S. Treiber and R. B. Hawkins, (1993), "Nonlinear Optimization of Hydrocracker Fraction Plant", *Computers and Chemical Engineering*, Vol. 17, No. 2, p.123-128.

Barlow, R. J., 1989, *Statistics - A Guide to the Use of Statistical Methods in the Physical Sciences*, John Wiley & Sons, New York.

Basta, N., (1996), "Process Simulation: New Mountains to Conquer," *Chemical Engineering*, Vol. 103, No. 5, p. 149-152.

Brooks, A., D. Kendric, A. Meerhaus, and R. Raman, 1998, *GAMS: A Users Guide*, GAMS Development Corporation, Washington, D.C.

Chen, Xueyu, 1998, *The Optimal Implementation of On-Line Optimization for Refinery and Chemical Processes*, Ph.D. dissertation, Louisiana State University, Baton Rouge, Louisiana.

Brown, P. R. and R. R. Rhinehart, 2000, Automated Steady-State Identification in Multivariable Systems, Hydrocarbon Processing, September 2000.

Crowe, C.M., (1989), "Test of Maximum Power for Detection of Gross Errors in Process Constraints," *AIChE Journal*, Vol. 35, No. 5, pp. 869-872.

Crowe, C.M., (1992), "The Maximum Power Test for Gross Errors in the Original Constraints in Data Reconciliation," *The Canadian Journal of Chemical Engineering*, Vol. 70, pp. 1030-1036.

Darby, M. L., and D.C. White, (1988), "On-Line Optimization of Complex Process Units" Chemical *Engineering Progress*, Vol. 84, No. 8 p. 51-59.

Fatora, F.C., G.B. Gochenour, B.G. Houk, and D.N. Kelly, (1992), "Closed-Loop Real-Time Optimization and Control of a World Scale Olefin Plant," *AIChE 1992 Spring National Meeting*, Houston, Texas.

Fatora, F. C. and J. S. Ayala, "Successful Closed Loop Real Time Optimization," reprinted from June 1992 *Hydrocarbon Processing*.

Fourier, R. D. M. Gay and B. W. Kernighan, 1993, *AMPL: A Modeling Language for Mathematical Programming*, The Scientific Press, San Francisco, California.

Gott, J., C. Roubidoux and R. Heersink, (1991), "On-Line Optimization for Smart FCC Controls", *National Petroleum Refineries Association (NPRA) Computer Conference*, Paper No. CC-91-130, Houston, November 11-13.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, (1986), Robust Statistics - the Approach Based on Influence Functions, John Wiley & Son, New York.

Hertwig, T., 1997, Private Communication.

Huber, P.J., (1981), *Robust Statistics*, John Wiley & Sons, New York.

Johnston, L.P.M. and M.A. Kramer, (1995), "Maximum Likelihood Data Rectification: Steady-State Systems," *AIChE Journal*, Vol. 41, No. 11, p. 2415-26.

Kelly, D. N., F. C. Fatora, and S. L. Davenport, (1996), “Implementation of a Closed Loop Real-Time Optimization System on a Large-Scale Ethylene Plant,” Private Communication.
Krist, J.H.A., M.R. Lapere, S. Groot Wassink, R. Neyts, and J.L.A. Koolen, (1994),” Generic System for On-Line Optimization and the Implementation in a Benzene Plant,” *Computers Chem. Engng*, Vol. 18, Suppl., pp. S517-S524.
Larsen, R. J. And M. L. Marx, 1986, *An Introduction to Mathematical Statistics and Its Applications*, Prentice-Hall, New Jersey.
Lauks, U. E., R. J. Vanbinder, P. J. Valkenburg and C. van Leeuwen, (1992), "On-Line Optimization of an Ethylene Plant" *European Symposium on Computer Aided Process Engineering, ESCAPE-1, Supplement to Computers Chem. Engng*., Vol 16, Supp., p. S213--S220.
Madron, Frantisek, 1992, Process Plant Performance: Measurement and Data Processing for Optimization and Retrofits, Ellis Horwood Ltd, Simon and Schuster, England,
Mah, R. S. H., (1990), *Chemical Process Structures and Information Flow*, Butterworth Publishers, Stoneham, MA.
Mah, R.S.H. and A.C. Tamhane, (1982), "Detection of Gross Errors in Process Data," *AIChE J.*, Vol. 28, No. 5, pp. 828-830.
Mah, R.S.H., G.M. Stanley, and D.M. Downing, (1976), “Reconciliation and Rectification of Process Flow and Inventory Data,” *I & EC Proc. Des. Dev*., 15, 175-183.
Moore, R. D. and A. B. Corripio, (1991), "On-Line Optimization of Distillation Columns in Series," *Chem. Eng. Comm*., Vol. 106, p. 71-86.
Narasimhan S., and C. Jordache, 2000, *Data Reconciliation and Gross Error Detection*, Gulf Publishing Company, Houston, Texas.
Narasimhan, S. and R. S. H. Mah, (1987), "Generalized Likelihood Ratio Method for Gross Errors Identification," *AIChE J.*, Vol. 33, No. 9, pp.1514-1521.
Ozyurt, D., R. W. Pike, F. C. Knopf, M. K. Rich, J. R. Hopper and C. L. Yaws, 2003, “Integrated Approach to Unit Optimization,” *Petroleum Technology Quarterly*, Vol. 8 No. 5, p. 47-51 (Autumn, 2003).
Ozyurt, D. And R. W. Pike, 2004, “Theory and Practice of Simultaneous Data Reconciliation and Gross Error Detection for Chemical Processes,” *Computers & Chemical Engineering*, Vol. 28, No. 1, p. 381-402 (2004).
Pike, R. W., (1986), *Optimization for Engineering Systems*, Van Nostrand Reinhold Publishers, New York.
Pike, R. W., 2015, Essentials of Economic Decision Analysis for Chemical Engineering, Title ID: 5287746 ISBN-13: 978-1507771723, Createspace, Amazon.com (2015)
Rhemann, H., G. Schwarz, T.A. Badgwell, M.L. Darby, and D.C. White, (1989), “On-Line FCCU Advanced Control and Optimization,” *Hydrocarbon Processing*, No. 6, p. 64-71.
Rollins, D.K., and J.F. Davis, (1992), "Unbiased Estimation of Gross Error in Process Measurements," *AIChE J.*, Vol. 38, No. 4, pp. 563-572.
Romagnoli, J. A. and M. C. Sanchez, (2000), *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press, New York
Saha, L. E., A. J. Chontos and D. R. Hatch, (1990), "Optimization at Wyoming Gas Plant Improves Profitability," *Oil and Gas Journal*, No. 5, p. 49-60.
Sanders, Fred F., (1995), “Watch out for Instrument Errors,” *Chemical Engineering Progress*, No. 7, p 62-66.

Serth, R. W. and W. A. Heenan, 1986, Gross Error Detection and Data Reconciliation in Steam-Metering Systems, AIChE Journal, Vol. 32, No. 5, May 1986
Scott, M.D., S.L. Mullick, and J.M. Thiessen, (1995), “Rigorous On-Line Model for Optimization of a Multi-Unit Hydrotreater-Reformer Complex,” *AIChE 1995 Spring National Meeting, First International Plant Operations and Design Conference.*
Scott, M.D., J.M. Thiessen, and S.L. Mullick, (1994), “Reactor Integrated Rigorous On-Line Model (ROM) for a Multi-Unit Hydrotreater-Catalytic Reformer Complex Optimization,” *National Petroleum refineries Association (NPRA) Computer Conference*, Anaheim.
Sourander, M. L., M.Kolari, J. C. Cugini, J. B. Poje and D. C. White, (1984), "Control and Optimization of Olefin-Cracking Heaters" *Hydrocarbon Processing*, No. 6, p. 63-69.
Tamhane, A.C., (1982), "A Note on the Use of Residuals for Detecting an Outlier in Linear Regression," *Biometrka*, Vol. 69, pp.488.
Tjoa, I. B. and L. T. Biegler, (1991), "Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems," *Computers Chem. Engng*., Vol. 15, No. 10, p.679-90.
Tong, H. and C.M. Crowe, (1994), "The Application of Principal Component Analysis to Tests for Gross Errors in Data Reconciliation," PSE'94, Kyongju, Korea.
Tong, H. and C.M. Crowe, (1995), “Detection of Gross Errors in Data Reconciliation by Principal Component Analysis,” *AIChE Journal*, Vol. 41, No. 7, pp. 1712-1722.
Van Wijk, R. A. and M. R. Pope, (1992), "Advanced Process Control and On-Line Optimization in Shell Refineries’," *European Symposium on Computer Aided Process Engineering, ESCAPE-1, Supplement to Computers Chem. Engng*., Vol 16, Supp., p. S69-S80.
Veverka, V. And F. Madron, 1997, *Material and Energy Balancing in the Process Industries, Elsevier*, Science B.V., Amsterdam, The Netherlands.
Willsky, A.S., and H.L. Jones, (1974), “A Generalized Likelihood Ratio Approach to State Estimation in Linear System Subject to Abrupt Changes,” *Proc. IEEE Conf. Decision and Control*, pp.846-853.
Xu, S., M. Baldea, T. F. Edgar, W. Wojsniz, T. Bivins and M. Nixon, 2015, “An Improved Methodology for Outlier Detection in Dynamic Datasets,” AIChE Journal, Vol. 61, No. 2, 2015.
Yamamura, K, M. Nakajima, and H. Matsuyama, (1988), "Detection of Gross Errors in Process Data Using Mass and Energy Balances," *International Chemical Engineering*, Vol. 28, No. 1, pp.91-98.
Zhang, Zhengjiang and Chen Junghui, 2015, “Correntropy Based Data Reconciliation and Gross Error Detection and Identification for Nonlinear Dynamic Processes,” Computers and Chemical Engineering 75 (2015) 120–134