3
Process Monitoring and Data Reconciliation

Georges Heyen and Boris Kalitventzeff

3.1
Introduction

Measurements are needed to monitor process efficiency and equipment condition, but also to take care that operating conditions remain within an acceptable range to ensure good product quality and to avoid equipment failure and any hazardous conditions. Recent progress in automatic data collection and archiving has solved part of the problem, at least for modern, well-instrumented plants. Operators are now faced with a lot of data, but they have little means to extract and fully exploit the relevant information it contains.

Furthermore, plant operators recognize that measurements and laboratory analysis are never error-free. Using these measurements without any correction yields inconsistencies when generating plant balances or estimating performance indicators. Even careful installation and maintenance of the hardware can not completely eliminate this problem.

Model-based statistical methods, such as data reconciliation, have been developed to analyze and validate plant measurements. The objective of these techniques is to remove errors from available measurements and to yield complete estimates of all the process state variables as well as of unmeasured process parameters.

This chapter constitutes a tutorial on process monitoring and data reconciliation. First, the key concepts and issues underlying a plant data validation, sources of error and redundancy considerations are introduced. Then, the data reconciliation problem is formulated for simple stready-state linear systems and extended further to consider nonlinear cases. The role of sensibility analysis is also introduced. Dynamic data reconciliation, which is still a subject of major research interest, is treated next. The chapter concludes with a section devoted to the optimal design of the measurement system. Detailed algorithms and supporting software are presented along with the solution of some motivating examples.

Computer Aided Process and Product Engineering. Edited by Luis Puigjaner and Georges Heyen

ISBN: 3-527-30804-0

3.2
Introductory Concepts for Validation of Plant Data

Data validation makes use of a plant model in order to identify measurement errors and to reduce their average magnitude. It provides estimates of all process state variables, whether directly measured or not, with the lowest possible uncertainty. It allows one to assess the value of key performance indicators, which are target values for process operation, or is used as a soft sensor to provide estimates of some unmeasured variables, as in inferential control applications.

Especially in a framework of real-time optimal control, where model fidelity is of paramount importance, data validation is a recommended step before fine-tuning model parameters: there is no incentive in seeking to optimize a model when it does not match the actual behavior of the real plant.

Data validation can also help in gross error detection, meaning either process faults (such as leaks) or instrument faults (such as identification of instrument bias and automatic instrument recalibration).

Long an academic research topic, data validation is currently attracting more and more interest, since the amount of measured data collected by Digital Control Systems (DCS) and archived in process information management systems,exceeds what can be handled by operators and plant managers. Real-time applications, such as optimal control, also require frequent parameter updates, in order to ensure fidelity of the plant model. The economic value of extracting consistent information from raw data is recognized. Data validation thus plays a key role in providing coherent and error-free information to decision makers.

3.2.1
Sources of Error

Some sources of errors in the balances depend on the sensors themselves:

- Intrinsic sensor precision is limited, especially for online equipment, where robustness is usually considered more important than accuracy.
- Sensor calibration is seldom performed as often as desired, since this is a costly and time-consuming procedure requiring competent manpower.
- Signal converters and transmission add noise to the original measurement.
- Synchronization of measurements may also pose a problem, especially for chemical analysis, where a significant delay exists between sampling and result availability.

Other errors arise from the sensor location or the influence of external effects. For instance, the measurement of gas temperature at the exit of a furnace can be influenced by radiation from the hot wall in the furnace. Inhomogeneous flow can also cause sampling problems. A local measurement is not representative of an average bulk property.

A second source of error when calculating plant balances is the small instabilities of the plant operation and the fact that samples and measurements are not taken at

exactly the same time. Using time averages for plant data partly reduces this problem.

3.2.2 Redundancy

Besides safety considerations, the ultimate goal in performing measurements is to assess the plant performance and to take actions in order to optimize the operating conditions. However, most performance indicators can not be directly measured and must be inferred from some measurements using a model. For instance, the extent of a reaction in a continuous reactor can be calculated from a flow rate and two composition measurements. In general terms, model equations that relate unmeasured variables to a sufficient number of available measurements are used.

However, in some cases, more measurements are available than are strictly needed, and the same performance indicator can be calculated in several ways using different subsets of measurements. For instance, the conversion in an adiabatic reactor where a single reaction takes place is directly related to the temperature variation. Thus the extent of the reaction can be inferred from a flow rate and two temperature measurements using the energy balance equation. In practice, all estimates of performance indicators will be different, which makes life difficult and can lead to endless discussions about "best practice."

Measurement redundancy should not be viewed as a source of trouble, but as an opportunity to perform extensive checking. When redundant measurements are available, they allow one not only to detect and quantify errors, but also to reduce the uncertainty using procedures known as data validation.

3.2.3 Data Validation

The data validation procedure comprises several steps. The first is the measurement collection. Nowadays, in well-instrumented plants, this is performed routinely by automated equipment.

The second step is conditioning and filtering: not all measurements are available simultaneously, and synchronization might be required. Some data are acquired at higher frequency and filtering or averaging can be justified.

The third step is to verify the process condition and the adequacy of the model. For instance, if a steady-state model is to be used for data reconciliation, the time series of raw measurements should be analyzed to detect any significant transient behavior.

The fourth step is gross error detection: the data reconciliation procedure to be applied later is meant to correct small random errors. Thus, large systematic errors that could result from complete sensor failure should be detected first. This is usually done by verifying that all raw data remain within the upper and lower bounds.

More advanced statistical techniques, such as principal component analysis (PCA), can also be applied at this stage. *Ad hoc* procedures are applied in case some measured value is found inadequate or missing: it can be replaced by a default value or by the previous one available.

The fifth step checks the feasibility of data reconciliation. The model equations are analyzed and the variables are sorted. Measured variables are redundant (and can thus be validated) or just determined; unmeasured variables are determinable or not. When all variables are either measured or observable, the data reconciliation problem can be solved to provide an estimate for all state variables.

The sixth step is the solution of the data reconciliation problem. The mathematical formulation of this problem will be presented in more detail later.

Each measurement is corrected as slightly as possible in such a way that the corrected measurements match all the constraints (or balances) of the process model. Unmeasured variables can be calculated from reconciled values using some model equations.

In the seventh step the systems perform a result analysis. The magnitude of the correction for each measurement is compared to its standard deviation. Large corrections are flagged as suspected gross errors.

In the final step, results are edited and may be archived in the plant information management system. Customized reports can be edited and forwarded to various users (e.g., list of suspect sensors sent to maintenance, performance indicators sent to the operators, daily balance and validated environmental figures to site management).

3.3 Formulation

Data reconciliation is based on measurement redundancy. This concept is not limited to the case where the same variable is measured simultaneously by several sensors. It is generalized with the concept of spatial redundancy, where a single variable can be estimated in several independent ways from separate sets of measurements. For instance, the outlet of a mixer can be directly measured or estimated by summing the measurements of all inlet flow rates. For dynamic systems, temporal redundancy is also available, by which repeated observations of the same variables are obtained. More generally, plant structure is additional information that can be exploited to correct measurements.

Variables describing the state of a process are related by some constraints. The basic laws of nature must be verified: mass balance, energy balance, some equilibrium constraints. Data reconciliation uses information redundancy and conservation laws to correct measurements and convert them into accurate and reliable knowledge.

Kuehn and Davidson (1961) were the first to explore the problem of data reconciliation in the process industry. Vaclavek (1968, 1969) also addressed the problem of variable classification, and the formulation of the reconciliation model. Mah et al.

(1976) proposed a variable classification procedure based on graph theory, while Crowe (1989) based an analysis on a projection matrix approach to obtain a reduced system. Joris and Kalitventzeff (1987) proposed a classification algorithm for general nonlinear equation systems, comprising mass and energy balances, phase equilibrium and nonlinear link equations. A thorough review of classification methods is available in Veverka and Madron (1996) and in Romagnoli and Sanchez (2000). A historical perspective of the main contributions on data reconciliation can also be found in Narasimhan and Jordache (2000).

3.3.1
Steady-State Linear System

The simplest data reconciliation problem deals with steady state mass balances, assuming all variables are measured, and results in a linear problem. In this case $\mathbf{x}$ is the vector of n state variables, while $\mathbf{y}$ is the vector of measurements. We assume that random errors $\mathbf{e}=\mathbf{y}-\mathbf{x}$ follow a multivariate normal distribution with zero mean.

The state variables are linked by a set of m linear constraints:

$$\mathbf{A}\mathbf{x} - \mathbf{d} = 0 \tag{1}$$

The data reconciliation problem consists of identifying the state variables $\mathbf{x}$ that verify the set of constraints and are close to the measured values in the least square sense, which results in the following objective function:

$$\min_{x}(\mathbf{y} - \mathbf{x})^{\mathrm{T}}\mathbf{W}(\mathbf{y} - \mathbf{x}) \tag{2}$$

where $\mathbf{W}$ is a weight matrix.

The method of Lagrange multipliers allows one to obtain an analytical solution:

$$\hat{\mathbf{x}} = \mathbf{y} - \mathbf{W}^{-1}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{W}^{-1}\mathbf{A}^{\mathrm{T}}\right)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{d}) \tag{3a}$$

It is assumed that there are no linearly dependent constraints.

In order to solve practical problems and obtain physically meaningful solutions, it may be necessary to take into account inequality constraints on some variables (e.g., flow rate should be positive). However, this makes the solution more complex, and the constrained problem can not be solved analytically.

It can be shown that $\hat{\mathbf{x}}$ is the maximum likelihood estimate of the state variables if the measurement errors are normally distributed with zero mean, and if the weight matrix $\mathbf{W}$ corresponds to the inverse of the error covariance matrix $\mathbf{C}$. Equation (3) then becomes:

$$\hat{\mathbf{x}} = \mathbf{y} - \mathbf{C}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{C}\mathbf{A}^{\mathrm{T}}\right)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{d}) = \left[\mathbf{I} - \mathbf{C}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{C}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{A}\right]\mathbf{y} + \mathbf{C}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{C}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{d} \tag{3b}$$

$$\hat{\mathbf{x}} = \mathbf{M}\mathbf{y} + \mathbf{e}$$

The estimates are thus related to the measured values by a linear transformation. They are therefore normally distributed with the average value and covariance matrix obtained by calculating the expected values:

$$E(\hat{\mathbf{x}}) = \mathbf{M}E(\mathbf{y}) = \mathbf{x}$$
$$\mathrm{Cov}(\hat{\mathbf{x}}) = E\left[(\mathbf{My})(\mathbf{My})^{\mathrm{T}}\right] = \mathbf{MCM}^{\mathrm{T}} \tag{4}$$

This shows that the estimated state variables are unbiased. Furthermore, the accuracy of the estimates can easily be obtained from the measurement accuracy (covariance matrix **C**) and from the model equations (matrix **A**).

3.3.2 Steady-State Nonlinear System

The data reconciliation problem can be extended to nonlinear steady-state models and to cases where some variables **z** are not measured. This is expressed by:

$$\min_{\mathbf{x},\mathbf{z}} (\mathbf{y}-\mathbf{x})^{\mathrm{T}}\mathbf{W}(\mathbf{y}-\mathbf{x})$$
$$\text{s.t. } \mathbf{f}(\mathbf{x},\mathbf{z}) = 0 \tag{5}$$

where the model equations are mass and component balance equations, energy balance, equilibrium conditions, and link equations relating measured values to state variables (e.g., conversion from mass fractions to partial molar flow rates).

Usually the use of performance equations is not recommended, unless the performance parameters (such as compressor efficiency and overall heat transfer coefficients or fouling factors for heat exchangers) remain unmeasured and will thus be estimated by solving the data reconciliation problem. It would be difficult to justify correcting measurements using an empirical correlation, e.g., by correcting the outlet temperatures of a compressor by enforcing the value of the isentropic efficiency. The main purpose of data reconciliation is to allow monitoring of those efficiency parameters and to detect their degradation.

Equation (5) takes the form of a nonlinear constrained minimization problem. It can be transformed into an unconstrained problem using Lagrange multipliers $\boldsymbol{\Lambda}$ and the augmented objective function L has to be minimized:

$$L(\mathbf{x},\mathbf{z},\Lambda) = \left\{\frac{1}{2}(\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{y}) + \Lambda^{\mathrm{T}}\cdot\mathbf{f}(\mathbf{x},\mathbf{z})\right\}$$
$$\min_{\mathbf{x},\mathbf{y},\Lambda} L(\mathbf{x},\mathbf{y},\Lambda) \tag{6}$$

The solution must verify the necessary optimality conditions i.e., the first derivatives of the objective function with respect to all independent variables must vanish. Thus one has to solve the system of normal equations:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{C}^{-1}(\mathbf{x}-\mathbf{y}) + \mathbf{A}^{\mathrm{T}}\cdot\boldsymbol{\Lambda} = 0$$
$$\frac{\partial L}{\partial \mathbf{z}} = \mathbf{B}^{\mathrm{T}}\cdot\boldsymbol{\Lambda} = 0 \tag{6}$$
$$\frac{\partial L}{\partial \boldsymbol{\Lambda}} = \mathbf{f}(\mathbf{x},\mathbf{z}) = 0$$

This last equation can be linearized as:

$$\frac{\partial L}{\partial \Lambda} = \mathbf{A} \cdot \mathbf{x} + \mathbf{B} \cdot \mathbf{z} + \mathbf{d} = 0 \tag{7}$$

where **A** and **B** are partial Jacobian matrices of the model equation system:

$$\begin{aligned} \mathbf{A} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \\ \mathbf{B} &= \frac{\partial \mathbf{f}}{\partial \mathbf{z}} \end{aligned} \tag{8}$$

The system of normal equations in Eq. (6) is nonlinear and has to be solved iteratively. Initial guesses for measured values are straightforward to obtain. Process knowledge usually estimates good initial values for unmeasured variables. No obvious initial values exist for Lagrange multipliers, but solution algorithms are not too demanding in that respect (Kalitventzeff et al., 1978). The Newton-Raphson method is suitable for small problems and requires a solution of successive linearizations of the original problem Eq. (6):

$$\begin{vmatrix} \mathbf{X} \\ \mathbf{Z} \\ \Lambda \end{vmatrix} = \mathbf{J}^{-1} \begin{vmatrix} \mathbf{C}^{-1}\mathbf{Y} \\ 0 \\ -\mathbf{d} \end{vmatrix} \tag{9}$$

where the Jacobian matrix **J** of the equation system has the following structure:

$$\mathbf{J} = \begin{vmatrix} \mathbf{C}^{-1} & 0 & \mathbf{A}^{\mathrm{T}} \\ 0 & 0 & \mathbf{B}^{\mathrm{T}} \\ \mathbf{A} & \mathbf{B} & 0 \end{vmatrix} \tag{10}$$

Numerical algorithms embedding a step size control, such as Powell's dogleg algorithm (Chen and Stadtherr 1981) are quite successful for larger problems.

When solving very large problems, it is necessary to exploit the sparsity of the Jacobian matrix and use appropriate solution algorithms, such as those described by Chen and Stadtherr (1984a). It is common to assume that measurements are independent, which reduces the weight matrix $\mathbf{C}^{-1}$ to a diagonal. Ideally, the elements of matrices **A** and **B** should be evaluated analytically. This is straightforward for the elements corresponding to mass balance equations, which are linear, but can be difficult when the equations involve physical properties obtained from an independent physical property package.

The solution strategy exposed above does not allow one to handle inequality constraints. This justifies the use of alternative algorithms to solve directly the nonlinear programming (NLP) problem defined by Eq. (6). Sequential quadratic programming (SQP) is the method of choice (Chen and Stadtherr 1984a; Kyriakopoulou and Kalitventzeff 1996, 1997). At each iteration, an approximation of the original problem is solved: the original objective function being quadratic is retained and the model constraints are linearized around the current estimate of the solution.

Before solving the NLP problem, some variable classification and preanalysis is needed to identify unobservable variables, parameters, and nonredundant measurements. Measured variables can be classified as *redundant* (if the measurement is absent or detected as a gross error, the variable can still be estimated from the model) or *nonredundant*. Likewise, unmeasured variables are classified as *observable* (estimated uniquely from the model) or *unobservable*. The reconciliation algorithm will correct only redundant variables. If some variables are not observable, the program will either request additional measurements (and possibly suggest a feasible set) or solve a smaller subproblem involving only observable variables. The preliminary analysis should also detect *overspecified variables* (particularly those set to constants) and *trivial redundancy*, where the measured variable does not depend at all upon its measured value but is inferred directly from the model. Finally, it should also identify model equations that do not influence the reconciliation, but are merely used to calculate some unmeasured variables. Such preliminary tests are extremely important, especially when the data reconciliation runs as an automated process. In particular, if some measurements are eliminated as gross errors due to sensor failure, nonredundant measurements can lead to unobservable values and nonunique solutions, rendering the estimates and fitted values useless. As a result, these cases need to be detected in advance through variable classification. Moreover, under these conditions, the NLP may be harder to converge.

3.3.3 Sensitivity Analysis

Solving the data reconciliation problem provides more than validated measurements. A sensitivity analysis can also be carried out. It is based on the linearization of equation system in Eq. (9), possibly augmented to take into account active inequality constraints.

Equation (9) shows that reconciled values of process variables $\mathbf{x}$ and $\mathbf{z}$, and of Lagrange multipliers $\boldsymbol{\Lambda}$ are linear combinations of the measurements. Thus their covariance matrix is directly derived from the measurements covariance matrix (Heyen et al. 1996).

Knowing the variance of validated variables allows one to detect the respective importance of all measurements in the state identification problem. In particular, some measurements might appear to have little effect on the result and might thus be discarded from analysis. Some measurements may appear to have a very high impact on key validated variables and on their variance: these measurements should be carried out with special caution, and it may prove wise to duplicate the corresponding sensors.

The standard deviation of validated values can be compared to the standard deviation of the raw measurement. Their ratio measures the improvement in confidence brought by the validation procedure. A nonredundant measurement will not be improved by validation. The reliability of the estimates for unmeasured observable variables is also quantified.

The sensitivity analysis also allows one to identify all state variables dependent on a given measurement, as well as the contribution of the measurement variance to the variance of the reconciled value. This information helps locate critical sensors, whose failure may lead to troubles in monitoring the process.

A similar analysis can be carried out for all state variables, whether measured or not. For each variable, a list of all measurements used to estimate its reconciled value is obtained. The standard deviation of the reconciled variable is calculated, but also its sensitivity with respect to the measurement's standard deviation. This allows one to locate sensors whose accuracy should be improved in order to reduce the uncertainty affecting the major process performance indicators.

3.3.4
Dynamic Data Reconciliation

The algorithm described above is suitable for analyzing steady-state processes. In practice it is also used to handle measurements obtained from processes operated close to steady state, with small disturbances. Measurements are collected over a period of time and average values are treated with the steady state algorithm. This approach is acceptable when the goal is to monitor some performance parameters that vary slowly with time, such as the fouling coefficient in heat exchangers. It is also useful when validated data are needed, to fine tune a steady-state simulation model, e.g., before optimizing set point values that are updated once every few hours.

However, a different approach is required when the transient behavior needs to be monitored accurately. This is the case for regulatory control applications, where data validation has to treat data obtained with a much shorter sampling interval. Dynamic material and energy balance relationships must then be considered as a constraint.

The earliest algorithm was proposed by Kalman (1960) for the linear time-invariant system model.

The general nonlinear process model describes the evolution of the state variables $\mathbf{x}$ by a set of ordinary differential equations (ODE):

$$\dot{\mathbf{x}} = f(t, \mathbf{x}, \mathbf{u}) + w(t) \tag{11}$$

where $\mathbf{x}$ are state variables, $\mathbf{u}$ are process inputs, and $w(t)$ is white noise with zero mean and covariance matrix $\mathbf{R}(t)$.

To model the measurement process, one usually considers sampling at discrete times $t = k\mathrm{T}$, and measurements related to state variables by:

$$\mathbf{y}_k = h(\mathbf{x}_k) + v_k \tag{12}$$

where measurement errors are normally distributed random variables with zero mean and covariance matrix $\mathbf{Q}_k$. One usually considers that process noise w and measurement noise v are not correlated.

By linearizing Eqs. 11 and 12 at each time step around the current state estimates, an *extended Kalman filter* can be built (see, for instance, Narasimhan and Jordache

2000). It allows one to propagate an initial estimate of the states and the associated error covariance, and to update them at discrete time intervals using the measurement innovation (the difference between the measured values and the predictions obtained by integrating the process model from the previous time step).

An alternative approach relies on NLP techniques. As proposed by Liebman et al. (1992), the problem can be formulated as

$$\min_{\mathbf{x}} \frac{1}{2} \sum_{j=t_0}^{t_N} \left(\mathbf{y}_j - \mathbf{x}(\mathbf{t}_j)\right)^{\mathrm{T}} \mathbf{Q}_j^{-1} \left(\mathbf{y}_j - \mathbf{x}(t_j)\right) \tag{13}$$

subject to

$$\mathbf{f}\left(\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t}, \mathbf{x}(t)\right) = 0; \quad \mathbf{x}(t_0) = \hat{\mathbf{x}}_0 \tag{14}$$

$$h\left(\mathbf{x}(t)\right) = 0 \tag{15}$$

$$g\left(\mathbf{x}(t)\right) \leq 0 \tag{16}$$

In this formulation, we expect that all state variables can be measured. When some measurements are not available, this can be handled by introducing null elements in the weight matrix **Q**. Besides enforcing process specific constraints, the equalities in Eq. (15) can also be used to define nonlinear relationships between state variables and some measurements.

All measurements pertaining to a given time horizon $[t_0 \ldots t_N]$ are reconciled simultaneously. Obviously, the calculation effort increases with the length of the time horizon, and thus with the number of measurements. A tradeoff exists between calculation effort and data consistency. If measurements are repeated N times in the horizon interval, each measured value will be reconciled N times with different neighboring measurements, as long as it is part of the moving horizon. Which set of reconciled values is the "best" and should be considered for archiving is an open question. The value corresponding to the latest time t_N will probably be selected for online control application, while a value taken in the middle of the time window might be preferred for archiving or offline calculations.

Two solution strategies can be considered. The sequential solution and optimization combines an optimization algorithm such as SQP with an ODE solver. Optimization variables are the initial conditions for the ODE system. Each time the optimizer sets a new value for the optimization variables, the differential equations are solved numerically and the objective function Eq. (13) is evaluated. This method is straightforward, but not very efficient: accurate solutions of the ODE system are required repeatedly and handling the constraints Eqs. (15) and (16) might require a lot of trial and error. An implementation of this approach in a MATLAB environment is described by Romagnoli and Sanchez (2000).

Simultaneous solution and optimization is considered more efficient. The differential constraints are approximated by a set of algebraic equations using a weighted residuals method, such as orthogonal collocation. Predicted values of the state vari-

ables are thus obtained by solving the resulting set of algebraic equations, supplemented by the algebraic constraints of Eqs. (15) and (16). With this transformation, the distinction between dynamic data reconciliation and steady state data reconciliation vanishes. However this formulation requires solving a large NLP problem. This approach was first proposed by Liebman et al. (1992).

3.4 Software Solution

Data reconciliation is a functionality that is now embedded in many process analysis and simulation packages or is proposed as a standalone software solution. Bagajewicz and Rollins (2002) present a review of eight commercial and one academic data reconciliation packages. Most of them are limited to material and component balances.

More advanced features are only available in a few packages: direct connectivity to DCS systems for online applications, access to an extensive physical property library, handling pseudocomponents (petroleum fractions), simultaneous data validation, and identification of process performance indicators, sensitivity analysis, automatic gross error detection and correction, a model library for major process unit modules, handling of rigorous energy balances and phase equilibrium constraints, evaluation of confidence limits for all estimates. The packages offering the larger sets of features are Datacon (Invensys) (2004) and Vali (Belsim) (2004).

Dynamic data reconciliation is still an active research topic (Binder et al., 1998). It is used in combination with some real-time optimization applications, usually in the form of custom-developed extended Kalman Filters (see, for instance, Musch et al. (2004)), but dedicated commercial packages have yet to reach the market.

3.5 Integration in the Process Decision Chain

Data reconciliation is just one step – although an important step – in the data processing chain. Several operations, collectively known as data validation, are executed sequentially:

- Raw measurements are filtered to eliminate some random noise. When data is collected at high frequency, a moving average might be calculated to reduce the signal variance.
- If steady state data reconciliation is foreseen, the steady state has to be detected.
- Measurements are screened in order to detect outliers, or truly abnormal values (out of feasible range, e.g., negative flow rate).
- The state of the process might be identified when the plant can operate in different regimes or with a different set of operating units. Principal Component Analysis (PCA) analysis is typically used for that purpose, and allows one to select a reference case and to assign the right model structure to the available data set. This

step also allows some gross error detection (measurement set deviates significantly from all characterized normal sets).

- Variable classification takes place in order to verify that redundancy is present in the data set and that all state variables can be observed.
- The data reconciliation problem is solved.
- A global Chi-square test can detect the presence of gross errors.
- *A posteriori* uncertainty is calculated for all variables, and corrections are compared to the measurement standard deviation. In an attempt to identify gross errors, sequential elimination of suspect measurements (those with large corrections) can possibly identify suspect sensors. Alternatively, looking at subsystems of equations linking variables with large corrections allows one to pinpoint suspect units or operations in the plant.
- Key performance indicators and their confidence limits are evaluated and made available for reporting.
- Model parameters are tuned based on reconciled measurements and made available to process optimizers.

3.6 Optimal Design of Measurement System

The quality of validated data obviously depends on the quality of the measurement. Recent studies have paid more attention to this topic. The goal is to design measurement systems allowing one to achieve a prescribed accuracy in the estimates of some key process parameters, and to secure enough redundancy to make the monitoring process resilient with respect to sensor failures. Some preliminary results have been published, but no general solution can be found addressing large-scale nonlinear systems or dynamics.

Madron (1972) solved the linear mass balance case using a graph-oriented method. Meyer et al. (1994) proposed an alternative minimum-cost design method based on a similar approach. Bagajewicz (1997) analyzed the problem for mass balance networks, where all constraint equations are linear. Bagajewicz and Sanchez (1999) also analyze reallocation of existing sensors. The design and retrofit of a sensor network was also analyzed by Benqlilou et al. (2004) who discussed both the strategy and tools structure.

3.6.1 Sensor Placement based on Genetic Algorithm

A model-based sensor location tool, making use of a genetic algorithm to minimize the investment cost of the measurement system has been proposed by Heyen et al. (2002) and further developed by Gerkens and Heyen (2004).

They propose a general mathematical formulation of the sensor selection and location problem in order to reduce the cost of the measurement system while providing

estimates of all specified key process parameters within a prescribed accuracy. The goal is to extend the capability of previously published algorithms and to address a broader problem, not being restricted to flow measurements and linear constraints.

The set of constraint equations is obtained by linearizing the process model at the nominal operating conditions, assuming steady state. The process model is complemented with link equations that relate the state variables to any accepted measurements, or to key process parameters whose values should be estimated from the set of measurements. In our case, the set of state variables for process streams comprises all stream temperatures, pressures and partial molar flow rates. In order to handle total flow rate measurements, the link equation describing the mass flow rate as the sum of all partial molar flow rates weighted by the component's molar mass has to be defined. Similarly, link equations relating the molar or mass fractions to the partial molar flow rates have also to be added for any stream where an analytical sensor can be located.

Link equations also have also to be added to express key process parameters, such as heat transfer coefficients, reaction extents or compressor efficiencies.

In the optimization problem formulation, the major contribution to the objective function is the annualized operating cost of the measurement system. In the proposed approach, we will assume that all variables are measured; those that are actually unmeasured will be handled as measured variables with a large standard deviation. Data reconciliation requires a solution of the optimization problem described by Eq. (5). The weight matrix $\mathbf{W} = \mathbf{C}^{-1}$ is limited to diagonal terms, which are the inverse of the measurement variance. The constrained problem is transformed into an unconstrained one using the Lagrange formulation as previously shown.

Assuming all state variables are measured, the solution takes the following form:

$$\begin{bmatrix} \mathbf{X} \\ \boldsymbol{\Lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{W} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{A} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{WY} \\ -\mathbf{d} \end{bmatrix} = \mathbf{M}^{-1} \begin{bmatrix} \mathbf{C}^{-1}\mathbf{Y} \\ -\mathbf{d} \end{bmatrix} \tag{17}$$

The linear approximation of the constraints is easily obtained from the solution of the nonlinear model, since **A** is the Jacobian matrix of the nonlinear model evaluated at the solution.

Thus matrix **M** can be easily built, knowing the variance of measured variables appearing in submatrix **W** and the model Jacobian matrix **A** (which is constant). This matrix will be modified when assigning sensors to variables. Any diagonal element of matrix **W** will remain zero (corresponding to infinite variance) as long as a sensor is not assigned to the corresponding process variable; it will be computed from the sensor precision and the variable value when a sensor is assigned in Section 3.6.2.3. Equation (17) need not be solved, since measured values Y are not known. However the variances of the reconciled values X depend only on the variance of measurements as shown in Heyen et al. (1996):

$$\mathrm{var}(X_i) = \sum_{j=1}^{m} \frac{([\mathbf{M}^{-1}]_{ij})^2}{\mathrm{var}(Y_j)} \tag{18}$$

The elements of $\mathbf{M}^{-1}$ are obtained by calculating a lower and upper triangular (LU) factors of matrix **M**. In the case when matrix **M** is singular, we can conclude that the measurement set has to be rejected, since it does not allow observation of all variables. Row i of $\mathbf{M}^{-1}$ is obtained by back substitution using the LU factors, using a right-hand-side vector whose components are δ_{ij} (Kronecker factor: $\delta_{ij} = 1$ when $i = j$, $\delta_{ij} = 0$ otherwise).

In the summation of Eq. (18), only the variables Y_j that have been assigned a sensor are considered, since the variance of unmeasured variables has been set to infinity.

3.6.2
Detailed Implementation of the Algorithm

Solution of the sensor network problem is carried out in seven steps:

1. process model formulation and definition of link equations;
2. model solution for the nominal operating conditions and model linearization;
3. specification of the sensor database and related costs;
4. specification of the precision requirements for observed variables;
5. verification of problem feasibility;
6. optimization of the sensor network;
7. report generation.

Each of the steps is described in detail before presenting a test case.

3.6.2.1
Process Model Formulation and Definition of Link Equations

In the current implementation, the process model is generated using the model editor of the Vali 3 data validation software, which is used as the basis for this work (Belsim 2004). The model is formulated by drawing a flow sheet using icons representing the common unit operations, and linking them with material and energy streams. Physical and thermodynamic properties are selected from a range of physical property models. Any acceptable measurement of a quantity that is not a state variable (T, P, partial molar flow rate) requires the definition of an extra variable and the associated link equation, which is done automatically for standard measurement types (e.g., mass or volume flow rate, density, dew point, molar or mass fractions, etc.). Similarly, extra variables and link equations must be defined for any process parameter to be assessed from the plant measurements. A proper choice of extra variables is important, since we may note that many state variables can not be measured in practice (e.g., no device exists to directly measure a partial molar flow rate or an enthalpy flow).

In order to allow the model solution, enough variables need to be set by assigning them values corresponding to the nominal operating conditions. The set of specified variables must at least match the degrees of freedom of the model, but overspecifications are allowed, since a least square solution will be obtained by the data reconciliation algorithm.

3.6.2.2

Model Solution for the Nominal Operating Conditions and Model Linearization

The data reconciliation problem is solved either using a large-scale SQP solver, or the Lagrange multiplier approach. When the solution is found, the value of all state variables and extra variables is available, and the sensitivity analysis is carried out (Heyen et al. 1996). A dump file is generated, containing all variable values, and the nonzero coefficients of the Jacobian matrix of the model and link equations. All variables are identified by a unique tag name indicating its type (e.g., S32.T is the temperature of stream S32, E102.K is the overall heat transfer coefficient of heat exchanger E102, and S32.MFH2O is the molar fraction of component H2O in stream S32).

3.6.2.3

Specification of the Sensor Database and Related Costs

A data file must be prepared that defines for each acceptable sensor type the following parameters:

- the sensor name;
- the annualized cost of operating such a sensor;
- parameters a_i and b_i of the equation allowing to estimate the sensor accuracy from the measured value y_i, according to the relation: $\sigma_i = a_i + b_i y_i$;
- a character string pattern to match the name of any process variable that can be measured by the given sensor (e.g., a chromatograph will match any mole fraction, and will thus have the pattern MF*, while an oxygen analyzer will be characterized by the pattern MFO2).

3.6.2.4

Specification of the Precision Requirements for Observed Variables

A data file must be prepared that defines the precision requirements for the sensor network after processing the information using the validation procedure. The following information is to be provided for all specified key performance indicators or for any process variable to be assessed:

- the composite variable name (stream or unit name + parameter name);
- the required standard deviation σ_i^t, either as an absolute value, or as a percentage of the measured value.

3.6.2.5

Verification of Problem Feasibility

Before attempting to optimize the sensor network, the program first checks for the existence of a solution. It solves the linearized data reconciliation problem assuming all possible sensors have been implemented. In the case where several sensors are available for a given variable, the most precise one is adopted. This also provides an upper limit C_{max} for the cost of the sensor network.

A feasible solution is found when two conditions are met:

- the problem matrix **M** is not singular.
- the standard deviation σ_i of all selected reconciled variables is lower than the specified value σ_i^t.

When the second condition is not met, several options can be examined. One can extend the choice of sensors available in the sensor definition file by adding more precise instruments. One can also extend the choice of sensors by allowing measurement of other variable types. Finally, one can modify the process definition by adding extra variables and link equations, allowing more variables besides state variables to be measured.

3.6.2.6

Optimization of the Sensor Network

Knowing that a feasible solution exists, one can start a search for a lower cost configuration. The optimization problem as posed involves a large number of binary variables (in the order of number of streams × number of sensor types). The objective function is multimodal for most problems. However, identifying sets of suboptimal solutions is of interest, since criteria besides cost might influence the selection process. Since the problem is highly combinatorial and not differentiable, we attempted to solve it using a genetic algorithm (Goldberg 1989). The implementation we adopted is based on the freeware code developed by Carroll (1998). The selection scheme used involves tournament selection with a shuffling technique for choosing random pairs for mating. The evolution algorithm includes jump mutation, creep mutation, and the option for single-point or uniform crossover.

The sensor selection is represented by a long string (gene) of binary decision variables (chromosomes). In the problem analysis phase, all possible sensor allocations are identified by finding matches between variable names (see Section 3.6.2.2) and sensor definition strings (see Section 3.6.2.3). A decision variable is added each time a match is found. Multiple sensors with different performance and cost can be assigned to the same process variable.

The initial gene population is generated randomly. Since we know from the number of variables and the number of constraint equations the number of degrees of freedom of the problem, we can bias the initial sensor population by fixing a rather high probability of selection (typically 80 %) for each sensor. We found however that this parameter is not critical. The initial population count does not appear to be critical either. Problems with a few hundred binary variables were solved by following the evolution of populations of 10–40 genes, 20 being our most frequent choice.

Each time a population is generated, the fitness of its members must be evaluated. For each gene representing a sensor assignment, we can estimate the cost C of the network, by summing the individual costs of all selected sensors. We also have to build the corresponding matrix $\mathbf{M}$ (Eq. (3b)) and factorize it, which is done using a code exploiting the sparsity of the matrix.

The standard deviation σ_i of all process variables is then estimated using Eq. (18).

This allows calculating a penalty function P that takes into account the uncertainty affecting all observed variables. This penalty function sums penalty terms for all m target variables.

$$P = \sum_{i=1}^{m} P_i \tag{19}$$

where $P_i = \frac{\sigma_i}{\sigma_i^t}$ when $\sigma_i \leq \sigma_i^t$

and $P_i = 0.01 \min\left(10, \frac{\sigma_i}{\sigma_i^t}\right)^2$ when $\sigma_i > \sigma_i^t$

The fitness function F of the population is then evaluated as follows:

- if matrix **M** is singular, return $F = -C_{max}$
- otherwise return $F = -(C + P)$.

Penalty function Eq. (5) (slightly) increases the merit of a sensor network that performs better than specified. Penalty function Eq. (6) penalizes genes that do not meet the specified accuracy, but it does not reject them totally, since some of their chromosomes might code interesting sensor subnetworks.

The population is submitted to evolution according to the mating, crossover, and mutation strategy. Care is taken that the current best gene is always kept in the population, and is duplicated in case it should be submitted to mutation. After a specified number of generations, the value of the best member of the population is monitored. When no improvement is detected for a number of generations, the current best gene is accepted as a solution. There is no guarantee that this solution is an optimal one, but it is feasible and (much) better than the initial one.

3.6.2.7
Report Generation

The program reports the best obtained configurations as a list of sensors assigned to process variables to be measured. The predicted standard deviation for all process variables is also reported, as well as a comparison between the achieved and target accuracies for all key process parameters.

3.6.3
Perspectives

The software prototype described here has been further improved by allowing more flexibility in the sensor definition (e.g., defining acceptable application ranges for each sensor type) and by addressing retrofit problems by specifying an initial instrument layout. The capability of optimizing a network for several operating conditions has also been implemented. The solution time grows significantly with the number of potential sensors. In order to address this issue, the algorithm has been parallelized (Gerkens and Heyen 2004) and the efficiency of parallel processing remains good as long as the number of processors is a divisor of the chromosome population size. Full optimization of very complex processes remains a challenge, but suboptimal feasible solutions can be obtained by requiring observability for smaller subflowsheets.

The proposed method can be easily adapted to different objective functions besides cost to account for different design objectives. Possible objectives could address the

resiliency of the sensor network to equipment failures, or the capability to detect gross errors, in the line proposed by Bagajewicz (2001).

There is no guarantee that this solution found with the proposed method is an optimal one, but it is feasible and (much) better than the initial one.

3.7 An Example

A simplified ammonia synthesis loop illustrates the use of data validation, including sensitivity analysis and the design of sensor networks.

The process model for this plant is shown in Figure 3.1. The process involves a five-component mixture (N_2, H_2, NH_3, CH_4, Ar), 10 units, 14 process streams, and 4 utility streams (ammonia refrigerant, boiler feed water, and steam).

Feed stream f0 is compressed before entering the synthesis loop, where it is mixed with the reactor product f14. The mixture enters the recycle compressor C-2 and is chilled in exchanger E-1 by vaporizing ammonia. Separator F-1 allows one to recover liquid ammonia in f5, separated from the uncondensed stream f6. A purge f7 leaves

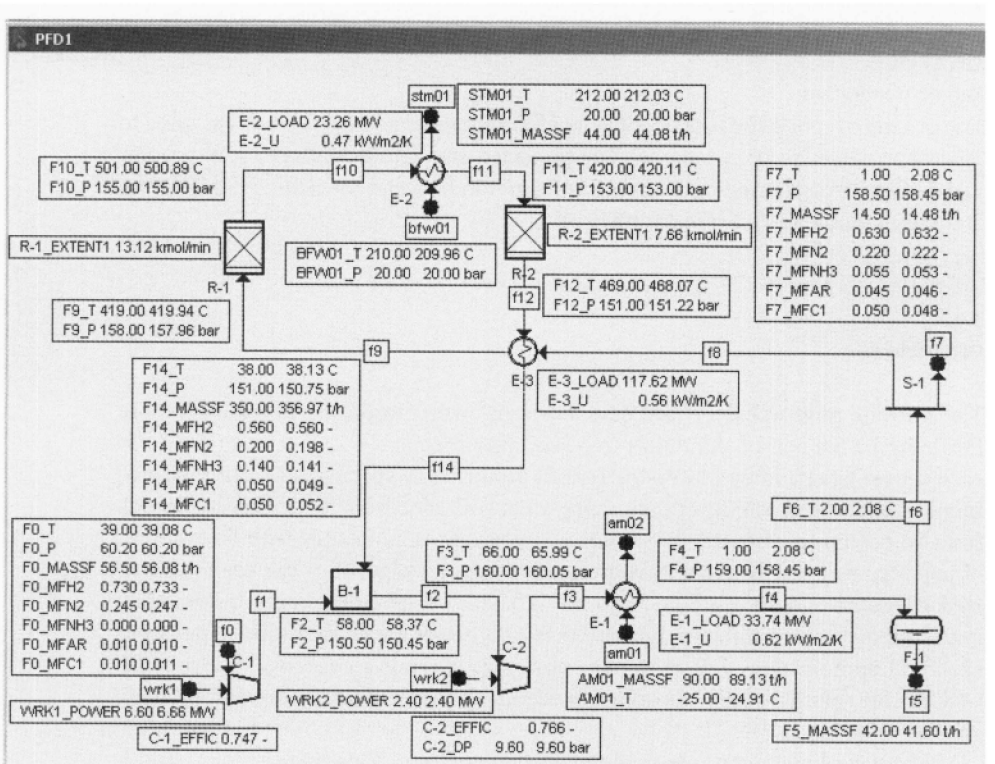

Figure 3.1 Data validation, base case. Measured and reconciled values are shown in result boxes as well as key performance indicators

the synthesis loop, while f8 enters the effluent to feed preheater E-3. The reaction takes place in two adiabatic reactors R-1 and R-2, with intermediate cooling in E-2, where steam is generated.

Energy balances and countercurrent heat transfer are considered in heat exchangers E-1, E-2, and E-3. Reactors R-1 and R-2 consider atomic balances and energy conservation. Compressors C-1 and C-2 take into account an isentropic efficiency factor (to be identified). Vapor-liquid equilibrium is verified in heat exchanger E-1 and in separator F-1.

The model comprises 160 variables, 89 being unmeasured. Overall, 118 equations have been written: 70 are balance equations and 48 are link equations relating the state variables (pressure, enthalpy and partial molar flow rates) either to variables that can be measured (temperature, molar fraction, and mass flow rate) or to performance indicators to be identified.

A set of measurements has been selected using engineering judgment. Values taken as measurements were obtained from a simulation model and disturbed by random errors.

The standard deviation assigned to the measurements was:

- 1 °C for temperatures below 100 °C, 2 °C for higher temperatures
- 1 % of measured value for pressures
- 2 % of measured values for flow rates
- 0.001 for molar fractions below 0.1, 1 % of measured value for higher compositions
- 3 % of the measured value for mechanical power.

Measured values are displayed in Figure 3.1, as are the validated results. The identified values of performance indicators are also displayed. These are the extent of the synthesis reaction in catalytic bed R-1 and R-2, the heat load and transfer coefficients in exchangers E-1, E-2 and E-3, and the isentropic efficiency of compressors C-1 and C-2.

Result analysis shows that all process variables can be observed. All measurement corrections are below 2σ, except for methane in stream f7.

The value of objective function Eq. (5) is 19.83, compared to a χ^2 threshold equal to 42.56. Thus, no gross error is suspected from the global test.

Sensitivity analysis reveals how the accuracy of some estimates could be improved. For instance, Table 3.1 shows the sensitivity analysis results for the heat transfer coefficient in unit E-1. The first line in the table reports the value, absolute accuracy and relative accuracy of this variable. The next rows in the table identify the measurements that have a significant influence on the validated value of the E-1 heat transfer coefficient. For instance, 77.57 % of the uncertainty on U comes from the uncertainty of variable AM01-T (temperature of stream am01). The derivative of U with respect to AM01-T is equal to 0.12784. Thus one can conclude that the uncertainty on the heat transfer coefficient could be reduced significantly if a more accurate measurement of a single temperature is available.

Table 3.2 shows that the reaction extent in reactor R-2 can be evaluated without resorting to precise chemical analysis. The uncertainty for this variable is 4.35 % of

Table 3.1 Sensitivity analysis for heat transfer coefficient in exchanger E-1

Variable		Tag Name	Value	Abs.Acc.	Rel.Acc.	Penal.	P.U.
K	U E-1	Computed	3.5950	0.14515	4.04 %		-
Measurement		**Tag Name**	**Contrib.**	**Der.Val.**	**Rel.Gain**	**Penal.**	**P.U.**
T	S AM01	AM01_T	77.57 %	0.12784	1.21 %	0.01	C
T	S AM02	AM02_T	5.75 %	-0.34800E-01	0.21 %	0.00	C
MFNH3	R F6	F7_MFNH3	4.33 %	-30.216	34.29 %	3.67	-
MASSF	R AM01	AM01_MASSF	4.05 %	0.16227E-01	46.50 %	0.23	t/h
MASSF	R F12	F14_MASSF	1.75 %	-0.27455E-02	33.79 %	0.99	t/h
T	S F7	F7_T	1.50 %	-0.17794E-01	62.36 %	1.16	C
T	S F6	F6_T	1.50 %	-0.17794E-01	62.36 %	0.01	C
T	S F4	F4_T	1.50 %	-0.17794E-01	62.36 %	1.16	C

Table 3.2 Sensitivity analysis for reaction extent in reactor R-2

Variable		Tag Name	Value	Abs.Acc.	Rel.Acc.	Penal.	P.U.
EXTENT1	U R-2	Computed	7.6642	0.33372	4.35 %		kmol min^{-1}
Measurement		**Tag Name**	**Contrib.**	**Der.Val.**	**Rel.Gain**	**Penal.**	**P.U.**
T	S F11	F11_T	26.82 %	-0.86410E-01	21.85 %	0.00	C
T	S F12	F12_T	25.13 %	0.83640E-01	26.78 %	0.22	C
T	S F9	F9_T	21.52 %	0.77397E-01	27.69 %	0.22	C
T	S F10	F10_T	19.95 %	-0.74532E-01	22.02 %	0.00	C
MASSF	R F5	F5_MASSF	1.56 %	0.49680E-01	49.64 %	0.23	t/h
MASSF	R BFW01	STM01_MASSF	1.51 %	0.46591E-01	35.39 %	0.01	t/h
MASSF	R AM01	AM01_MASSF	0.81 %	0.16647E-01	46.50 %	0.23	t/h
MASSF	R F0	F0_MASSF	0.77 %	0.25907E-01	58.25 %	0.14	t/h
MFNH3	R F12	F14_MFNH3	0.58 %	18.215	29.41 %	0.15	-

the estimated value and results mainly from the uncertainty in four temperature measurements. Better temperature sensors for streams f9, f10, f11 and f12 would allow one to better estimate the reaction extent.

This sensor network provides acceptable estimates for all process variables.

However the application of the sensor placement optimization using a genetic algorithm can identify a cheaper alternative.

Table 3.3 Cost, accuracy, and range for available sensors

Measured Variable	Relative cost	Standard deviation σ	Acceptable range
T	1	1 °C	$T < 150$ °C
T	1	2 °C	$T > 150$ °C
P	1	1 %	1–300 bar
Flow rate	5	2 %	1–100 kg s^{-1}
Power	1	3 %	1–10,000 kW
Molar composition (all components in stream)	20	0.001 1 %	$x_i < 0.1$ $x_i > 0.1$

A simplified sensor data base has been used for the example. Only six sensor types were defined, with accuracies and cost as defined in Table 3.3.

Accuracy targets are specified for seven variables:

- two compressor efficiencies, target σ = 4 % of estimated value
- three heat transfer coefficients, target σ = 5 % of estimated value
- two reaction extents, target σ = 5 % of estimated value.

The program detects that up to 59 sensors could be installed. When all of them are selected, the cost is 196 units, compared to 42 sensors and 123 cost units for our initial guess shown in Figure 3.1. Thus the solution space involves $2^{59} = 5.76 \times 10^{17}$ solutions (most of them being unfeasible).

We let the search algorithm operate with a population of 20 chromosomes, and iterate until no improvement is noticed for 200 consecutive generations. This requires a total of 507 generations and 10,161 evaluations of the fitness function, which runs in 90 s on a laptop PC (1 GHz Intel Pentium III processor, program compiled with Compaq FORTRAN compiler, local optimization only). Figure 3.2 shows

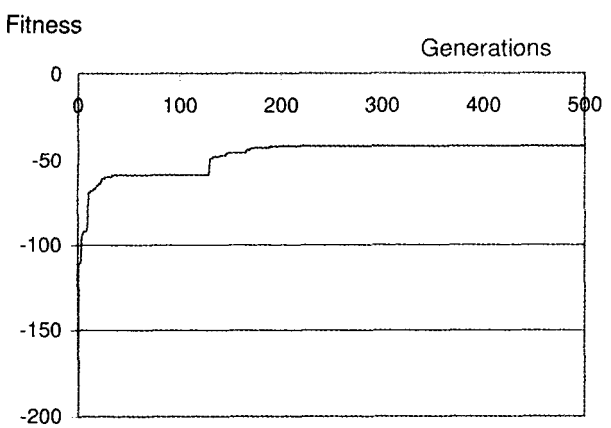

Figure 3.2 Evolution of fitness function with number of generations

that the fitness function value varies sharply in the first generations and later improves only marginally. A solution with a cost similar to the final one is obtained after 40 % of the calculation time.

The proposed solution involves only 26 sensors, for a total cost reduced to 53 cost units. The number of sensors is reduced from 16 to 11 for T, from 15 to 12 for P, from 6 to 2 for flow, and from 3 to 1 for composition. Thus the algorithm has been able to identify a solution satisfying all requirements with a considerable cost reduction.

3.8 Conclusions

Efficient and safe plant operation can only be achieved if the operators are able to monitor key process variables. These are the variables that either contribute to the process economy (e.g., yield of an operation) or are linked to the equipment quality (fouling in a heat exchanger, activity of a catalyst), to safety limits (departure from detonation limit), or to environmental considerations (amount of pollutant rejected).

Most performance parameters are not directly measured and are evaluated by a calculation based on several experimental data. Random errors that always affect any measurement also propagate in the estimation of performance parameters. When redundant measurements are available, they allow one to estimate the performance parameters based on several data sets, leading to different estimates, which may lead to confusion.

Data reconciliation allows one to address the state estimation and measurement correction problems in a global way by exploiting the measurement redundancy. Redundancy is no longer a problem, but an asset. The reconciled values exhibit a lower variance compared to original raw measurements; this allows process operation closer to limits (when this results in improved economy).

Benefits from data reconciliation are numerous and include:

- improvement of measurement layout;
- decrease of number of routine analyses;
- reduced frequency of sensor calibration: only faulty sensors need to be calibrated;
- removal of systematic measurement errors;
- systematic improvement of process data;
- clear picture of plant operating condition and reduced measurement noise in trends of key variables;
- early detection of sensor deviation and of equipment performance degradation;
- actual plant balances for accounting and performance follow-up;
- safe operation closer to the limits;
- quality at process level.

Current developments aim at combining online data acquisition with data reconciliation. Reconciled data are displayed in control rooms in parallel with raw measurements. Departures between reconciled and measured data can trigger alarms. Analy-

sis of time variation of those corrections can draw attention to drifting sensors that need recalibration.

Data reconciliation can also be viewed as a virtual instrument; this approach is particularly developed in biochemical processes, where direct measurement of the key process variables (population of microorganisms and yield in valuable by-products) is estimated from variables that are directly measured online, such as effluent gas composition.

Current research aims at easing the development of data reconciliation models by employing libraries of predefined unit operations, automatic equation generation for typical measurement types, analyses of redundancy and observability, analyses of error distribution of reconciled values, interfaces to online data collection systems and archival data bases, and developing specific graphical user interfaces.

References

1 *Albuquerque J. S. Biegler L. T.* Data reconciliation and gross-error detection for dynamic systems. AIChE J. 42 (1996) p. 2841–2856

2 *Bagajewicz M. J.* Process Plant Instrumentation: Design and Upgrade. Chap. 6, Technomic Publishing Company, Lancaster PA (USA) (1997)

3 *Bagajewicz M. J.* Design and retrofit of sensor networks in process plants. AIChE J. 43(9) (2001) p. 2300–2306

4 *Bagajewicz M. J. Rollins D. K.* Data reconciliation. In: B.G. Liptak (ed.) Instrument Engineers' Handbook (3rd edn.), Vol. 3: Process Software and Digital Networks. CRC, Taylor and Francis, Boca Raton, FL (USA) (2002).

5 *Bagajewicz M. J. Sanchez M. C.* Design and upgrade of nonredundant and redundant linear sensor networks. AIChE J. 45(9) (1999) p. 1927–1938

6 *Belsim,* VALI 4 User's Guide. Belsim Saint-Georges-sur-Meuse, Belgium (2004).

7 *Benqlilou C., Graells M. Puigjaner L.* Decision-making strategy and tools for sensor networks design and retrofit. Ind. Eng. Chem. Res. 43 (2004) p. 1711–1722

8 *Binder T. Blank L. Dahmen W. Marquardt W.* Towards multiscale dynamic data reconciliation. In: R. Berber (ed.) Nonlinear Model-Based Process Control. NATO ASI series, Kluwer, Dordrecht (1998)

9 *Carroll D. L.* FORTRAN Genetic Algorithm Driver, version 1.7, http://www.staff.uiuc.edu/carroll/ga.html (1998) accessed July 2001.

10 *Chen H. S. Stadherr M. A.* A modification of Powell's dogleg algorithm for solving systems of non-linear equations. Comput. Chem. Eng. 5(3) (1981) p. 143–150

11 *Chen H. S. Stadherr M. A.* On solving large sparse nonlinear equation systems. Comput. Chem. Eng. 8(1) (1984a) p. 1–6

12 *Chen H. S. Stadherr M. A.* Enhancements of Han-Powell method for successive quadratic programming. Comput. Chem. Eng. 8(3/4) (1984b) p. 299–234 (1984b)

13 *Crowe C. M.* Observability and redundancy of process data for steady state reconciliation. Chem. Eng. Sci. 44 (1989) p. 2909–2917

14 *Crowe C. M.* Data reconciliation – progress and challenges. J. Process Control. (6) (1996) p. 89–98

15 *Datacon Invensys, http://www.simsci.com/products/datacon.stm, cited 3 May (2004)*

16 *Gerkens C. Heyen G.* Use of Parallel Computers in Rational Design of Redundant Sensor Networks. 14th European Symposium on Computer Aided Process Engineering, Lisbon (2004)

17 *Goldberg D. E.* Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA (USA) (1989)

18 *Heyen G. Dumont M. N. Kalitventzeff B.* Computer aided design of redundant sensor networks. In: Grievnik J. Dumont M. N. Kalitventzeff B. 12th European symposium on Computer Aided Process Engineering The Hague, Elsevier Science, Amsterdam (2002)

19 *Heyen G. Maréchal E. Kalitventzeff B.* Sensitivity calculations and variance analysis in plant measurement reconciliation. Comput. Chem. Eng. 20S (1996) p. 539–544

20 *Joris P. Kalitventzeff B.* Process measurements analysis and validation. In: Proceedings Chemical Engineering Fundamentals Conference (CEF'87): Use of Computers in Chemical Engineering, Italy, pp. 41–46 (1987)

21 *Kalitventzeff B., Laval P. Gosset R. Heyen G.* The validation of industrial measurements, a necessary step before the parameter identification of the simulation model for large chemical engineering systems. In: Proceedings of International Congress "Contribution des calculateurs électroniques au développement du génie chimique," Société de Chimie Industrielle, Paris (1978)

22 *Kalman R. E.* A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 82D (1960) p. 35–45

23 *Kuehn D. R. Davidson H.* Computer control: mathematics of control. Chem. Eng. Prog. 57 (1961) p. 44–47

24 *Kyriakopoulou D. J. Kalitventzeff B.* Data reconciliation using an interior point sqp, ESCAPE-6, Rhodes, Greece, 26–29 May, Pergamon Press (1996)

25 *Kyriakopoulou D. J. Kalitventzeff B.* Reduced Hessian interior point SQP for large-scale process optimization. First European Congress on Chemical Engineering, Florence, Italy, 4–7 May AIDIC Milano (1997)

26 *Liebman M. J. Edgar T. F. Lasdon L. S.* Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. Comput. Chem. Eng. 16 (1992) p. 963–986

27 *Madron F.* Process Plant Performance: Measurement and Data Processing for Optimization and Retrofits. Ellis Horwood, London (1992)

28 *Mah R. S. H. Stanley G. M. Downing D. W.* Reconciliation and Rectification of Process Flow and Inventory Data. Ind. and Eng. Chem. Proc. Des. Dev. 15 (1976) p. 175–183

29 *Meyer M. Le Lann J. M. L. Koehert B. Enjalbert M.* Optimal selection of sensor location on a complex plant using graph-oriented approach. In: Moser F. Schnitzer H. Bart H.-J. (eds.) European Symposium on Cumputer-Aided Process Engineering-3, Graz, Austria (1993). Suppl. to Computers and Chemical Engineering, Pergamon Press, Oxford (1993)

30 *Musch H. List T. Dempf D. Heyen G.* On-line Estimation of Reactor Key Performance Indicators: an Industrial Case Study. (2004)

29 *Narasimhan S. Jordache C.*, Data Reconciliation and Gross Error Detection, an Intelligent use of Process Data. Gulf, Publishing Company, Houston, TX (USA) (2000)

30 *Romagnoli J. A. Sanchez M. C.* Data Processing and Reconciliation for Chemical Process Operations. Academic Press, San Diego, CA (USA) (2000)

31 *Vaclavek V.* Studies on system engineering: on the application of the calculus of observations in calculations of chemical engineering balances. Coll. Czech. Chem. Commun. 34 (1968) p. 3653–3660

32 *Vaclavek V.* Studies on System Engineering: Optimal Choice of the Balance Measurements in Complicated Chemical Engineering Systems. Chem. Eng. Sci. 24 (1969) p. 947–955

33 *Vali Belsim* http://www.belsim.com/Products-main.htm, cited 3 May 2004

34 *Veverka V. V. Madron F.* Material and energy balancing in the process industries. From microscopic balances to large plants, Computer-Aided Chemical Engineering. Elsevier Science, Amsterdam (1996)