DATA RECONCILIATION AND INSTRUMENTATION UPGRADE. OVERVIEW AND CHALLENGES.

Miguel J. Bagajewicz
University of Oklahoma, 100 E. Boyd T-335, Norman, OK 73072

Abstract

This paper discusses the state of the art in data reconciliation and instrumentation upgrade. In the field of data reconciliation, several new directions of research and recommendations for software vendors as well as practitioners are made. In the case of the emerging field of instrumentation upgrade, aside from also pinpointing research and development directions, practical economical aspects are discussed.

Keywords

Data Reconciliation, Instrumentation Design, Instrumentation Upgrade.

Introduction

This article focuses on the relationship between data reconciliation, a statistically-based technique to obtain estimators of process variables and the problem of determining how to place instrumentation throughout the process so that data reconciliation performs following certain pre-specified performance goals. The article is organized as follows: data reconciliation is reviewed first and the problem of instrumentation design/ upgrade is discussed afterwards. In both cases, the state of the art in academia and industry some of the existing challenges are discussed.

Data Reconciliation

Data filtering and reconciliation has been used for several years as means of obtaining accurate and consistent data in process plants. Early work in dynamic data reconciliation is rooted in the problem of process state estimation using the concept of filtering. Lately, the problem is addressed using model based data smoothing.

Depending on the data used, there are three types of estimation problems for the state of the system at time t (figure 1). Kalman filtering (1960) deals with the problem of filtering, that is producing estimates at time t. Data Reconciliation deals with the problem of smoothing, using the condition that variables are connected through a model.

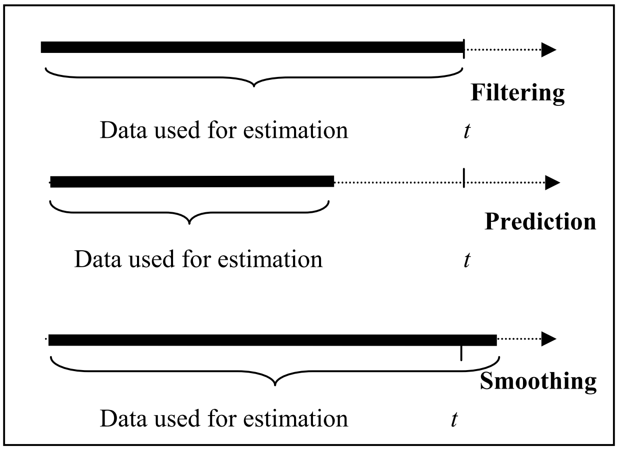


Figure 1. Types of state estimation (Gelb, 1974)

In other words, in data reconciliation plant measurements (flowrates, temperatures, pressures, concentrations, etc) are used to obtain estimators that conform to a certain chosen model, typically a set of differential algebraic equations (DAE) (equations (1-2)

$$\frac{dx_1}{dt} = g_1(x_1, x_2) \tag{1}$$

$$g_2(x_1, x_2) = 0 \tag{2}$$

where the variables are divided in two sets, x_1 being the set corresponding to variables participating in accumulation terms and x_2 the rest. Thus, the general data reconciliation problem is stated as follows:

> *Given a set of measurement values of a subset of state variables* $z_M = (z_{M,1}, z_{M,2})$ *it is desired to obtain the best estimators of these measured state variables* $\widetilde{x}_M$ *and as many of the unmeasured variables* $\widetilde{x}_U$ *as possible.*

When estimates $\widetilde{x}_M$ at N instances of time at which the measurements were made are sought, and the variance of the measurements (Q) is known, the following optimization problem is used:

$$\left.\begin{array}{l} Min \sum_{k=0}^{N} [a_{M,k}]^T Q^{-1} [a_{M,k}] \\ \text{s.t.} \\ \dfrac{d\widetilde{x}_1}{dt} = g_1(\widetilde{x}_1, \widetilde{x}_2) \\ g_2(\widetilde{x}_1, \widetilde{x}_2) = 0 \\ a_{M,k} = \widetilde{x}_M(t_k) - z_{M,k} \end{array}\right\} \quad (3)$$

This least square problem can be derived from Bayesian theory using the assumption that the distribution of errors is normal (Johnston and Kramer, 1995). In addition, Crowe (1996) showed that the same result can be derived using information theory. For the case where steady state is assumed, only one measurement is used, usually an average of several measurements, and using special procedures to include accumulation term. Three excellent books describe the subject (Madron, 1992; Romagnoli J. and M. Sánchez, 1999; Narasimhan S. and C. Jordache, 2000), which cover in detail some material here included.

Steady State Linear Data Reconciliation

Consider the case where no hold-up change takes place or negligible hold up is assumed. In such case only flowrates are estimated. Therefore, we write the problem as follows

$$\left.\begin{array}{l} Min\, [\widetilde{f}_M - f_M^+]^T Q^{-1} [\widetilde{f}_M - f_M^+] \\ \text{s.t.} \\ D\widetilde{f} = 0 \end{array}\right\} \quad (4)$$

where $\widetilde{f}$ is the vector of reconciled flows, which contains measured and unmeasured flows. $\widetilde{f} = [\widetilde{f}_M \ \widetilde{f}_U]^T$. Various techniques were developed to convert the above problem into one containing measured variables only: Matrix Projection (Crowe *et al.*, 1983),Q-R decomposition (Swartz, 1989; Sánchez and Romagnoli, 1996) and Gauss Jordan rearrangement of D (Madron, 1992).

Non-linear versions of the steady state data reconciliation problem including the reconciliation of temperatures and even pressure measurements exist and are very successful and robust in practice. Dynamic reconciliation is still eluding engineering practice. Several commercial packages exist (Datacon form Simsci, Sigmafine from OSI, Adviser from Aspentech, etc) all of them using steady state data reconciliation.

Because processes are never truly at "steady-state", one of the criticisms that this model, linear and nonlinear, has endured is that not only random errors, but also process variations are included in the averaging of data, smearing the results in an unpredictable manner. However, Bagajewicz and Jiang (2000) and Bagajewicz and Gonzales (2001) proved that this is not true for linear systems without hold-ups, and with proper data handling, the results are the same. Moreover, they showed that when holdups exist, the deviations can be small. For non-linear cases, some deviations of this are, however, possible. Finally, another unresolved problem is the assessment of the variance-covariance matrix Q. Only a few articles deal with it (Almasy and Mah, 1984; Darouach et al., 1989; Keller et al., 1992; Chen et al., 1997)

Dynamic data reconciliation was not developed by the software industry based on the claim that it is too computationally intensive, and on doubts that non-linear models would be robust and user friendly. The reason may be that the commercial cycle for steady state data reconciliation is not over, and that there are more burning problems to overcome, gross error handling being one.

Gross Error Detection

Two central issues are of concern: proper location of gross errors (instrument biases and leaks) and estimation of their sizes. Thus, the challenging task is to

- Identify the existence of gross errors
- Identify the gross errors location
- Identify the gross error type
- Determine the size of the gross error.

After the gross errors are identified, two responses are possible and/or desired:

- Eliminate the measurement with the bias, or
- Correct the model (case of a leak) and run the reconciliation again.

The first alternative is the one implemented in commercial software, which only considers biases.

Test for Gross Error Presence/Location

Hypothesis testing is used for this task. We here present the three most popular tests.

Global Test: The null hypothesis H_0 is that there is no gross error. Let r be the vector or residuals of the material balances, that is $r=C_R z$, where z are the flowrate measurements. Then, the expected value of r *is* $E(r)=0$, and the covariance matrix of r *is* $Cov(r)= C_R Q_R C_R^T$, where Q_R is the covariance matrix of random measurement errors. In the absence of gross errors, the following variable

$$\chi_m^2 = N_M r^T (C_R Q_R C_R^T)^{-1} r \quad (5)$$

follows a Chi-squared distribution with m degrees of freedom ($\chi^2_{m,\alpha}$), where m is the number of rows of C_R. This number can be obtained before even performing data reconciliation. If it falls within the interval of confidence, that is, if it is lower than a certain critical value, then the null hypothesis cannot be ruled out. On the other hand, if it is larger than the critical value, it is said that a gross error has been detected, that is the null hypothesis cannot be accepted. Note that the global test cannot determine where are the gross errors, or how many are there.

Nodal Test *(Mah and Tamhane, 1982):* In the absence of gross errors the constraint residuals r follow a m-variate normal distribution (m is the rank of C_R). Therefore

$$\phi_i = N_M^{1/2} \frac{r_i}{\sqrt{(C_R Q_R C_R^T)_{ii}}} \quad (6)$$

follows a standard normal distribution, *N(0,1),* under H_0. If ϕ_i is larger than the critical value based on a confidence level α, then one concludes that there is at least one gross error in the set of measurement that participates in the corresponding node balance. Rollins et al. (1996) proposed a strategy using this test on linear combination of nodes.

Measurement Test: It is based on the vector of measurement adjustments (or corrections) $a = F_R^+ - \widetilde{F}_R$, where a is the vector of measurement adjustments. The test is based on the assumption that the random errors for measurements are independently and normally distributed with zero mean. Under the null hypothesis, H_0, the expected value of a is $E(a) = 0$ and the covariance matrix of a is $Cov(a) = \hat{Q}_R$. Thus, the following variable is expected to follow a normal distribution *N(0,1)*.

$$\eta_i = \frac{a_i}{\sqrt{(\hat{Q}_R)_{ii}}} \quad (7)$$

Thus, if no gross error is present the above value should be lower than a critical value. If α is the confidence level, then this critical value $x_{\alpha/2}$. is obtained directly from the normal distribution tables $x_{\alpha/2}$. Several studies, modifications and improvements have been proposed for this test (Mah and Tamhane, 1982; Crowe *et al*., 1983). In a recent development, Bagajewicz and Rollins (2002) discuss the consistency of this test, warning that, even under deterministic conditions it may point to the wrong variable.

Other tests used specifically for gross error identification exist. Among the most popular, are the generalized likelihood ratio (Narasimhan and Mah, 1986, 1987, 1988), principal component tests (Tong and Crowe, 1995) and the Bonferroni tests on unbiased estimators (Rollins and Davis, 1992).

Multiple Gross Error Identification

The tests described above are suitable for the detection of one gross error. However, when more gross errors exist, strategies are needed to identify them. One of the first strategies proposed is serial elimination (Ripps, 1965), which consists of coupling a certain test with an elimination strategy. If the test failed, then a strategy is proposed to identify one or more variables, which are the "most suspected ones". The measurements of these variables are eliminated and the test is run again. Commercial versions of this procedure (Datacon, Sigmafine) eliminate one measurement at a time and use the measurement test or similar. Several variations of this scheme were proposed (Romagnoli and Stephanopoulos, 1980, Rosenberg *et al*., 1987; Iordache *et al*., 1985).

Gross Error Size Estimation

Once the gross errors have been identified, it is desired to determine their size. There are several methods that have been developed in recent years to perform this. When one gross error is present, Madron (1982) proposed an expression based on the statistical properties of $r(C_R Q_R C_R^T)^{-1} r$. However, in the presence of multiple gross errors these formulas do not apply.

Serial Compensation (Narasimhan and Mah, 1987) identifies one gross error at a time and estimates its size, compensates the measurement and continues until no error is found. This technique has proven to be relatively efficient when one gross error is present, but not in the presence of multiple gross errors. Serial elimination does not address leaks (Mah, 1990). Serial Compensation is applicable to all types of gross errors and can maintain redundancy during the procedure but its results are completely dependant on the accuracy of estimation for the size of gross errors (Rollins and Davis, 1992). To improve these methods, simultaneous or collective compensation proposes the estimation of all gross errors simultaneously. Rollins and Davis (1992) proposed an unbiased estimation technique (UBET), which relies heavily on the identification of candidate gross errors performed by other methods. Keller *et al*. (1994) proposed the successive application of the generalized likelihood ratio with collective compensation of all candidates at each step; Kim *et al*. (1997) proposed a modified Iterative Measurement Test (MIMT) using nonlinear programming (NLP) techniques; Sánchez and Romagnoli (1994) proposed a combinatorial approach to pick candidates and use them in a compensation model based on the use of the global test; Bagajewicz and Jiang (1998) proposed the successive use of a statistical test to identify one gross error at a time and independently developed a compensation model which is identical to the one presented by Sánchez and Romagnoli (1994).

Some alternative objective functions that are capable of handling the gross errors in the data simultaneously with data reconciliation have been proposed. We cite only those that have reached publicly offered software. Tjoa and Biegler (1991 proposed a mixture distribution as the objective (likelihood) function. A test follows the reconciliation to determine the gross error presence. Albuquerque and Biegler (1996) proposed the use the Fair function and Johnston and Kramer (1995) proposed to use the Lorentzian distribution.
Of most of the methods that have been developed, three have been identified as efficient (Bagajewicz et al., 1999).

- UBET (Rollins and Davis, 1992), as modified by Bagajewicz *et al*. (1999).
- SICC (Jiang and Bagajewicz, 1999)
- MSEGE (Sánchez *et al*., 1999)

These techniques cannot however overcome a limitation that is inherent to the problem, which is the uncertainty of gross error location. This is explained next.

Equivalency Theory

This theory states that *two sets of gross errors are equivalent when they have the same effect in data reconciliation, that is, when simulating either one in a compensation model, leads to the same value of objective* (Bagajewicz and Jiang, 1998). Therefore, the equivalent sets of gross errors are theoretically undistinguishable. In other words, when a set of gross errors is identified, there exists an equal possibility that the true locations of gross errors are in one of its equivalent sets. From the view of graph theory, equivalent sets exist when candidate stream/leaks form a loop in an augmented graph consisting of the original graph representing the flowsheet with the addition of environmental node.

For example, consider the process of Figure 2 and assume that all streams are measured (the oval represents the environment node). As shown in Table 1, a bias of (-2) in S_4 and a bias of (+1) in S_5 (Case 1) can be represented by two alternative sets of two gross errors (Cases 2 and 3). *By applying this theory, one can see that any proposed set of gross error candidates cannot form a loop. Otherwise the size of these gross errors is indeterminate, a condition that leads to singularities.* This explains why combinations of introduced gross errors like S_1-S_6, S_2-S_4, S_1-S_3 leads so easily to singularities: The addition of just one stream to these sets can lead to a loop. Take for example the combination of simulated gross errors S_1-S_6: just the addition of stream S_2 will form a set with a loop through the environmental node. The addition of stream S_5 to the set {S_2, S_4} also forms a loop.

Degenerate Cases: The equivalencies above are built in the assumption that the number of gross errors identified is equal to the real number of gross errors. However, there are examples where the actual number of gross errors can be larger than the number of gross errors identified.

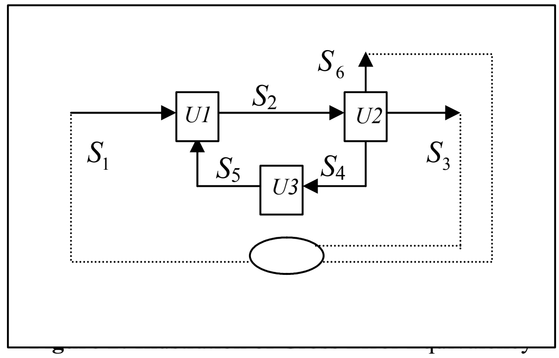


Table 1: Illustration of Equivalent Sets in {S_2, S_4, S_5}

		S_1	S_2	S_3	S_4	S_5	S_6
Case / Measurements		12	18	10	4	7	2
(1) Bias in S_4, S_5	Reconciled	12	18	10	6	6	2
	Biases				-2	1	
(2) Bias in S_2,S_4	Reconciled	12	19	10	7	7	2
	Biases		-1		-3		
(3) Bias in S_2,S_5	Reconciled	12	16	10	4	4	2
	Biases		2			3	

Singularities
Almost all gross error detection techniques rely on the building of a gross error candidate list, and the solution of a set of equations to estimate their sizes. If a subset of these gross error candidates participates in a loop in the process graph, these equations will contain matrices that will become singular. Several gross error methods that are prone to fail because of this problem can be easily modified to overcome it (Bagajewicz *et al*., 1999). Essentially, the modification consists of eliminating one element of such loop.

Uncertainties
According to the equivalency theory, one can avoid the singularities. However, when picking the candidate set one is picking a subset of some equivalent set. Thus, there is always the possibility that the gross errors are not located in the set identified, but in some other. When one is assessing the power of a method, this is easy to pinpoint and will be discussed below. However, when one is using a certain gross error detection scheme, one is left with a set of identified gross errors for which several alternatives exist. Indeed, the gross errors identified could be part of

equivalent sets of increasing number of streams, that is, sets containing one, two, three or more streams in addition to the ones that have been identified. All these sets form loops in the process graph, which can be easily identified. For example of Figure 2, if one finds two gross errors, say S_2 and S_4 one equivalent set that can be constructed adding only one more stream is $\{S_2, S_4 S_5\}$. However, if the gross error identified is S_2 no equivalent set (loop) can be formed adding one stream. One needs two streams to form either $\{S_2, S_4 S_5\}$ or $\{S_2, S_3 S_1\}$ or $\{S_2, S_6 S_1\}$. As one can see, the possibilities are endless.

One should realize that in the case where one forms an equivalent set adding one stream, one can pick out of this set any subset with the same number of gross errors as those found. Formulas to recalculate the sizes of these new gross errors have been derived, but presented elsewhere (Jiang and Bagajewicz, 1999). For example, if one finds two gross errors in S_4 and S_5 of size -2 and +1, one can consider that this is a basic set of a general equivalent set $\{S_2, S_4, S_5\}$. Thus, the three possibilities depicted in Table 1 are equivalent. In this case, one would say: "Gross errors have been identified in S_4 and S_5 of size -2 and +1. To obtain equivalent lists, pick any two gross errors of the following list S_2, S_4, S_5". Degenerate cases imply gross errors of the same size (unless there is an overlap of loops). Thus, they are less likely to occur.

Leaks

If one considers a leak just as another stream, a leak forms at least one loop with some streams or other leaks in the augmented graph. Therefore it will be represented with at least one equivalent set of biases identified with the model. Let us illustrate this issue with two examples. Consider the process in Figure 3. Assume that there is a leak with the size of +5 and the measurements for flowrates of S1 and S_2 are 100 and 95 respectively. The *SICC* strategy identifies S_1 with a bias of 5. The equivalent sets are a leak with the size of 5 and S_2 with a bias of –5.

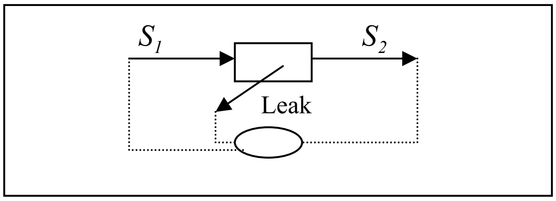


Figure 3: A Simple Process with a Leak

Thus, equivalent sets can be formed with streams and leaks of any loop. Thus, any leak is equivalent to a set of equal size biases in a set of streams connecting to the unit where the leak occurs and the environment. This is an important result and leads to the following conclusion:

Since a leak is equivalent to a set of biased streams, any steady state method that contains a test to detect biased instruments can be used in conjunction with the equivalency theory to assess the existence of leaks.

Finally, Bagajewicz and Jiang (2000) showed that:

1) Any spanning tree of the graph of the process can be used in an estimation scheme to capture all gross errors. In fact, the UBET method (Rollins and Davis, 1992) modified by Bagajewicz *et al*. (1999) is successful because it utilizes one such spanning tree.

2) One can propose a MILP procedure to exploit equivalencies and obtain a set of minimum cardinality.

Industrial Practice and Software

Several issues emerge from the status of industrial practice and the software available. We briefly discuss some of them hoping a more lively discussion will take place at the conference

1) Should the current steady state paradigm be replaced by more sophisticated dynamic data reconciliation software? My answer to this is that this is not needed for the time being, since more pressing problems abound (see below)
2) The current power for gross error handling (detection, elimination, estimation) by the software in the market is insufficient. Although some attempts have been made to improve over serial elimination, this continues to dominate. My suggestion to practitioners is to increase the pressure on vendors to improve in this regard. My apologies to the vendors for not making your life easier.
3) In a very much related issue to the need of improve gross error handling, I propose to implement equivalency theory, or any other methodology to handle uncertainty.
4) Research should extend equivalency theory to non-linear cases.
5) Variance estimation has deserved the attention of a handful of researchers. No significant implementation in practice is known to me.
6) Many times I was challenged by enthusiasts of the use of PCA monitoring methods that PCA can perform better gross error detection. This matter is worth exploring.

INSTRUMENTATION UPGRADE

The field of instrumentation design/upgrade oriented towards the use of data reconciliation and some fault detection techniques is reviewed here. The field of sensor location for control and on-line optimization is purposely not covered. Sensor networks should be able to handle gross errors effectively, that is, detect them when they are too large and avoid large corruption of data when they are not detected. A robust sensor network should:

a) ***Be accurate:*** Accuracy of key variables obtained through data reconciliation should satisfy the needs of monitoring, control and production accounting.
b) ***Detect gross errors efficiently:*** This is connected to the redundancy of the system. The more measurements one adds the larger is the ability to detect gross errors.
c) ***Retain certain accuracy when gross errors are eliminated:*** Once serial elimination is used the remaining sensors should guarantee certain accuracy for key important variables.
d) ***Have Resilience or Control Smearing:*** When gross errors are not detected, the smearing should be low.
e) ***Be reliable:*** Instruments that break often are a frequent source of gross errors.

The instrumentation upgrade problem consists of *achieving certain desired network robustness at a minimum cost.* Madron (1992) reviews some of the earlier methods, which focus on accuracy only. Bagajewicz (1997) presents the first method that minimizes cost and focuses on robustness. Bagajewicz and Sanchez, (1999a,b,c; 2000a,b,c) discussed grassroots design and retrofit. The form of the grassroots problem is

Minimize {*Total Cost*}
s.t.
{ *Desired level of Precision of key variables*
Desired level of Reliability of key variables
Desired level of Gross-Error Robustness

where the total cost includes the maintenance cost, which regulates the availability of variables, a concept that substitutes reliability when the system is repairable. Precision is defined as the variance of the estimators. Availability is defined as the probability of a sensor not being in a failed state at a given time, whereas reliability is the same probability, but for the period from zero to the time in question. In the case of retrofit, the total cost represents the cost of relocating instruments, as well as adding new ones. In addition, some constraints regarding the total number of instruments per variable as well as restriction on which instruments can be relocated are added. Bagajewicz (2000) included all this material in his recent book. This procedure does take into account the ability of a sensor network to detect process faults and a logic for alarm systems. Even though important attempts were made (Tsai and Chang, 1997) a model-based on cost-efficient alarm design is yet to be produced. Likewise, the direct incorporation of control performance measures as additional constraints to this cost-optimal model has not been fully investigated yet. In turn, from the exclusive point of view of fault detection, the problem of the design of instrumentation is:

Minimize {*Total Cost*}
subject to
- *Desired Observability of Faults*
- *Desired Level of Resolution of Faults*
- *Desired level of Reliability of Fault Observation*
- *Desired level of Gross-Error Robustness in the Sensor Network*

The combination of both goals, that is, the design of a sensor network capable of performing estimation of key variables for monitoring, production accounting, and parameter estimation for on-line optimization as well as for fault detection, diagnosis and alarm is emerging.

A cost-benefit analysis needs to be performed to determine the thresholds of all the properties that are required from the upgraded network. For example, in the case of production accounting, precision and gross error robustness, can be easily related to revenue, whereas in the case of quality control, precision, reliability and gross error robustness can also be related to quality standards and ultimately to lost revenue. A connection of this sort can be established for just about all scenarios of design and upgrade. This is discussed in more detail below.

Cost Optimal and Precise Sensor Networks

Consider just the constraints of precision, that is,

Minimize {*Total Cost*}
subject to
- *Desired Precision of Key Variables*

Consider the process flow diagram of figure 4. Assume that flow meters of precision 3%, 2% and 1% are available at costs 800, 1500 and 2500 respectively. When precision is only required for variables F_1 and F_4, with $\sigma_1^* = 1.5\%$ and $\sigma_4^* = 2.0\%$, two solutions are obtained. Although precision is achieved using, in this case, a non-redundant network, biases are impossible to detect. Therefore, if at least one degree of redundancy is requested, that is, at least two ways of estimating each key variable, then there are two solutions with a cost of C=3100. These solutions are: (F_1=3%, F_2=3%, F_3=3%) and (F_1=3%, F_2=3%, F_4=3%). Bagajewicz and Sánchez, (1999c) showed how this can be formally requested.

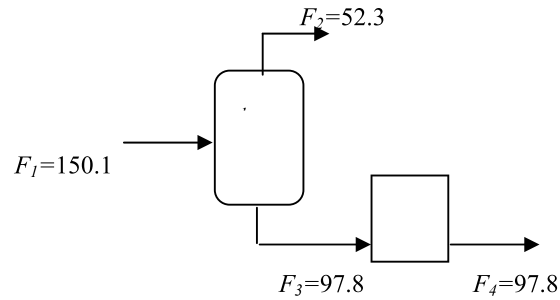


Figure 4.

To solve this problem, integer variables are used to denote whether a variable is measured (q_i =1) or not (q_i =0).

Thus, the investment cost can be represented by a summation of the type $\sum_{\forall i} c_i q_i$, where c_i is the cost of the instrument in variable *i*. The precision constraint is mathematically represented by requiring the standard deviation of the estimates σ_i, obtained using data reconciliation is smaller than a threshold value σ_j^*. This problem was studied by the author (Bagajewicz, 1997) and it is assumed that there is only one potential measuring device with associated cost c_i for each variable (that is, no hardware redundancy), but this condition can be relaxed (Bagajewicz, 1997, 2000).

Bagajewicz (1997) proposed a tree enumeration procedure and recently, Chmielewski *et al*. (1999) proposed an alternative formulation and Bagajewicz and Cabrera (2001) presented a mixed integer linear programming formulation. Finally, a series of techniques to introduce constraints that can force redundancy were introduced by Bagajewicz and Sánchez (1999c) for this case and were extended to bilinear systems by Bagajewicz (2000).

Design for Maximum Precision

Madron and Veverka (1992) proposed to design sensor networks minimizing the mean square error of the required quantities. This problem was efficiently solved using graph theory (Madron, 1992). A problem maximizing the precision of only one variable was proposed by a team of the BP and University College, London (Alhéritière et al., 1998), who unfortunately did not use integer variables. The generalized maximum precision problem (Bagajewicz, 2000) considers the minimization of a weighted sum of the precision of the parameters.

$$\begin{aligned} &Min \quad \sum_{j \in M_P} a_j \sigma_j^2(q) \\ &subject \quad to \\ &\qquad\qquad Cost \le c_T \end{aligned}$$

where c_T is the total resource allocated to all sensors. The result of this generalized problem is a design for multiple parameter estimation; it is more realistic due to the discrete variables, and takes into account redundancy as well as all possible forms of obtaining the parameters. A mathematical connection between the maximum precision and the minimum cost representations of the problem exists (Bagajewicz and Sánchez, 1999a). More precisely, the solution of one problem is one solution of the other and vice versa. Bhushan and Rengaswami (2002) refined this statement and proposed an improvement to this.

Parameter Estimation and Precision Upgrade

Considerable attention is being put nowadays to the issue of parameter estimation, especially in the context of the increasing popularity of on-line optimization. The practice, however, has been around long before this concept became popular. For example, flow rate and temperature measurements are used to determine the level of fouling in heat exchangers, by simply calculating the heat transfer coefficients. These heat transfer coefficients are used in a simulation model to schedule cleaning. The issue is elusive if the data contains too many gross errors. Therefore, data reconciliation and bias detection is almost a must if this task is to be done efficiently. However, in many cases an upgrade of the instrumentation is needed to make this possible or to improve the precision of these estimates. Other examples of parameter estimation to feed simulation and optimization models are the determination of the column efficiencies, or reactor parameters.

There are three possible ways of performing the upgrade of a sensor network: 1) by the addition of new instruments, 2) by the substitution of existing instruments by new ones, and 3) by relocation of existing instruments.

Typically, addition of new instruments has been the response first considered. Kretsovalis and Mah (1987) proposed a combinatorial strategy to incorporate measurements one at a time, to an observable system, but no constraints were considered. Alhéritière *et al*., (1998) plotted the increased precision as a function of investment. Although this approach is intuitive, its importance relies on the possibility of visualizing the effect of instrumentation cost.

Krishnan *et al*. (1992a,b) relies on a screening procedure to position instrumentation that involves three steps. 1) A first step performs a structural analysis (singular value decomposition), which disregards measurements with little or no effect on the parameters, 2) A second step disregards measurements that have insignificant effect on the axis length of the confidence region of the parameter estimates, 3) The last step determines the interaction between the parameter estimates. Unfortunately, this method does not take into account cost and does not offer a systematic procedure to make a final selection of the "best" set. In contrast, Bagajewicz (2000) discusses linearization techniques in the context of cost-based minimization design procedures.

Aside from the addition of instrumentation, there are two other ways of upgrading a sensor network: by the substitution of existing instruments by new ones, and by relocation of existing instruments. One example is the substitution of thermocouples by thermoresistances or their relocation. Another is the case of laboratory analysis.

Bagajewicz and Sánchez (2000b) proposed a minimum cost model as follows:

Minimize {*Total Cost of New Instrumentation*}
subject to

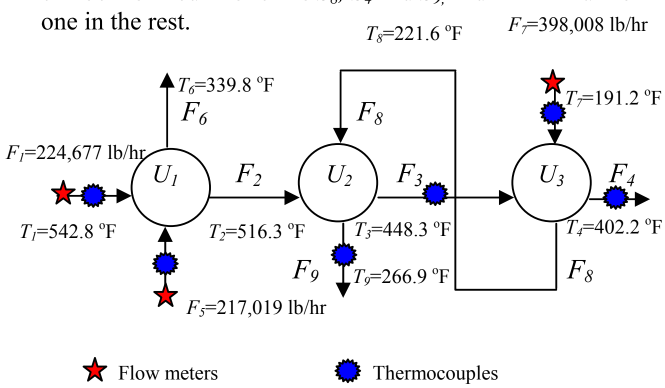
- *Desired Precision of Key Variables*
- *Limitations on the number of new instruments.*

The mathematical model makes use of binary variables and is linear. The second constraint establishes an upper bound on the number of sensors used to measure each variable. This number is usually one for the case of flow rates, but it can be larger in the case of laboratory measurements of concentrations.

Addition of Instrumentation for Precision Upgrade in Parameter Estimation

Figure 5 depicts a set of heat exchangers where crude is heated up using hot gas-oil coming from a column. The heat transfer coefficients for the heat exchangers are estimated using of temperature and flow rate measurements. The existing instrumentation is indicated in the figure; their corresponding precision is 3% for flow meters and 2 ^oF for the thermocouples. The standard deviations of heat transfer coefficients calculated using the installed set of instruments are [12.27 2.96 3.06] (BTU/h ft^2 ^oF). In order to enhance the precision of the parameter, new instruments should be added. In this example, hardware redundancy is considered. Furthermore, different types of new instruments are available to measure some temperatures.

New flow meters have a 3% precision and a cost of 2250. Two thermocouples are available for streams S_1, S_4 and S_9. These thermocouples have a standard deviation of 2 ^oF and 0.2 ^oF and a cost of 500 and 1500, respectively. In the rest of the variables only one thermocouple of 2 ^oF of standard deviation and a cost of 500 can be installed. Only one flow meter is allowed to be added in all streams except S_6 and S_8. No thermocouple can be installed in stream S_8, a maximum of two additional thermocouples can be installed in streams S_8, S_4 and S_9, and a maximum of one in the rest.



Heat Exchanger	*Area* (ft^2)	F_T	Cp_h (BTU/lb °F)	Cp_c (BTU/lb °F)
U_1	500	0.997	0.6656	0.5689
U_2	1100	0.991	0.6380	0.5415
U_3	700	0.995	0.6095	0.52

Figure 5: Industrial Heat Exchanger Network (Adapted from Bagajewicz, 2000)

The following table presents three different solutions of three different costs (one with three alternative solutions).

Table 2

σ_{U_1}	σ_{U_2}	σ_{U_3}	***Cost***	***Optimal Set***
3.62	1.97	2.71	500	T_6 (1)
2.78	1.69	2.38	1500	T_2 (1) T_4 (1) T_6 (1)
2.72	1.50	2.28	6500	F_2 F_3 T_2 (1) T_4(1) T_6(1) T_9(1)
				F_2 F_4 T_2 (1) T_4(1) T_6(1) T_9(1)
				F_3 F_4 T_2 (1) T_4(1) T_6(1) T_9(1)

When there are two possible instruments to measure a variable, the type of instrument is indicated between parentheses in the optimal solution set. Thus for example, T_4 (1) indicates that the first instrument available to measure the temperature in stream S_4 is selected, in this case a thermocouple with a precision of 2 ^oF is selected. As the requirements of precision increase more instruments are added and in some cases (Case 3) alternative solutions exist.

Resource and Instrumentation Reallocation

In many cases, measurements can be easily transferred at small or negligible cost from one stream to another (concentration measurements performed in the laboratory, pressure gauges, thermocouples, etc.). However, flow meters are probably an exception. Even in the case where cost is not considered, one would like to minimize the number of changes. In addition, these reallocation costs may overcome the simple addition of new instrumentation. Therefore, any reallocation and upgrade program should consider the trade off between all these decisions. A mathematical programming representation has been developed (Bagajewicz and Sánchez, 2000b), which determines the appropriate trade-off between these decisions. A simple example to illustrate the technique is described next.

For the flash tank of figure 6, it is desired to reallocate instrumentation to improve the estimation of its vaporization efficiency. The flash is well instrumented and hardware redundancy on the feed flow rate is available. The precision of these flow meters is 2.5 both instruments on F_1, 1.515 for the one in F_2 and 1.418 for the one in F_3. In turn, the precision of the concentration measurements are 0.015 and 0.01 for y_1, respectively, and 0.01 for y_2 and y_3. Finally the pressure gage has a precision of 14. New flow meters are available. Their cost and precision are 350, 350, 400 and 2, 1.48 and 1.38, for F_1, F_2 and F_3, respectively. In turn, new composition measurements of cost 2700 and 0.01 precision are available. Finally, a new pressure gage of the same precision and cost of 100 can be installed.

Flow meters can be exchanged from F_1 and F_2 with a cost of 80 and vice versa but not transferred to F_3 and only the concentration measurement of y_1 can be transferred to y_2 at no cost and to y_3 at a cost of 50. The first row of the table of results represents the case for the existing instrumentation. A reduction of the standard deviation from 0.00438 to 0.00347 results if the laboratory analysis for the feed stream is relocated to the liquid stream and a pressure sensor is added. The cost of this case is 100. Higher precision (case 3) is obtained by means of the reallocation and addition of another measurement. When more precision is required (σ^* = 0.0031), no reallocation and instrument addition can achieve this goal.

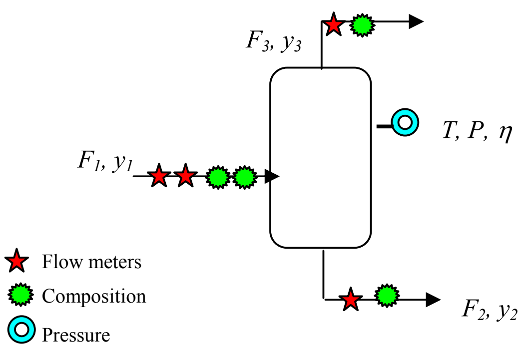


Figure 6. (Following Bagajewicz, 1997)

Table 3

σ^*	σ	*Cost*	*Reallocations*	*New Instruments*
∞	0.00438	-	-	-
0.0038	0.00352	100	-	*P*
	0.00347	100	y_1 to y_2	*P*
0.0033	0.00329	2800	y_1 to y_2	y_3 *P*

Reliable and Repairable Sensor Networks

Bagajewicz (2000) discusses the concepts of availability and reliability for the estimates of key variables and distinguishes them from the same concepts applied to instruments. The former are called *estimation* reliability/availability whereas the latter are called *service* reliability/availability. Ali and Narasimhan (1993) proposed to use an objective function defined as the minimum estimation reliability throughout the whole network. In other words, the reliability of the system is maximized by maximizing its weakest element. To address the fact that their representation does not control cost they proposed to limit the number of sensors to the minimum possible that will still guarantee observability. Although their procedure does not guarantee global optimality, it produces good results. Bilinear networks are also discussed by Ali and Narasimhan (1996) in detail. While this work used methods based on graph theory, genetic algorithms were successfully used by Sen *et al.* (1998) not only for reliable networks, but also a variety of other objective functions.

The cost based representation for the design of the sensor network subject to reliability constraints is (Bagajewicz and Sánchez, 2000a):

Minimize {*Total Cost of New Instrumentation*}
subject to
- *Desired reliability of Key variables*

The reliability of each variable is calculated using the failure probabilities of all the sensors participating in the corresponding material balances. If all sensors have the same cost one obtains a problem where the number of sensors is minimized. Finally, the representation due to Ali and Narasimhan (1993) can be put in the form of a minimum cost problem. The details of such equivalency can be found in the article by Bagajewicz and Sánchez, (2000a), where examples are shown. Finally, a single model containing precision and reliability constraints can be constructed.

When repairs are not performed, the service availability of a sensor is equal to its service reliability. In addition, the failure rate has been considered in a simplified way as a constant. However, in the presence of repairs, failure is no longer an event that depends on how many hours the sensor survived from the time it has been put in service. It is also conditioned by the fact that due to preventive or corrective maintenance, the sensor has been repaired at a certain time after being put in service. These events condition the failure rate. We thus distinguish unconditional from conditional events in failure and repair. These concepts are important because sensor maintenance cost accounts for nearly 20% of all maintenance cost (Masterson, 1999). Its reduction or containment is therefore essential. The connection between failure rate, repair rate and the expected number of repairs as well as illustrations of the impact of maintenance on sensor network design are described by Sánchez and Bagajewicz (2000).

Consider the simplified ammonia network (figure 7), for which flow meters for each stream may be selected from a set of three instruments with different precision, purchase cost and failure rate. Three flow meters are available at a cost of 350, 250 and 200, respectively. Their precision and failure rate are 1.5%, 2.5%, 3% and 0.3, 0.6, 0.7 failures/yr, respectively.

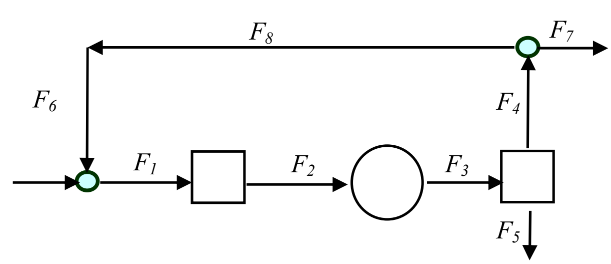


Figure 7: Simplified Ammonia Plant Network

Maintenance corrective costs include spare part and manpower cost of 10 and 40, respectively, and a life cycle of 5 years as well as an annual interest of 6% is used. The limits on the requirements of precision, residual precision and availability (probability of being running properly) are included for two selected flow rates in the next table:

Table 4. Design Requirements

Flow	Precision Requirements	Availability Requirements
F_1	-	0.9
F_2	1.5%	
F_5	2.5%	
F_7	-	0.9

The repair rate of instruments, a parameter that is a characteristic of the plant in consideration, has been varied between 1 and 20. The results of the optimization problem are presented for each case in the next table (Sánchez and Bagajewicz, 2000). (Prec. stands for precision and Avail. stands for Availability).

Table 5. *Results*

Repair Rate	*Measured Variables*	*Instrument Precision*	*Cost*	*Prec. (%) (F_2 F_5)*	*Avail. (F_1 F_7)*
1	F_1 F_4 F_5 F_6 F_7 F_8	3 % 1 % 1% 1% 3% 2%	2040	0.8067 1.2893	0.9021 0.9021
2	F_4 F_5 F_6 F_7 F_8	3 % 3 % 1% 3% 1%	1670	0.9283 1.9928	0.9222 0.9062
4	F_4 F_5 F_6 F_7 F_8	3 % 3 % 1% 3% 3%	1684	1.2313 1.9963	0.9636 0.9511
20	F_4 F_5 F_6 F_7 F_8	3 % 3 % 1% 3% 3%	1775	1.2313 1.9963	0.9983 0.9969

In the first case, the repair rate is comparatively low. Consequently the availability of instruments in the life cycle is also relatively low. To satisfy the availability of key variables, the optimal solution includes a set of six instruments. Three of these instruments are of type 1, which is a flow meter of low failure rate, high precision and the highest cost. For this reason precision and residual precision are better than the required values. When the repair rate is 2, an optimal solution exists that consists of five instruments. Two of these instruments are of type 1 and the rest are of type 3. Consequently, the total instrumentation cost decreases. A lower instrumentation cost is obtained for a repair rate equal to 4. In this case, although sensors are located on the same streams as in the previous case, one sensor of higher failure rate is installed to measure F_8. The results of the last case show that the influence of availability constraints decreases for high repair rates. The cost increases because of the effect of increasing the repair rate μ (from 4 to 20).

As a conclusion, the repair rate has a direct influence on the availability of a variable. If the repair rate is high, the design follows the requirements of precision and residual precision constraints. Thus, the availability of a variable is likely to force the design to have lower repair rates while cost may increase because to incorporate more instruments to calculate the variable by alternative ways.

Robust Sensor Networks

As it was described above, a robust sensor network provides meaningful values of precision, residual precision, variable availability, error detectability and resilience (Bagajewicz, 2000) These five properties encompass all the most desired features of a network. It is expected that more properties will be added to define robustness in the future. For example, when neural networks, wavelet analysis, principal component analysis (PCA), partial least squares (PLS) and other techniques for process monitoring are used, different definitions of robustness are warranted. A conceptual example showing the effect of robustness constraints are illustrated next

We now add residual precision capabilities to the example of figure 2. Consider now that residual precision of order one (precision left after one measurement is deleted from the set) is added to flows F_1 and F_4 as follows. The requirements are that precision should not drop below 1.5% and 3% respectively, when one measurement is lost. The solution is to put sensors of precision 2%, 3%, 3%, 3% in F_1 through F_4, respectively at a cost of 3,900. Assume now that residual precision is requested to the same level as precision (1.5% and 2% respectively). Then two alternative solutions with a cost of 5,500 are obtained as indicated in table 6. Not only the cost is higher but also there is one more degree of redundancy. For larger problems, the number of alternatives will increase, requiring new criteria to further screen alternatives.

Table 6. *Effect of Residual Precision*

S_1	S_2	S_3	S_4
1%	2%	2%	--
1%	2%	--	2%

We now turn to adding error detectability. As the capability of detecting smaller gross errors in the data increases, so does the precision of the sensor network. However, if the requirement is too stringent, no network may be able to satisfy it. If an error detectability level of 3.9 is required, the resulting network will be able to detect gross errors of 3.9 times the precision of the sensors. Using a statistical power of 50% is assumed for the detection algorithm, two solutions from a set of only 4 feasible solutions are found with a cost of 4,800.

Table 7. *Effect of error detectability*

S_1	S_2	S_3	S_4
1%	3%	--	2%
1%	3%	2%	--

If an error detectability level of 3.4 is requested, the problem has only one solution, namely flow meters with precision of 1%, 3%, 1%, 1% for F_1 through F_4, respectively, with a of cost of 8,300. Finally, we illustrate a resilience requirement, which limits the smearing effect of gross errors of a certain size when they are undetected. If a level of 3 times the standard deviation for all measurements is required, then the solution is to put sensors with precision 1%, 3%, 1%, 1% in F_1 through F_4, respectively, with a cost of 8,300. Relaxing (increasing) the resilience levels and maintaining the error detectability at the same level may actually lead to solutions of higher cost, even to infeasibility. Thus, robustness has a cost.

Multiobjective and Unconstrained Methods

One of the difficulties of the present approach (Bagajewicz, 1997; Chmielewski et al, 1999; Bagajewicz and Cabrera, 2001) is that the model is based on thresholds of different instrumentation network properties (such as residual precision, gross error detectability, resilience, etc). However, the process engineer has no feeling for what are the values of many of these properties that should be requested. To address the aforementioned limitations, a multicriteria approach to the problem in which all the network features are alternative objectives together with the cost has been proposed by various researchers: (Viswanath and Narasimhan, 2001; Carnero *et al*, 2001a,b, Bagajewicz and Cabrera, 2001). An alternative is to express every property of the instrumentation network in terms of its monetary value. Such an approach was done for precision related to quality (Bagajewicz, 2002) and for precision in flowrates for balances (Bagajewicz and Markowski, 2002). Since the last reference is part of this conference, we refer the reader to this article. We believe this is the future of this field.

Design of Sensor Networks for Process Fault Diagnosis

Process faults, which typically are rooted in some unit, propagate throughout the process, altering the readings of instruments (pressures, temperatures, flow rates, etc.). The task of detecting and identifying faults is different from that of gross error detection, which concentrates on instrument malfunction. As a consequence, the discrimination between instrument malfunction and process fault is an additional task of the alarm system. Therefore, the problem of designing an alarm system consists of determining the cost-optimal position of sensors, such that all process faults, single or multiple and simultaneous, can be detected and distinguished from instrument malfunction (biases). In addition, alarm settings need to be determined. We will concentrate now on the qualitative task of identifying the faults exclusively.

The first attempt to present a technique to locate sensors was done by Lambert (1977), where fault trees are used based on failure probabilities. Since fault trees cannot handle cycles the technique has not been developed further. Raghuraj *et al*. (1999) proposed an algorithm for one and multiple fault observability and resolution. They used directed graphs (DG), that is, graphs without signs. The arcs of the DG represent a "will cause" relationship, that is, an arc from node *A* to node *B* implies that *A* is a sufficient condition for *B*, which in general is not true for a Signed DG, where an arc represents a "can cause" relationship. The strategy used to solve the problem is based on identifying directed paths from root nodes where faults can occur to nodes where effects can be measured, called the observability set. Of all these paths, the objective is to choose the minimal subset of sensors from the observability set that would have at least one directed path from every root node.

Maximum fault resolution is a property that guarantees that the location and number of faults can be always achieved. Therefore, a *sensor network for maximum fault resolution is such that each fault has one and only one set of nodes from which it is observable.*

Bagajewicz and Fuxman (2001) presented a cost optimal model for fault resolution, which is illustrated next using a CSTR example introduced by Bhushan and Rengaswami (2000) (Figure 6). We reproduce here recent results obtained by Bagajewicz and Fuxman (2001).

The results of running a mathematical programming model for four cases are shown in the table below. These cases correspond to single and double fault detection capabilities. The table is organized as follows: the costs used in each case are first given followed by one or two columns depicting the solutions obtained using a cross to indicate that the corresponding node should be measured. Case 1 uses the same cost for all sensors, which is equivalent to minimize the number of sensors used. This set is an alternative set to the one obtained by Bhushan and Rengaswami (2000) (column 4). All cases have 5 sensors, except Case 4, which has 6.

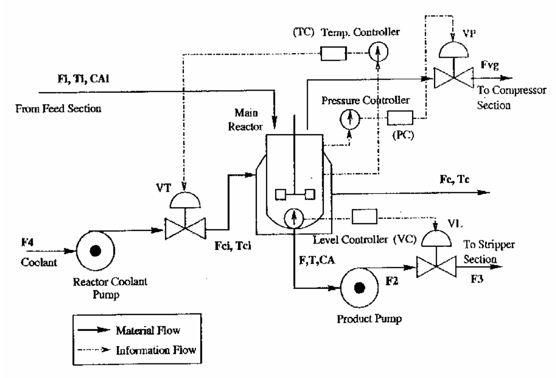


Figure 6. CSTR (From Bhushan and Rengaswami, 2000).

If one assumes that the cost of measuring CA is too high and, in addition, one does not want to use the controlled variables sensors as valid sensors for the problem of fault sensors searching, then one can alter the costs accordingly (Case 2) obtaining the result of column 6. This solution suggests that *N,* the number of moles in the gaseous phase, should be measured. If one wants to avoid this measurement one can assign *N* a high cost (case 3).

Furthermore, for this case we assigned a high cost to controllers and valves to avoid them as well. The result is shown in column 8. Case 4 is the case where controlled variable sensors can be used together with process variables (Case 4).

More recently, Bhushan and Rengaswami (2001, 2002) presented a minimum cost model based on process and sensor failure probabilities. An extension of these problems to include normal monitoring goals simultaneously with fault detection and resolution goals is warranted because of the expected synergistic effects.

Table 8. Effect of error detectability

Node name	Costs Case 1	q^*	q^*	Costs Case 2	q^*	Costs Case 3	q^*	Costs Case 4	q^*
CA	100	x	x	100		100		100	
TS	100			100		100		1	x
TC	100			1	x	100		100	
VT	100	x		1		100		100	
Fc	100		x	1		1	x	1	x
F4	100			1		1		1	
Tc	100	x	x	1	x	1	x	1	x
N	100			1	x	100		100	
PS	100			100		100		1	x
PC	100			1		100		100	
VP	100	x	x	1	x	100	x	100	x
Fvg	100			1		1	x	1	
VS	100			100		100		1	
VC	100	x		1	x	100		100	
VL	100			1		100		100	
F3	100			1		1	x	1	x
F2	100			1		1		1	
F	100		x	1		1		1	

Industrial Practice and Challenges

Instrumentation design, and even more importantly, instrumentation upgrade is less mature than data reconciliation and consequently it has not reached application in industry. It follows naturally, that as industry uses more and more data reconciliation, instrumentation upgrade will prevail as means of improving quality of monitoring. Several theoretical challenges are still unexplored.

Even though important attempts are being made to address the issue of alarms (Tsai and Chang, 1997) a model-based on cost-efficient alarm design is yet to be produced. Likewise, the direct incorporation of control performance measures as additional constraints of a cost-optimal representation has not been fully investigated yet. Finally, methods for cost-optimal instrumentation design corresponding to the implementation of other monitoring procedures, like principal component analysis (PCA), projection to latent structures (PLS), wavelet analysis, and neural networks, among others, have not been yet proposed. Robustness for example will need to be redefined when these other techniques are used.

The combination of all goals, that is, the design of a sensor network capable of performing estimation of key variables for control production accounting, parameter estimation for on-line optimization as well as fault detection, diagnosis and alarm (all of this in the context of detecting and assessing instrument biases and process leaks) is perhaps the long term goal of this field.

No software is available commercially, although the author has one prototype. We expect this to change rapidly, because the technology is now mature enough on the conceptual side and it is ready to face the challenges of computational efficiency. Practitioners will then provide new feedback for additional research directions.

Conclusions

Steady State data reconciliation and gross error detection is a mature field with certain challenges remaining. One of these challenges is the elimination of the uncertainties on the location of the gross errors, uncertainties that independent of the method of detection and compensation. Sensor Location is another emerging field, for which efficient techniques are being developed. This field is ready to be embraced by practitioners.

References

Albuquerque J. S. and L. T. Biegler. *Data Reconciliation and Gross-Error Detection for Dynamic Systems*. AIChE J., Vol. 42, No. 10, p. 2841, (1996).

Alhéritière C., N. Thornhill, S. Fraser and M. Knight. *Cost Benefit Analysis of Refinery Process Data: Case Study*. Comp. Chem. Eng., 22, Suppl., pp. S1031-S1034, (1998).

Ali, Y. and S. Narasimhan. *Sensor Network Design for Maximizing Reliability of Linear Processes*. AIChE J., 39, 5, pp. 2237-2249, (1993).

Ali Y. and S. Narasimhan. *Sensor Network Design for Maximizing Reliability of Bilinear Processes*. AIChE J., 42, 9, pp. 2563-2575, (1996).

Almasy, G. A. and R. S. H. Mah, *Estimation of Measurement Error Variances from Process Data*, Ind. Eng. Chem. Proc. Des. Dev., 23, pp. 779, (1984).

Bagajewicz M. *Optimal Sensor Location in Process Plants*. AIChE Journal. Vol. 43, No. 9, 2300, September (1997).

Bagajewicz, M. *Design and Upgrade of Process Plant Instrumentation.* Technomic. (2000)

Bagajewicz M. *An Unconstrained Approach for Instrumentation Network Design and Upgrade.* Proceedings of ESCAPE 12, May 26-29, The Hague, Netherlands (2002).

Bagajewicz M**.** and E. Cabrera**.** *A New MILP Formulation for Instrumentation Network Design and Upgrade.*

Proceedings of the 4th IFAC Workshop on On-Line Fault Detection & Supervision in the Chemical Process Industries, June 8-9, Seoul, Korea (2001). To appear in AIChE Journal.

Bagajewicz M. and A. Fuxman. *An MILP Model For Cost Optimal Instrumentation Network Design And Upgrade For Fault Detection.* Proceedings of the 4th IFAC Workshop on On-Line Fault Detection & Supervision in the Chemical Process Industries, June 8-9, Seoul, Korea (2001).

Bagajewicz M. and M.C. Gonzales. *Is the Practice of Using Unsteady Data to Perform Steady State Reconciliation Correct?,* AIChE Spring Meeting. Houston, (2001).

Bagajewicz, M. and Q. Jiang, *Gross Error Modeling and Detection in Plant Linear Dynamic Reconciliation*. Computers and Chemical Engineering, 22, 12, 1789-1810 (1998).

Bagajewicz M. and Q. Jiang. *A Mixed Integer Linear Programming-Based Technique for the Estimation of Multiple Gross Errors in Process Measurements.* Vol. 177, pp. 139-155, (2000).

Bagajewicz M. and Q. Jiang. *Comparison of Steady State and Integral Dynamic Data Reconciliation.* Computers & Chemical Engineering. Vol. 24, No 11, pp. 2367-2518 (2000).

Bagajewicz M. and M. Markowski. *Instrumentation Design and Upgrade Using an Unconstrained Method with Pure Economical Objectives.* Proceedings of FOCAPO. Coral Springs, FL (2003).

Bagajewicz, M. and D. Rollins. *On the Consistency of the Measuremnat and GLR tests for Gross Error Detection*. Submitted to Comp. & Chem. Eng. (2002).

Bagajewicz M. and M. Sánchez. *Sensor Network Design and Upgrade for Plant Parameter Estimation*. Computers and Chemical Engineering. Vol 23., Supp., pp. S593-S596 (1999a).

Bagajewicz M. and M. Sánchez. *Duality of Sensor Network Design Models for Parameter Estimation*. AIChE J., 45, 3, pp. 661-664 (1999b)

Bagajewicz M. and M. Sánchez. *Design and Upgrade of Nonredundant and Redundant Linear Sensor Networks*. AIChE J. Vol. 45, No. 9., pp. 1927-1939. (1999c)

Bagajewicz M. and M. Sánchez. *On the Impact of Corrective Maintenance in the Design of Sensor Networks.* Industrial and Engineering Chemistry Research. Vol. 39, no. 4, pp. 977-981 (2000a).

Bagajewicz M. and M. Sánchez. *Reallocation and Upgrade of Instrumentation in Process Plants*. Computers and Chemical Engineering. 24, 8, pp. 1961-1980 (2000b).

Bagajewicz M. and M. Sánchez. *Cost-Optimal Design of Reliable Sensor Networks.* Computers and Chemical Engineering. Vol. 23, 11/12, pp. 1757-1762 (2000c).

Bagajewicz, M., Q. Jiang, and M. Sanchez. *Removing and Assessing Uncertainties in Two Efficient Gross Error Collective Compensation Methods.* To Appear, Chem. Eng. Comm. (1999).

Bhushan; M. and Rengaswamy, R. *Design of Sensor Network Based on the Signed Directed Graph of the Process for Efficient Fault Diagnosis.* Ind. Eng. Chem. Res. 39(4) pp. 999-1019, **(**2000**).**

Bhushan and R. Rengaswami M**.** *Design of Sensor Networks Based on the Signed Directed Graph of the Process for Efficient Fault Diagnosisi.* Ind & Eng. Chem. Res., 39, pp. 999-1019 (2001).

Bhushan; M. and Rengaswamy, R. *Comprehensive Design of a Sensor Network for Chemical Plants Based on Various Diagnosability and Reliability Criteria. 1. Framework***.** Ind. Eng. Chem. Res. 41(7) pp 1826 – 1839, **(**2002**).**

Bhushan; M. and Rengaswamy, R. *Comprehensive Design of a Sensor Network for Chemical Plants Based on Various Diagnosability and Reliability Criteria. 1. Applications.* Ind. Eng. Chem. Res. 41(7) pp 1840 – 1860, **(**2002**).**

Carnero, M, J. Hernandez, M. Sanchez and A Bandoni. Multiobjective Sensor Network Design. Proceedings of EMPROMER 2001, Volume I, 325-330, (2001)

Chen J., A. Bandoni and J. A. Romagnoli. *Robust Estimation of Measurement Error Variance/ Covariance from Process Sampling Data.* Computers Chem. Eng. 21, 6, pp. 593-600, (1997).

Chmielewski D., T. E. Palmer and V. Manousiouthakis. *Cost Optimal Retrofit of Sensor Networks with Loss Estimation Accuracy.* AIChE Annual Meeting, Dallas, (1999).

Crowe C. M. *Formulation of Linear Data Reconciliation using Information Theory*. Computers Chem. Eng. 51,12, 3359-3366 (1996).

Crowe C, Y. Garcia Campos and A. Hrymak. *Reconciliation of Process Flow Rates by Matrix Projection. I. The Linear Ca*se. AIChE J., 29,818, (1983).

Darouach M., R. Ragot, M. Zasadzinski and G. Krzakala. *Maximum Likelihood Estimator of Measurement Error Variances in Data Reconciliation.* IFAC. AIPAC Symp. 2, pp. 135-139, (1989).

Gelb A., Editor. *Applied Optimal Estimation*. The M.I.T. Press, Cambridge, Massachusetts, (1974).

Iordache C., R. Mah and A. Tamhane. *Performance Studies of the Measurement Test for detection of Gross Errors in Process Data,* AIChE J., 31, 1187, (1985).

Jiang and M. Bagajewicz. *On A Strategy of Serial Identification with Collective Compensation for Multiple Gross Error Estimation in Linear Steady State Reconciliation*. Ind. & Eng. Chem. Research, 38, 5,2119-2128 (1999).

Johnston L.P.M. and M. A. Kramer. *Maximum Likelihood Data Rectification. Steady State Systems*. AIChE J., 41, 11 (1995).

Kalman, R. E., *New Approach to Linear Filtering and Prediction Problems*, J. Basic Eng., ASME, 82D, 35 (1960).

Keller, J. Y, M. Darouach and G. Krzakala, *Fault detection of Multiple Biases or process leaks in Linear Steady State Systems.* Comp. & Chem. Eng., 18, 1001(1994).

Keller, J. Y, M. Zasadzinski and M. Darouach*, Analytical Estimator of Measurement Error Variances in Data Reconciliation*, Comp. & Chem. Eng. 16, 185, (1992).

Kim I., M. S. Kang, S. Park and T. Edgar, *Robust Data Reconciliation and Gross Error Detection: The Modified MIMT using NLP*. Comp. & Chem. Eng., 21, 775(1997).

Kretsovalis, A. and R. S. H. Mah. *Effect of Redundancy on Estimation Accuracy in Process Data Reconciliation.* Chem. Eng. Sci., 42, 2115, (1987a)

Krishnan S., G. Barton and J. Perkins. *Robust Parameter Estimation in On-line Optimization – Part I. Methodology and Simulated Case Study*. Comp. Chem. Eng, 16**,** pp. 545-562, (1992a).

Krishnan S., G. Barton and J. Perkins. *Robust Parameter Estimation in On-line Optimization – Part II. Application to an Industrial Process.* Comp. Chem. Eng., 17, pp. 663-669, (1992b).

Lambert H.E. *Fault trees for Locating Sensors in Process Systems.* CEP, August, pp. 81-85 (1977)

Madron F., *Process Plant Performance, Measurement Data Processing for Optimization and Retrofits*. Ellis Horwood, West Sussex, England (1992).

Madron F. and V. Veverka. *Optimal Selection of Measuring Points in Complex Plants by Linear Models.* AIChE J., 38, 2, pp. 227, (1992).

Mah, R. S. H., *Chemical Process Structures and Information Flows.* Butterworths, Stoneham, MA, USA (1990).

Mah, R. S. H. and A. C. Tamhane, *Detection of Gross Errors in Process Data.* AIChE J., 28, 828 (1982).

Masterson J. S. *Reduce Maintenance Costs with Smart Field Devices.* Hydrocarbon Processing, January, (1999).

Narasimhan S. and R. S. H. Mah *Maximum Power Test for Gross Error Detection Using Generalized Likelihood Ratios*. AIChE J. 32, 1409 (1986).

Narasimhan, S. and R. S. H. Mah. *Generalized Likelihood Ratio Method for Gross Error Detection.* AIChE J., 33, 1514(1987).

Narasimhan, S. and R. S. H. Mah, *Generalized Likelihood Ratios for Gross Error Identification in Dynamic Processes*, AIChE J., 34, 1321(1988).

Narasimhan S. and C. Jordache. *Data Reconciliation & Gross Error Detection. An Intelligent Use of Process Data*. Gulf Publishing Company, Houston (2000).

Raghuraj R., M. Bhushan and R. Rengaswamy *Locating Sensors in Complex Chemical Plants based on Fault Diagnostic Observability Criteria".* AIChE J., 45, 2, pp. 310-322 (1999).

Ripps D. L., *Adjustment of Experimental Data*. Chem. Eng. Prog. Symp. Ser., 61, 8–13 (1965).

Rollins D. K., Y. Cheng and S. Devanathan. *Intelligent Selection of Hypothesis tests to Enhance Gross Error Identification.* Comp. and Chem. Eng., 20, 5, pp. 517530 (1996).

Rollins, D. K. and J. F. Davis. *Unbiased Estimation of Gross Errors in Process Measurements.* AIChE J., 38,563(1992).

Romagnoli J. and M. Sánchez. *Data Processing and Reconciliation for Chemical Processes Operations.* Academic Press, (1999).

Romagnoli J. and G. Stephanopoulos. *On the Rectification of Measurement Errors for Complex Chemical Plants*. Chem. Eng. Sci. 35, 5, 1067-1081, (1980).

Rosenberg, J., R. S. H. Mah and C. Iordache. *Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data.* Ind. Eng. Chem. Res., 26, 555(1987).

Sánchez M. and M. Bagajewicz. *On the Impact of Corrective Maintenance in the Design of Sensor Networks.* Industrial and Engineering Chemistry Research. Vol. 39, no. 4, pp. 977. (2000).

Sánchez M. and J. Romagnoli. *Monitoreo de Procesos Continuos: Análisis Comparativo de Técnicas de Identificación y Cálculo de Bias en los Sensores.* AADECA 94 - XIV Simposio Nacional de Control Automático (1994).

Sánchez M. and J. Romagnoli. *Use of Orthogonal Transformations n Classification/Data Reconciliation*. Comp. Chem. Engng, 20, 483- 493 (1996).

Sánchez, M., J. Romagnoli, Qiyou Jiang and M. Bagajewicz. *Simultaneous Estimation of Biases and Leaks in Process Plants.* Computers and Chemical Engineering. 23, 7, 841-858 (1999)

Sen S., S. Narasimhan and K. Deb. *Sensor Network Design of Linear Processes using Genetic Algorithms*. Comput. Chem. Eng, 22, 3, pp. 385-390, (1998).

Swartz C. L. E., *Data Reconciliation for Generalized Flowsheet Applications.* American Chemical Society of National Meeting. Dallas, TX (1989).

Tjoa I. B. and L. T. Biegler. *Simultaneous Strategies for Data Reconciliation and Gross Error detection of Nonlinear Systems*. Comp. and Chem. Eng., 15, 10, pp. 679-690 (1991).

Tong H. and C. M. Crowe*. Detection of Gross Errors in Data Reconciliation by Principal Component Analysis*. AIChE J., 41, 7, 1712, (1995)

Tsai C. S. and C. T. Chang. *Optimal Alarm Logic Design for Mass-Flow Networks*. AIChE J., 43,11, pp. 3021, (1997).

Viswanath, A. and S. Narasimhan. Multiobjective Sensor Network Design Using Genetic Algorithms. Proceedings of the 4th IFAC Workshop. Jejudo (Chejudo) Island, Korea, (2001).