DATA RECONCILIATION AND INSTRUMENTATION UPGRADE. OVERVIEW AND CHALLENGES

PASI 2005.

Cataratas del Iguazu, Argentina

Miguel Bagajewicz

University of Oklahoma

OUTLINE

- **A LARGE NUMBER OF PEOPLE FROM ACADEMIA AND INDUSTRY HAVE CONTRIBUTED TO THE AREA OF DATA RECONCILIATION.**
- **HUNDREDS OF ARTICLES AND THREE BOOKS HAVE BEEN WRITTEN.**
- **MORE THAN 5 COMMERCIAL SOFTWARE EXIST.**

- **ALTHOUGH A LITTLE YOUNGER, THE AREA OF INSTRUMENATION UPGRADE IS EQUALLY MATURE**
- **ONE BOOK HAS BEEN WRITTEN**

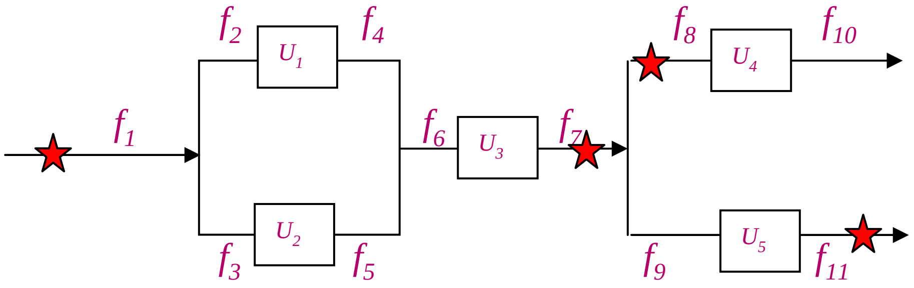
OUTLINE

- **OBSERVABILITY AND REDUNDANCY**
- **DIFFERENT TYPES OF DATA RECONCILIATION**
  - **- Steady State vs. Dynamic**
  - **- Linear vs. Nonlinear**

- **GROSS ERRORS**
  - **- Biased instrumentation, model mismatch and outliers**
  - **- Detection, identification and size estimation**

- **INSTRUMENTATION UPGRADE**
- **SOME EXISTING CHALLENGES**
- **INDUSTRIAL PRACTICE**

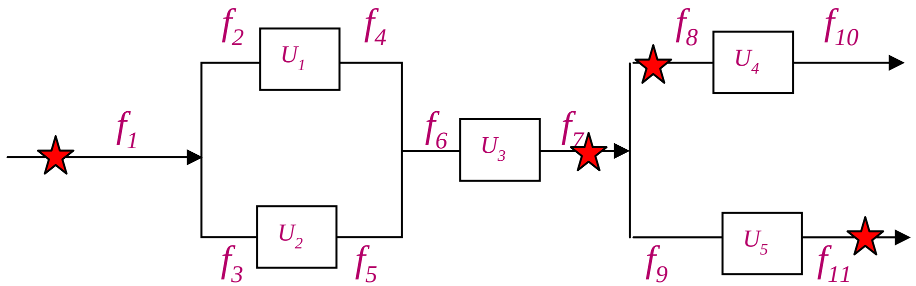
Simple Process Model of Mass Conservation



$$\left.\begin{aligned}
f_1 - f_2 - f_3 &= 0\\
f_2 - f_4 &= 0\\
f_3 - f_5 &= 0\\
f_4 + f_5 - f_6 &= 0\\
f_6 - f_7 &= 0\\
f_7 - f_8 - f_9 &= 0\\
f_8 - f_{10} &= 0\\
f_9 - f_{11} &= 0
\end{aligned}\right\}$$

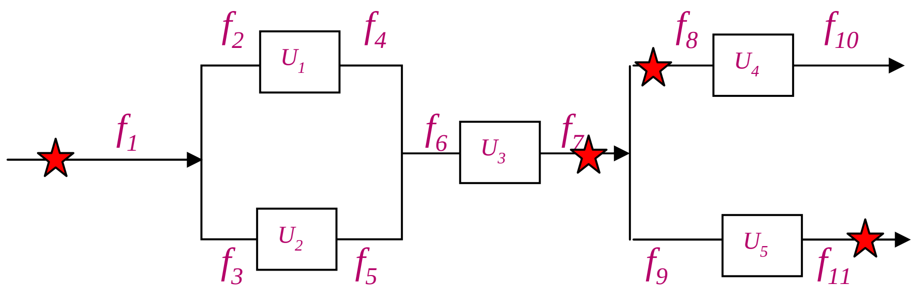
Material Balance Equations

Variable Classification



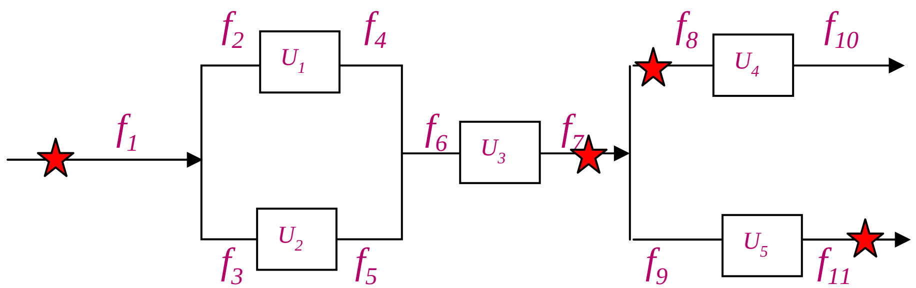
Variables
- ***Measured (M)***
- ***Unmeasured (UM)***
  - ***Observable (O)***
  - ***Unobservable (UO)***

Variable Classification



Variables
- ***Measured (M)***
  - ***Redundant (R)***
  - ***Non-redundant (NR)***
- ***Unmeasured (UM)***
  - ***Observable (O)***
  - ***Unobservable (UO)***

Conflict among Redundant Variables



$$\left.\begin{aligned} f_1 - f_7 &= 0 \\ f_1 - f_8 - f_{11} &= 0 \end{aligned}\right\}$$

Material Balance Equations

Conflict Resolution

$$\left.\begin{aligned}&\boldsymbol{Min}\ [\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} - \boldsymbol{f}_{\boldsymbol{R}}^{+}\ ]^{T}\boldsymbol{Q}_{\boldsymbol{R}}^{-1}[\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} - \boldsymbol{f}_{\boldsymbol{R}}^{+}\ ]\\ &\quad s.t.\\ &\quad \boldsymbol{E}_{\boldsymbol{R}}\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} = 0\end{aligned}\right\}$$

Data reconciliation in

its simplest form

Analytical Solution

$$\tilde{\boldsymbol{f}}_{\boldsymbol{R}} = \left[\boldsymbol{I} - \boldsymbol{Q}_{\boldsymbol{R}}\boldsymbol{E}_{\boldsymbol{R}}^{T}\left(\boldsymbol{E}_{\boldsymbol{R}}\boldsymbol{Q}_{\boldsymbol{R}}\boldsymbol{E}_{\boldsymbol{R}}^{T}\right)^{-1}\boldsymbol{E}_{\boldsymbol{R}}\right]\boldsymbol{f}_{\boldsymbol{R}}^{+}$$

Precision of Estimates

If $z = \Gamma x$, and the variance of x is Q, then the variance of z is given by: $\tilde{Q} = \Gamma Q \Gamma^T$

$$\tilde{f}_R = \left[I - Q_R E_R^T \left(E_R Q_R E_R^T \right)^{-1} E_R \right] f_R^+ \longrightarrow \tilde{Q}_{R,F} = Q_{R,F} - Q_{R,F} C_R^T \left(C_R Q_{R,F} C_R^T \right)^{-1} C_R Q_{R,F}$$

$$\tilde{f}_{NR} = f_{NR}^+ \qquad \tilde{Q}_{NR,F} = Q_{NR,F}$$

$$\tilde{f}_O = C_{RO} \tilde{f}_R + C_{NRO} \tilde{f}_{NR} \qquad \tilde{Q}_O = [C_{RO} \; C_{SRO}] \begin{bmatrix} \tilde{Q}_R \\ Q_{NR} \end{bmatrix} [C_{RO} \; C_{SRO}]^T$$

Some Practical Difficulties

- Variance-Covariance matrix is not Known
- Process plants have a usually a large number of Tanks
- Plants are not usually at Steady State
- How many measurements is enough?

Estimation of the Variance-Covariance Matrix.

•Direct Approach

$$\begin{cases} \bar{f}_{R,i} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \tilde{f}_{R,i}\big|_k \\ Cov\left(\tilde{f}_{R,i}, \tilde{f}_{R,j}\right) = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}\left(\tilde{f}_{R,i}\big|_k - \bar{f}_{R,i}\right)\left(\tilde{f}_{R,j}\big|_k - \bar{f}_{R,j}\right) \end{cases}$$

•Indirect Approach

Almasy and Mah (1984), Darouach et al., (1989) and Keller et al (1992)

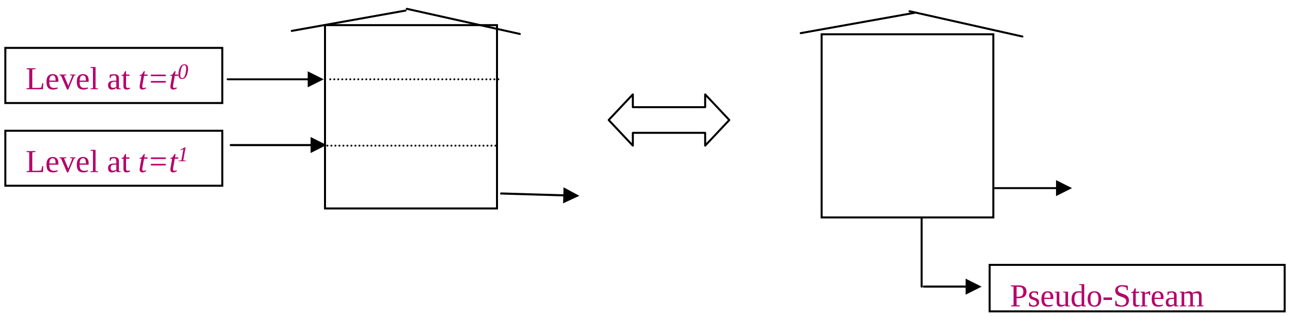
$$E_R\, f_R^+ = r \longrightarrow Cov(r) = E_R Q_R E_R^T$$

1) Obtain r

2) Maximum likelihood estimate Q_R

However, this procedure is not good if outliers are present. Robust estimators have been proposed (Chen et al, 1997)

Tank Hold Up Measurements



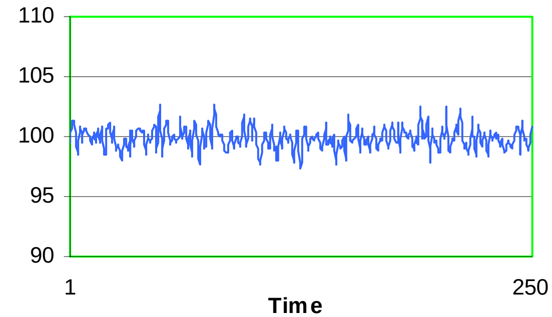
Steady State formulations are used

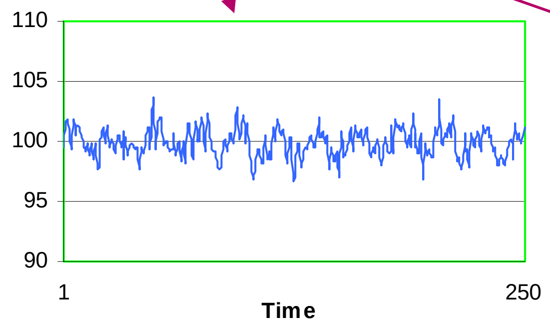
The procedure is based on the following assumptions:

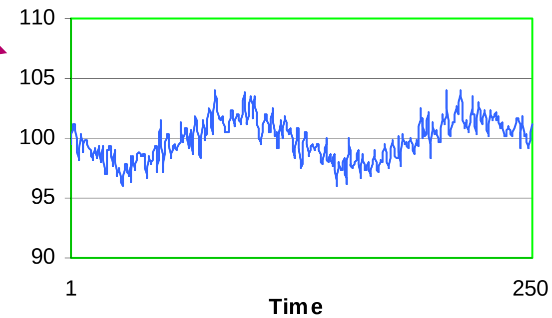
$$\left.\begin{array}{l} \boldsymbol{Min}\ [\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} - \boldsymbol{f}_{\boldsymbol{R}}^{+}\ ]^{T}\ \boldsymbol{Q}_{\boldsymbol{R}}^{-1}\ [\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} - \boldsymbol{f}_{\boldsymbol{R}}^{+}\ ] \\ \\ \boldsymbol{E}_{\boldsymbol{R}}\ \tilde{\boldsymbol{f}}_{\boldsymbol{R}} = 0 \end{array}\right\}$$

a) A normal distribution of measurement errors.
b) A single value per variable.
c) A "steady-state" system.

a) Substantiated by the central limit theorem.
b) Also valid for means.
c) No plant is truly at "steady-state".
Process oscillations occur. Therefore, it is said that it is valid for a "pseudo-steady state" system"







Reconciliation of averages is equal to the average of reconciled values using dynamic data reconciliation (Bagajewicz and Jiang, 2000; Bagajewicz and Gonzales,2001).

That is, there is no need to adjust the variance-covariance matrix for process variations.

Dynamic Data Reconciliation

Linear Case(after cooptation):

$$Min \sum_{\forall i} \left\{ [\tilde{f}_{Ri} - f_{Ri}^{+}]^T Q_{Rf}^{-1} [\tilde{f}_{Ri} - f_{Ri}^{+}] + [\tilde{V}_{Ri} - V_{Ri}^{+}]^T Q_{RV}^{-1} [\tilde{V}_{Ri} - V_{Ri}^{+}] \right\}$$

$$B \frac{d\tilde{V}_R}{dt} = A\tilde{f}_R$$
$$C\tilde{f}_R = 0$$

When B=I, the Kalman filter can be used.

Dynamic Data Reconciliation

$$Min \sum_{\forall i} \left\{ [\tilde{f}_{Ri} - f_{Ri}^{+}]^T Q_{Rf}^{-1} [\tilde{f}_{Ri} - f_{Ri}^{+}] + [\tilde{V}_{Ri} - V_{Ri}^{+}]^T Q_{RV}^{-1} [\tilde{V}_{Ri} - V_{Ri}^{+}] \right\}$$

$$B \frac{d\tilde{V}_R}{dt} = A\tilde{f}_R$$

$$C\tilde{f}_R = 0$$

Difference Approach: Darouach, M. and M. Zasadzinski, 1991, Rollins, D. K. and S. Devanathan, 1993.

$$B\left(\tilde{V}_{R,i} - \tilde{V}_{R,i}\right) = A\tilde{F}_{R,i} \qquad C\tilde{F}_{R,i} = 0$$

An algebraic system of equations follows.

Integral Approach: Jiang and Bagajewicz, 1997.

$$f_R \approx \sum_{k=0}^{s} \alpha_k^R t^k \qquad B_R[V_R - V_{R0}] = B_R \sum_{k=0}^{s} \omega_{k+1}^R t^{k+1} = A_R \sum_{k=0}^{s} \frac{\alpha_k^R}{k+1} t^{k+1}$$

The technique estimates the coefficients of polynomials.

Nonlinear Data Reconciliation

$$Min \sum_{k=0}^{N} [\tilde{x}_M(t_k) - z_{M,k}]^T Q^{-1} [\tilde{x}_M(t_k) - z_{M,k}]$$

$$\frac{d\tilde{x}_1}{dt} = g_1(\tilde{x}_1, \tilde{x}_2)$$

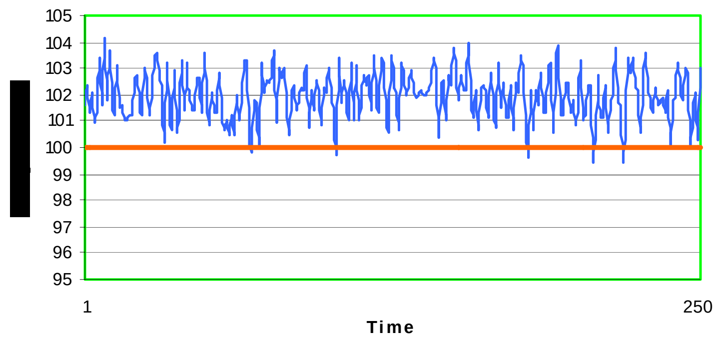
$$g_2(\tilde{x}_1, \tilde{x}_2) = 0$$

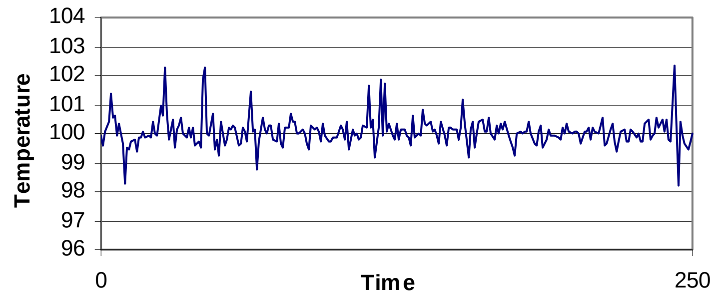
Applied in practice to <u>steady state</u> models with material, component and energy balances. In the dynamic case, orthogonal collocation was used (Liebmann et al, 1992) or linearization (Ramamurthi et al.,1993) or use of DAE (Albuquerque and Biegler, 1996).

Gross Errors

Types of Gross Errors

- Biases
- Leaks (Model departures)
- True outliers





Hypothesis Testing

Global Test (Detection)

$$\gamma = r(E_R Q_R E_R^T)^{-1} r$$

$$\begin{cases} H_0 : \mu_R^T (E_R Q_R E_R^T) \mu_R = 0 \\ H_1 : \mu_R^T (E_R Q_R E_R^T) \mu_R \neq 0 \end{cases}$$

$$Distribution : Chi - Squared$$

Nodal Test (Detection and Identification)

$$Z_i^N = n^{1/2} \frac{|r_i|}{\sqrt{(E_R Q_R E_R^T)_{ii}}}$$

$$\begin{cases} H_0 : \mu_r = 0 \\ H_1 : \mu_r \neq 0 \end{cases}$$

$$Distribution : Normal$$

Maximum Power versions of this test were also developed. Rollins et al (1996) proposed an intelligent combination of nodes technique

Hypothesis Testing

Principal Component (Tong and Crowe, 1995)

W_r : *matrix of eigenvectors of* $(E_R Q_R E_R^T)$

Λ_r : *matrix of eigenvalues of* $(E_R Q_R E_R^T)$

$p_r = W^T r$ $\quad\quad p_r \sim N(0, I)$ $\longrightarrow$

$$\begin{cases} H_0 : W^T \mu_r = 0 \\ H_1 : W^T \mu_r \neq 0 \end{cases}$$

$Distribution : Normal$

Hypothesis Testing

Measurement Test

$$Z_{\chi}^{LCT} = \frac{\tilde{f}_i - f_i^{+}}{\sqrt{(\tilde{Q}_R)_{ii}}}$$

$$\begin{cases} H_0 : \phi_i - f_i = 0 \\ H_1 : \phi_i - f_i \neq 0 \end{cases}$$

$$Distribution : Normal$$

This test is inadmissible. Under deterministic conditions it may point to the wrong location.

Hypothesis Testing

Generalized Likelihood ratio

$$\lambda = \sup_{\forall i} \frac{\Pr\{r|H_1\}}{\Pr\{r|H_0\}}$$

$$\begin{cases} H_0 : \mu_r = 0 \\ H_1 : \mu_r = bAe_i \end{cases}$$

$$Distribution : Chi - Squared$$

$$\downarrow$$

$$\lambda = \sup_{\forall b,i} \frac{\exp\{-0.5(r - bAe_i)^T (E_R Q E_R^T)^{-1} (r - bAe_i)\}}{\exp\{-0.5 r^T (E_R Q E_R^T)^{-1} r\}}$$

Leaks can also be tested.

Multiple Error Detection

Serial Elimination

Apply recursively the test and eliminate the measurement

Serial Compensation

Apply recursively the test, determine the size of the gross error and adjust the measurement

Serial Collective Compensation

Apply recursively the test, determine the sizes of all gross error and adjust the measurements

Multiple Error Detection

Unbiased Estimation

One shot collective information of all possible errors followed by hypothesis testing. Bagajewicz and Jiang, 2000, proposed an MILP strategy based on this.

Two distributions approach

Assume that gross error have a distribution with larger variance and use maximum likelihood methods (Romagnoli et al., 1981) (Tjoa and Biegler, 1991) (Ragot et al., 1992)

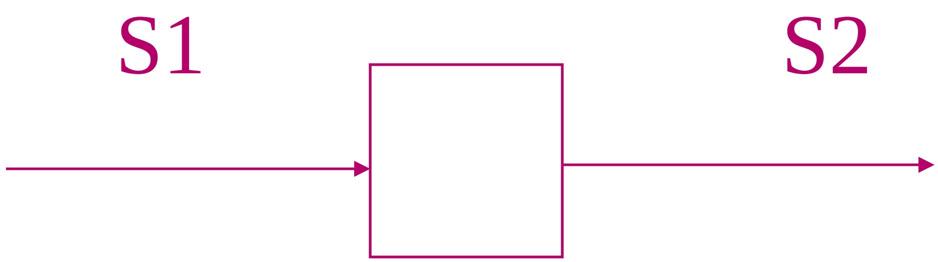
Multiscale Bayesian approach. Bakshi et al (2001).

EQUIVALENCY THEORY

- EXACT LOCATION DETERMINATION IS NOT ALWAYS POSSIBLE, REGARDLESS OF THE METHOD USED.
- MANY SETS OF GROSS ERRORS ARE EQUIVALENT, THAT IS, THEY HAVE THE SAME EFFECT IN DATA RECONCILIATION WHEN THEY ARE COMPENSATED.

BASIC EQUIVALENCIES

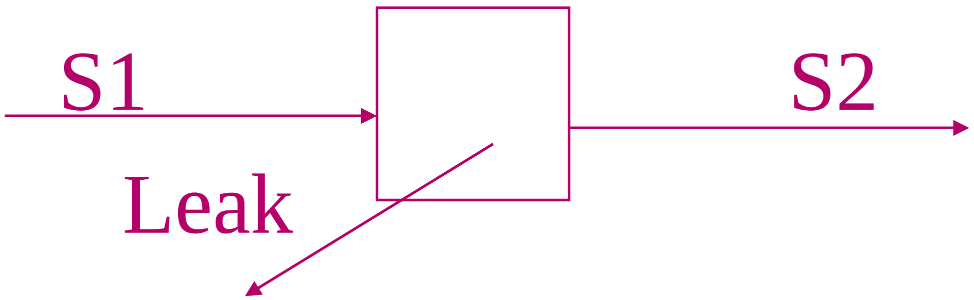
In a single unit a bias in an inlet stream is equivalent to a bias in an output stream.



		S_1	S_2
Measurement		4	3
Case 1	Reconciled data	3	3
	Estimated bias	1	
Case 2	Reconciled data	4	4
	Estimated bias		-1

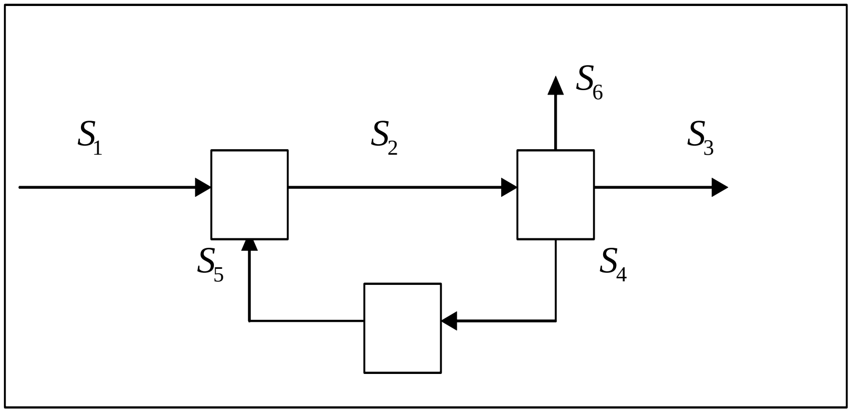
BASIC EQUIVALENCIES

In a single unit a bias in a stream is equivalent to a leak



		S_1	S_2	*Leak*
Measurement		4	3	
Case1	Reconciled data	4	3	
	Estimated bias/leak			1
Case2	Reconciled data	4	4	
	Estimated bias/leak		-1	

EQUIVALENCY THEORY



For the set $\Lambda=\{S_3, S_6\}$ a gross error in one of them can be alternatively placed in the other without change in the result of the reconciliation. We say that this set has Gross Error Cardinality $\Gamma(\Lambda)=1$. ONE GROSS ERROR CAN REPRESENT ALL POSSIBLE GROSS ERRORS IN THE SET.

GROSS ERROR DETECTION

TWO SUCCESFUL IDENTIFICATIONS:

- Exact location
- Equivalent location

THIS MEANS THAT THE CONCEPT OF POWER IN LINEAR DATA RECONCILIATION SHOULD BE REVISITED TO INCLUDE EQUIVALENCIES

COMMERCIAL CODES

Package	Nature	Offered by
IOO (Interactive On-Line Opt.)	Academic	Louisiana State University (USA)
DATACON	Commercial	Simulation Sciences (USA)
SIGMAFINE	Commercial	OSI (USA)
VALI	Commercial	Belsim (Belgium)
ADVISOR	Commercial	Aspentech (USA)
RECONCILER	Commercial	Resolution Integration Solutions (USA)
PRODUCTION BALANCE	Commercial	Honeywell (USA)
RECON	Commercial	Chemplant Technologies (Czech Republic)

While the data reconciliation in all these packages is good, gross error detection has not caught with developments in the last 10 years.

Global test and Serial Elimination using the measurement test seem to be the gross error detection and identification of choice.

INSTRUMENTATION UPGRADE
(The inverse engineering problem)

Given

Data Reconciliation (or other) monitoring Objectives

Obtain:

Sensor Locations

(number and type)

INSTRUMENTATION DESIGN

Minimize Cost (Investment + Maintenance)

s.t.

- -Desired precision of estimates
- -Desired gross error robustness
  - Detectability, Residual Precision, Resilience.
- -Desired reliability/availability

Design of Repairable Networks

EXAMPLE: Ammonia Plant

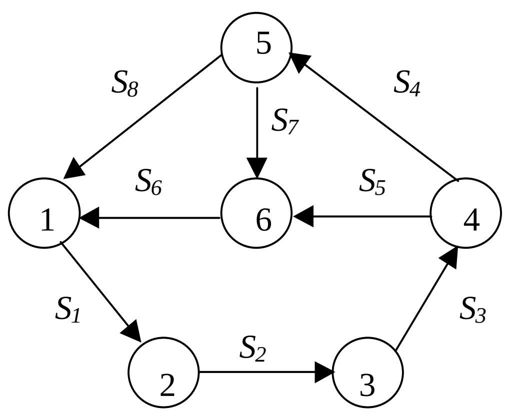


Table 3: Optimization results for the simplified ammonia process flowsheet

Repair Rate	*Measured Variables*	*Instrument Precision (%)*	*Cost*	*Precision(%) (S2) (S5)*	*Precision Availability(%) (S2) (S5)*	*Availability (S1) (S7)*
1	S1 S4 S5 S6 S7 S8	3 1 1 1 3 2	2040.2	0.8067 1.2893	0.9841 1.2937	0.9021 0.9021
2	S4 S5 S6 S7 S8	3 3 1 3 1	1699.8	0.9283 1.9928	1.9712 2.0086	0.9222 0.9062
4	S4 S5 S6 S7 S8	3 3 1 3 3	1683.7	1.2313 1.9963	1.9712 2.0086	0.9636 0.9511
20	S4 S5 S6 S7 S8	3 3 1 3 3	1775.2	1.2313 1.9963	1.9712 2.0086	0.9983 0.9969

- ***There is a minimum in cost as a function of the repair rate. This allows the design of maintenance policies.***

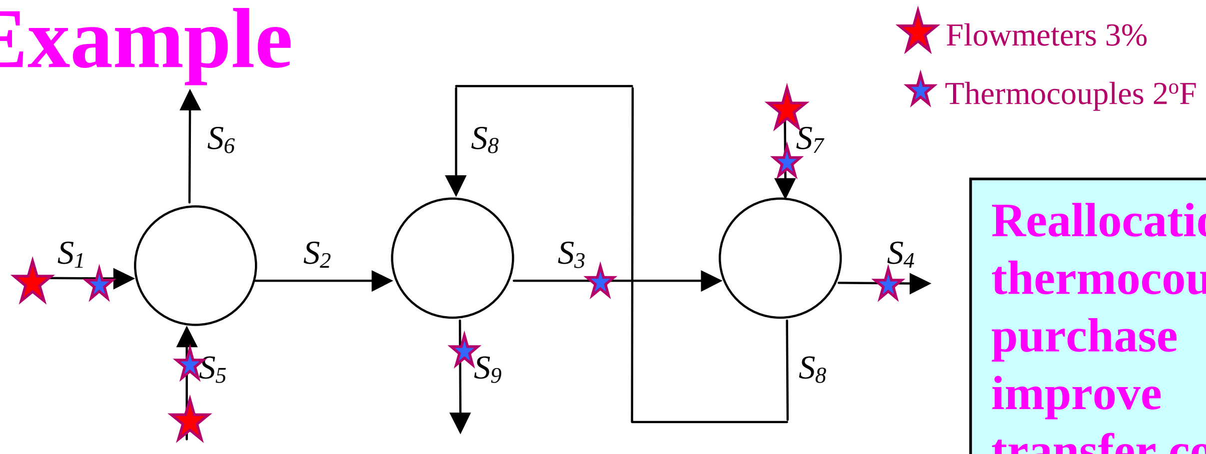
Upgrade

Upgrade consists of any combination of :

- **Adding instrumentation.**
- **Replacing instruments.**
- **Relocating instruments (thermocouples, sampling places, etc).**

Upgrade

Example



Reallocation and/or addition of thermocouples as well as a purchase of a new flowmeter improve the precision of heat transfer coefficients

Case	$\sigma^*_{U_1}$	$\sigma^*_{U_2}$	$\sigma^*_{U_3}$	σ_{U_1}	σ_{U_2}	σ_{U_3}	c	Reallocations	New Instruments
1	4.00	4.00	4.00	3.2826	1.9254	2.2168	100	$u_{T_1,2,T_6}$	-
2	2.00	2.00	2.00	-	-	-	-	-	-
3	2.00	2.00	2.20	1.3891	1.5148	2.1935	3000	-	T_2, T_6
4	1.50	1.50	2.20	1.3492	1.3664	2.1125	5250	-	F_4, T_2, T_6
5	2.40	2.30	2.20	2.0587	1.8174	2.1938	1500	-	T_6
6	2.20	1.80	2.40	1.7890	1.6827	2.2014	1600	$u_{T_1,2,T_2}$	T_6

Latest Trends

+ Multiobjective Optimization (Narasimhan and Sen, 2001, Sanchez et al, 2000): Pareto optimal solutions (cost vs. precision of estimates are build)

+ Unconstrained Optimization (Bagajewicz 2002, Bagajewicz and Markowski 2003): Reduce everything to cost, that is find the economic value of precision and accuracy.

Unconstrained Optimization

Let SN_0 be an existing network, then an upgrade to network SN has a Value defined as:

$$\text{Value}\,(SN) = \text{Profit}\,(SN) - \text{Profit}\,(SN_0)$$

Then the upgrade SND problem is defined as:

$$\textbf{Maximize} \quad \{\, \text{Value}\,(SN) - \text{Cost}\,(SN) \,\}$$

Integrated Approach

$$\textbf{Maximize} \quad \sum_{i=1}^{3} \{ V_i\,(SN) \} - \text{Cost}\,(SN)$$

where $V_i\,(SN)$ are the Value functions from the three perspectives

i=1 Control Systems

i=2 Material Accounting

i=3 Fault Diagnosis

Material Accounting Perspective

Given an distribution $g\left(\xi, \hat{\sigma}_{p,m}(SN_0)\right)$ one can calculate the probability that target production is not met.

This is quantified as the Downside Expected Production Loss:

$$DEPL(\hat{\sigma}_p) = T\int_{-\infty}^{m^*}(m^* - \xi)g(\xi, \hat{\sigma}_p)d\xi \approx 0.2T\hat{\sigma}_p$$

The above expression assumes process variability $(\sigma_p) <<< \hat{\sigma}_p$

Material Accounting Perspective

In the presence of biases we have:

Portion of time in each state

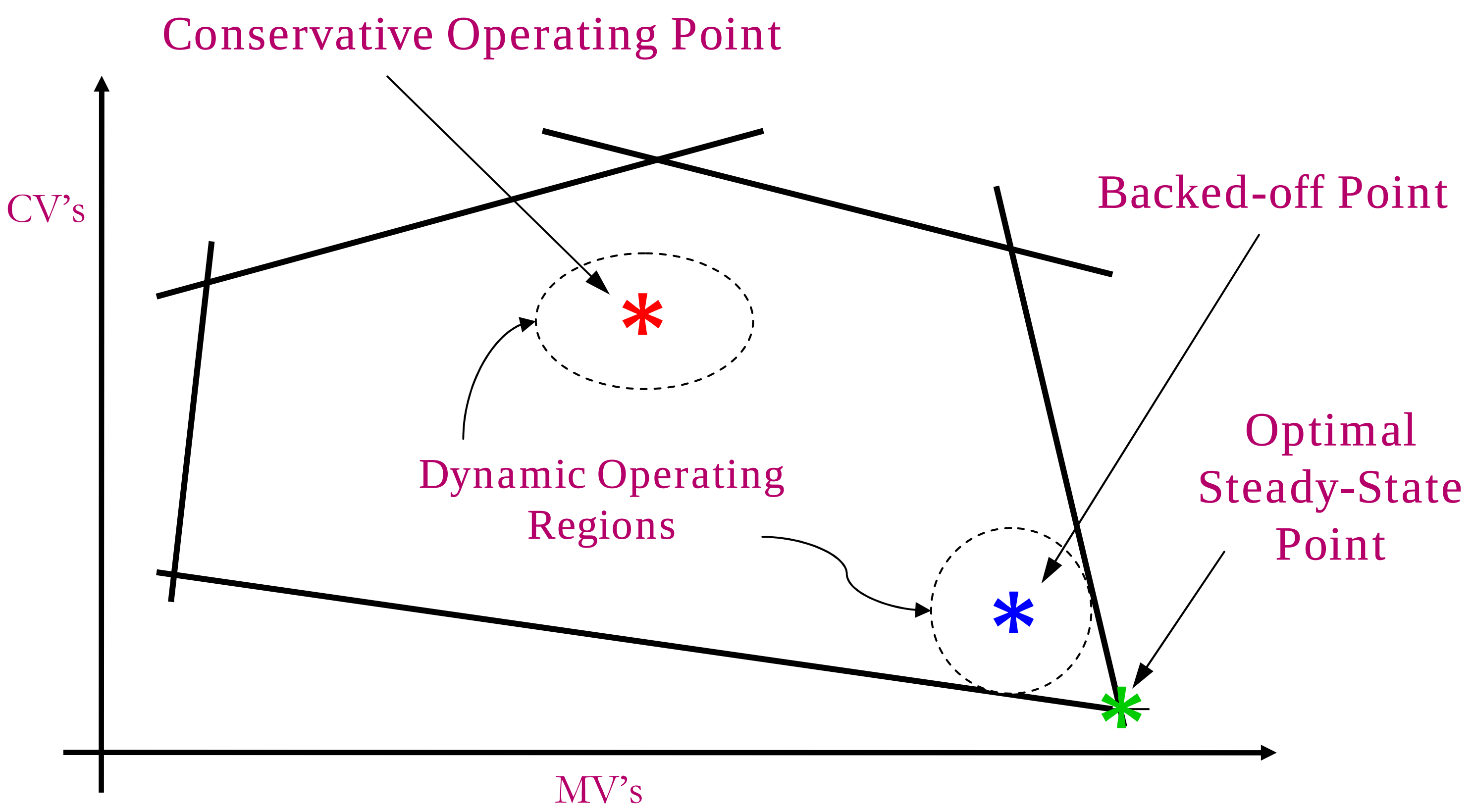
$$DEFL = \Psi^0 DEFL^0 + \sum_i \Psi_i^1 \, DEFL^1 \big| i + \sum_{i1,i2} \Psi_{i1,i2}^2 . DEFL^2 \big| i1, i2 + \dots$$

Financial loss when **no** gross error is present

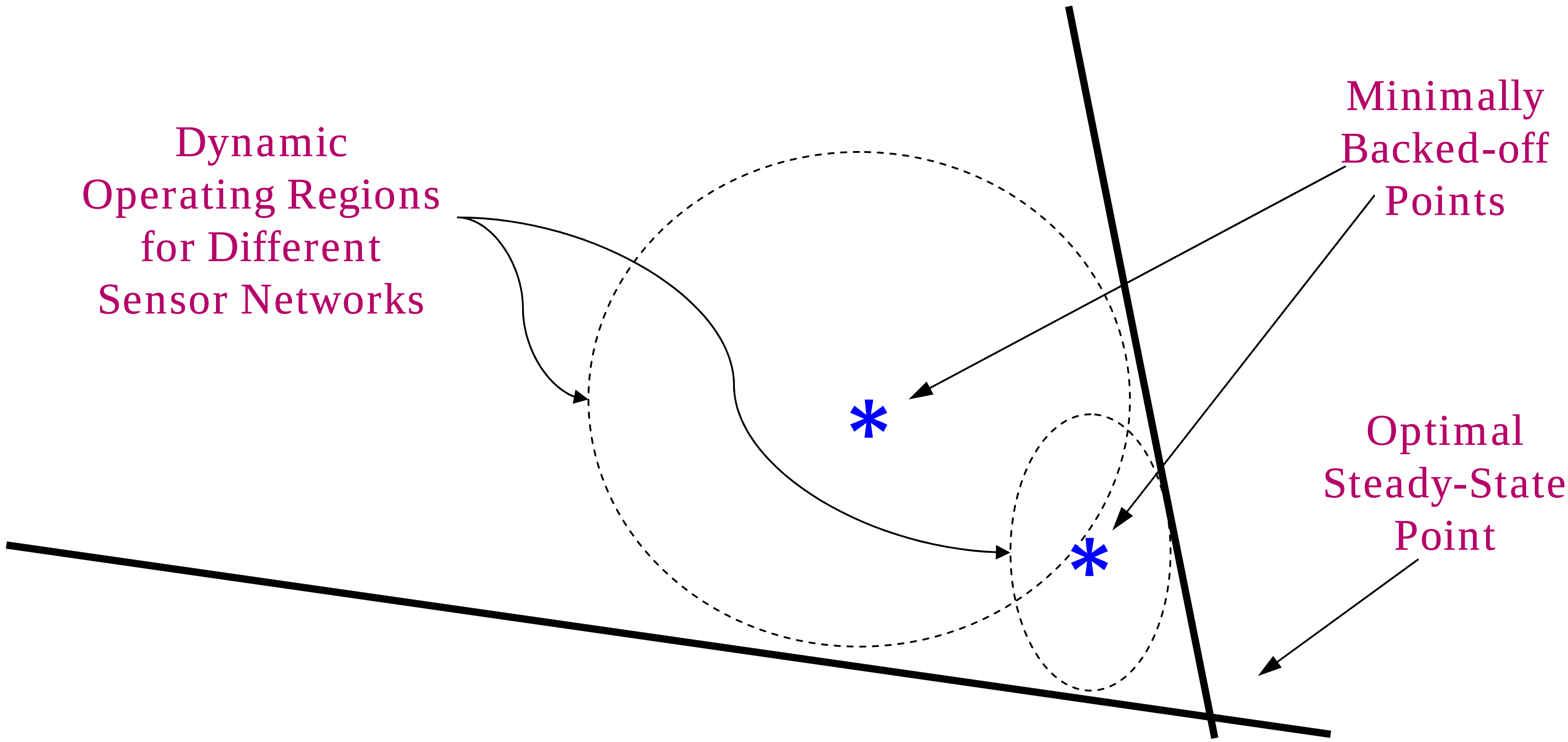
Financial loss when **one** gross error is present

Financial loss when **two** gross error are present

Control Perspective



Control Perspective



Faults Perspective

- Consider a set $\boldsymbol{F}$ of possible faults $\boldsymbol{F}=\{\, \boldsymbol{f_i} \,\}$.

Define a set $\boldsymbol{A_i}(SN)$ as the set of sensors in SN that can observe fault $\boldsymbol{f_i}$.

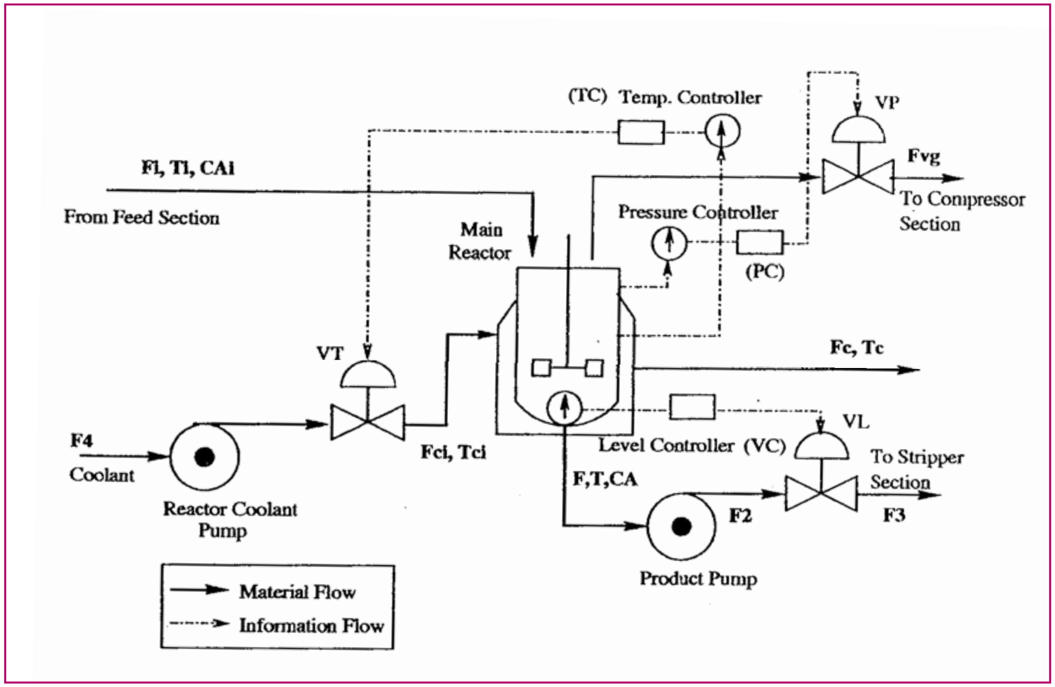
If $\boldsymbol{A_i}(SN)$ is not empty then $\boldsymbol{f_i}$ can be detected.

Assume immediate correction occurs for detected faults.

If all faults in $\boldsymbol{F}$ can be detected, then no production losses or safety incidents will be expected.

Example

CSTR Process (A→B+C)



Assume the current network (SN_0) consists of 6 sensors located at

C_{Ai}, C_A, T, V, F, P

each having a precision of 2%.

Results (Control)

No	New Sensors	Value ($/yr)	Sensor Costs ($/yr)	Value - Sensor Costs ($/yr)
1	C_A, P	7,060	2,000	5,060
2	C_A, T_c, P	7,630	3,000	4,630
3	P	4,500	1,000	3,500
4	T, P	5,420	2,000	3,420
5	C_A, T, T_c, V, P	8,090	5,000	3,080
6	T, T_c, V, P	6,150	4,000	2,140
7	none	0	0	0

Results (Material Accounting)

No	New Sensors	Value ($/yr)	Sensor Costs ($/yr)	Value - Sensor Costs ($/yr)
1	C_{Ai}	131	1000	-868
2	F_{vg}	127	1000	-872
3	F_2	77	1000	-922
4	All sensors	355	13,000	-12,644

In all cases the cost of adding sensors far exceeds the profit retuned in the form of Upgrade Value.

Results (Faults)

No	New Sensors	Value ($/yr)	Sensor Costs ($/yr)	Value - Sensor Costs ($/yr)
1	T_{ci}, T_i	7,810	2,000	5,810
2	F_c, T_c, T_i	7,810	3,000	4,810
3	T_c, T_{ci}	4,720	2,000	2,720
4	F_c, T_c	4,720	2,000	2,720

INTEGRATED PERSPECTIVE

No	New Sensors	Value ($/yr)	Sensor Costs ($/yr)	Value - Sensor Costs ($/yr)
1	C_A, P, T_{ci}, T_i	14,930	4,000	10,930
2	C_A, P, F_c, T_c, T_i	15,525	5,000	10,525

Case 1: union of best networks from individual perspectives.

Case 2: union of second best networks.

- These are the best combinations given the tables presented.
- Exhaustive enumeration search is underway.

CHALLENGES

- **Academic:** Multiple Gross Error Identification

  Gross Errors for Nonlinear Systems.

  Unconstrained Methods. Solution Procedures

- **Industrial:** Dynamic data reconciliation.

  Gross Error Handling.

  Sensor Upgrades

CONCLUSIONS

- Data Reconciliation is an academically mature field.
- It is a must when parameter estimation (mainly for on-line optimization) is desired.
- Commercial codes are robust but lack of up to date gross error detection/location techniques.
- Instrumentation Upgrade methodologies have reach maturity
- Industry understands the need for upgrading, but academic efforts have not yet reached commercial status. They will, soon.