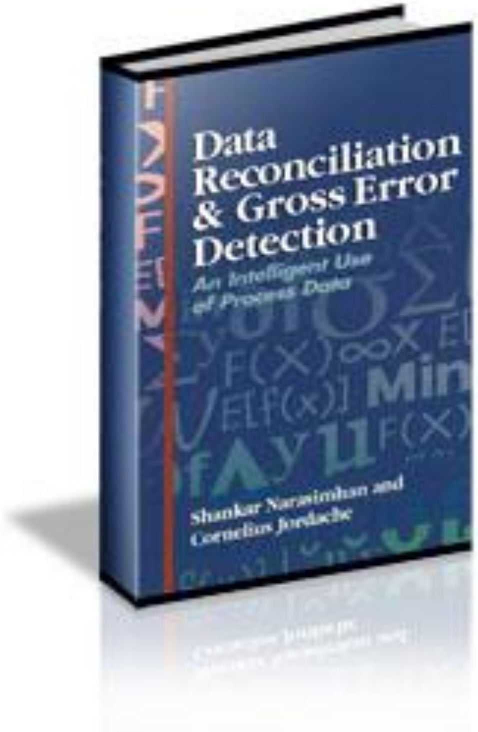


Data Reconciliation

Davide Manca

Lesson 7 of "Process Systems Engineering" – Master Degree in Chemical Engineering – Politecnico di Milano

Hierarchical approach to process optimization

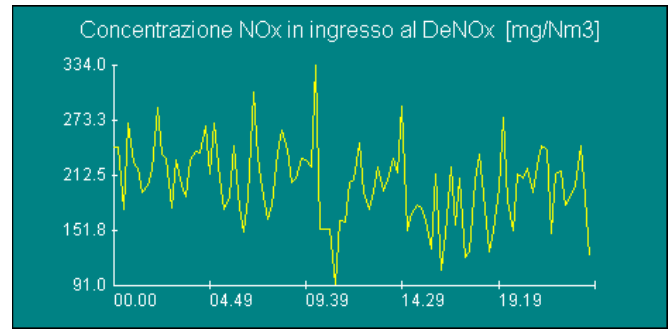
- Supervision
- Optimization
- Process control
- Data reconciliation
- Process simulation
- Statistical analysis
- Online data collection

Process

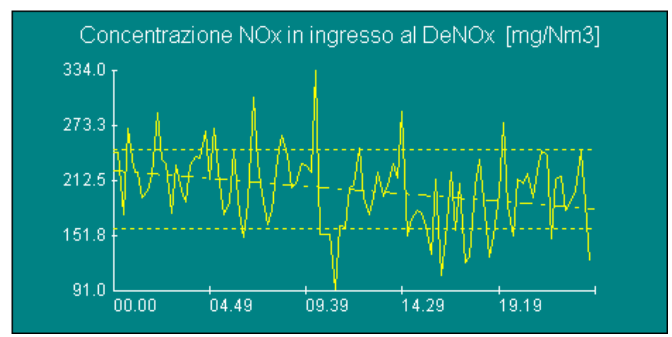
DCS

Supervisor system

Data collection



Statistical analysis

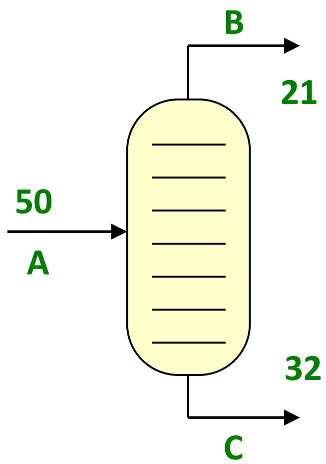


IN → **Process model** → OUT

Data reconciliation

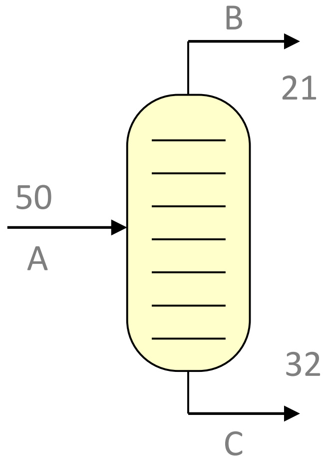
About data reconciliation…

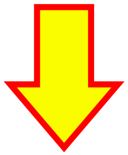
- **Classical reconciliation of measures**



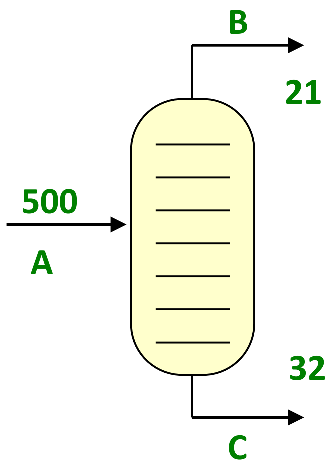
About data reconciliation…

- Classical reconciliation of measures



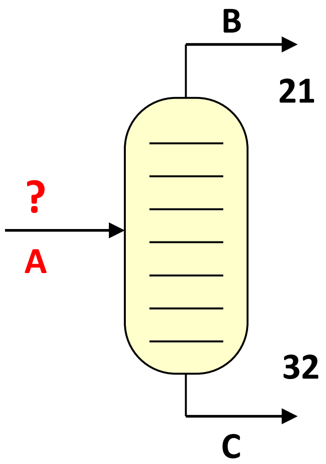


- **Gross error detection**



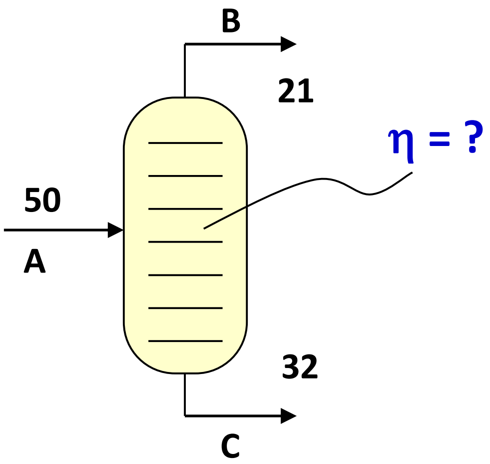
About data reconciliation…

- Coaptation



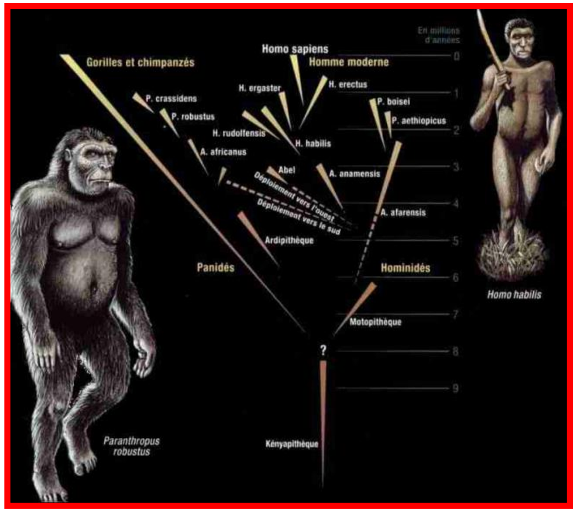
About data reconciliation…

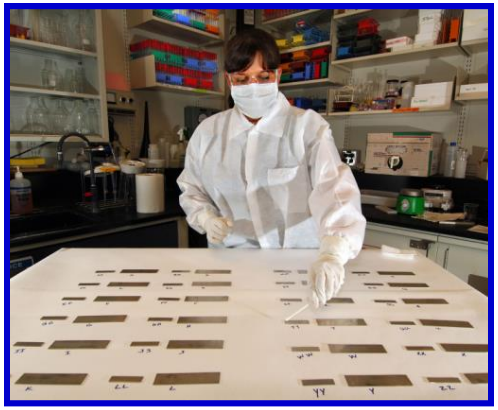
- **Model identification**



Introduction

- The **Data Reconciliation** methodology can be divided into **three distinct phases** (Romagnoli e Sanchez, 2000):
  - **Classification** of process variables and decomposition of the problem;
  - **Detection**, **identification** and **estimation** of gross errors;
  - **Estimation** of **process variables not measured** or not measurable.





Measurement classification

- Because of costs, convenience, and technical reasons, **not all the process variables are measured**.
- By assuming that the process is working in steady-state conditions, some unmeasured variables can be **estimated** using other measured variables and calculations based on mass and energy balances.
- The estimation of **not measured variables** depends on the **process layout** and on the in-the-field **instrumentation**.
- In general, the process instrumentation is incomplete (it does not measure all the process variables).

The **unmeasured** variables can be divided into:

– **Predictable variables** (determinable)

– **Unpredictable variables** (undeterminable)

Measurement classification

- Furthemore, **measures** can be classified into:
  - **redundant**
  - **nonredundant**

- A measure is **redundant** if it remains **determinable** when the observation is removed.
- The classification of the variables is an essential tool to design and revamp monitoring systems.
- A **robust classification** of variables leads to **significant savings** linked to the selection of instrumentation for field installation.
- An incorrect classification of variables leads to the introduction of unnecessary instrumentation involving higher investment costs.

Measurement classification

- An **unmeasured** variable is **determinable** if it can be calculated using the available measures and balance equations.
- An **unmeasured** variable is **indeterminable** if it cannot be calculated using the available measures and balance equations.

- A **measured** process **variable** is **redundant** (overdetermined) if it can also be calculated using the remaining measures and balance equations.
- A **measured** process **variable** is **nonredundant** if it cannot be calculated using the remaining measures and balance equations.

Measurement classification

- Once the variables are classified, we have a significant amount of information concerning the process topology.
- It is now possible to solve the following problems:
  - **Select the set of measured variables** which must be corrected (reconciled) in order to increase the accuracy of the measured and unmeasured process variables.
  - **Select the minimum number of measures** so that all the unmeasured variables can be determined.

Process model

- The **process model** is a mathematical formulation that describes its behaviour under either **STEADY STATE** or **DYNAMIC** conditions.
- The process model is used at several levels:
  - To infer unmeasurable parameters
  - To reconcile measures
  - To identify measures affected by gross errors
  - To determine the optimal control action
    - Model based control (for example: Model Predictive control)
    - Feedforward control
  - For process optimization
  - For process supervision
- The process can be described by either linear or nonlinear models: ARX, NARX, ARMAX, NARMAX, Laplace transforms, Regressions, Artificial neural networks (ANN), deterministic and phenomenological models (First Principles), …

Process model

- The detail of the process model must be related to the requested description. We can distinguish between:
  - Stationary and dynamic model
  - Linear or nonlinear model
  - Robust or efficient model
  - Simplified or detailed model
- In the most complex situation the model is detailed, nonlinear and dynamic. We must write, for the equipment and the streams of the process, the material, energy, and momentum balances. The resulting system will contain differential algebraic equations and possibly partial differential. There are suitable numerical routines to integrate these systems.
  Even the use of modern computers, with extremely fast CPUs, requires a good amount of time for simulation (*e.g.*, model predictive control, optimization), which can be greater than the maximum acceptable time (horizon control). In this case it is recommended to adopt/implement more simplified models to reduce the CPU time (*e.g.*, ARX or ANN models).

Solution methodology

❑ Equation oriented

This approach is based on material and energy balances applied to the connections of the plant, used as equality constraints to be satisfied by finding the minimum. The output variables of the procedure correspond to the input ones. The difference between calculated and measured values is due to a measurement error. To estimate the degrees of freedom of the plant, we must have new and different measures as accurate as possible, distributed through the process.

❑ Black box

We have a process simulator that calculates the output variables to be reconciled respect to the given input variables. The output variables are: streams and/or compositions unknown and non measurable process parameters. The simulation program is called iteratively by a non-linear regression routine which determines the degrees of freedom in order to minimize the *distance* between the measured and the calculated values.

Redundancy

- Romagnoli and Sanchez (2000) define a **system** as being **redundant** when the whole collection of data/information available exceeds the minimum required amount for a univocal determination of the independent variables that describe the selected model.
- Since the data are obtained from process measurements affected by probabilistic fluctuations, **redundant data are generally inconsistent** thus every data subset provides different results from other subsets.
- In order to obtain a **consistent solution** to the problem of determining the measures, it is therefore necessary to introduce an **additional criterion**.

Redundancy

Redundancy of the system

In case of Black Box approach we define the **redundancy** as the difference between the number of measured variables and the number of degrees of freedom:

Redundancy = # measured variables – # dof = NY – NPAR

The **system** describing numerically the reconciliation problem is **OVERDETERMINED**. There are more equations than unknowns.

$$\begin{cases} y_{exper}(1) - y_{calc}^{1}\left(x_1, x_2, \ldots, x_{NPAR}\right) = 0 \\ y_{exper}(2) - y_{calc}^{2}\left(x_1, x_2, \ldots, x_{NPAR}\right) = 0 \\ \ldots \\ y_{exper}(NY) - y_{calc}^{NY}\left(x_1, x_2, \ldots, x_{NPAR}\right) = 0 \end{cases}$$

The **overdetermination** of the system leads to the impossibility of completely satisfying it. Conversely, it is possible to minimize the sum of squares of the equations by solving a minimization problem with a non-linear regression in the parameters, $\mathbf{x}$.

Object function

- The reconciliation procedure has to **minimize** the following **objective function**:

$$\underset{\mathbf{x}}{Min}\ f = \sum_{i=1}^{NY} \frac{\left[y_{exper}(i) - y_i^{calc}(\mathbf{x}) \right]^2}{s^2(i)}$$

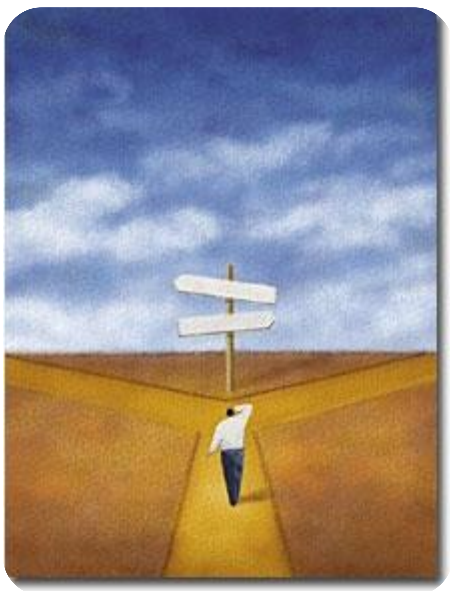
By introducing the incidence matrix $\mathbf{M_I}$, it is possible to check if a *dof* does NOT affect any measure (column-wise) or if a measure is NOT affected by any *dof* (row-wise).

If two columns are linearly dependent then there is a high functional dependency between those degrees of freedom.

$$\mathbf{M_I} = \begin{bmatrix} \frac{\partial y_{calc}^1}{\partial x_1} & \frac{\partial y_{calc}^1}{\partial x_2} & \cdots & \frac{\partial y_{calc}^1}{\partial x_{NPAR}} \\ \frac{\partial y_{calc}^2}{\partial x_1} & \frac{\partial y_{calc}^2}{\partial x_2} & \cdots & \frac{\partial y_{calc}^2}{\partial x_{NPAR}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_{calc}^{NY}}{\partial x_1} & \frac{\partial y_{calc}^{NY}}{\partial x_2} & \cdots & \frac{\partial y_{calc}^{NY}}{\partial x_{NPAR}} \end{bmatrix}$$

Solution of the reconciliation problem

- The Reconciliation problem can be solved if we have:
    - **Positive redundancy**
    - **Independent degrees of freedom**
    - A **robust numerical algorithm** especially if we work online



Solution of the reconciliation problem

- The **basic assumptions** are as follows:
  1. The **process model** is able to properly represent the system under consideration (model validation);
  2. The **measures** are subject to an ε error that is **normally distributed** with average equal to zero and variance σ known (or that can be computed);
  3. The measures come from a **stationary process**.
- The failure of Reconciliation (once hypothesis **1** is verified) is due to points **2** and **3**. There may be measures affected by **gross error** that have a non-zero averaged error ε:

$$E(\varepsilon)=\int_{-\infty}^{+\infty}\varepsilon\, p(\varepsilon)\,d\varepsilon=\int_{-\infty}^{+\infty}\frac{\varepsilon}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)d\varepsilon\neq 0$$

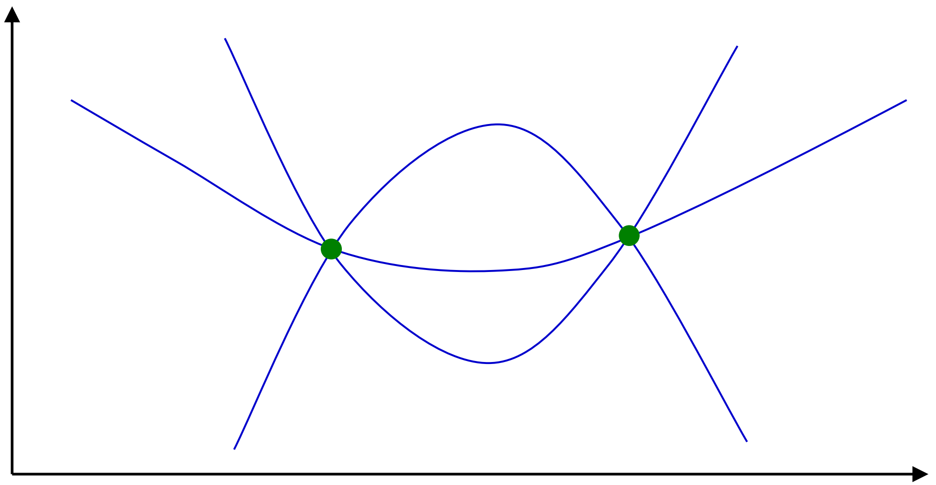
- Possible causes of gross errors are: unreliable instruments, non-homogeneous conditions around the instrument, process instability, accidents, transcription errors, communication failures, non-stationary conditions.

Statistical analysis

- To perform the Data Reconciliation procedure we must start from the averaged measured values (measured in the field at a given time when the process is mildly stationary).
- At this regard, we have the expected value $\mu(i)$ and variance $\sigma(i)$ of the measure.
- It is possible to distinguish between efficient and robust estimators:
  - **ROBUST estimators**
    - For $\mu(i)$ we use the Median: it is the central value of the population in ascending order. In the case of an even number of terms we do the arithmetic mean of the two central values.
    - For $\sigma(i)$ we use the MAD (Median Absolute Deviation)
      MAD(i) = 1.4826 * Median(|y_{exper}(i,k)–Median(y_{exper}(i,k)||)
  - **EFFICIENT estimators**
    - Arithmetic mean: $y_s(i) = \sum_{k=1}^{NS} y_{exper}(i,k) / NS$
    - Standard deviation or mean square deviation: $\sigma(i) = \sqrt{\sum_{k=1}^{NS} \frac{\left[y_{exper}(i,k) - y_s(i)\right]^2}{NS - 1}}$

Model identification

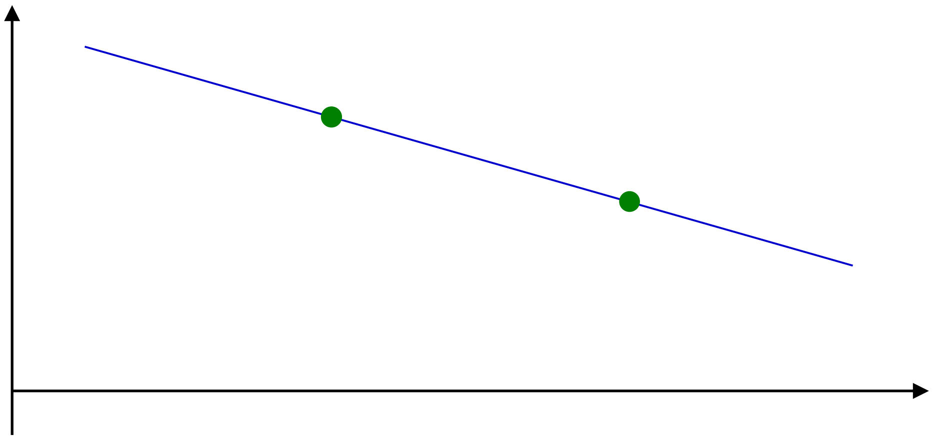
- Once we have defined **NY = number of measures** and **NPAR = number of degrees of freedom** (parameters) it is possible to distinguish the following cases:
- **NPAR > NY (NEGATIVE redundancy)**



For instance: the proposed model $y = ax^2 + bx + c$ comprises three parameters (NPAR = 3) whilst the experimental points are just two. There is an infinite number of parabolas that match exactly the experimental data. It is not possible to identify any Gross Errors.

Model identification

- **NPAR = NY (NO redundancy)**

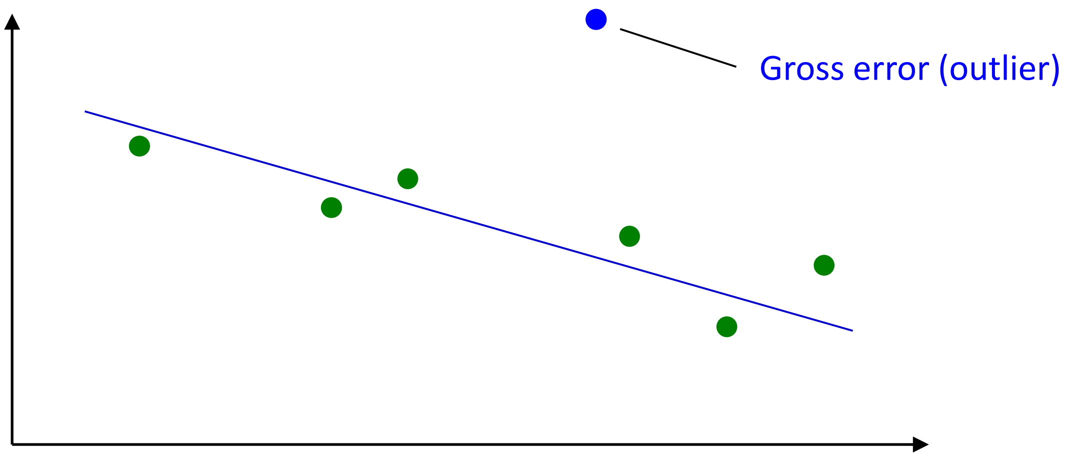


In this case, there is only one curve passing through the NY points. It is worth observing that, in this case, the model is a straight line ($y = ax + b$) depending on two parameters.

We have: NPAR = 2 and NY = 2. The redundancy is zero and it is NOT possible to detect any Gross Errors.

Model identification

- **NPAR < NY (POSITIVE redundancy)**



In this case, the proposed model is still a straight line (NPAR = 2) while the number of experimental points is seven: NY = 7. There is NOT a model that simultaneously satisfies all the experimental data. It is then necessary/advisable to minimize the error by minimizing the distance between the model and the measured data.

We can also detect NGE potential gross errors: NGE = NY – NPAR = R = Redundancy

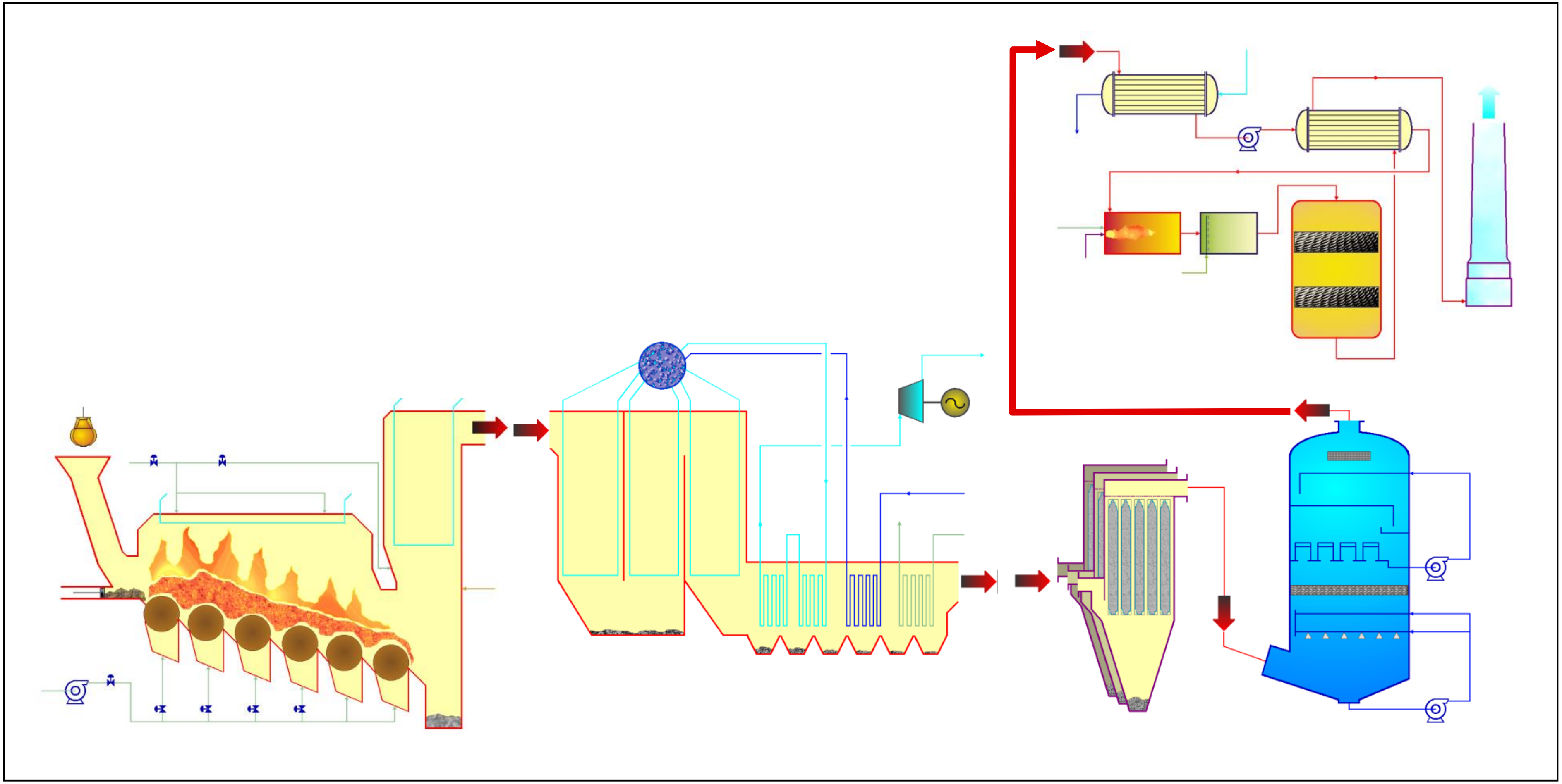
N.B.: if we identify a gross error it is possible to eliminate it or compensate it with the value that has been just reconciled. In this case we do not decrease the redundancy.

Case-study

On-line data reconciliation of an incineration plant

Case study: incineration plant

Waste to energy plant with DeNOx catalytic section



Case study: incineration plant

- **Specifications required**
  - Evaluate the **consistent value** of the **measurements** from the field
  - Identify measurements affected by gross error
  - Real-time knowledge of the characteristics of the incoming waste in terms of **elemental composition** and **heat of combustion**
  - Estimation of the inlet streams unmeasurable or not available:
    - **Air leakages**
    - **Methane flowrate** in the postcombustion chamber
  - Evaluation of the **operating parameters**:
    - Bag filter efficiency
    - Catalyst efficiency
    - Heat exchangers fouling factor

Case study: incineration plant

- **Problem definition**

Objective function:

$$\underset{\mathbf{x}}{Min}\ f = \sum_{i=1}^{NY} \frac{\left[y_{exper}(i) - y_i^{calc}(\mathbf{x}) \right]^2}{s^2(i)}$$

Measures to be reconciled: 24

T gas postcomb.	CO out postcomb.	O2 out postcomb.
T out gas radiative zone	T out gas superheater T	out gas economizer
T out gas preheater	T air combustion	Gas entering washing column
Gas out washing column	T out gas heater	Tout gas heat exchanger gas-gas
T gas stack	T in gas DeNOx	T out gas DeNOx
NOx entering DeNOx	NOx exiting DeNOx	Ammonia flow rate
HCl to the stack	SO2 to the stack	CO to the stack
Soot to the stack	O2 to the stack	Steam flowrate

Degrees of freedom (parameters of reconciliation): 23

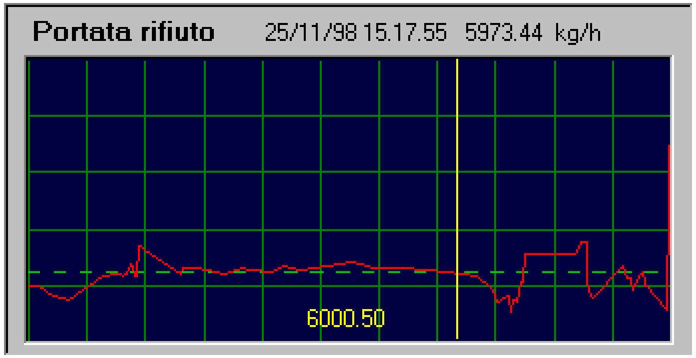
Waste flow rate	Ash fraction in the waste	Cl fraction in the waste
S fraction in the waste	N fraction in the waste	C fraction in the waste
Kiln air leakage	Bypass gas fraction in the furnace	Methane flow rate afterburner
Losses in the boiler	Corr. heat exch. coeff. rad. zone	Corr. heat exch. coeff. superheater
Corr. fact. economiz.	Corr. fact. Preheater	Bag filter efficiency
Acid wash efficiency	Basic wash efficiency	Corr. fact. Steam heater
Corr. fact. exch. gas-gas	Preheater air flow rate	Methane flow rate burner DeNOx
Catalyst efficiency DeNOx	Air leakages after postcombust.	

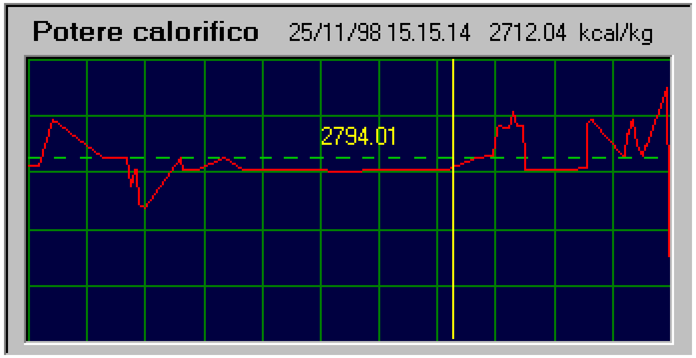
Case study: incineration plant

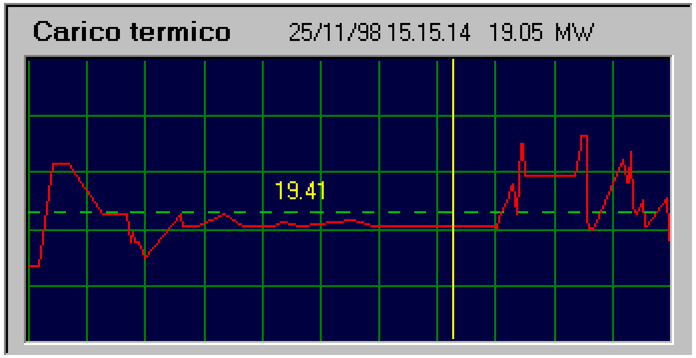
- **Problem solution**
  - We need a nonlinear regression routine to minimize the objective function.
  - We must have a detailed model of the process that simulates the measurements (*i.e.* calculates the reconciled values of the acquired measurements) whenever the regression routine suggests a new vector of degrees of freedom.
  - If the reconciliation procedure is NOT able to minimize the objective function to the required precision it means that the material, energy, and momentum balances describing the process "do not close". In this case we can assume the presence of a gross error and remove the measure respect to which there is the larger deviation (or better replace the measured value with the estimated one). The procedure continues until we reach the required accuracy. If the assumed replaced measurement affected by gross error does NOT make the procedure successful, we reintroduce the original removed measure and eliminate the next one featuring the greatest deviation. In this case study the redundancy is equal to one, consequently it is possible to identify just one gross error.

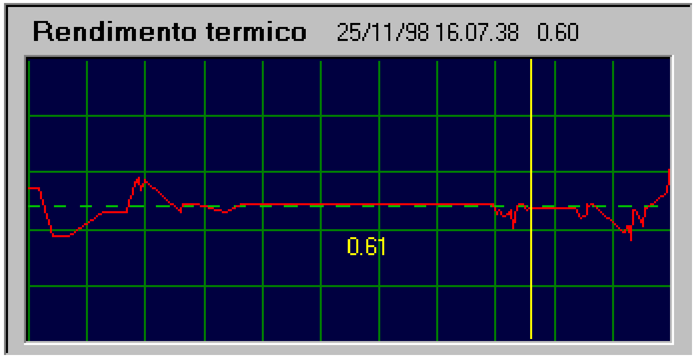
Case study: incineration plant

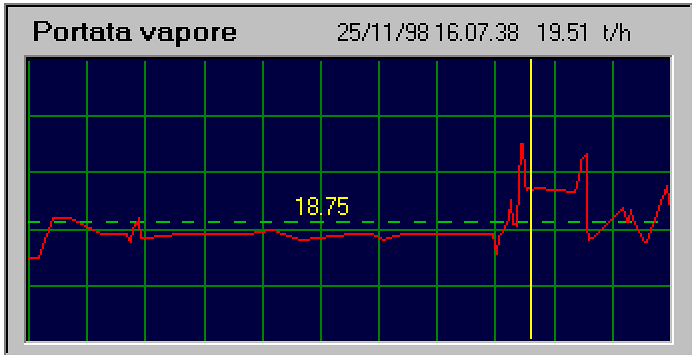
- **The results…**











Case study: incineration plant

- **The results…**

Riconciliazione.ris - Blocco note

File Modifica Cerca ?

```
 -------------------------------------------------------------
  DEMO CYCOM Plus : PROCEDURA DI RICONCILIAZIONE  -  RIEPILOGO
 -------------------------------------------------------------
  Impianto Rifiuti Solidi Urbani   18-May-01
 -------------------------------------------------------------

  Funzione obiettivo.........................    1298.264000
  MAX errore.................................   4.989020E-02

  Valori finali dei gradi di liberta` attivati

  Portata di rifiuto [kg/h]..................:    5307.524000
  Fraz.massiva H2O nella corrente entrante...:   1.338146E-01
  Fraz.massiva Cl nella corrente entrante....:   4.894615E-04
  Fraz.massiva S  nella corrente entrante....:   1.582423E-04
  Fraz.massiva N  nella corrente entrante....:   7.860772E-04
  Portata aria tenute forno [kg/h]...........:   17069.880000
  Frazione di by-pass fumi forno.............:   4.302227E-02
  Dispersioni termiche caldaia [% duty]......:   2.801692E-01
  Fat. correttivo emissiv. gas zona radiante :       1.349454
  Fat. correttivo scambio termico surriscal. :       1.066711
  Fat. correttivo scambio termico economiz.  :   5.813368E-01
  Fat. correttivo scambio termico preriscal. :   4.862194E-01
  Efficienza termica colonna di lavag.basico :   5.000000E-01
  Fat. corret. scambio termico risc. a vapore:   3.835258E-01
  Fat. corret. scambio termico scam. gas-gas :   7.480029E-01
  portata aria preriscaldatore               :   15341.030000
  portata CH4 bruciatore di supporto DENOX...:     164.789700
  Efficienza catalizzatore DENOX.............:       1.051901
```

Case study: incineration plant

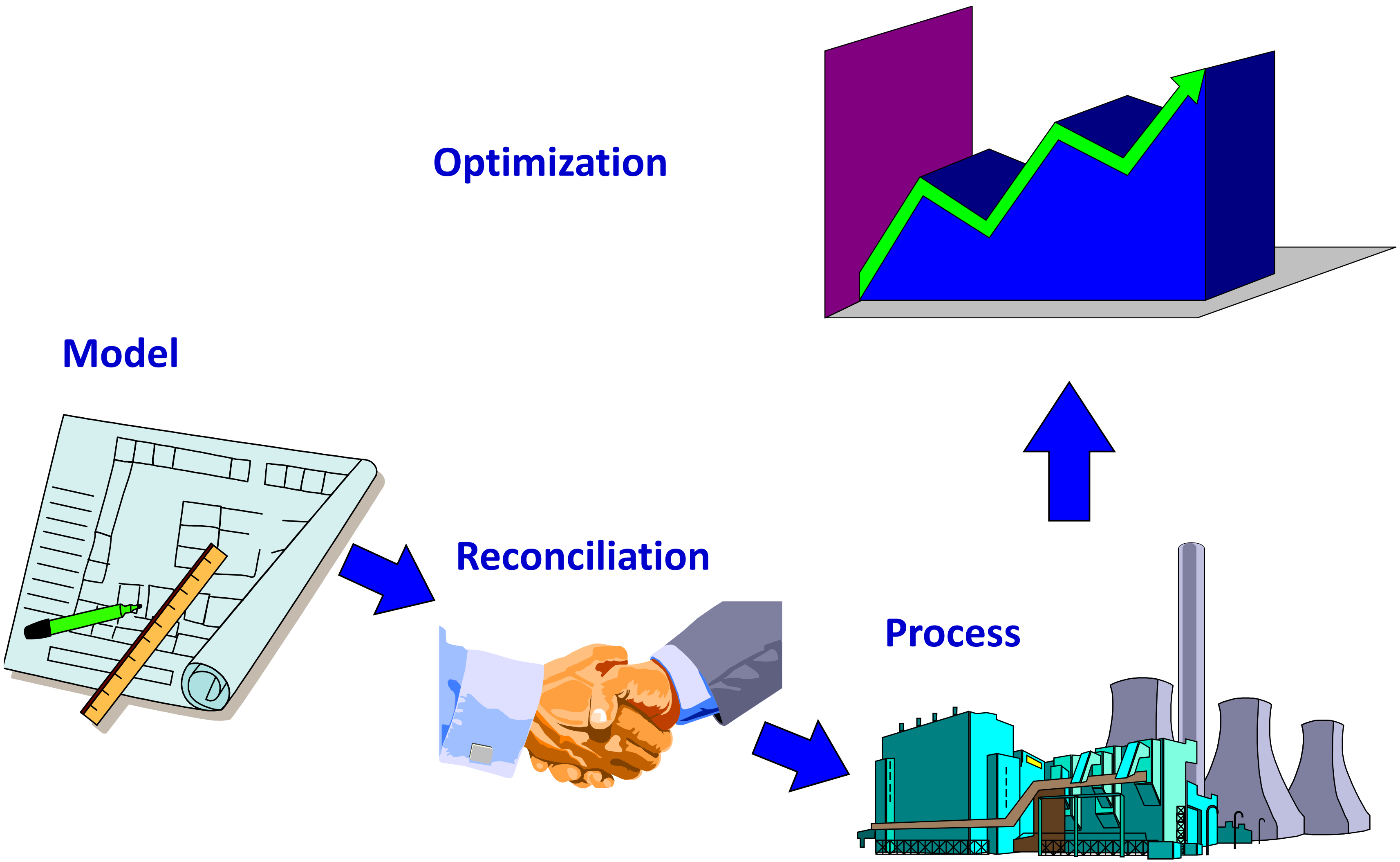
- The results…

Riconciliazione.ris - Blocco note

File Modifica Cerca ?

```
COMPOSIZIONE DELLA STREAM RICONCILIATA:

percentuale di C...........................:   3.431565E-01
percentuale di H...........................:   3.094497E-02
percentuale di O...........................:   1.405572E-01
percentuale di N...........................:   7.860772E-04
percentuale di S...........................:   1.582423E-04
percentuale di Cl..........................:   4.894615E-04
percentuale di H2O.........................:   1.338146E-01
percentuale di inerti......................:   3.500929E-01
potere calorifico [kcal/kg]................:    3160.306000


RIEPILOGO MISURE RICONCILIATE/DATI DI INPUT:

misura          val. sperim.       val. calcol.      errore % (1-100)

TIC1-1A          1037.72500         1010.50000          2.66
AI1-2               3.16731            3.04540          3.92
AIC1-1             10.56437           10.28697          2.66
TI1-2             489.27860          489.31060           .01
TI1-3             315.97500          316.26150           .09
TI1-4             240.67990          244.81750          1.70
TIC1-7            213.67680          214.63930           .45
TI1-6             185.37700          180.22980          2.82
FI1-14          47122.66000        47149.64000           .06
TI1-9              64.29519           62.79272          2.36
TI2-01            115.75740          115.60440           .13
TI2-03            260.80000          250.40330          4.07
TI2-08            293.80380          301.23770          2.50
AI2-01            202.84270          192.96910          4.99
AI2-02             18.03679           18.05212           .08
AI1-4               1.00000             .99891           .11
AI1-5               5.03523            5.01237           .46
FIC1-11            18.50559           18.49897           .04
----------------------------------------------------------
```

From data reconciliation to on-line optimization



Reconciliation and optimization

Reliable process data are the key to the efficient operation of chemical plants.

… it must be noted that errors in process data or inaccurate and unreliable methods of resolving these errors, can easily exceed or mask actual changes in process performance.

Romagnoli and Sanchez, 2000

- The incorrect knowledge of the operating conditions of the analyzed process leads to an erroneous representation and scope for improvement of it.

Reconciliation and optimization

- Think of a Ferrari that runs at 320 km/h:
  - If the uncertainty in measuring the lap time is 1 millisecond then we have a spatial uncertainty of 9 cm;
  - If the uncertainty in measuring the lap time is 1 second then we have a spatial uncertainty of 90 m.

Reconciliation and optimization

2006 Monaco Grand Prix — **Migliori tempi sul giro**

#	Nome	Cognome	Team	Tempo	Velocità media [km/h]	differenza %
1	Michael	Schumacher	Ferrari	01:15.1	160.014	
2	Kimi	Räikkönen	McLaren-Mercedes	01:15.3	159.628	0.241229
3	Fernando	Alonso	Renault	01:15.7	158.898	0.697439
4	Mark	Webber	Williams-Cosworth	01:15.7	158.879	0.709313
5	Giancarlo	Fisichella	Renault	01:15.9	158.379	1.021786
6	Juan Pablo	Montoya	McLaren-Mercedes	01:16.0	158.193	1.138025
7	Felipe	Massa	Ferrari	01:16.6	156.946	1.917332
8	Jarno	Trulli	Toyota	01:17.2	155.791	2.639144
9	Nico	Rosberg	Williams-Cosworth	01:17.2	155.696	2.698514
10	Jenson	Button	Honda	01:17.3	155.549	2.790381
11	Nick	Heidfeld	Sauber-BMW	01:17.3	155.511	2.814129
12	Rubens	Barrichello	Honda	01:17.3	155.509	2.815379
13	Tiago	Monteiro	MF1-Toyota	01:17.3	155.491	2.826628
14	Scott	Speed	STR-Cosworth	01:17.5	155.186	3.017236
15	Ralf	Schumacher	Toyota	01:17.5	155.068	3.090980
16	Christijan	Albers	MF1-Toyota	01:17.6	154.942	3.169723
17	Vitantonio	Liuzzi	STR-Cosworth	01:17.7	154.828	3.240966
18	Jacques	Villeneuve	Sauber-BMW	01:17.8	154.615	3.374080
19	David	Coulthard	Red Bull Racing	01:17.8	154.452	3.475946
20	Christian	Klien	Red Bull Racing	01:17.9	154.292	3.575937
21	Takuma	Sato	Super Aguri-Honda	01:18.8	152.602	4.632095
22	Franck	Montagny	Super Aguri-Honda	01:19.1	152.002	5.007062

5%

References

- Biegler L.T., J.J. Damiano, G.E. Blau, “Non linear parameter estimation: a Case-study comparison”, AIChE J., 32, 29-43, (1986)
- Crowe C.M., “Reconciliation of process flow rates by matrix projection. Part II: the non linear case”, AIChE J., 44, 2909-2917, (1986)
- Crowe C.M., “Data reconciliation – Progress and challenges”, J. Proc. Control, 6, 89-98, (1996)
- Kretsovalis A., R.S.H. Mah, “Observability and redundancy classification in generalized process networks. I: Theorems”, Comput. Chem. Eng., 12, 671-687, (1988a)
- Kretsovalis A., R.S.H. Mah, “Observability and redundancy classification in generalized process networks. II: Algorithms”, Comput. Chem. Eng., 12, 689-703, (1988b)
- Mah R.S.H., “Chemical process structures and Information flows”, Chem. Eng. Ser. Butterworth, Boston, (1990)
- Mah R.S.H., G. Stanley, D. Downing, “Reconciliation and rectification of process flows and inventory data”, Ind. Eng. Chem. Process Des. Dev., 15, 175-183, (1976)
- Manca D., M. Rovaglio, S. Pierucci, T. Faravelli, “Reconciliation, Optimization and Supervisory Control of Crude Oil Distillation Systems”, The First Conference on Chemical and Process Engineering, Firenze, 255-259, (1993)
- Manca D., M. Rovaglio, “Numerical Modeling of a Discontinuous Incineration Process With On Line Validation”, Ind. Eng. Chem. Res., 44, 3159-3177, (2005)

References

- **Narasimhan S., C. Jordache, “Data Reconciliation and Gross Error Detection : An Intelligent Use of Process Data”, Gulf Professional Publishing, Houston, (2000)**
- *Rovaglio M., D. Manca, M. Mariano, G. Nini, R. Gani, “Data Reconciliation and Process Optimization for Hazardous Waste Incineration Plants”, Proceedings of "PSE: Process Systems Engineering", pp. 613-621, ISBN 89 950005-1-1, Kyongju Korea, (1994)*
- *Rovaglio M., D. Manca, “Reconciliation and Model Identification as a First Step for On Line Optimization of Incineration Processes”, Proceedings of International Conference on Incineration & Thermal Treatment Technologies, IT3, S. Francisco, USA, pp. 587 596, (1997)*
- **Romagnoli J.A., M.C. Sanchez, Data Processing and Reconciliation for Chemical Process Operations, Academic press, New York, 2000**
- Stanley G., R.S.H. Mah, “Observability and redundancy in process data estimation”, Chem. Eng. Sci., 36, 259-272, (1981a)
- Stanley G., R.S.H. Mah, “Observability and redundancy classification in process networks – Theorems and algorithms”, Chem. Eng. Sci., 36, 1941-1954, (1981b)
- Tjoa I. B., L.T. Biegler, “Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems”, Comp. and Chem. Engineering, 15, 679–690, (1991)
- Vaclavek V., “Studies on system engineering. III. Optimal choice of the balance measurement in complicated chemical engineering systems”, Chem. Eng. Sci., 24, 947-955, (1969)