3.3 Data Reconciliation: What's the Real Operating Condition?

Prof. Flavio Manenti
Politecnico di Milano, Italy
Xi'an Jiao Tong Univ., China; Tomsk Polytechnic Univ., Russia
flavio.manenti@polimi.it

GRICU PhD School 2021

Digitalization Tools for the Chemical and Process Industries

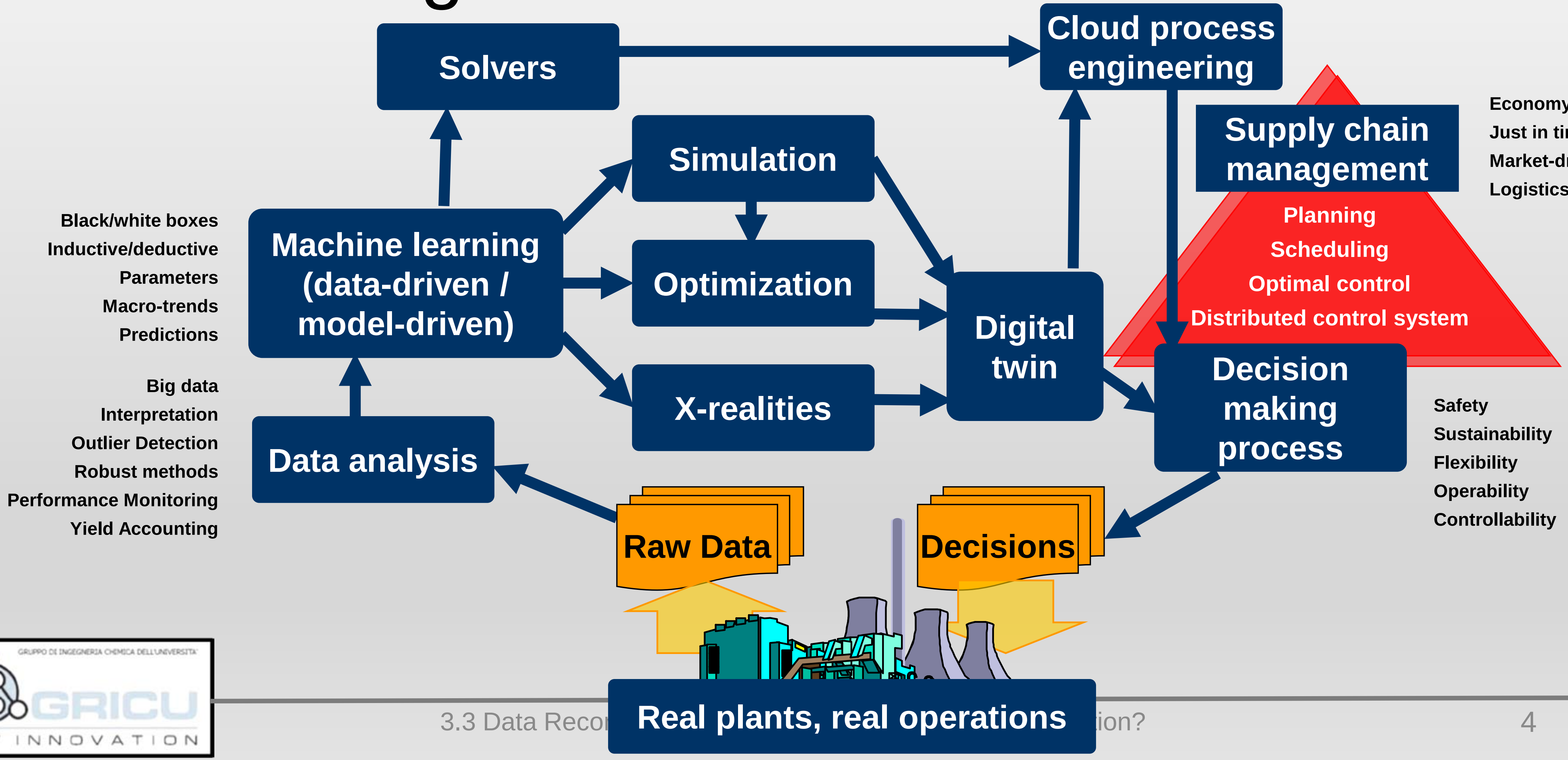
March 18, 2021

Outline

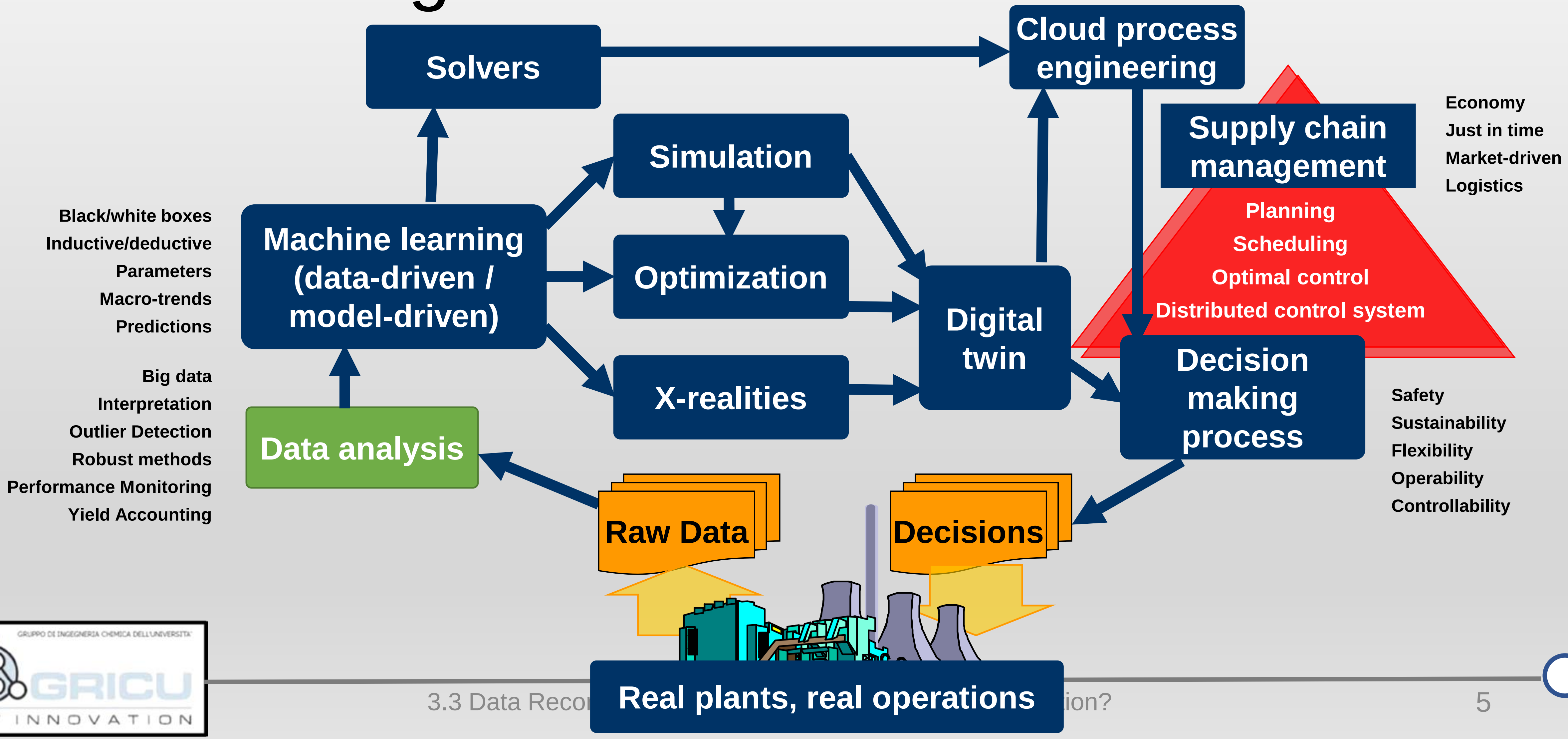
- A process engineering (personal) vision of digitalization
- Fundamentals of data reconciliation
- Mathematical formulation
- Exercise
- Gross errors
- Industrial case studies
  - Itelyum regeneration
  - Lukoil ltd

Digitalization
Which topics?

A process engineering (personal) vision of digitalization



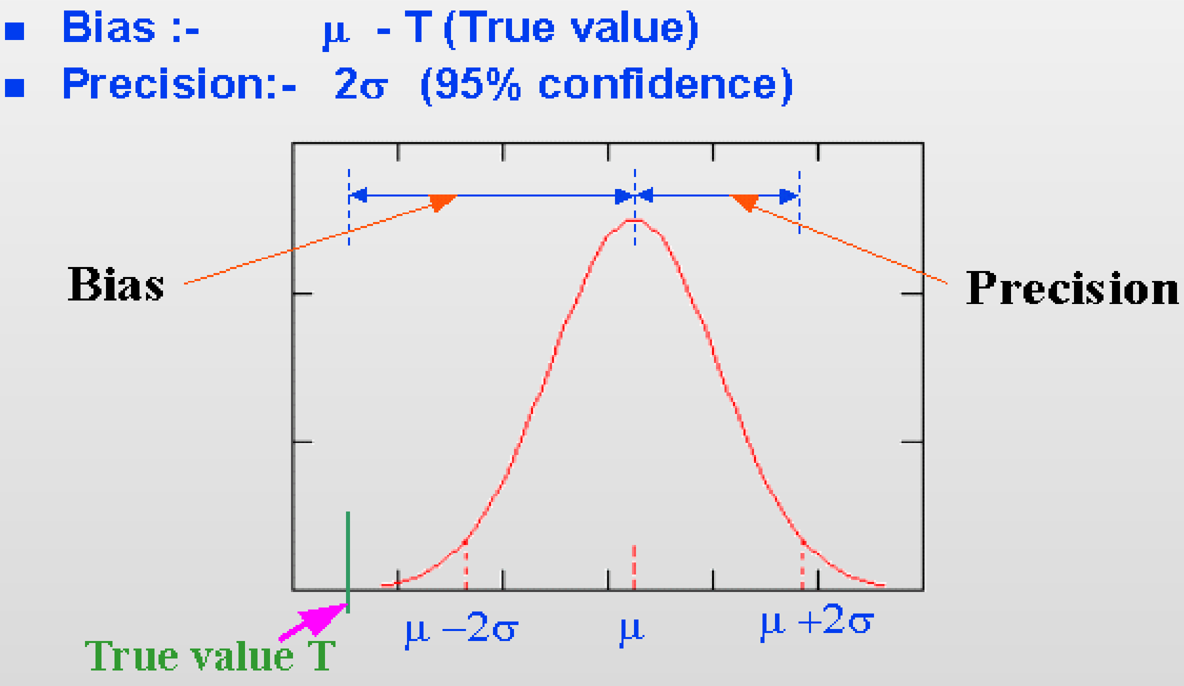
A process engineering (personal) vision of digitalization

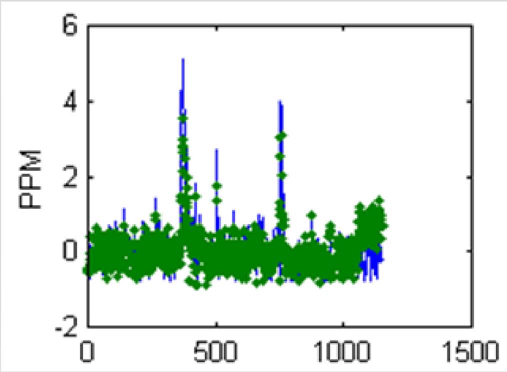


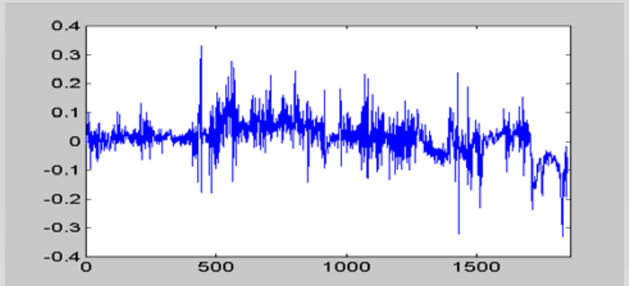
Data reconciliation
Fundamentals

Plant measurements

- All measurements are subject to ***uncertainty***
- ***Random Errors*** (unavoidable and hopefully small)
- ***Gross Errors*** (conceptually avoidable) and ***outliers***



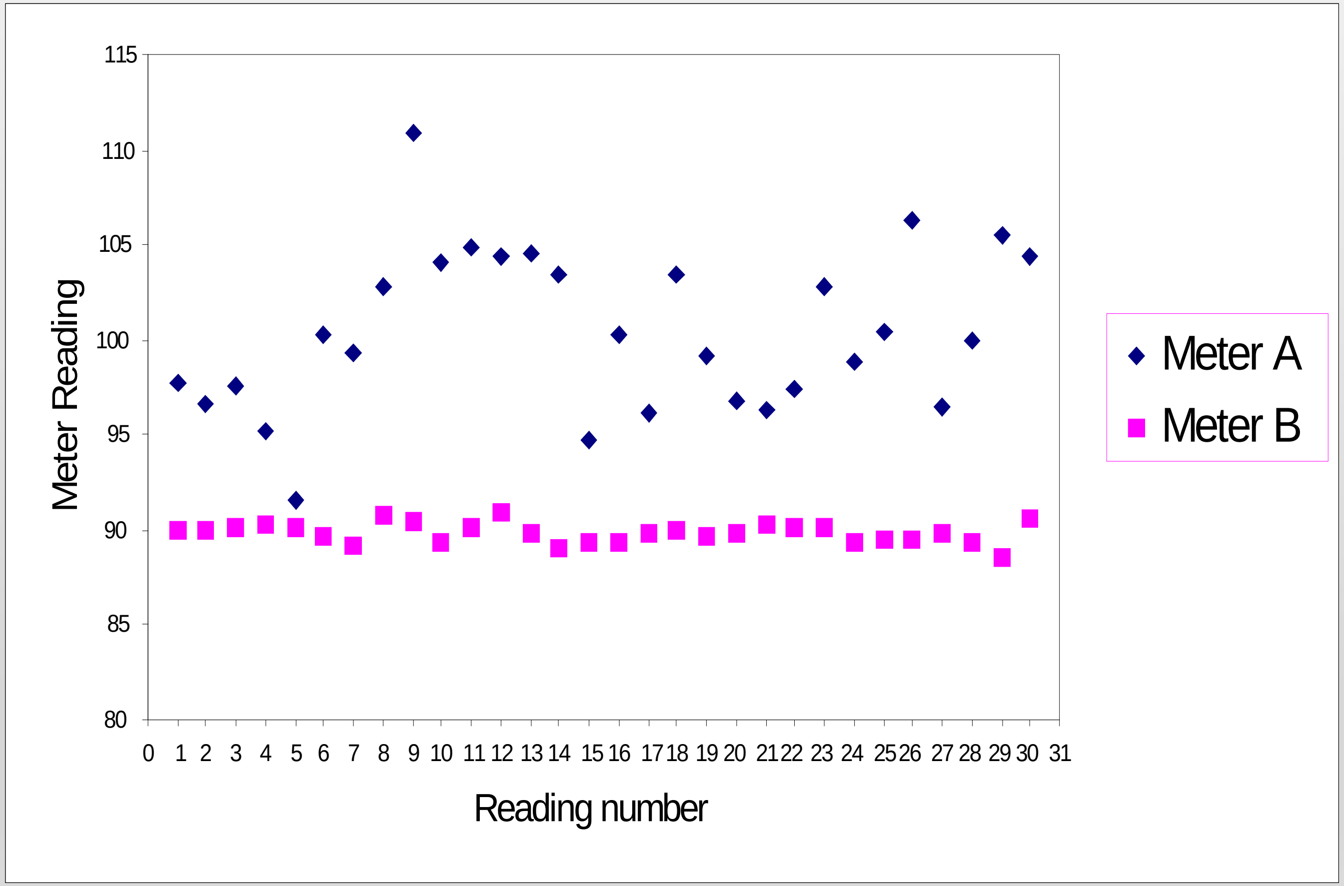




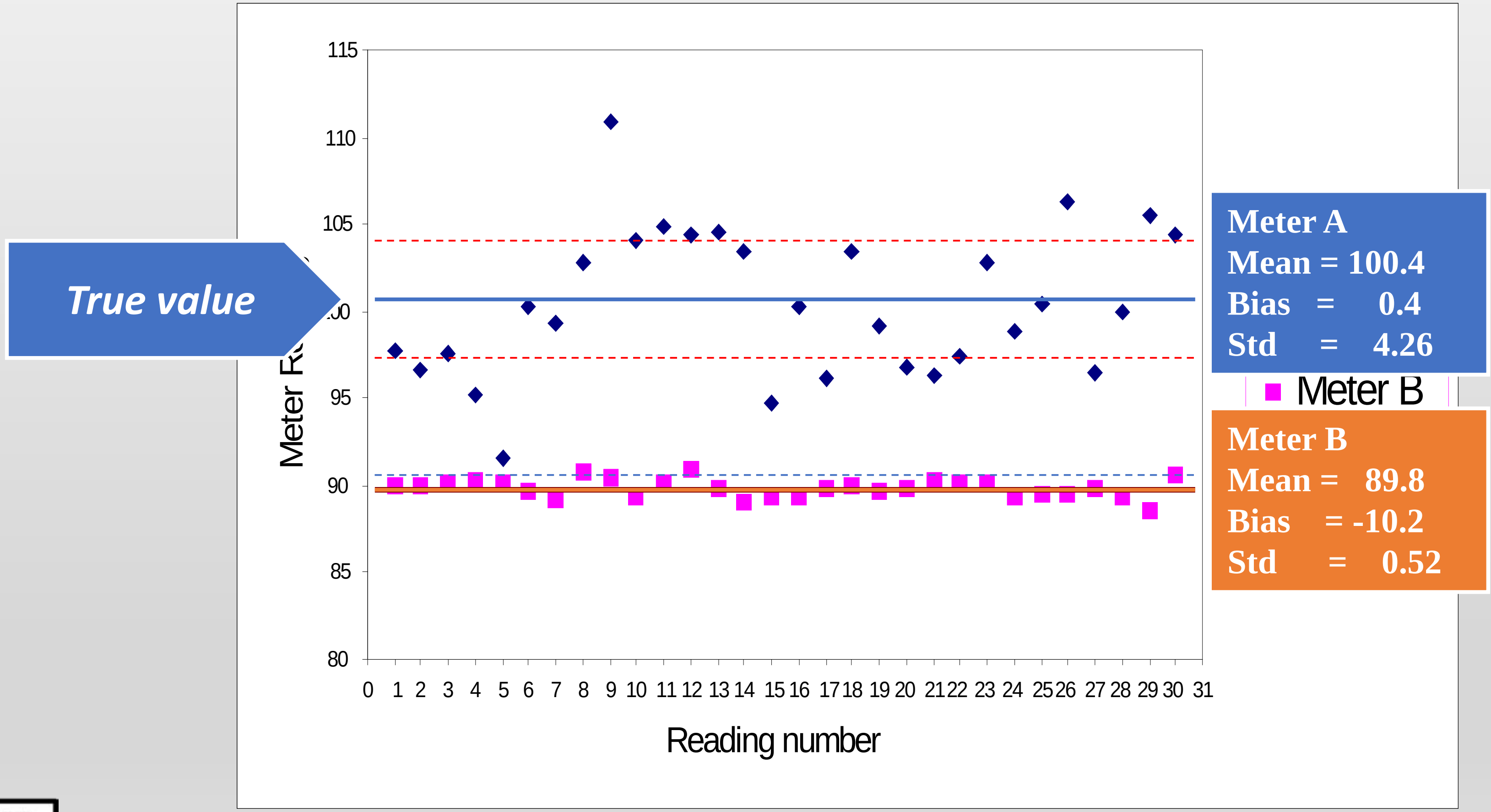
GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA'
GRICU
GET INNOVATION

Plant measurements

What's the good one?



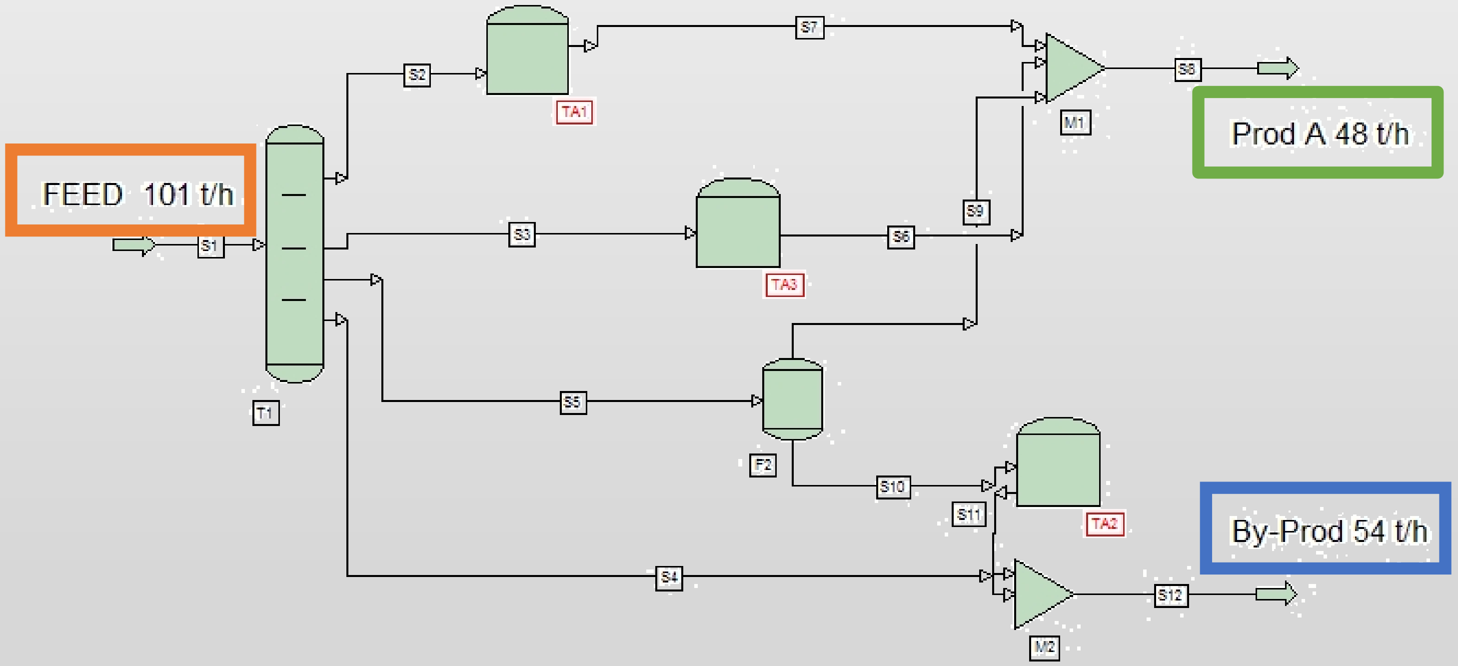
Plant measurements



Process data reconciliation
What and why

Error effects on process operations management

Because of errors, measures do not conserve mass and energy



Feed = 101 t/h ≠ *Prod = 48 t/h* + *By-pProd = 54 t/h*

Process data reconciliation

Process data reconciliation is ***NOT averaging***, but ***NLP***:

$$\min_{\mathbf{x}} \Phi = \sum_{i=1}^{NM} \omega_i \left(m_i - x_i\right)^2$$

$$s.t.: \quad g\left(\mathbf{x}\right) = 0$$

$$h\left(\mathbf{x}\right) \le 0$$

i = *ith measurement device*
NM = *total number of measurement devices*
M = *measured values*
x = *reconciled and it points out the degrees of freedom of the optimization problem*
elements of omega are the weights, usually: $\omega_i = 1/\sigma_i^2$
s.t. = *subject to*
g and h = *constraints of the optimization problem*
g(x) = *equality constraints*
h(x) = *inequality constraints*

Mass reconciliation

When the measured information is only mass flowrate (no composition measures):

$$\min_{\mathbf{w}^{rec}} \Phi = \sum_{i=1}^{NM} \omega_i \left(w_i - w_i^{rec} \right)^2$$

$$s.t.: \quad g(x) = 0$$

$$h(x) \leq 0$$

Assumption of no outliers

where **w** are the measured and reconciled mass flowrates, respectively

Feasible? It depends…

Degree of redundancy: data reconciliation is possible only if measures are ***enough*** and ***well distributed***:

$$DoR = Measures + Balances - Reconciled$$

DoR < 0	DoR = 0	DoR = 0	DoR > 0
Unsolvable, too many unknowns	*Coaptation (unsolvable data rec.)*	Inference (if the rest is reconciled)	Solvable, reconciliation (usually) possible

DoR

The larger, the better for the data reconciliation
The smaller, the better for instrumentation/CapEx

Illustrative case

Unit/Process box

Balance

Measure

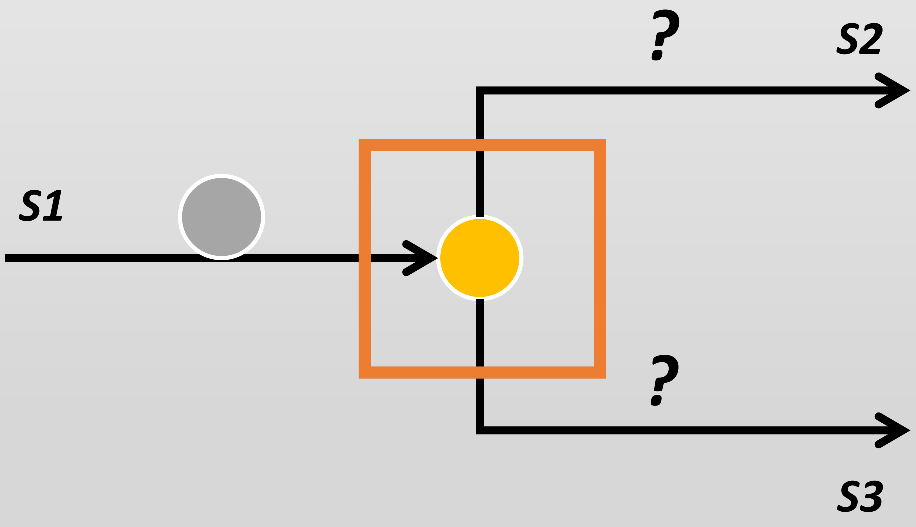
Coaptation

Inference

Reconciliation

DoR < 0

Unsolvable, too many unknowns



Illustrative case

Unit/Process box

Balance

Measure

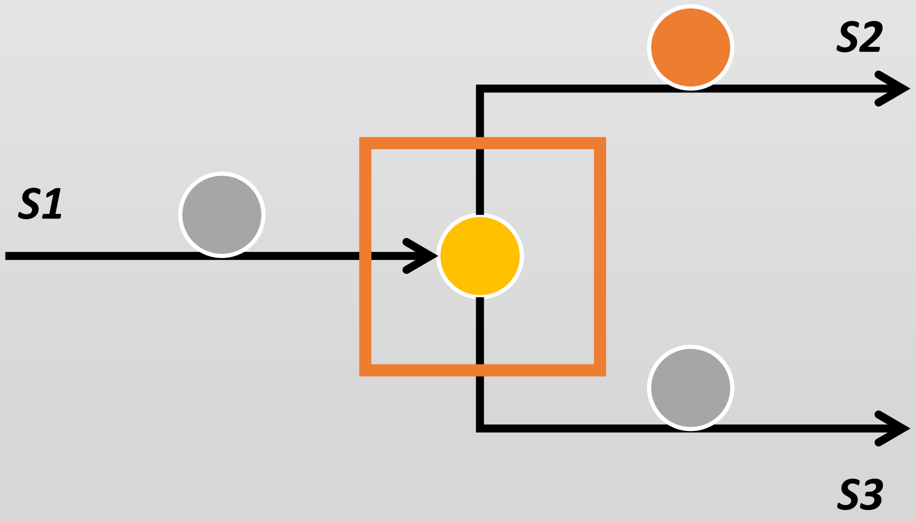
Coaptation

Inference

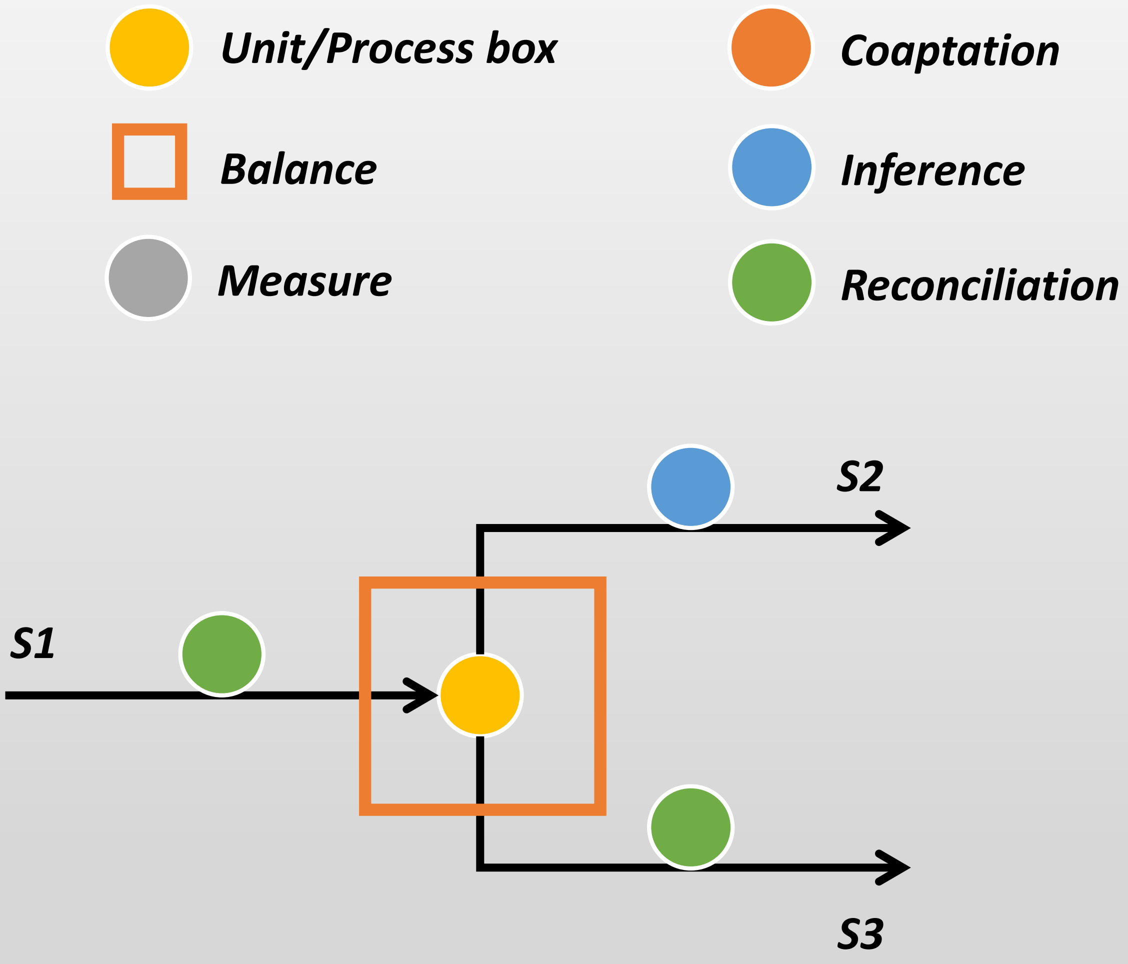
Reconciliation

DoR = 0

Coaptation (unsolvable data rec.)



Illustrative case



DoR = 0

Inference (if the rest is reconciled)

Illustrative case

Unit/Process box

Balance

Measure

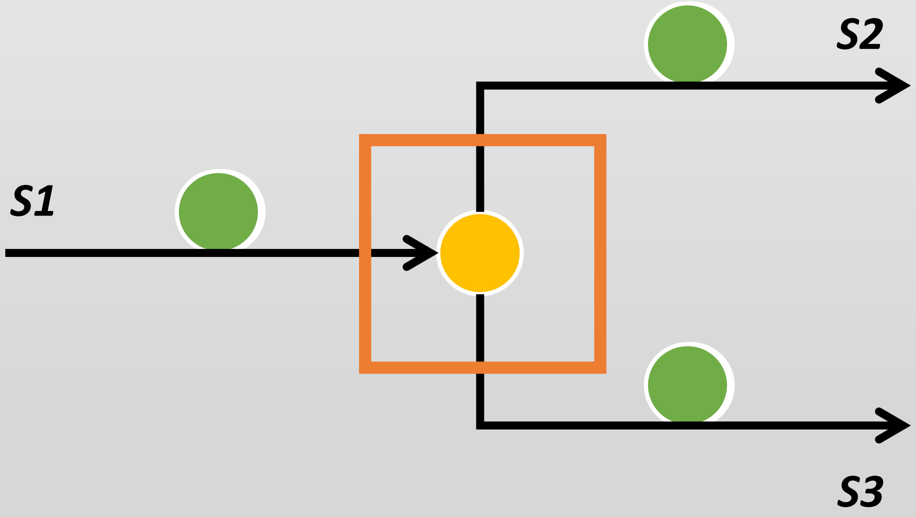
Coaptation

Inference

Reconciliation

DoR > 0

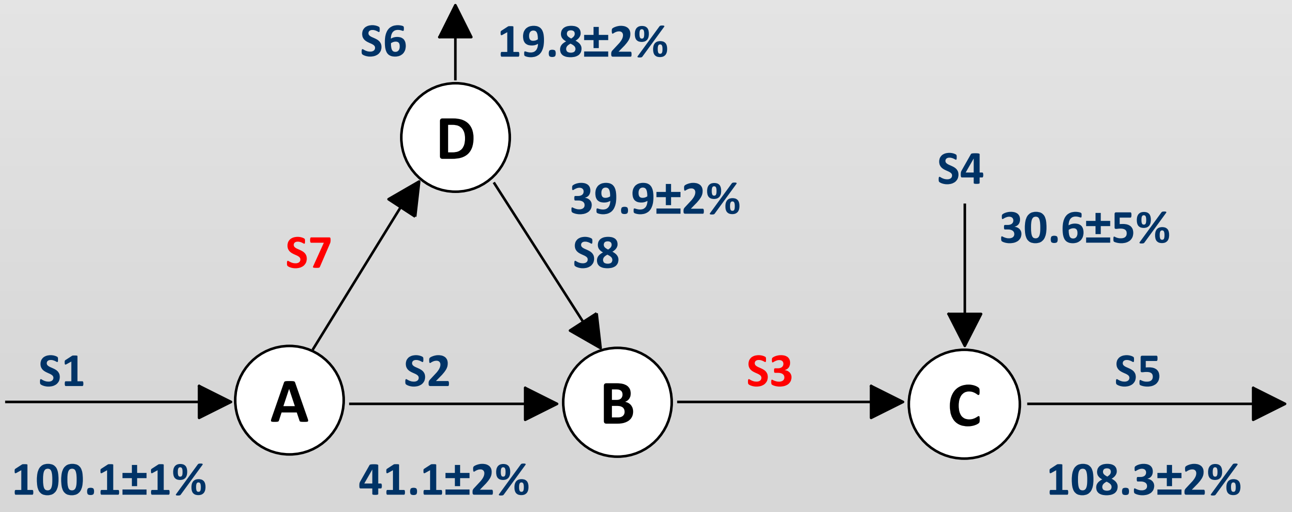
Solvable,
reconciliation
(usually) possible



Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

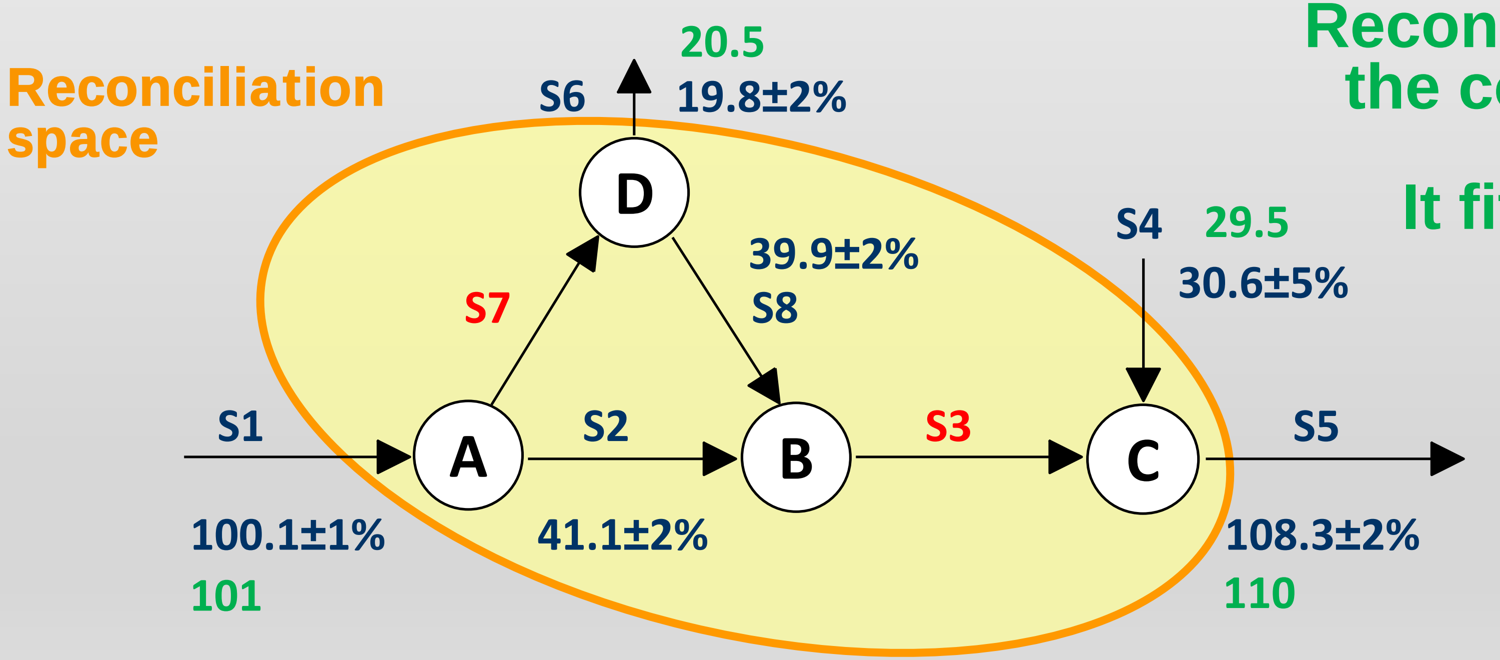
8 flowrates, 4 mass balances, 6 flow measurements



Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements



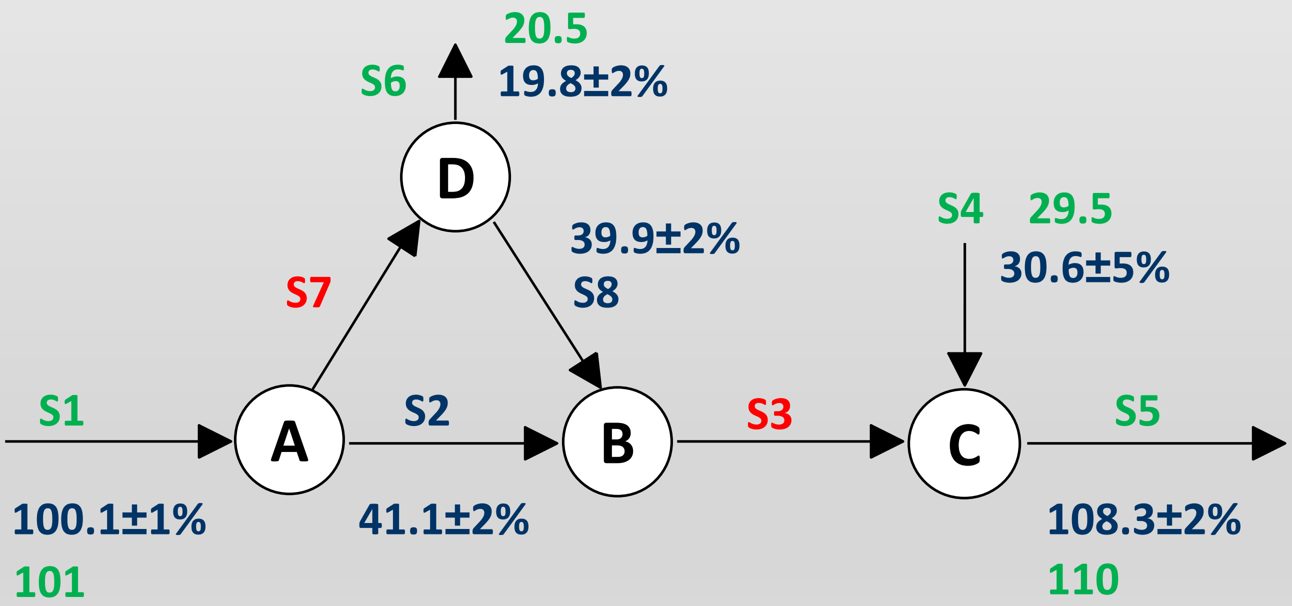
Reconciled data is the correct data

It fits the balances

Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

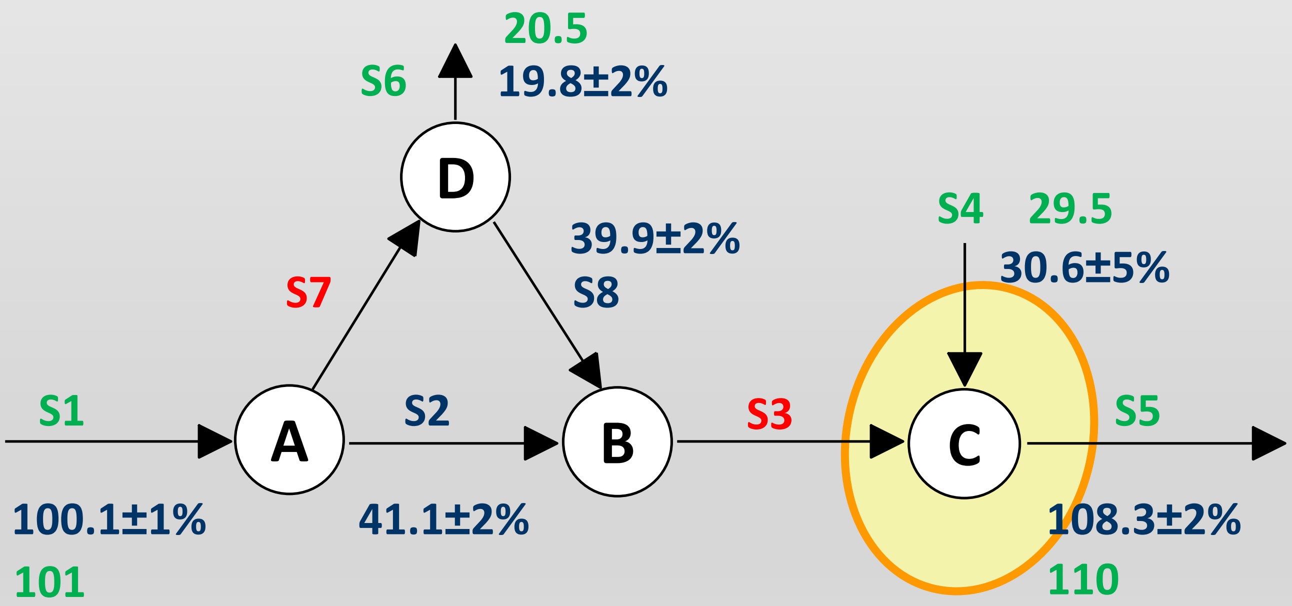
8 flowrates, 4 mass balances, 6 flow measurements



Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements



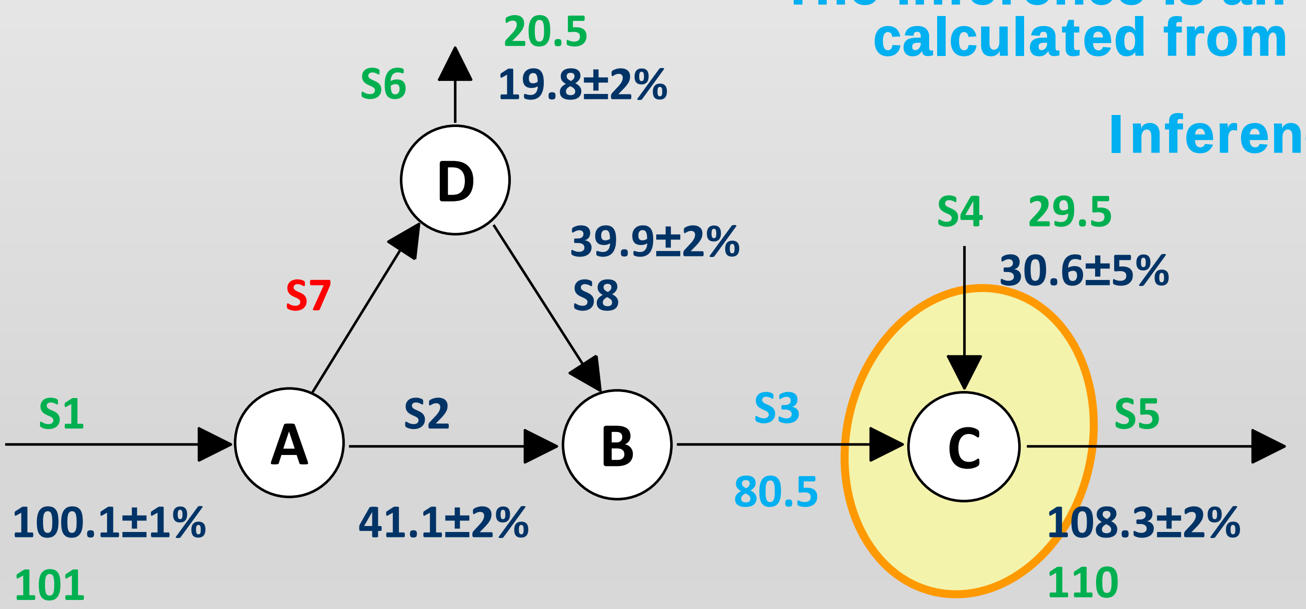
Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements

The inference is an unmeasured entity calculated from reconciled data

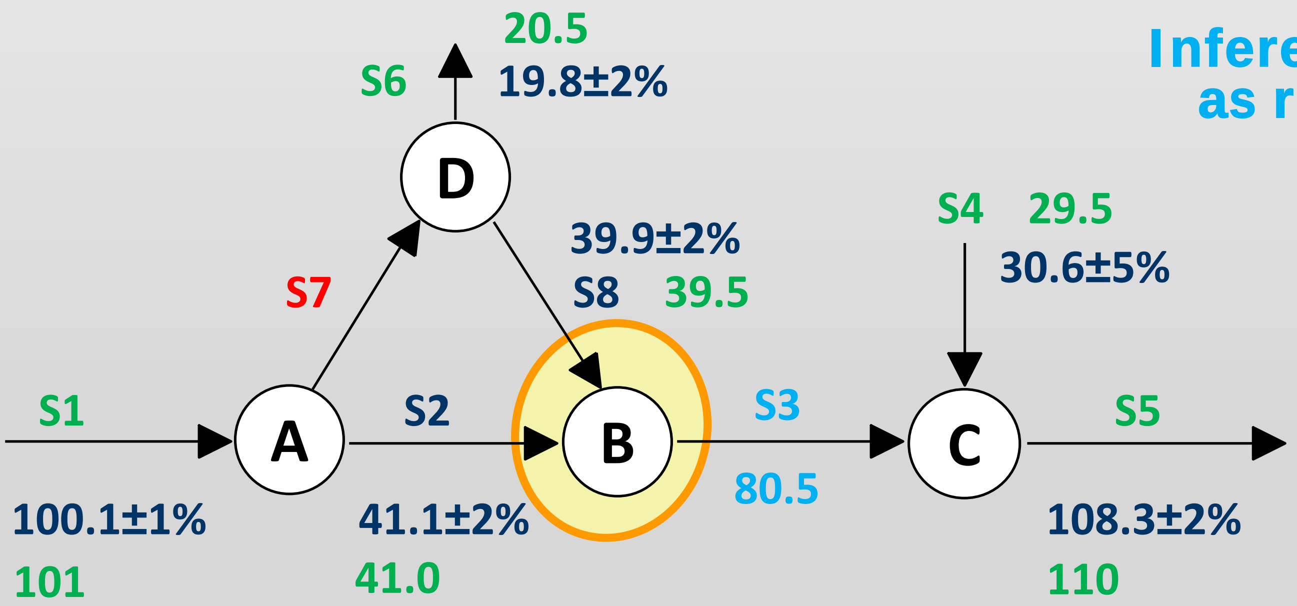
Inference is a safe number



Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements

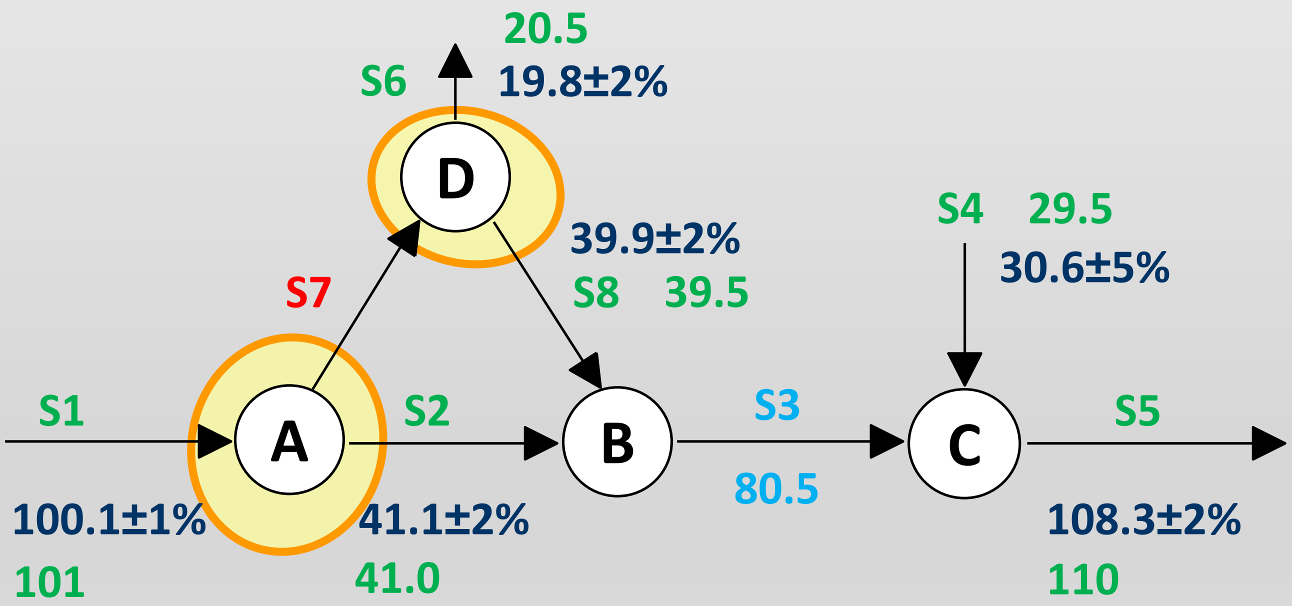


Inference can be used as reconciled data

Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements

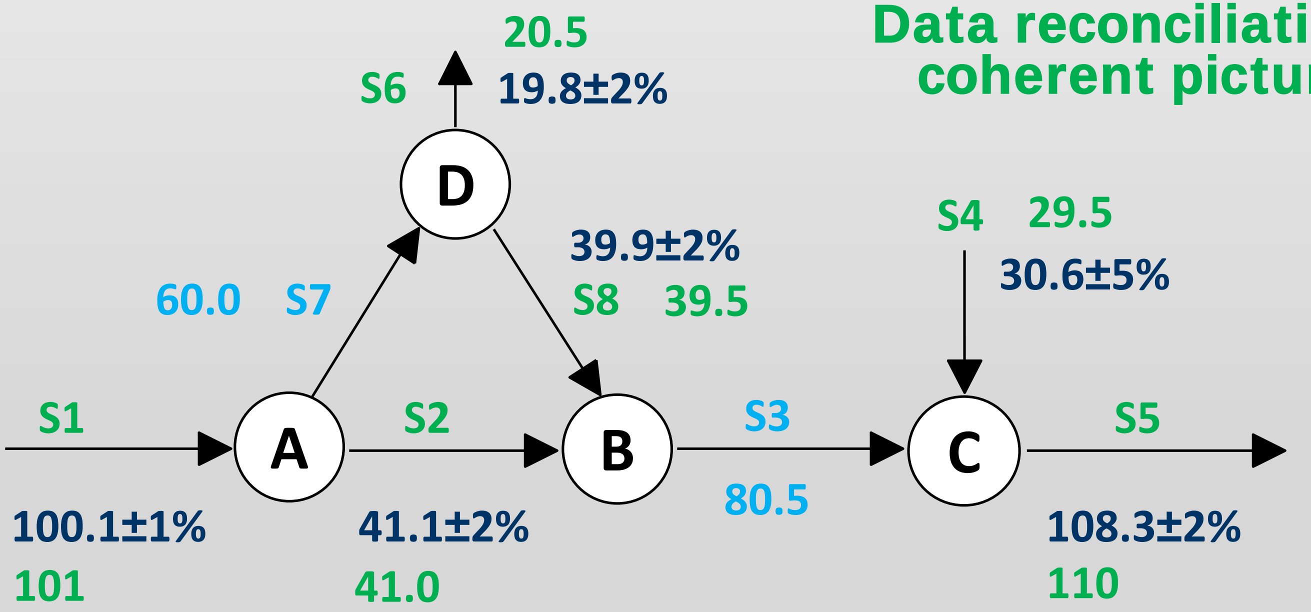


GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA'
GRICU
GET INNOVATION

Plantwide applications and exceptions

DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements



Data reconciliation provided the coherent picture of the plant

Main remarks

With data reconciliation:

The overall plant behavior is known

Operations are properly monitored

Decisions are therefore safe…

… but not necessarily wise!

Additional remark

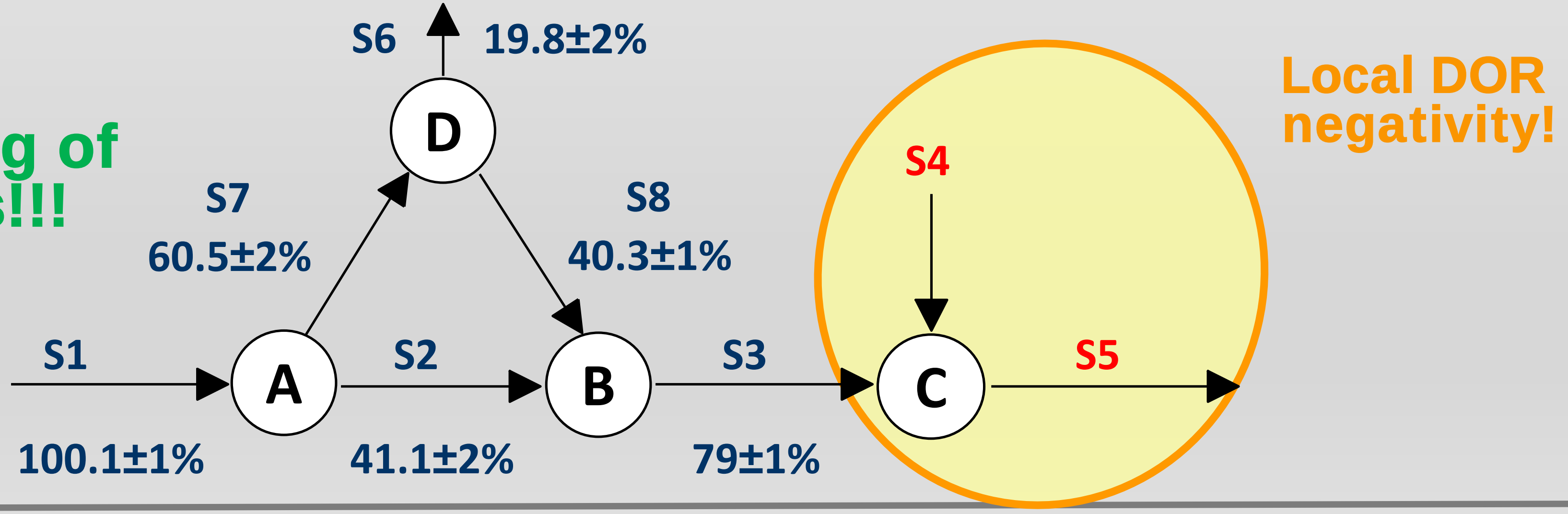
DOR = 6 + 4 – 8 = 2 > 0:

8 flowrates, 4 mass balances, 6 flow measurements

Pay attention: measurements must be cleverly positioned!

DOR > 0 is the necessary but not sufficient condition

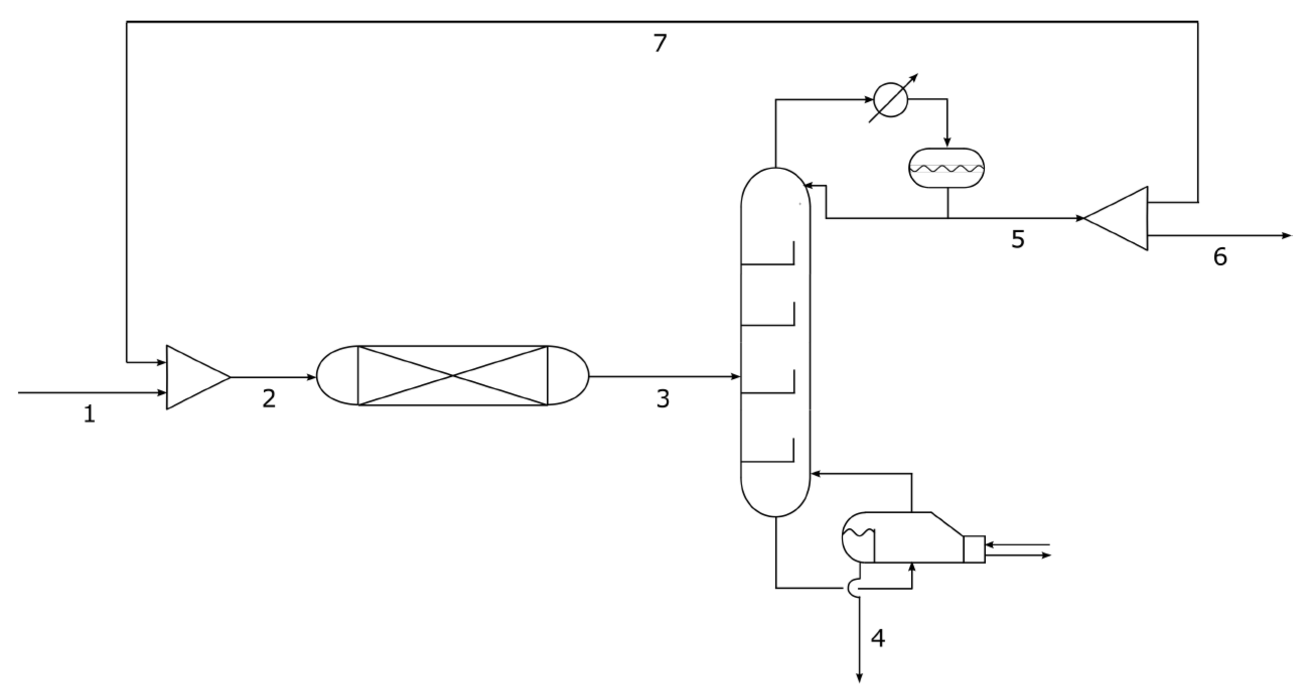
Clever positioning of plant measures!!!

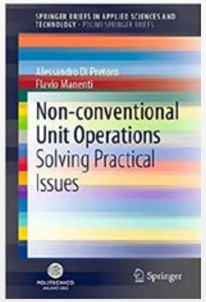


Exercise

Consider the following isomerization process. It consists of synthesis and purification/recycle sections. All the relevant mass flowrates are measured, but balances are not fulfilled.

(I) Reconcile the process data.





Process data

Stream #	Flowrate $[kg/h]$
1	95.00
2	170.00
3	175.00
4	75.00
5	103.00
6	15.00
7	82.00

Solution

Supplementary material (to be uploaded tomorrow)

Correction at 01:45 PM, March 12th, 2021

Outliers and gross errors

Reconciliation is nothing without robustness…

Definition
Gross error

A gross error is CLEARLY a bad point…

Definition
Outlier

Whatever data unsuitable for a model is called outlier for that model

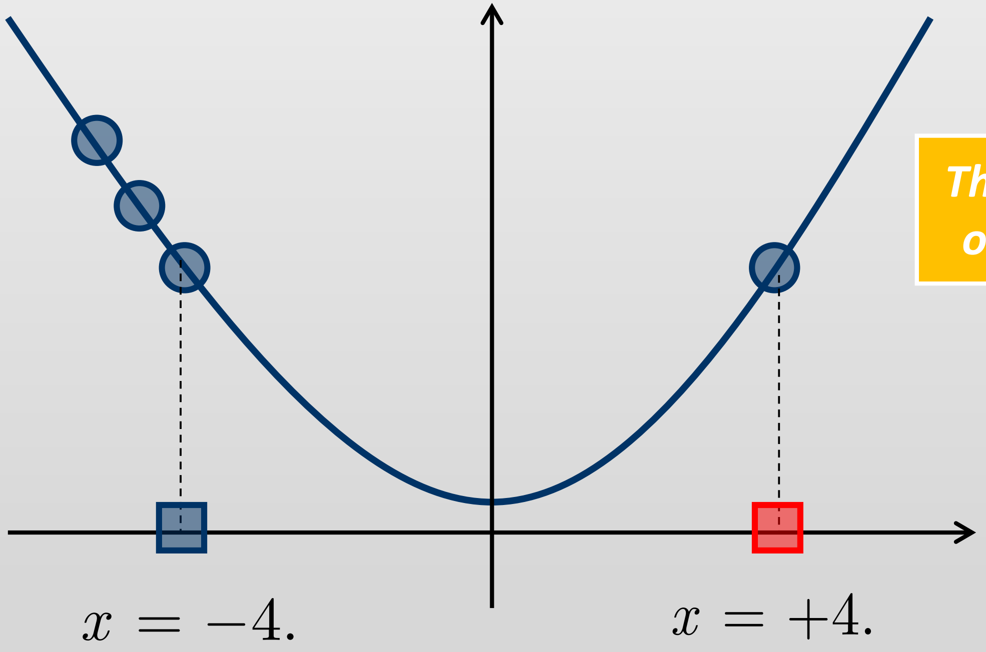
Whatever outlier is always joined to a specific model

Their (outlier and gross error) presence generates a significant increase for the mean square error (measure – true value).

$$MSE = \sum_{i=1}^{NM} \frac{\left(y_i - y_i^{rec}\right)^2}{NM} \qquad NM > 30$$

But…
Exceptions

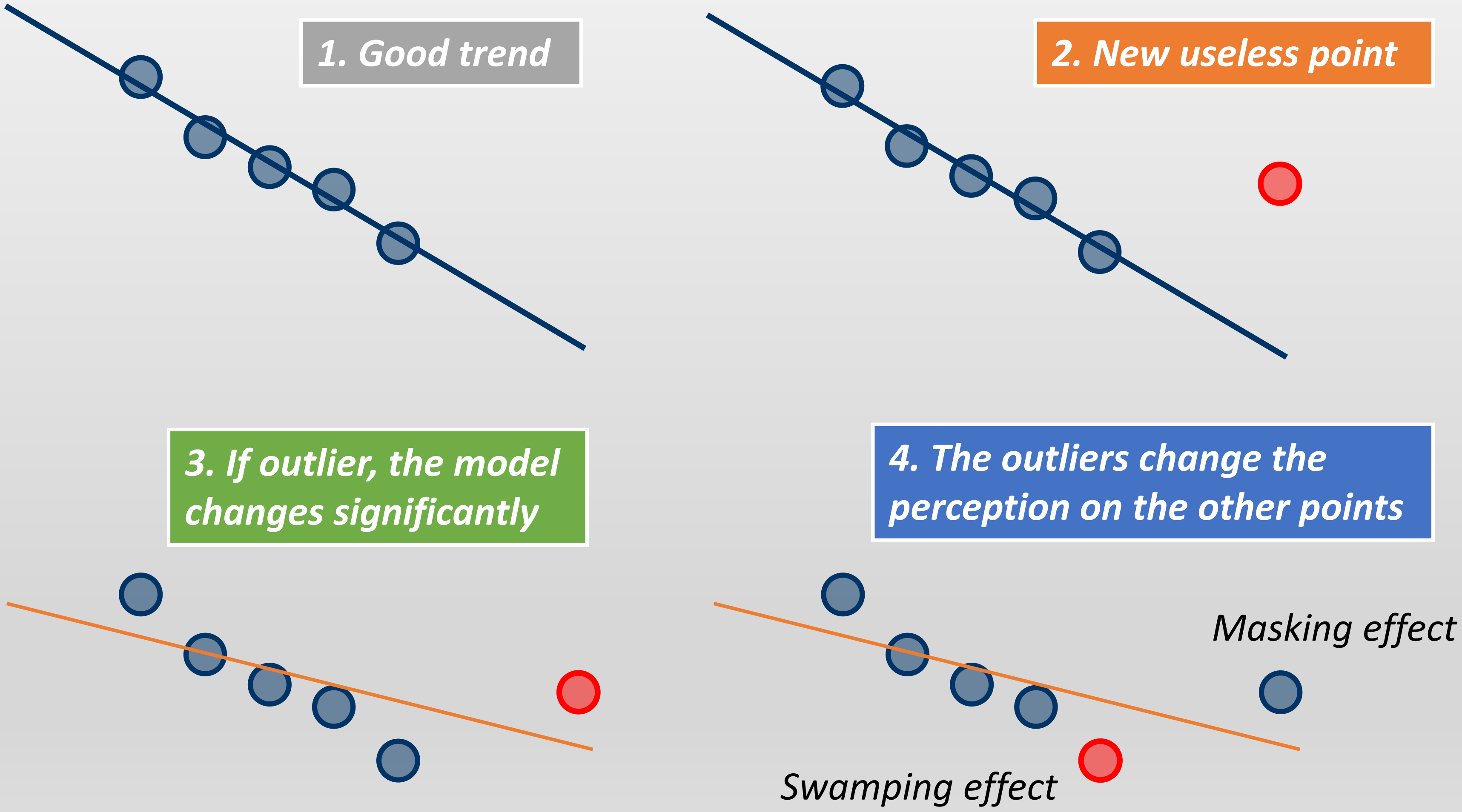
Data collection



The gross error is not an outlier for the Model!!

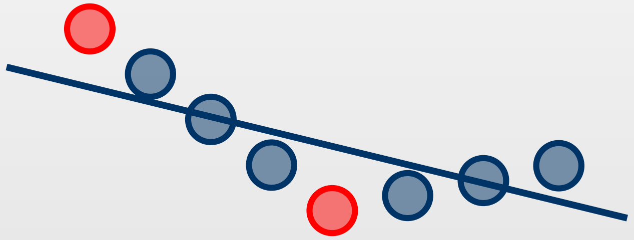
Gross error generation

Sources of outliers in big data sets

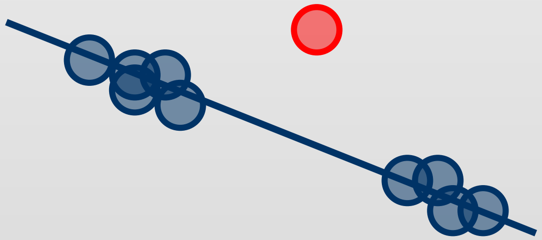


Sources of outliers in big data sets

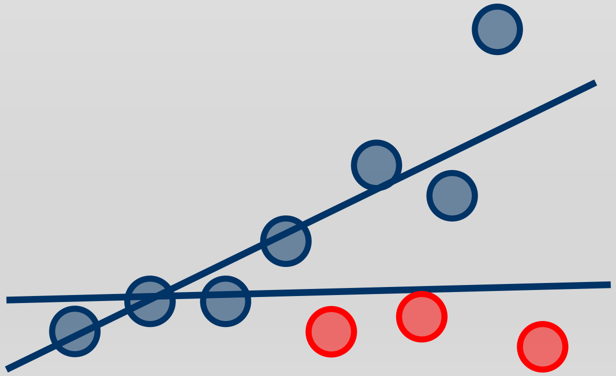
Inadequate models



Inadequate DoE or measures

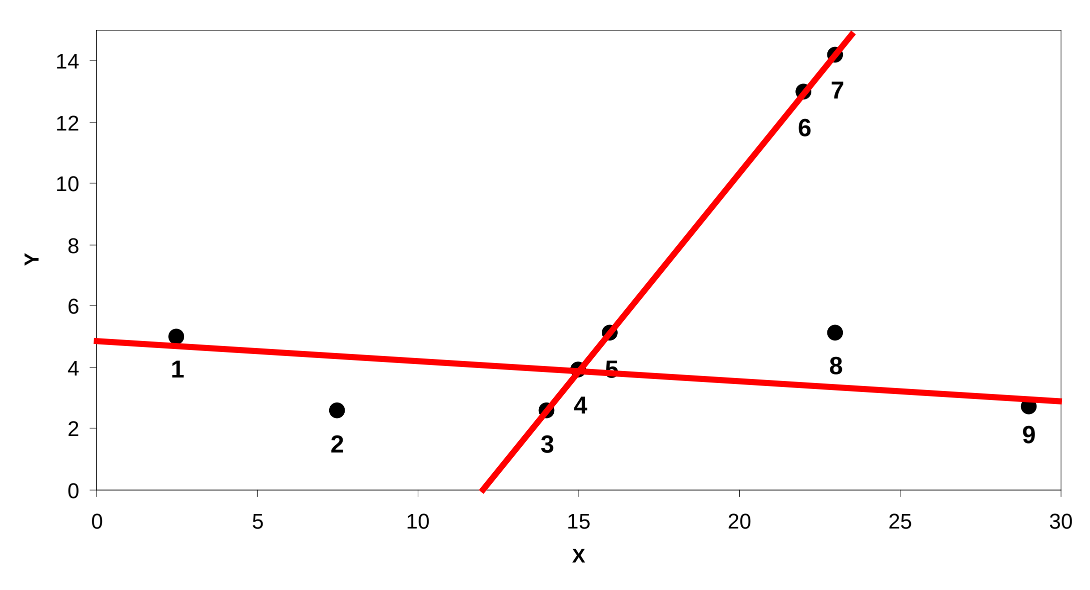


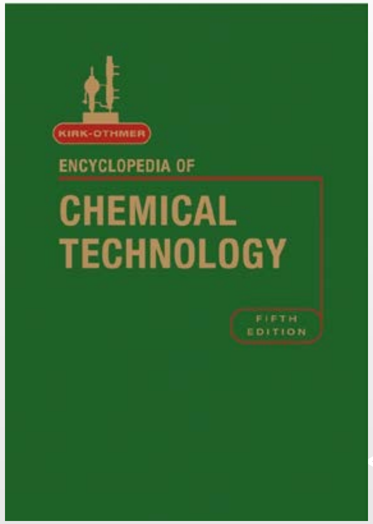
Violated homoscedasticity

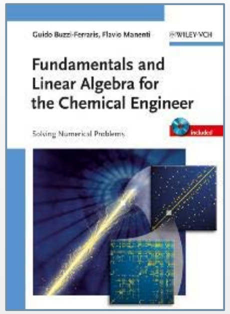


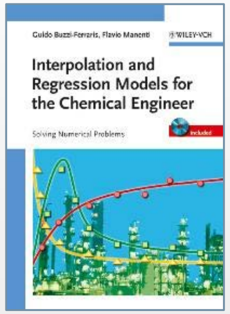
Literature example

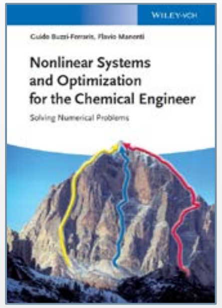
Ryan, T. P., Modern Regression Methods. 2nd Ed., Wiley, NJ, 2009

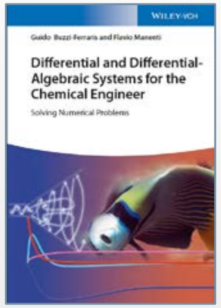












Kirk-Othmer Encyclopedia of Chemical Technology
Guido Buzzi-Ferraris, Flavio Manenti
Data Interpretation and Correlation, John Wiley & Sons, New York, USA

Identification of outliers
Robustness? Yes, but not too much

Computers and Chemical Engineering 35 (2011) 388–390

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Note

Outlier detection in large data sets

Guido Buzzi-Ferraris*, Flavio Manenti

Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Recommended 3-pages reading

GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA'
GRICU
GET INNOVATION

Clever mean and clever variance

n points $y_i\left(i=1,\ldots,n\right)$ with very large n

Symmetrical distribution with μ σ^2 and data reading/ordering on the DCS w.r.t. the error

Zeroth-order clever mean: $$cm_0=\frac{\sum_{i=1}^{n}y_i}{n}=\bar{y}$$

Zeroth-order clever variance: $$cv_0=\frac{\sum_{i=1}^{n}\left(y_i-cm_0\right)^2}{n-1}=s^2$$

Outliers at the boundaries

Assuming y*_1 as the first possible outlier, first-order clever mean and clever variance are:

$$cm_1 = \frac{\sum_{i=1}^{n} y_i - y_1^*}{n-1} \qquad cv_1 = \frac{\sum_{i=1}^{n} \left(y_i - cm_1\right)^2 - \left(y_1^* - cm_1\right)^2}{n-2}$$

Identification condition:

$$\left|cm_1 - y_1^*\right| > \delta \cdot \sqrt{cv_1}$$

If confirmed as outlier, the procedure is iterated until

$$\boxed{\left|cm_k - y_k^*\right| > \delta \cdot \sqrt{cv_k}} \qquad \boxed{\left|cm_{k+1} - y_{k+1}^*\right| < \delta \cdot \sqrt{cv_{k+1}}}$$

GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA' GRICU GET INNOVATION

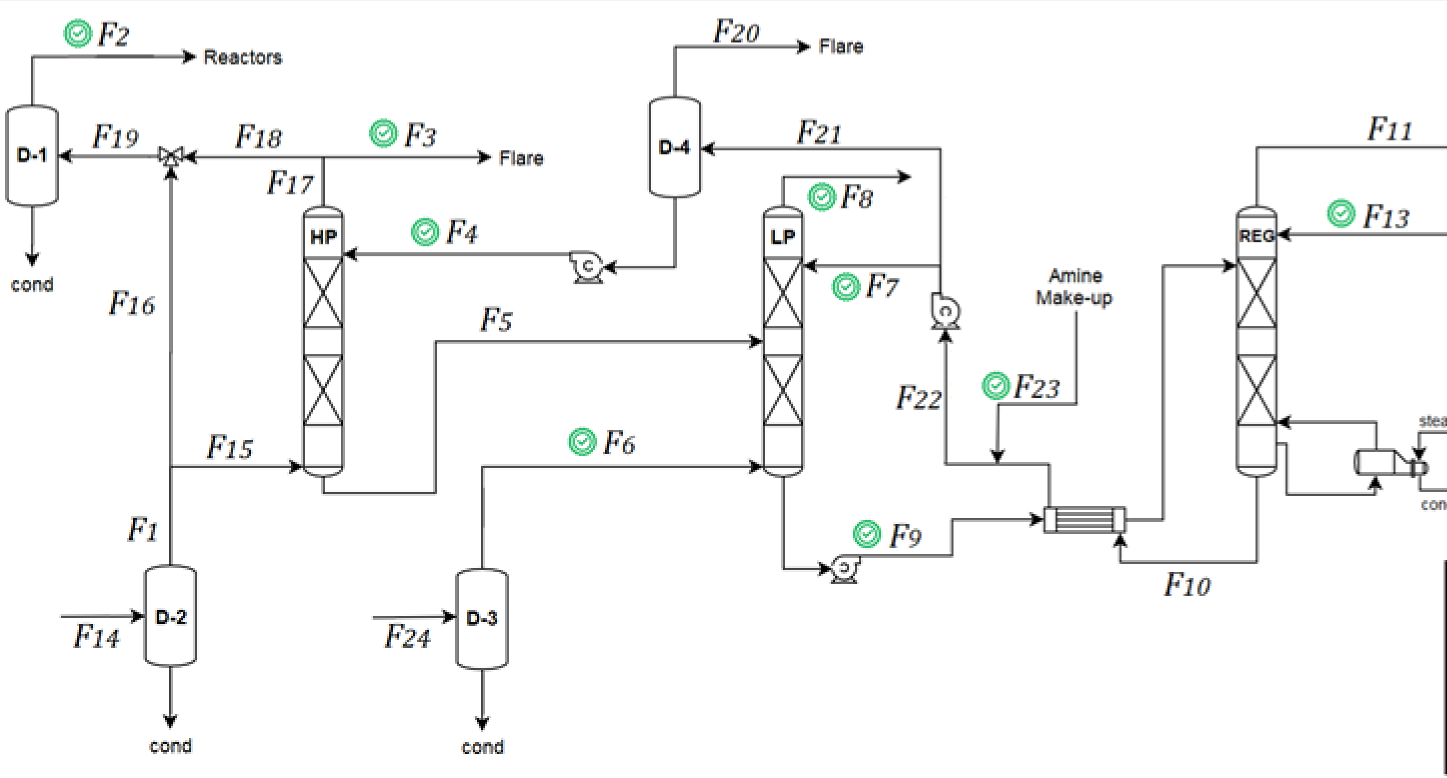
Industrial application
Big data analytics in process industry

CO_2 capture industrial plant

Courtesy of
Ing. Gallo

Process scheme and measures

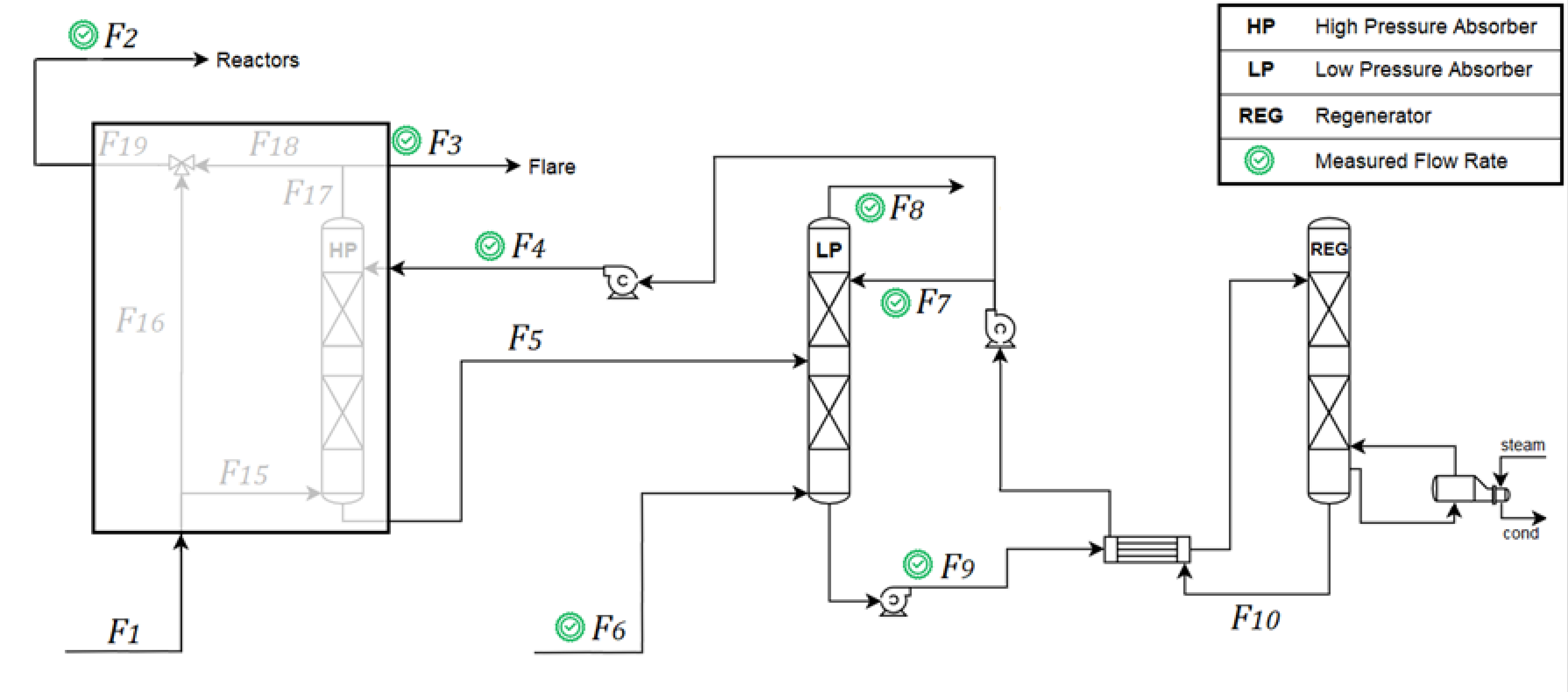
Redund… WHAT??? Ok… it's infeasibile…



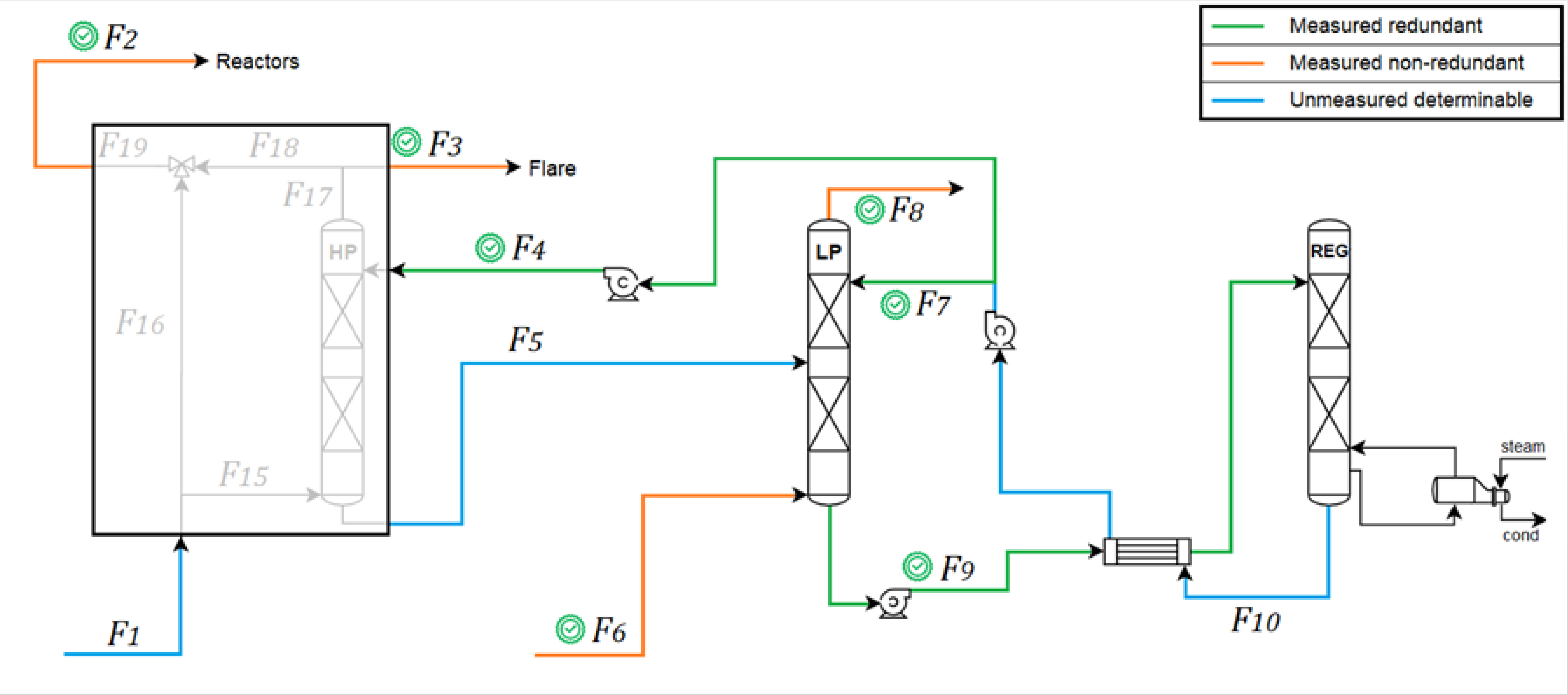
HP	High Pressure Absorber
LP	Low Pressure Absorber
REG	Regenerator
D-i	Separation Drum
	Measured Flow Rate

Are you sure? Simplify…

Remove NNF and introduce white boxes…



Now is feasible!



Digital architecture

Robust optimizer
- BzzMath™ Process Digitalization Library
- ROBOPT™
- Visual Studio

Middleware interface
- ASCII /e

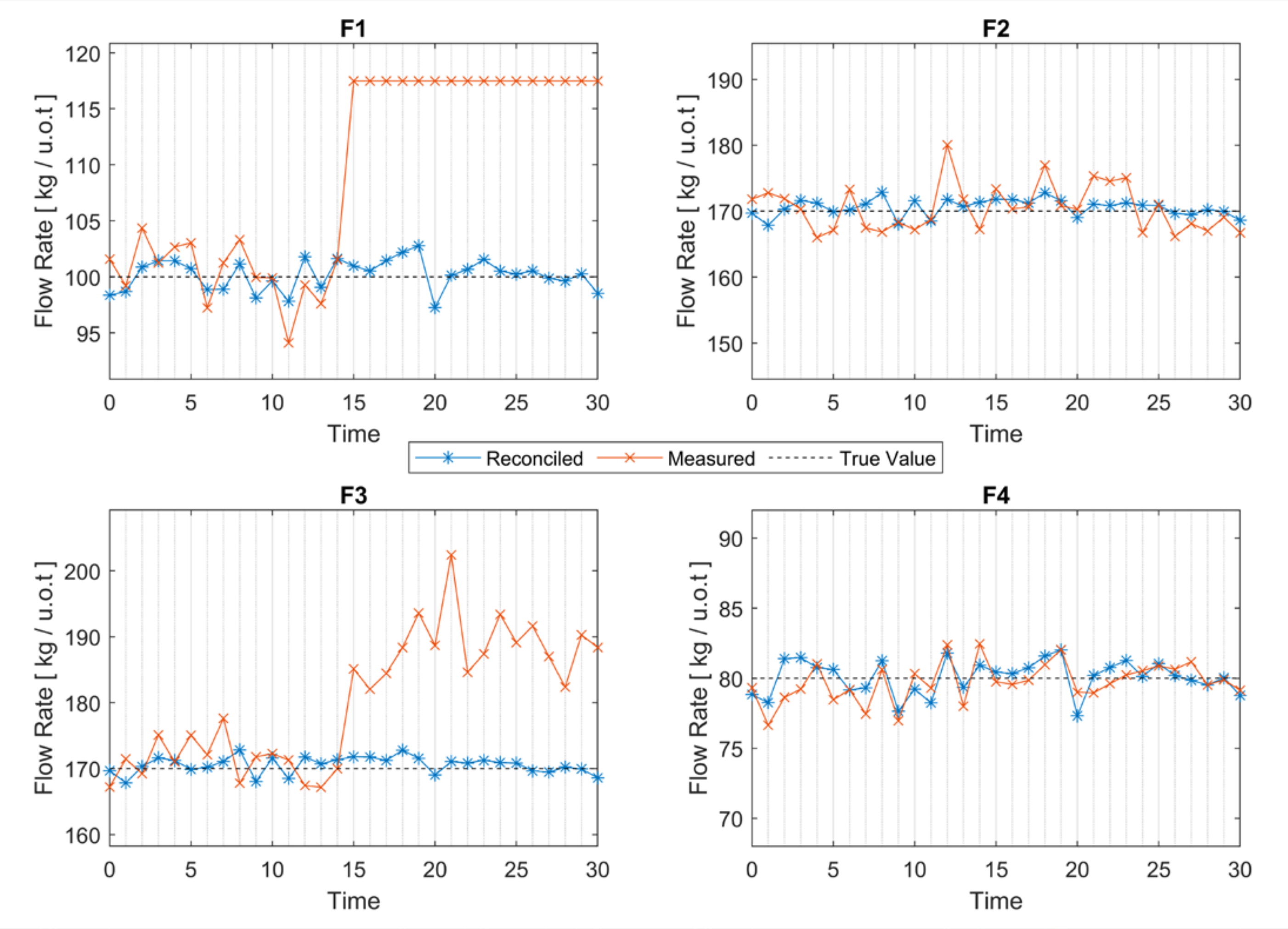
Process detailed simulation
- Schneider Electric — SimSci PRO/II
- aspentech — AspenHysys
- Honeywell — Unisim Design
- ...

Dedicated DCS
- YOKOGAWA
- Exaquantum

Exhausted oil refinery
- ITELYUM
- Pieve Fissiraga (LO, Italy)

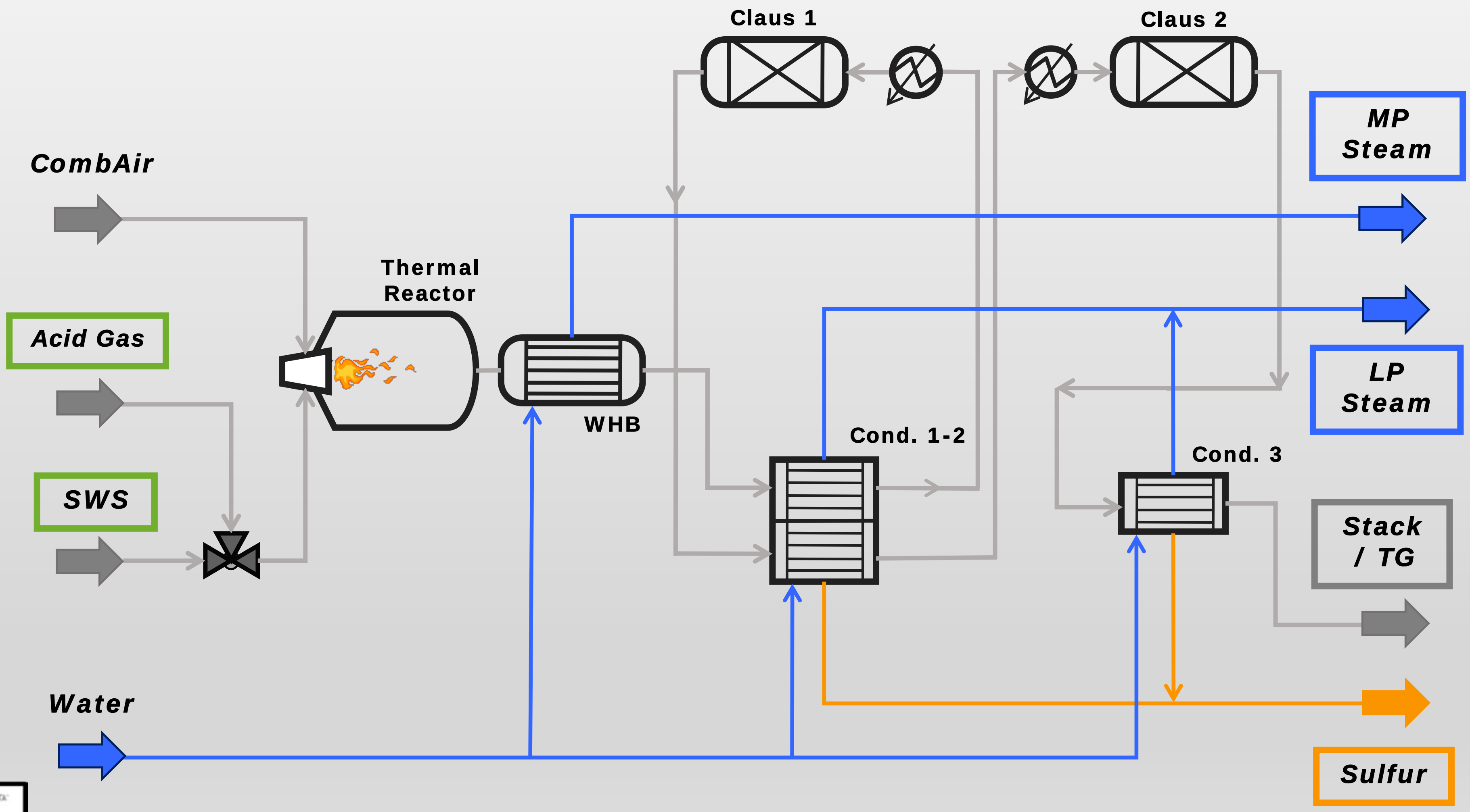
GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA' GRICU GET INNOVATION

Powerful of robust data reconciliation

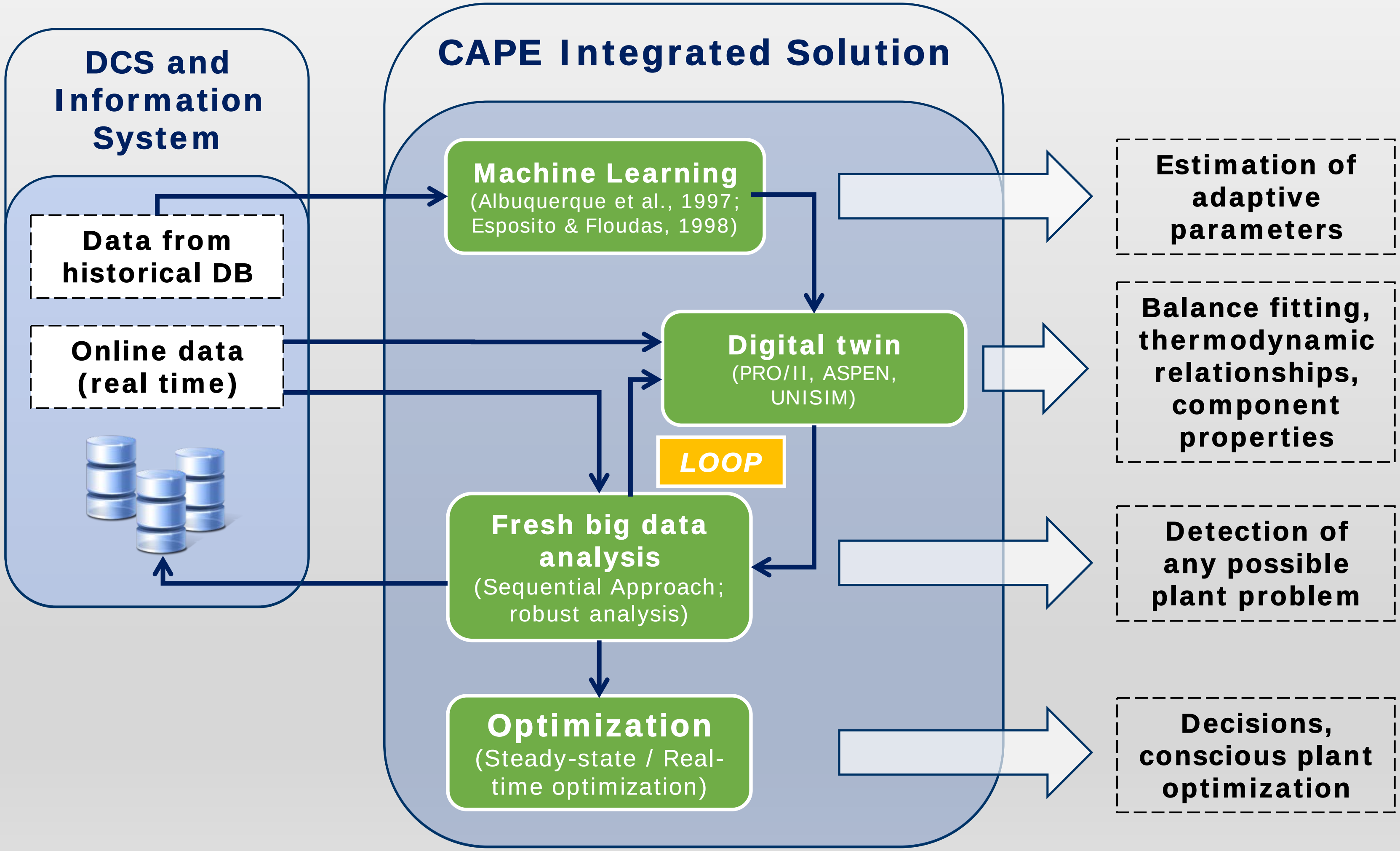


Industrial application
Robust data reconciliation in Lukoil

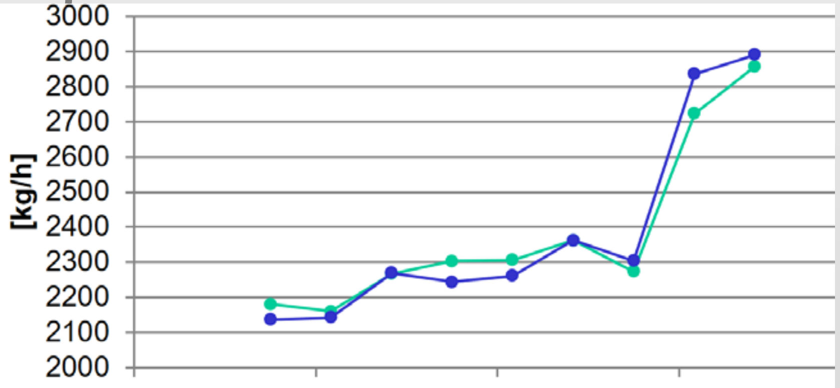
Sulfur recovery unit



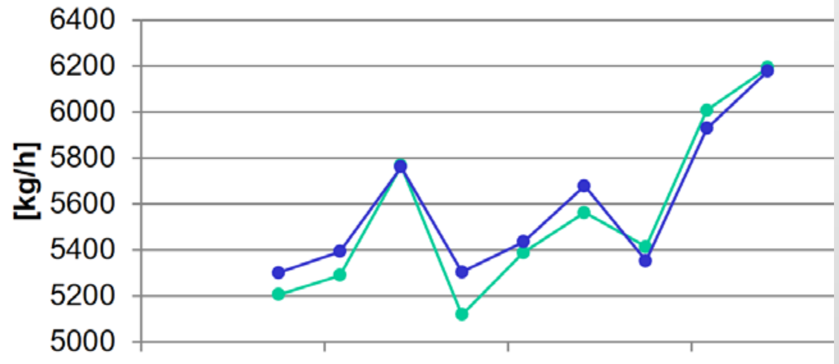
Digital architecture



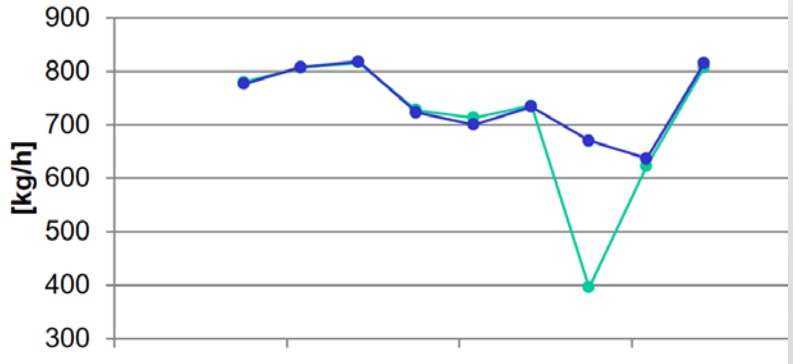
Gross error identification



Acid gas flowrate



Combustion air

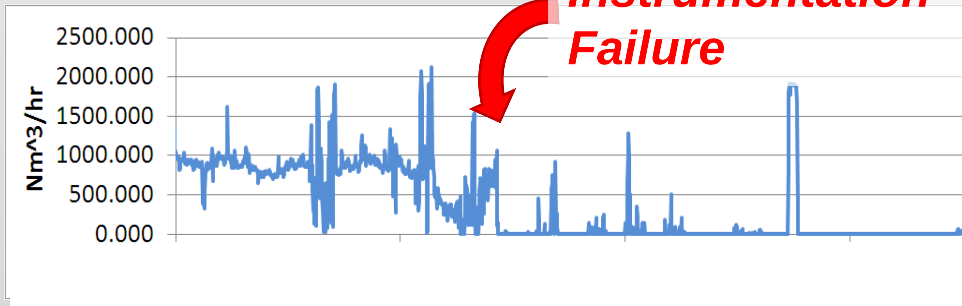


SWS flowrate

Effects of a misidentified gross error? Catastrophic ones…

Measures	Scenario #3							
	Measured		Nonrobust Rec		Res. [%]	Our CAPE Sol.		Res [%]
AAG Flow Rate	**2704.03**	kg/hr	3435.84	kg/hr	27.06	2702.40	kg/hr	-0.06
Acid Gas Temp	**52.66**	°C	53.72	°C	2.01	50.48	°C	-4.14
Sour Water Flow Rate	**0.00**	kg/hr	-	kg/hr	-	1179.36	kg/hr	-
Sour Water Temp	**84.99**	°C	-	°C	-	84.29	°C	-0.82
Combustion Air Flow Rate	**7352.21**	kg/hr	5721.63	kg/hr	-22.18	7466.47	kg/hr	1.55
Combustion Air Temp	**175.92**	°C	178.97	°C	1.73	187.47	°C	6.57
Furnace Temp	**1380.26**	°C	1264.32	°C	-8.40	1374.59	°C	-0.41
WHB Temp	**319.10**	°C	305.78	°C	-4.17	310.03	°C	-2.84
First Condenser Temp	**167.87**	°C	168.25	°C	0.23	165.15	°C	-1.62
First Claus Inlet Temp	**227.35**	°C	240.84	°C	5.93	228.13	°C	0.34
First Claus Outlet Temp	**290.26**	°C	280.87	°C	-3.24	290.30	°C	0.01
Second Condenser Temp	**163.65**	°C	164.67	°C	0.62	157.04	°C	-4.04
Second Claus Inlet Temp	**210.05**	°C	231.74	°C	10.33	221.26	°C	5.34
Second Claus Outlet Temp	**216.62**	°C	234.16	°C	8.10	228.36	°C	5.42
Steam Generated at WHB	**7152.15**	kg/hr	5473.00	kg/hr	-23.48	7198.00	kg/hr	0.64
H_2S at Tail Gas	**6330**	ppmv	4720.27	ppmv	-25.43	6148	ppmv	-2.88
SO_2 at Tail Gas	**2631**	ppmv	3029.00	ppmv	15.13	2621	ppmv	-0.38
H_2S/SO_2 at Tail Gas	**2.41**	-	1.56	-	-35.23	2.35	-	-2.50
Recovered Sulfur	**-**	[t/day]	65.14	t/day		62.14	t/day	
Final Value of Obj Func			0.296			0.0941		

Instrumentation Failure



GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITÀ
GRICU
GET INNOVATION

References

1. Crowe, C. M. (1996). Data reconciliation - Progress and challenges. Journal of Process Control, 6(2-3 SPEC. ISS.), 89–98. https://doi.org/10.1016/0959-1524(96)00012-1
2. Sánchez, M., Romagnoli, J., Jiang, Q., & Bagajewicz, M. (1999). Simultaneous estimation of biases and leaks in process plants. Computers and Chemical Engineering, 23(7), 841–857.
3. Narasimhan, S., & Jordache, C. (2000). Data reconciliation and gross error detection : an intelligent use of process data. In Annals of the New York Academy of Sciences (Vol. 195).
4. Buzzi-Ferraris, G., & Manenti, F. (2011). Outlier detection in large data sets. Computers and Chemical Engineering, 35(2), 388–390.
5. Llanos, C. E., Sanchéz, M. C., & Maronna, R. A. (2015). Robust Estimators for Data Reconciliation. Industrial and Engineering Chemistry Research, 54(18), 5096–5105. https://doi.org/10.1021/ie504735a
6. Loyola-Fuentes, J., & Smith, R. (2019). Data reconciliation and gross error detection in crude oil pre-heat trains undergoing shell-side and tube-side fouling deposition. Energy, 183, 368–384.
7. **Romagnoli, A. J.; Sanchez, M. C. (2000). Data Processing and Reconciliation for Chemical Process Operations, Volume 2. Academic Press.** BOOK
8. Buzzi-Ferraris, G., & Manenti, F. (2010). Interpolation and Regression Models for the Chemical Engineer: Solving Numerical Problems, Volume 2. Wiley-VCH, Weinheim, Germany BOOK
9. ***PROCESS DIGITALIZATION LIBRARY***: download section at https://super.chem.polimi.it TOOL

GRUPPO DI INGEGNERIA CHIMICA DELL'UNIVERSITA' GRICU GET INNOVATION