Data Reconciliation in the process industries

Prof. Cesar de Prada

Dpt. Systems Engineering and Automatic Control

University of Valladolid, Spain

prada@autom.uva.es

Outline

- ✓ Presentation
- ✓ Measurements and information
- ✓ Data reconciliation
- ✓ Gross errors
- ✓ Examples:
  - Sugar factory
  - Petrol refinery
- ✓ Conclusions

Today's process plants

More technology

More complex processes

More instrumentation and systems

More norms and regulations

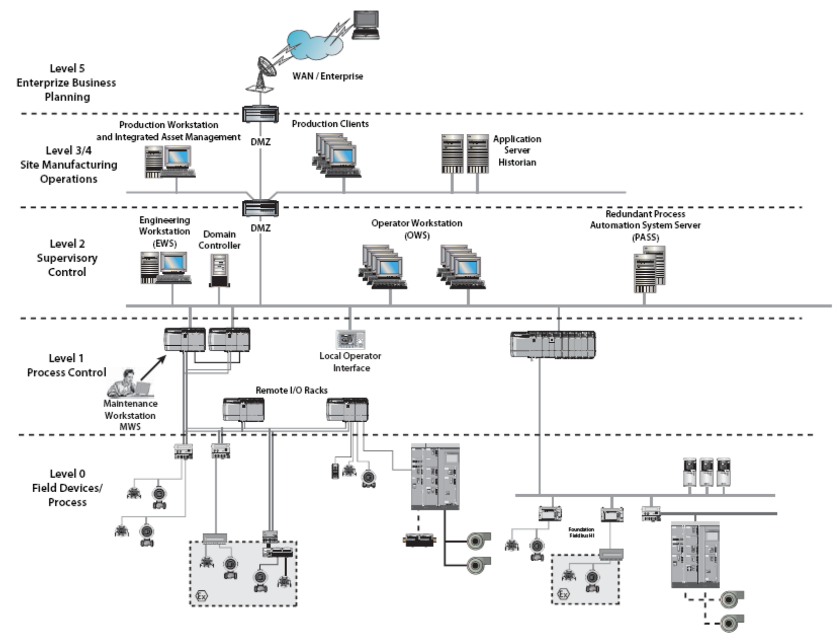
Reduced technical staff

More data than ever

Higher market pressures

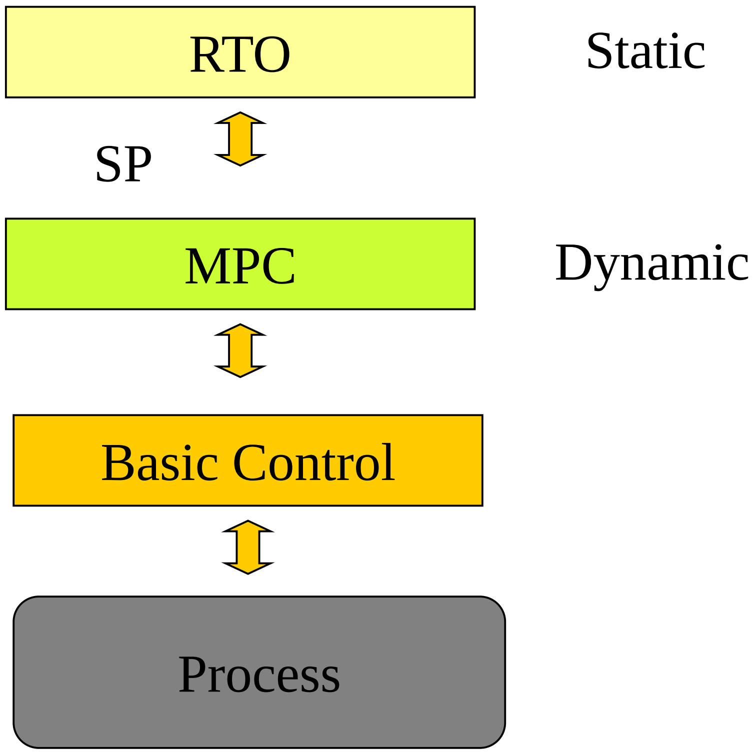
From data to knowledge

- ✓ Huge amount of data available in real time or historians.
- ✓ Better instrumentation and new sensors
- ✓ With less trained people in the control room or the technical teams, supporting tools are required for process safety, process behaviour predictions, help in Abnormal Situation Management,…
- ✓ Models and simulations, decision support systems, etc., are recognized as elements to condense knowledge
- ✓ The focus is on software applications at the MES level

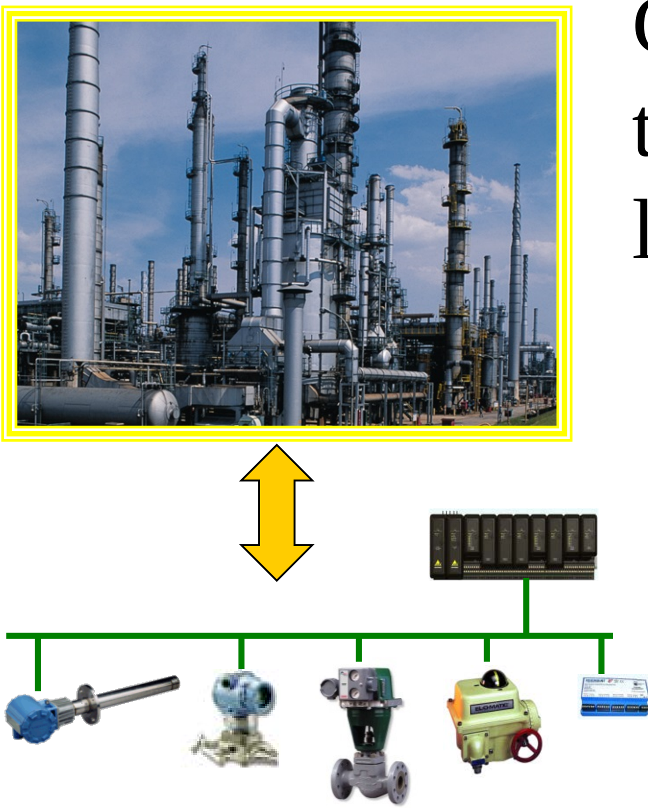


Models

- There is a lot of interest in the optimal (economic) operation of the processes
- Models play a key role in supporting the decision making process
- Advanced Control and Economic Optimization are the right tools
- Successful implementation requires suitable models and process information
- Few tools for estimating earnings and improvements



Data / Information



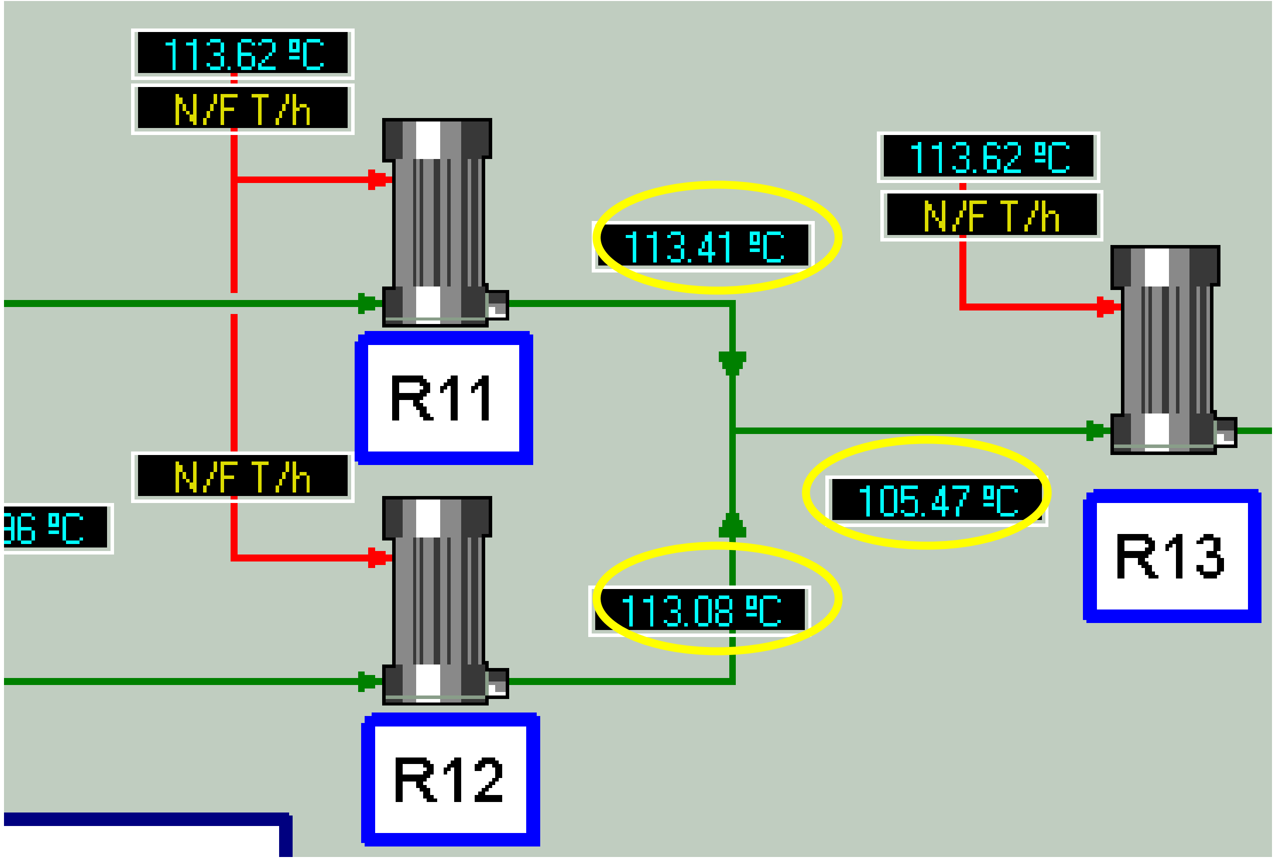
Complex decisions taken at different levels

From data to reliable and coherent information

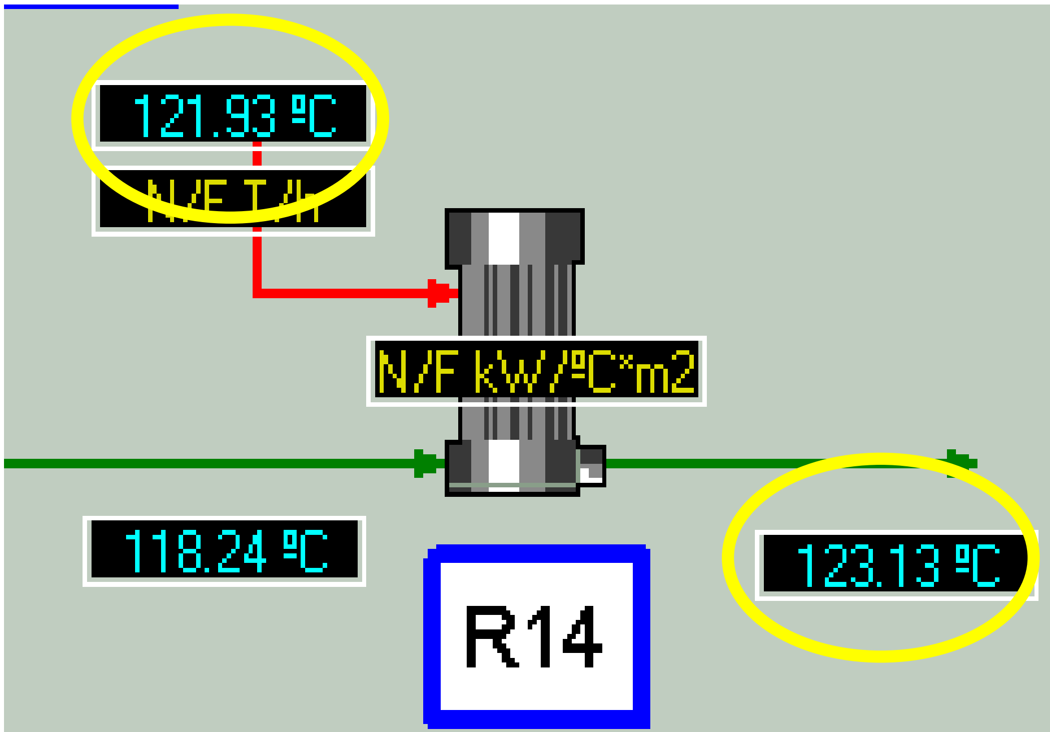
Plant data

- Some measurements are not consistent or unreliable
- There are many unmeasured variables
- Model parameters need to be estimated

Inconsistencies



Inconsistences



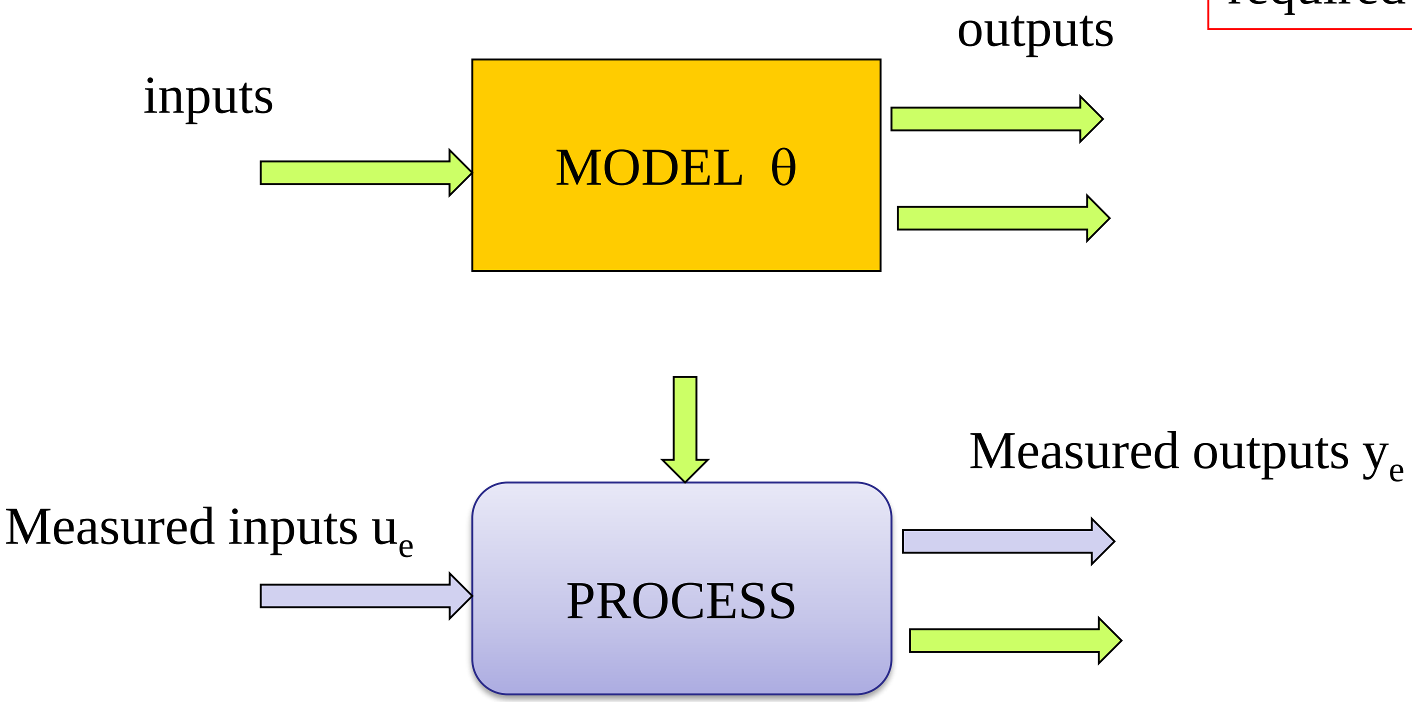
Data reconciliation

- ✓ Use plant/lab measurements and knowledge stored in the models to:
  - – Estimate the values of all variables and model parameters coherent with a process model and as close as possible to the measurements
  - – Detect and correct inconsistencies in the measurements
- ✓ Formulated as an optimization problem

Data reconciliation

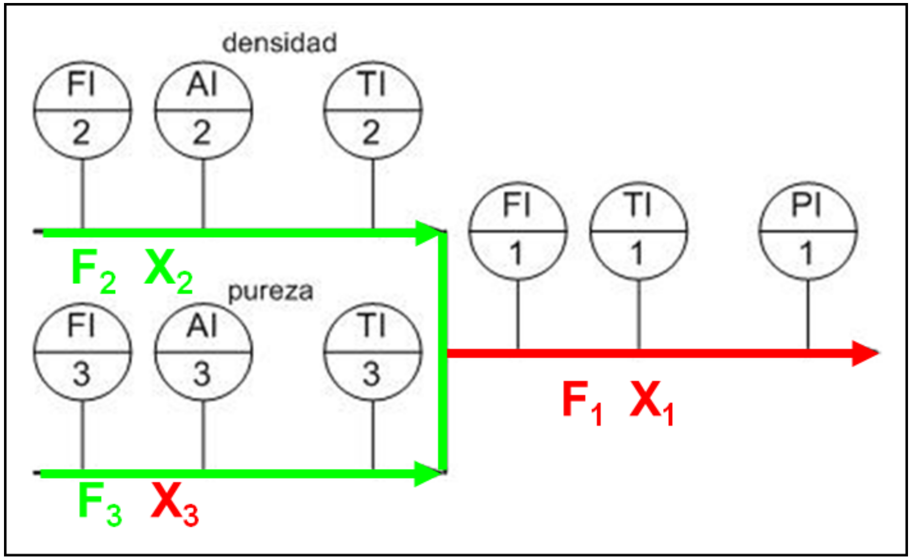
$$\frac{dx}{dt} = f(x,u,\theta) \qquad y = h(x,u,\theta)$$

A certain degree of redundancy in the measurements is required



Redundancy

Mass balances



$$F_1 = F_2 + F_3$$
$$F_1 X_1 = F_2 X_2 + F_3 X_3$$

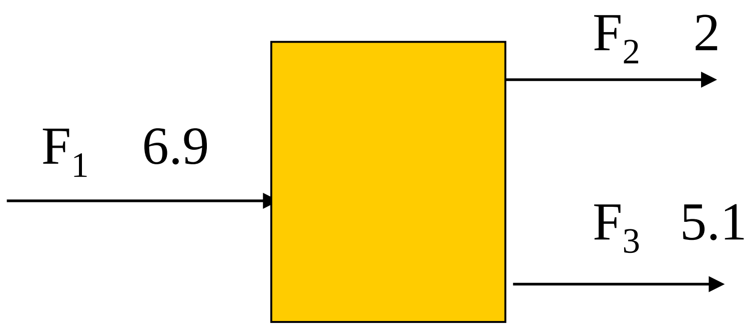
2 equations
6 variables

More than 4 measurements are required to avoid having a unique or multiple solutions

F flow
X composition

Data reconciliation

3 measurements, affected by noise, errors, etc.

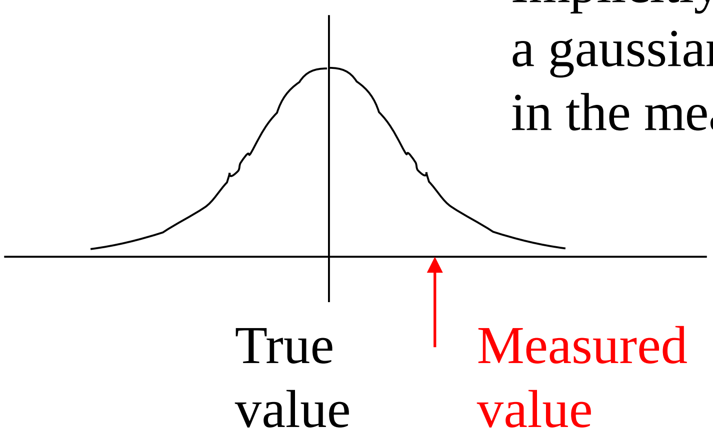


Redundant variables

$F_1 = F_2 + F_3$

Estimated values must satisfy the model

Implicitly, we assume a gaussian distribution in the measurements



Probalility of a measured value x_i around its true value (that verifies the model)

$$p_i(x_{mi}) = \frac{\exp\left[\frac{-(x_i - x_{mi})^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$

Data reconciliation

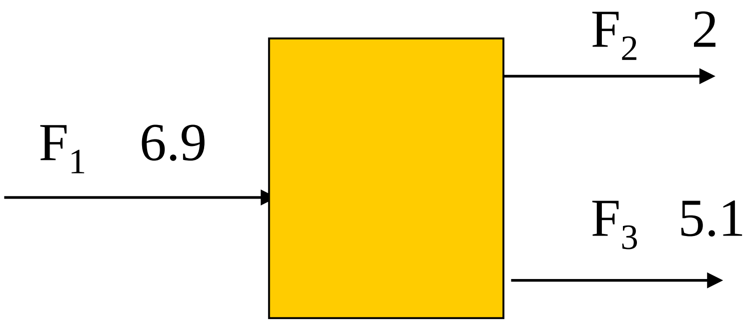
- Criterion (ML): Maximize the probability that the measured value of each variable x_m be equal to the true one, which verifies the model x (o minimize its negative log)

$$\prod_{i=1}^{N} p_i(x_{mi}) = \prod_{t=1}^{N} \frac{\exp\left[\frac{-(x_i - x_{mi})^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$

Assuming independent variables

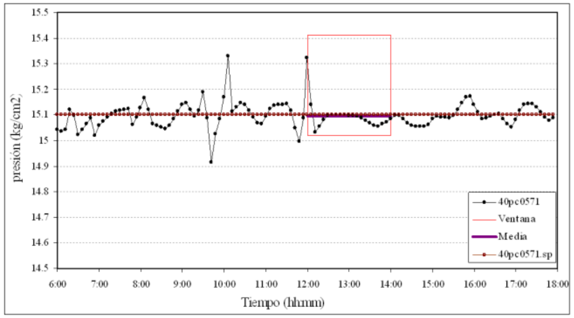
$$\min_{x_i,\theta}[-\log L(x_i)] = \min_{x_i,\theta} \sum_{i=1}^{N} \frac{(x_i - x_{mi}))^2}{2\sigma^2} + N \log \sigma\sqrt{2\pi}$$

Data reconciliation

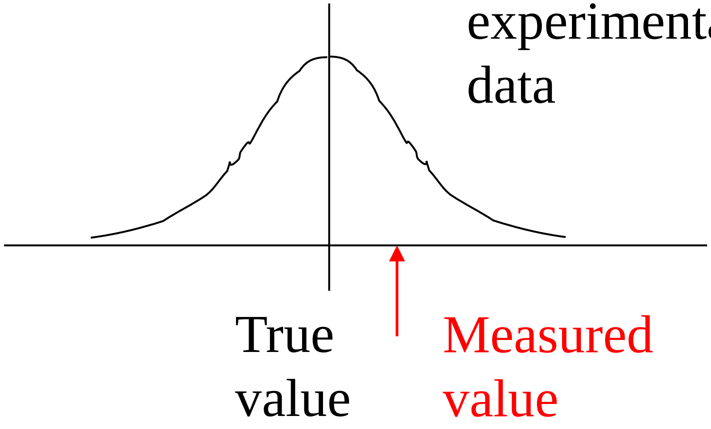


$$F_1 + F_2 + F_3 = 0$$

$$\boxed{\begin{array}{l} \min\limits_{x} \sum\limits_{j \in M} \left[\dfrac{F_j - F_{mj}}{\sigma_j} \right]^2 \\ f_i(F) = 0 \end{array}}$$



Compute variance from experimental data



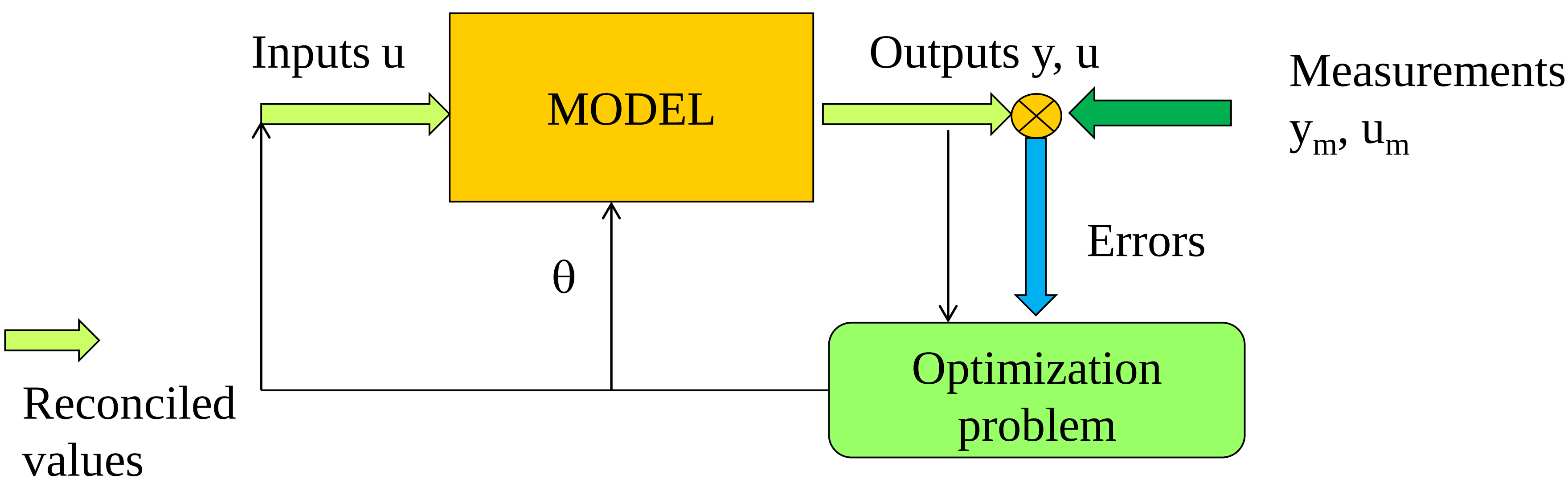
Data reconciliation

$$\min_{u,\theta} \sum_{i=1}^{N_{measured}} \alpha_i \left(y_i - y_{m,i}\right)^2 + \beta_i \left(u_i - u_{m,i}\right)^2$$

m measured values

$$\frac{dx}{dt} = f(x,u,\theta) \qquad y = h(x,u,\theta)$$

$$g(x,y,u,\theta) \le 0$$



Feasibility

$$\min_{u,\theta,\varepsilon} \sum_{i}^{meas} \frac{\alpha_i}{\sigma_i^2}\left(y_i - y_{m,i}\right)^2 + \sum_{j}^{meas} \frac{\beta_j}{\sigma_j^2}\left(u_j - u_{m,j}\right)^2 + \sum_{k}^{feas} \gamma_i \varepsilon_k^2$$

$$\frac{dx}{dt} = f(x,u,\theta) \qquad y = h(x,u,\theta)$$

$$g(x,y,u,\theta) \le \varepsilon \qquad \varepsilon \ge 0$$

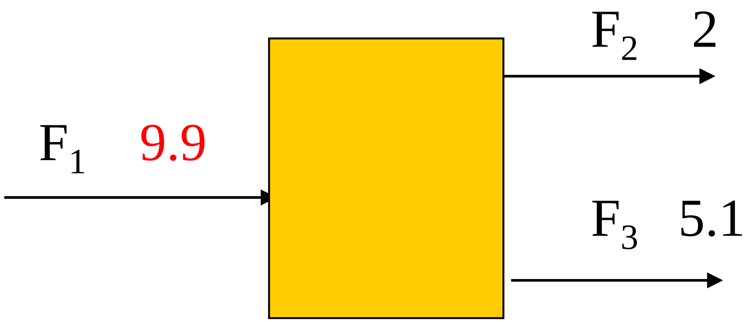
Normalization: span, variance, instrument precision,…

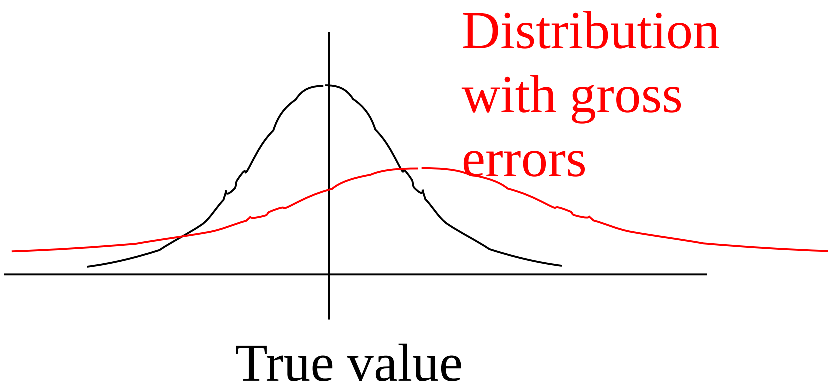
Feasibility: slack variables incorporated

α, β : relative importance of the variables and eliminate variables affected with gross errors

Identificability, regularization,…

Gross errors





$$F_1 + F_2 + F_3 = 0$$

$$\min_x \sum_{j \in M} \left[\frac{x_j - x_{mj}}{\sigma_j} \right]^2$$

$$f_i(x) = 0$$

Gross errors increase the dispersion and distort the solution

The errors are spread through all variables

Detecting gross errors

Two approaches:
- Gross errors detection and measurement removal
- Use of robust estimators

Analyse residuals with data without gross errors

Analyse residual of current data PCA

Test for significant differences and, in particular, for the largest ones and locate the variables that most contribute to them

Gross errors

$$\min_{u,\theta,\varepsilon} \sum_{i}^{meas} \frac{\alpha_i}{\sigma_i^2}\left(y_i - y_{m,i}\right)^2 + \sum_{j}^{meas} \frac{\beta_j}{\sigma_j^2}\left(u_j - u_{m,j}\right)^2 + \sum_{k}^{feas} \gamma_i \varepsilon_k^2$$

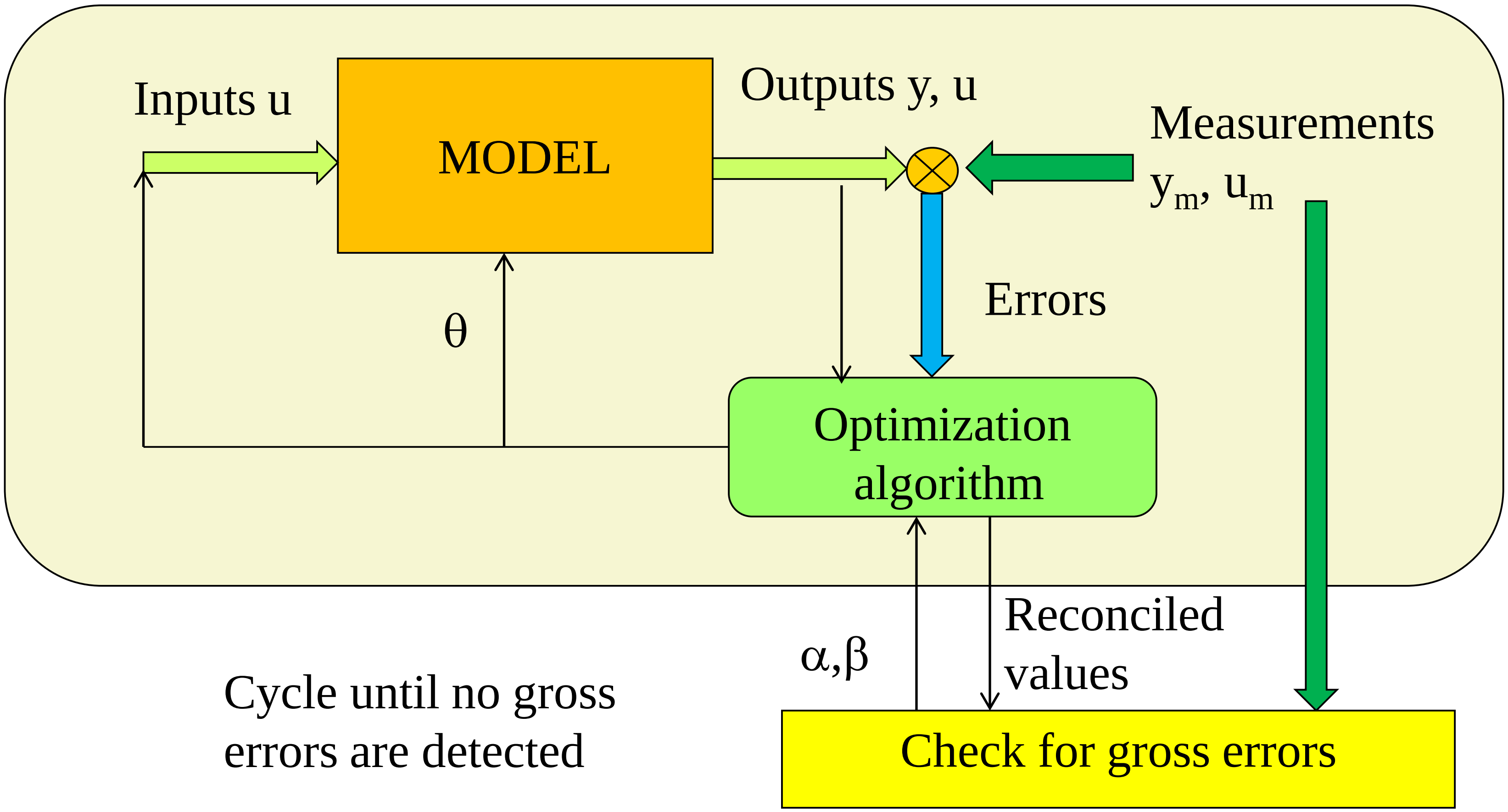
$$\frac{dx}{dt} = f(x,u,\theta) \qquad y = h(x,u,\theta)$$

$$g(x,y,u,\theta) \le \varepsilon \qquad \varepsilon \ge 0$$

In practice, gross errors can be detected by a combination of rule base and cyclic solution of the optimization problem. After an initial removal of a set of measurements from the cost function using rules, the solution is checked against the variance of the signal and those variables with measurements outside the 3σ band, are removed again.

Data Reconciliation- Gross errors

Two approaches:
- Gross errors detection and measurement removal
- Use of robust estimators



Cycle until no gross errors are detected

Robust Estimators

$$\min_{x} \sum_{j\in M}\left[\frac{x_j - x_{mj}}{\sigma_j}\right]^2 = \sum_{j\in M}\varepsilon_j^2$$
$$f_i(x) = 0$$

If the distribution of the measurement errors ε_j is non-Gaussian, as may happen if gross errors are present, the LS estimation may give incorrect results as it is not robust against deviations from the assumed Gaussian distribution.

The robustness of a ML-estimator against deviations from non-Gaussianity is measured by the influence function, which is proportional to the first derivative of the estimator. The estimator is robust if the influence function is bounded as the residuals go to infinity.

In particular, the LS estimator is not robust as the derivative

$$\frac{d\varepsilon_j^2}{d\varepsilon_j} = 2\varepsilon_j$$

is not bounded

Robust estimators

$$F_j = c^2\left[\frac{|\varepsilon_j|}{c} - \log\left(1+\frac{|\varepsilon_j|}{c}\right)\right]$$

Robust estimators use different cost functions, such as the Fair function F, that fulfils the robustness property:

$$\frac{dF_j}{d\varepsilon_j} = \frac{\varepsilon_j}{1+\frac{|\varepsilon_j|}{c}}$$

$$\min_x \sum_{j\in M} c^2\left[\frac{|\varepsilon_j|}{c} - \log\left(1+\frac{|\varepsilon_j|}{c}\right)\right]$$

$$\varepsilon_j = \frac{x_j - x_{mj}}{\sigma}$$

$$f_i(x) = 0$$

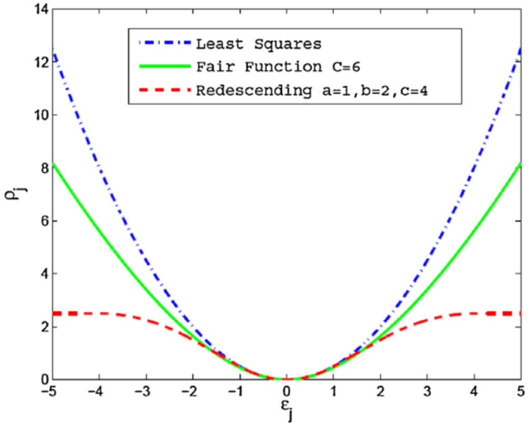
Robust data reconciliation formulation

Redescending Función

$$R_j = \begin{cases} 0.5\varepsilon_j^2 & 0 \le |\varepsilon_j| \le a \\ a|\varepsilon_j| - 0.5a^2 & a \le |\varepsilon_j| \le b \\ ab - 0.5a^2 + 0.5a(c-b)\left(1 - \left(\dfrac{c - |\varepsilon_j|}{c-b}\right)^2\right) & b \le |\varepsilon_j| \le c \\ ab - 0.5a^2 + 0.5a(c-b) & c \le |\varepsilon_j| \end{cases}$$

$c > b + 2a$

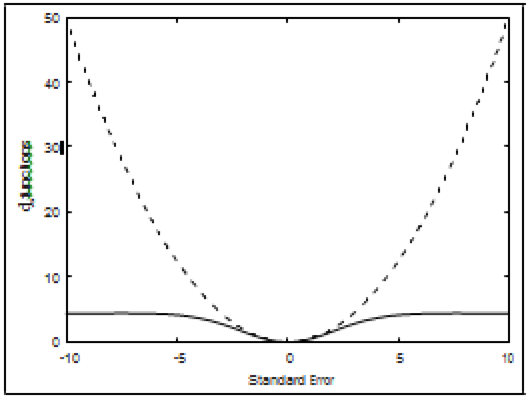
Hampel's redescending estimator

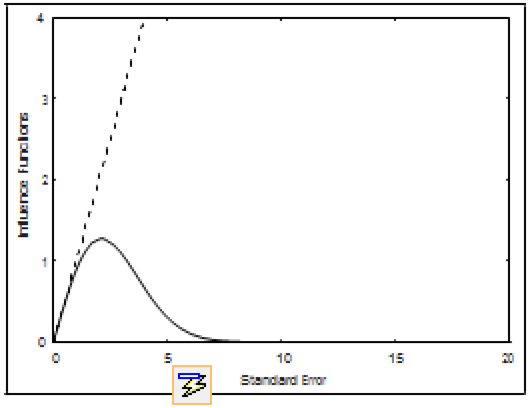


Welsch

$$W_j = \frac{c^2}{2}\left[1 - \exp\left(-\left(\frac{\varepsilon_j}{c}\right)^2\right)\right]$$

95% asymptotic efficiency on the standard normal distribution is obtained with the tuning constant c = 2.9846





Data reconciliation

Data reconciliation

Static

Dynamic

Batch

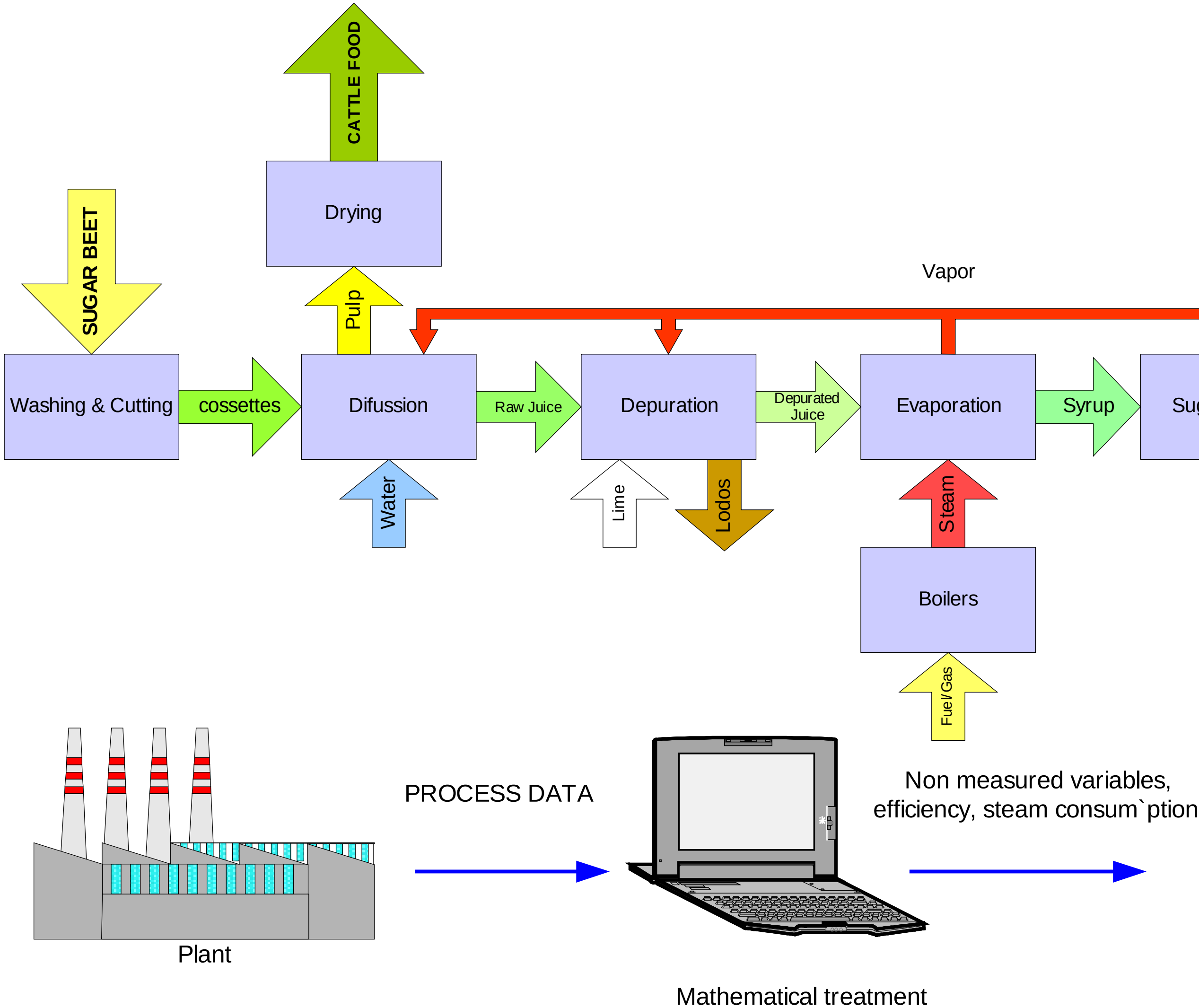
Steady state detector

Data averaged over a period of time

Dynamic optimization problem

Open field

Beet sugar factory



Information
KPIs

Sugar plant DR software

- Main elements:
- Periodic characterization of the plant status, using a steady model of the sugar plant.
- On line connection with the plan Distributed Control System (DCS) to obtain, the measured variables necessary for the balances and model identification.
- Data reconciliation, correcting measured variables in a way that the model is adjusted and calculating at the same time that unknown variables and model parameters.
- As a by-product of the data reconciliation, key performance indicators are estimated from calculated values in the reconciliation.

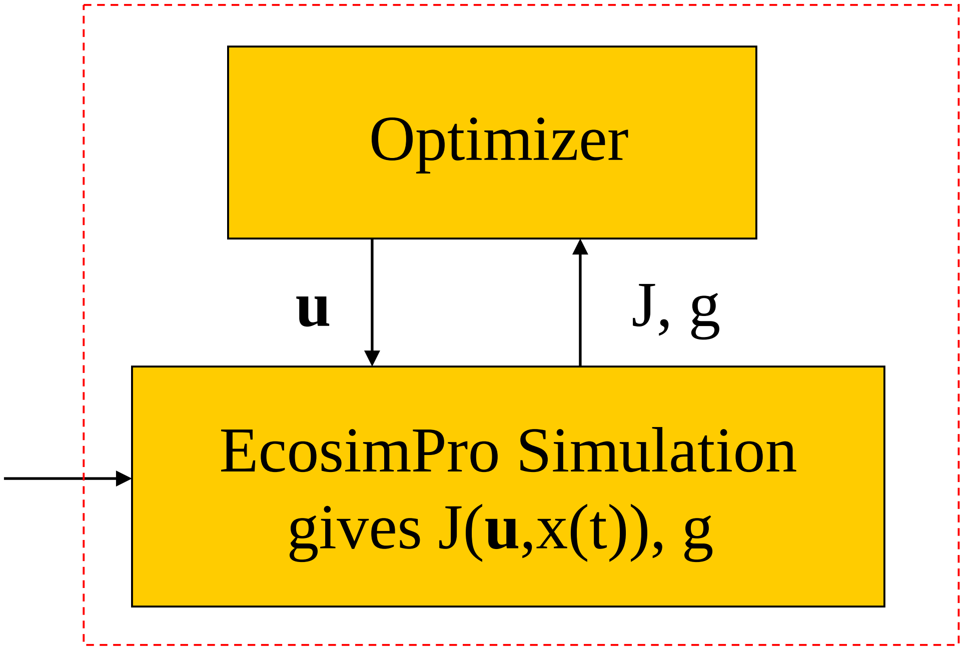
Models

- ✓ Static
- ✓ Mass energy balances
- ✓ Flows, pressures
- ✓ Equations and properties of the application domain
- ✓ Formulated in the EcosimPro environment
- ✓ Measurements averaged for a period of time
- ✓ Rules to eliminate bad measurements

Data reconciliation

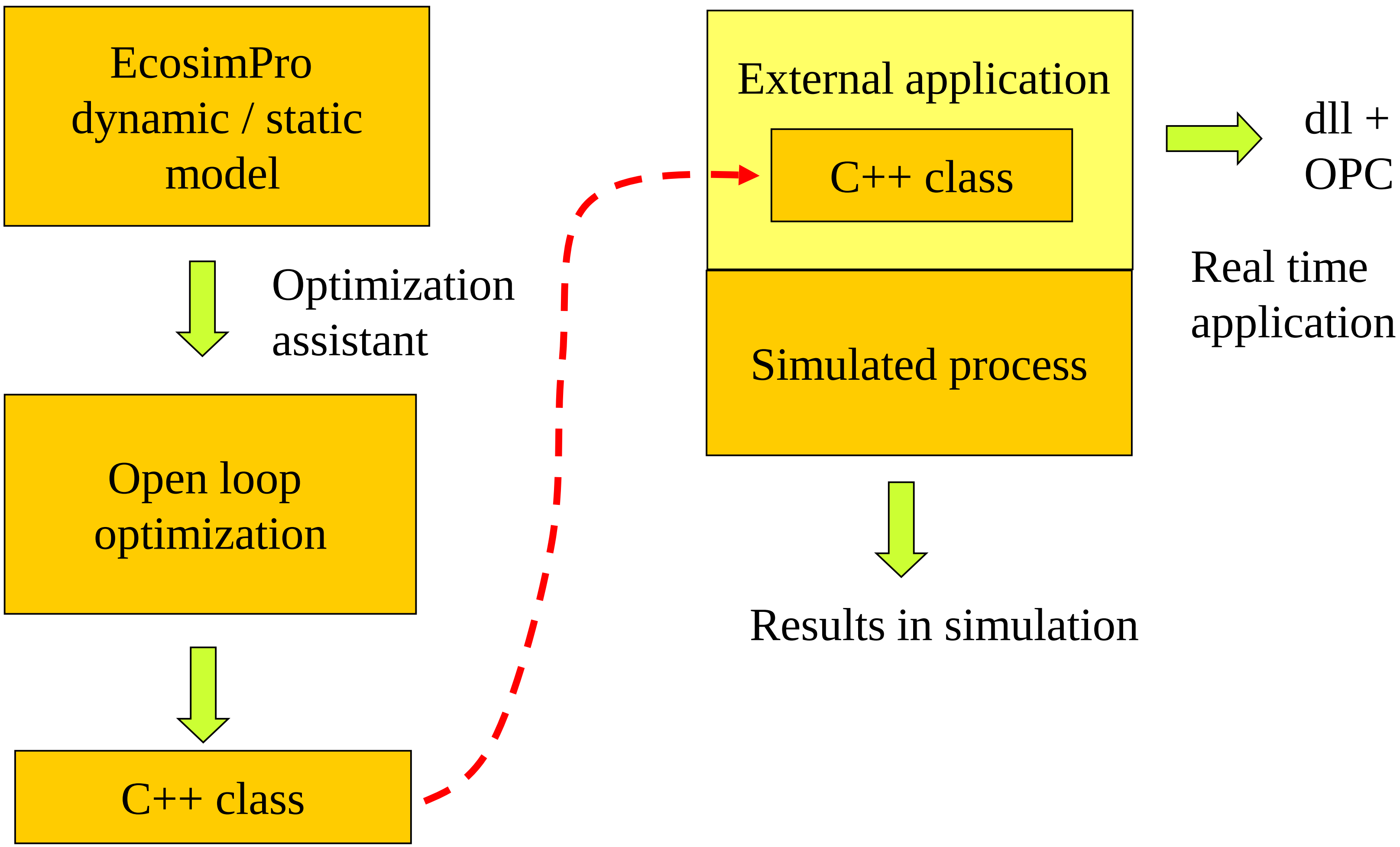
$$\min_{u,\theta,\varepsilon} \sum_{i}^{meas} \frac{\alpha_i}{\sigma_i^2}\left(y_i - y_{m,i}\right)^2 + \sum_{j}^{meas} \frac{\beta_j}{\sigma_j^2}\left(u_j - u_{m,j}\right)^2 + \sum_{k}^{feas} \gamma_i \varepsilon_k^2$$

$$y = h(x,u,\theta) \qquad g(x,y,u,\theta) \le \varepsilon \qquad \varepsilon \ge 0$$

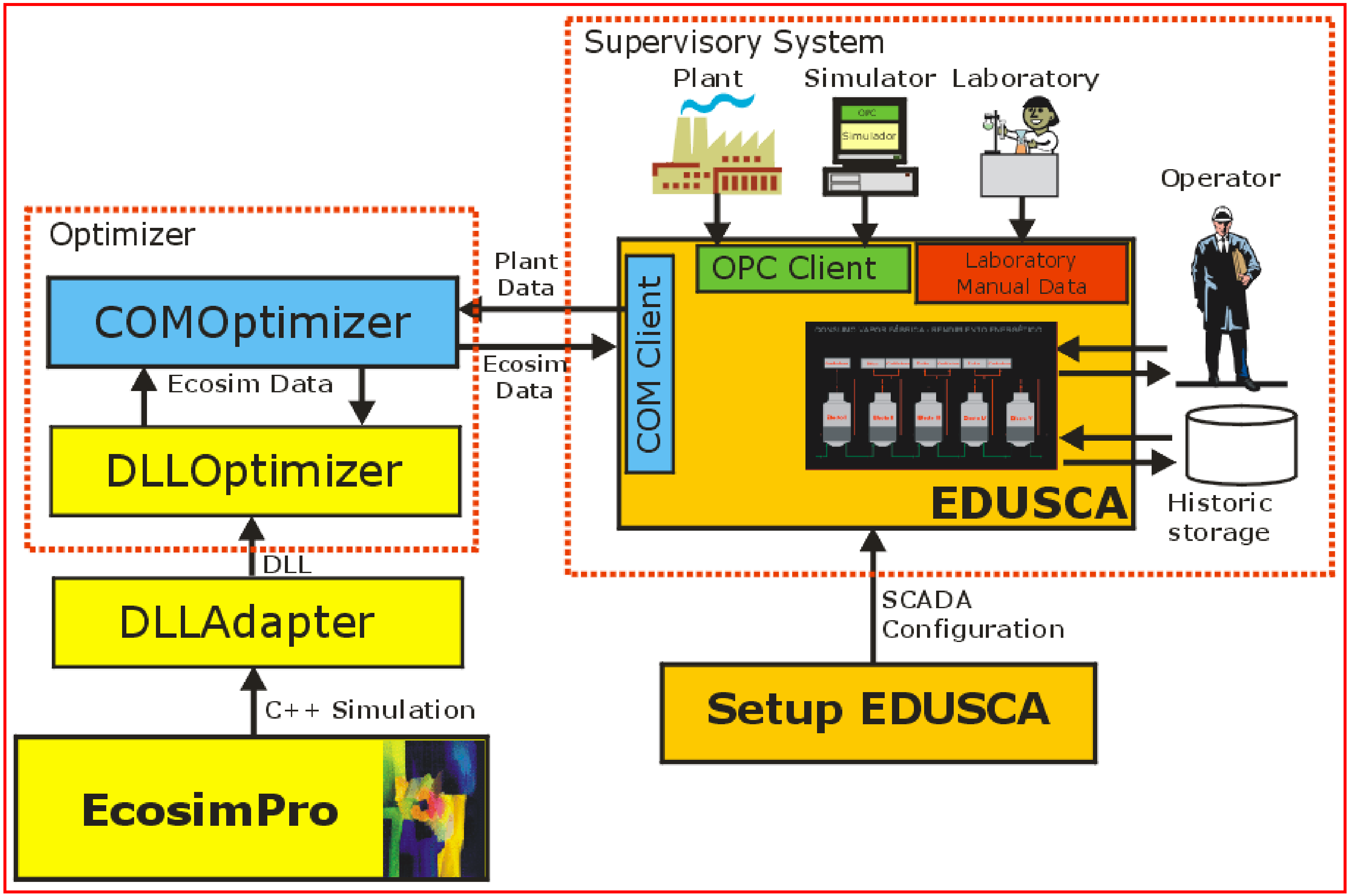


Solved with a sequential approach

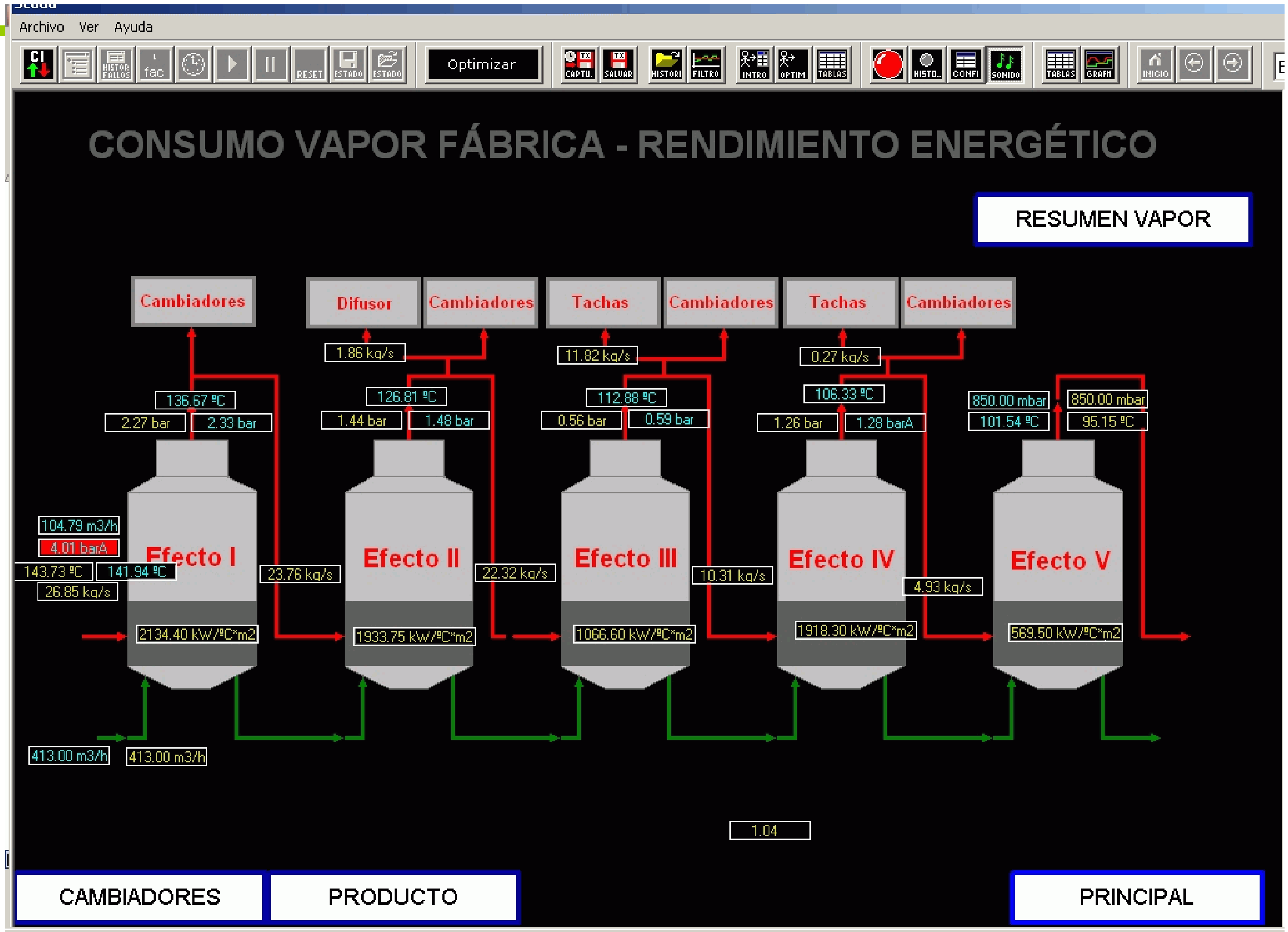
Implementation in EcosimPro



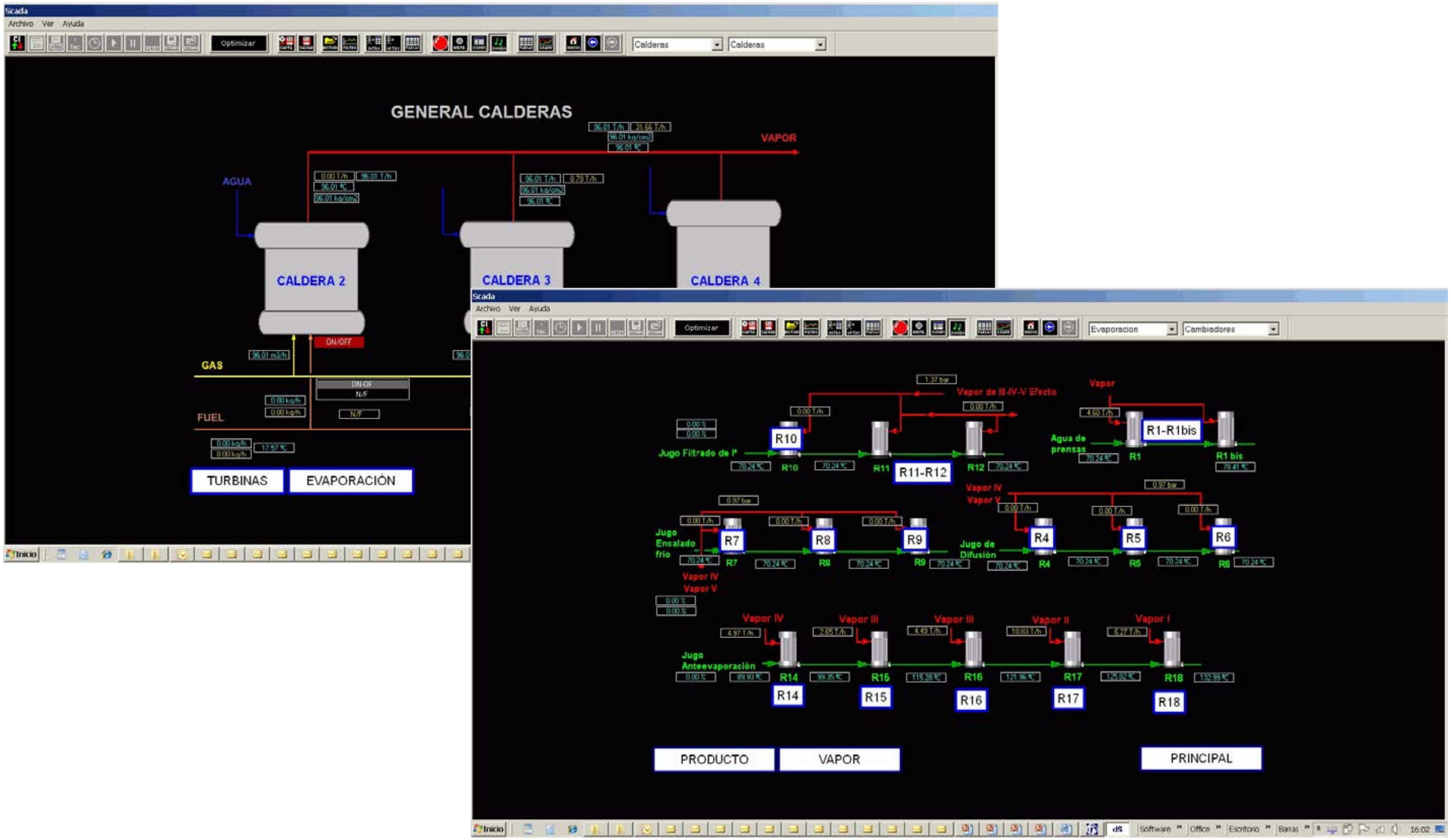
SCADA implementation



DR system / SCADA



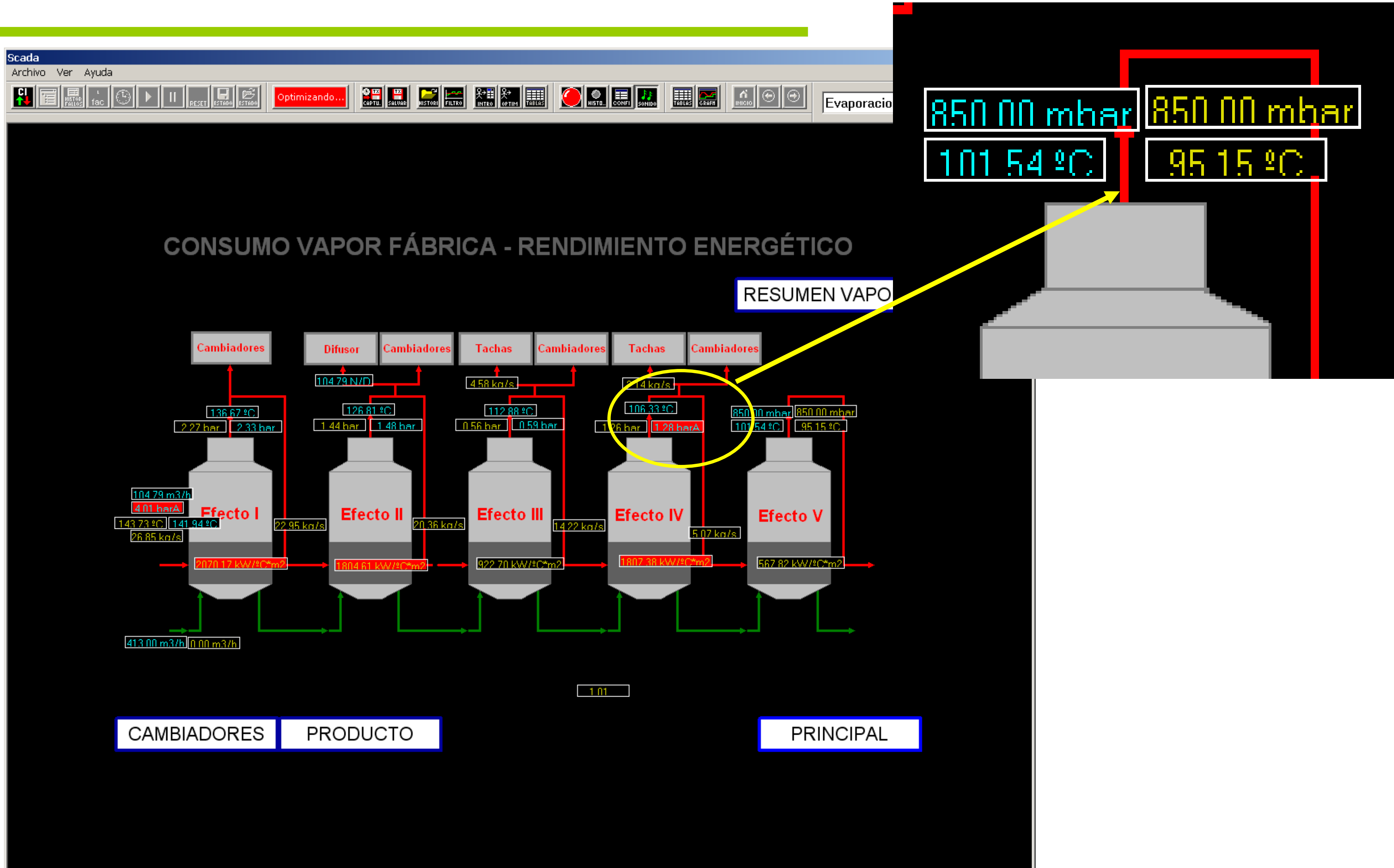
DR system SCADA



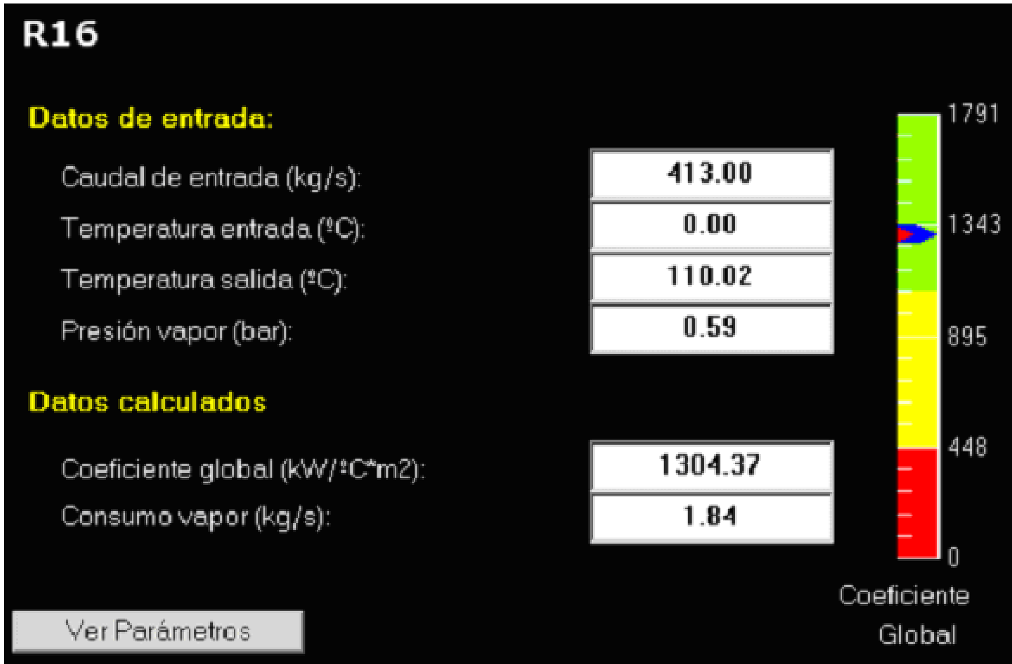
Information

- ✓ Detection of inconsistent measures. Help in fault detection.
- ✓ KPI: Evaluation of energetic behaviour indexes, efficiency, comparison between process heat transfer coefficients versus theoretical coefficients….
- ✓ Estimation of all unmeasured variables, some of them relevant for the energy evaluation such as steam consumptions.
- ✓ Keeping track of the time evolution of key variables during the sugar beet campaign, helping managers in locating malfunctions in the process or equipment fouling and planning maintenance

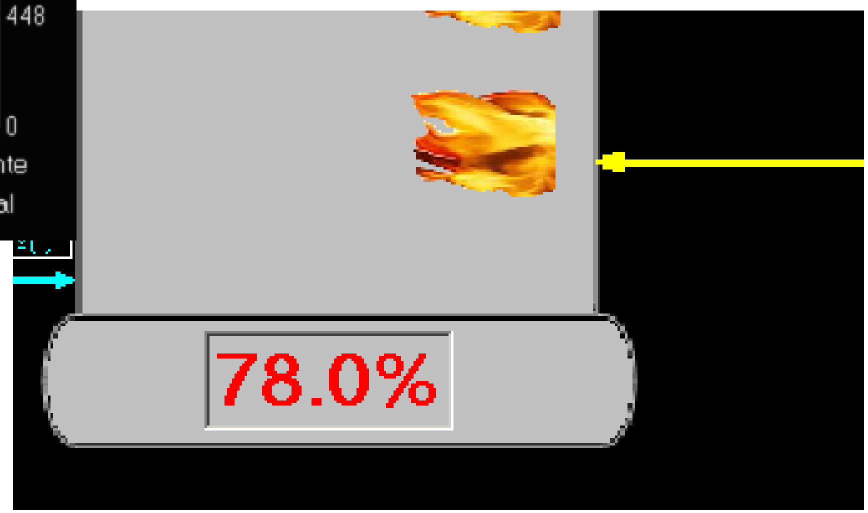
Inconsistencies



Key Performance Indicators KPI

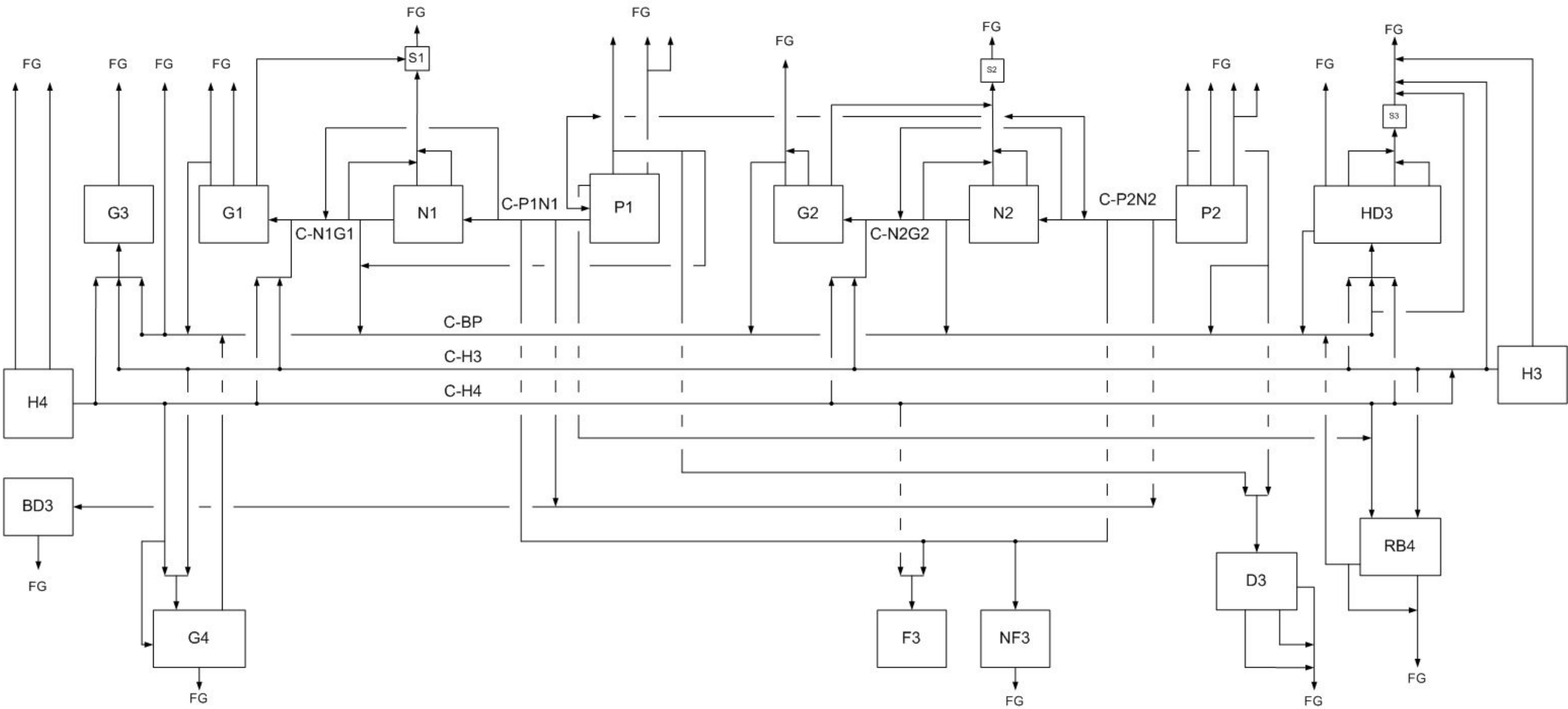


Heat exchanger coefficients

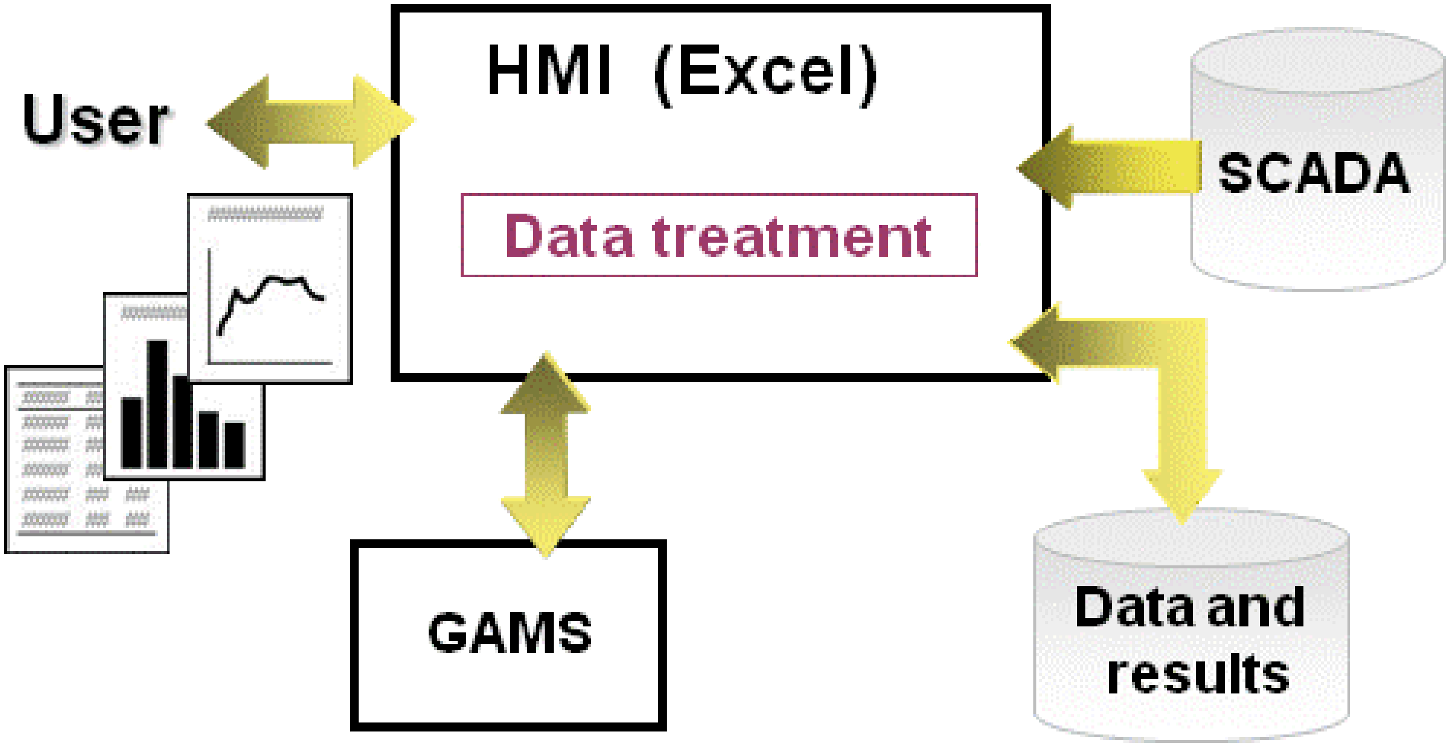


Boiler efficiency

Hydrogen network



Arquitecture

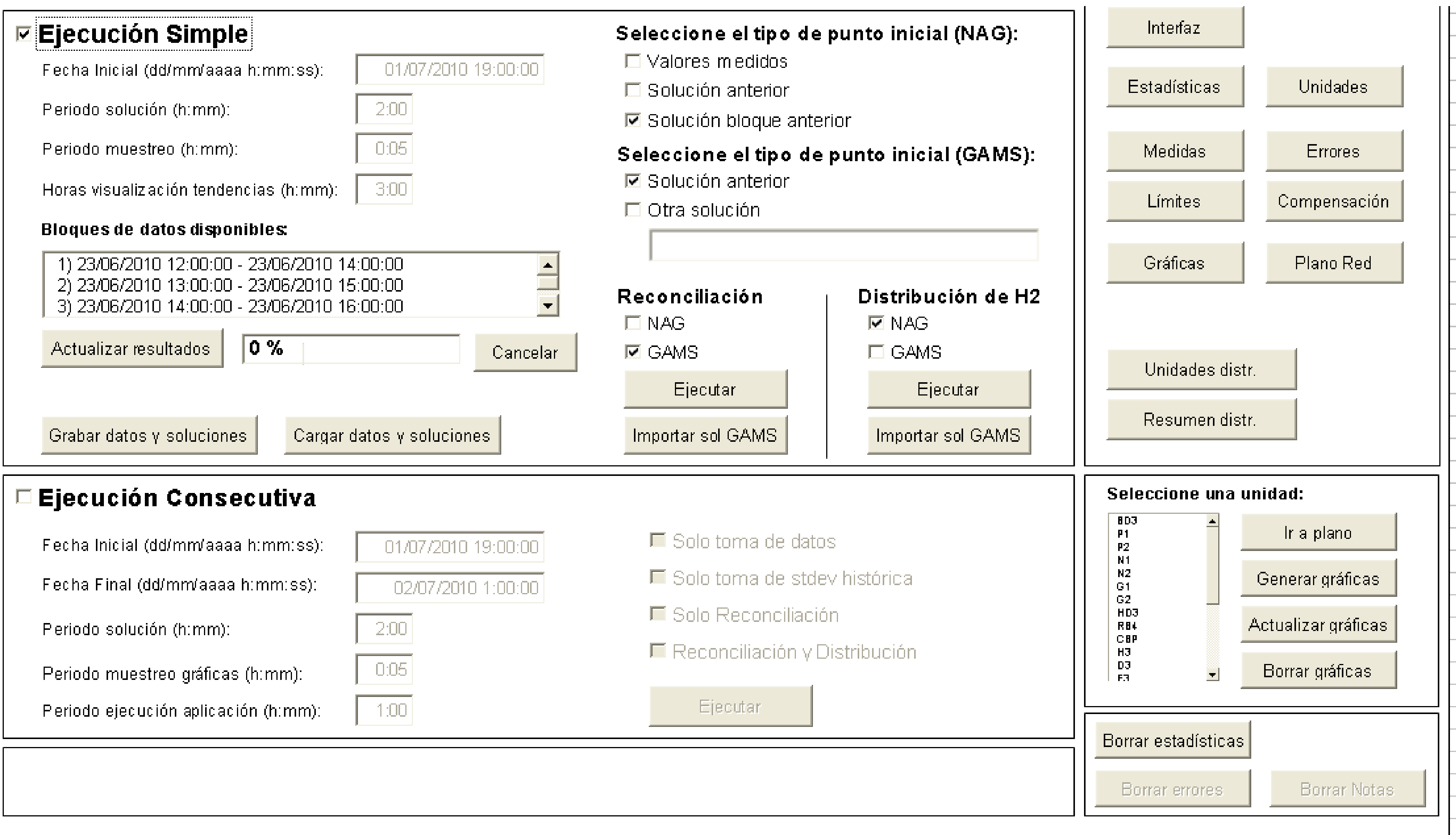


Data treatment

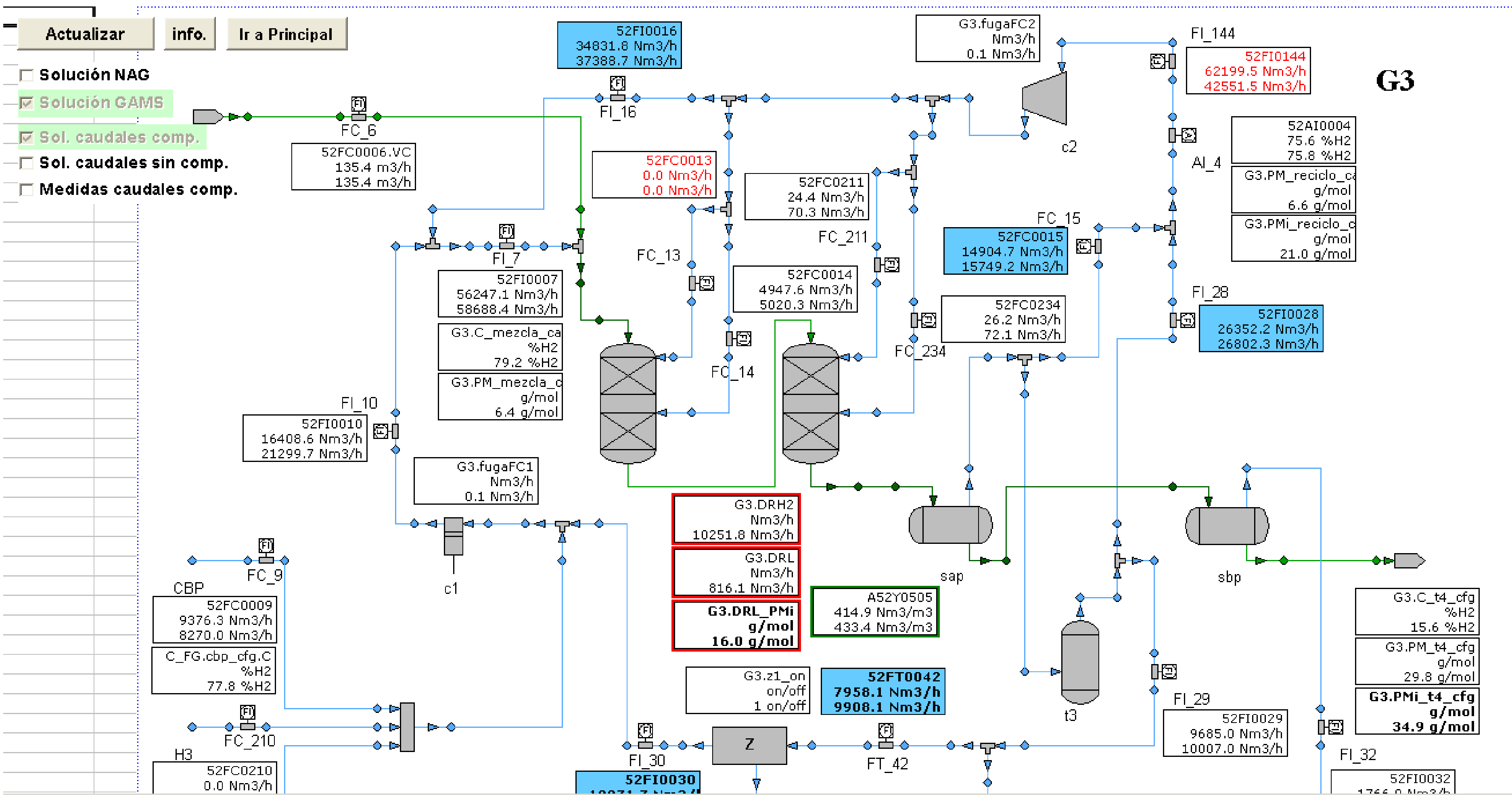
key role played by the data treatment in the success of the application in the refinery. If data from the SCADA system are not analyzed and filter previously to their use in the numerical methods, there are no chances to obtain good results. This layer is composed of a set of rules that detect faults and information inconsistences in the raw data and decides which options are the most adequate ones. For instance, detecting when a flow is actually zero, a plant is stopped, a measurement is out of range, etc. It has been developed for specific cases combining physical knowledge and heuristic rules.

As a result of these rules, the system adjust the model parameters and optimization weights, so that, e.g. a measurement can be eliminated from the data reconciliation cost function. Mayor changes take place when a plant is not operating. To deal with these cases, the network is formulated as a superstructure that allows to remove groups of equations depending on the value of binary variables that represent the state of the plants.

DR

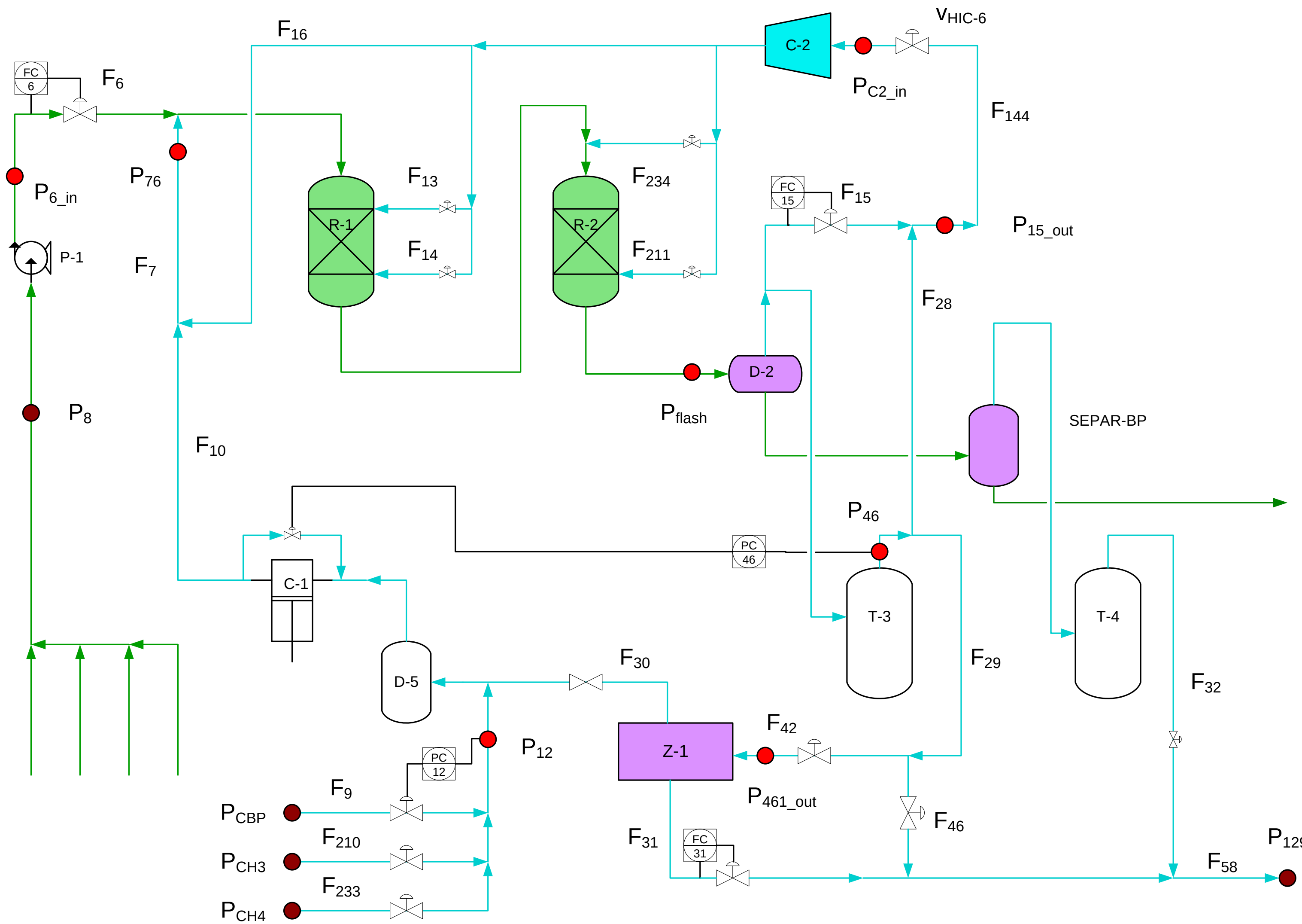

Ejecución Simple
Fecha Inicial (dd/mm/aaaa h:mm:ss): 01/07/2010 19:00:00
Periodo solución (h:mm): 2:00
Periodo muestreo (h:mm): 0:05
Horas visualización tendencias (h:mm): 3:00
Bloques de datos disponibles:
1) 23/06/2010 12:00:00 - 23/06/2010 14:00:00
2) 23/06/2010 13:00:00 - 23/06/2010 15:00:00
3) 23/06/2010 14:00:00 - 23/06/2010 16:00:00
Actualizar resultados
0 %
Cancelar
Grabar datos y soluciones
Cargar datos y soluciones
Seleccione el tipo de punto inicial (NAG):
Valores medidos
Solución anterior
Solución bloque anterior
Seleccione el tipo de punto inicial (GAMS):
Solución anterior
Otra solución
Reconciliación
NAG
GAMS
Ejecutar
Importar sol GAMS
Distribución de H2
NAG
GAMS
Ejecutar
Importar sol GAMS
Interfaz
Estadísticas
Unidades
Medidas
Errores
Límites
Compensación
Gráficas
Plano Red
Unidades distr.
Resumen distr.
Ejecución Consecutiva
Fecha Inicial (dd/mm/aaaa h:mm:ss): 01/07/2010 19:00:00
Fecha Final (dd/mm/aaaa h:mm:ss): 02/07/2010 1:00:00
Periodo solución (h:mm): 2:00
Periodo muestreo gráficas (h:mm): 0:05
Periodo ejecución aplicación (h:mm): 1:00
Solo toma de datos
Solo toma de stdev histórica
Solo Reconciliación
Reconciliación y Distribución
Ejecutar
Seleccione una unidad:
BD3
P1
P2
N1
N2
G1
G2
HD3
RB4
CBP
H3
D3
Ir a plano
Generar gráficas
Actualizar gráficas
Borrar gráficas
Borrar estadísticas
Borrar errores
Borrar Notas

DR



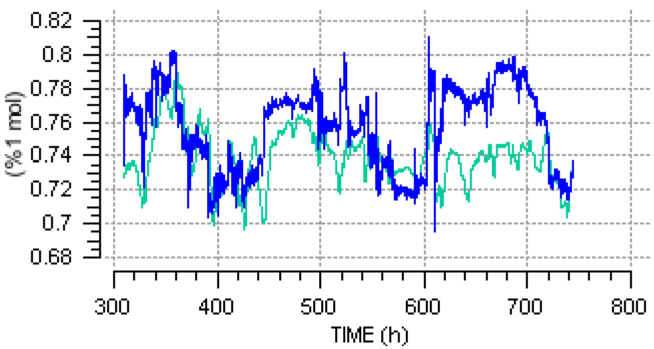
Actualizar
info.
Ir a Principal
Solución NAG
Solución GAMS
Sol. caudales comp.
Sol. caudales sin comp.
Medidas caudales comp.
G3
52FI0016
34831.8 Nm3/h
37388.7 Nm3/h
FI_16
G3.fugaFC2
Nm3/h
0.1 Nm3/h
FI_144
52FI0144
62199.5 Nm3/h
42551.5 Nm3/h
c2
FC_6
52FC0006.VC
135.4 m3/h
135.4 m3/h
52FC0013
0.0 Nm3/h
0.0 Nm3/h
52FC0211
24.4 Nm3/h
70.3 Nm3/h
52AI0004
75.6 %H2
75.8 %H2
AI_4
G3.PM_reciclo_ca
g/mol
6.6 g/mol
G3.PMi_reciclo_c
g/mol
21.0 g/mol
FC_15
52FC0015
14904.7 Nm3/h
15749.2 Nm3/h
FC_211
FI_7
FC_13
52FI0007
56247.1 Nm3/h
58688.4 Nm3/h
52FC0014
4947.6 Nm3/h
5020.3 Nm3/h
52FC0234
26.2 Nm3/h
72.1 Nm3/h
FI_28
52FI0028
26352.2 Nm3/h
26802.3 Nm3/h
G3.C_mezcla_ca
%H2
79.2 %H2
G3.PM_mezcla_c
g/mol
6.4 g/mol
FC_14
FC_234
FI_10
52FI0010
16408.6 Nm3/h
21299.7 Nm3/h
G3.fugaFC1
Nm3/h
0.1 Nm3/h
G3.DRH2
Nm3/h
10251.8 Nm3/h
G3.DRL
Nm3/h
816.1 Nm3/h
G3.DRL_PMi
g/mol
16.0 g/mol
A52Y0505
414.9 Nm3/m3
433.4 Nm3/m3
sap
sbp
c1
FC_9
CBP
52FC0009
9376.3 Nm3/h
8270.0 Nm3/h
C_FG.cbp_cfg.C
%H2
77.8 %H2
FC_210
H3
52FC0210
0.0 Nm3/h
G3.z1_on
on/off
1 on/off
52FT0042
7958.1 Nm3/h
9908.1 Nm3/h
t3
G3.C_t4_cfg
%H2
15.6 %H2
G3.PM_t4_cfg
g/mol
29.8 g/mol
G3.PMi_t4_cfg
g/mol
34.9 g/mol
FI_29
52FI0029
9685.0 Nm3/h
10007.0 Nm3/h
Z
FI_30
52FI0030
FT_42
FI_32
52FI0032

HDS

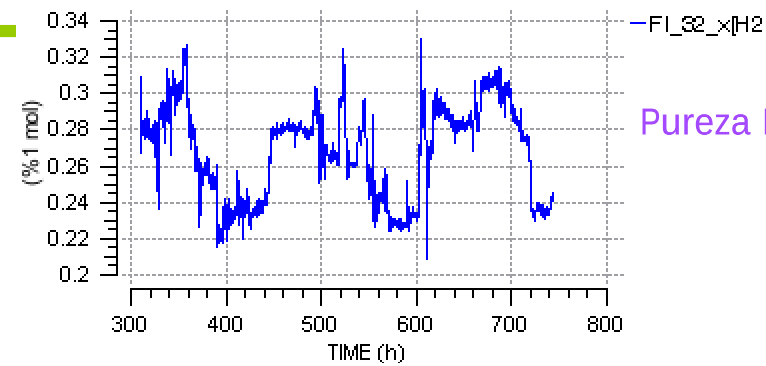


F_{16}
$V_{HIC\text{-}6}$
C-2
FC 6
F_6
P_{C2_in}
F_{144}
P_{76}
F_{13}
F_{234}
FC 15
F_{15}
P_{6_in}
R-1
R-2
P_{15_out}
P-1
F_7
F_{14}
F_{211}
F_{28}
P_8
D-2
P_{flash}
SEPAR-BP
F_{10}
P_{46}
C-1
PC 46
T-3
T-4
F_{30}
F_{29}
F_{32}
D-5
P_{12}
PC 12
Z-1
F_{42}
F_9
P_{461_out}
P_{CBP}
F_{46}
F_{31}
FC 31
P_{129}
F_{210}
F_{58}
P_{CH3}
F_{233}
P_{CH4}

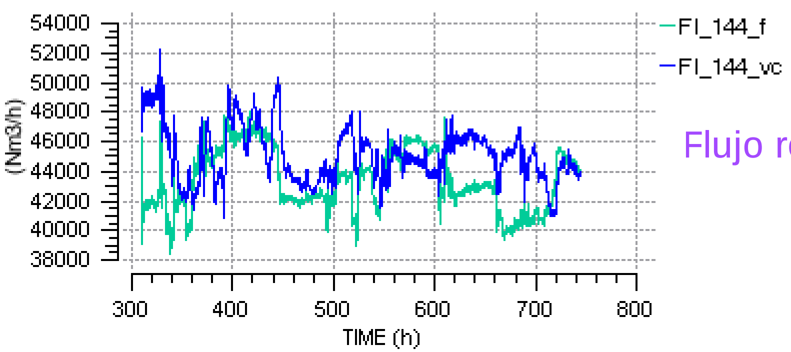
DR



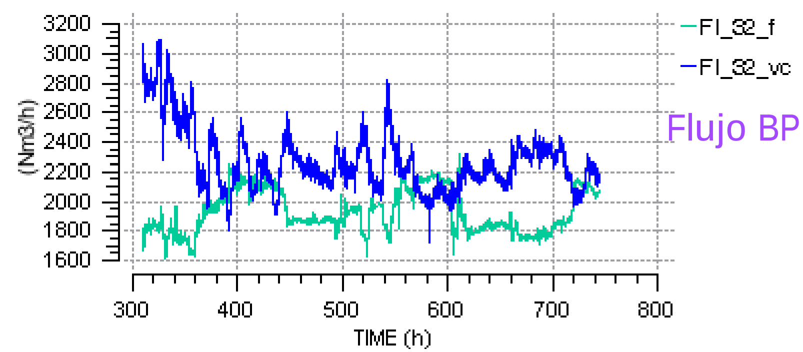
Pureza reciclo



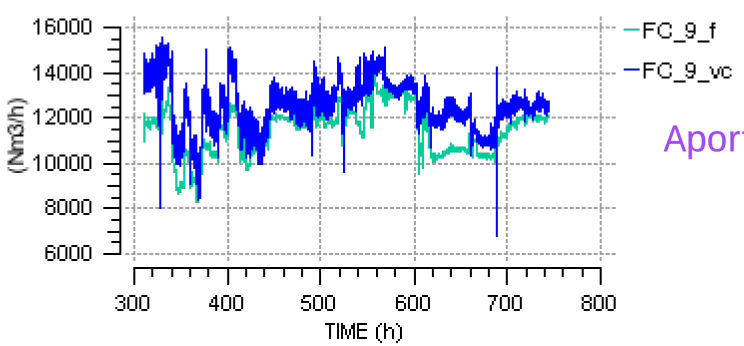
Pureza BP



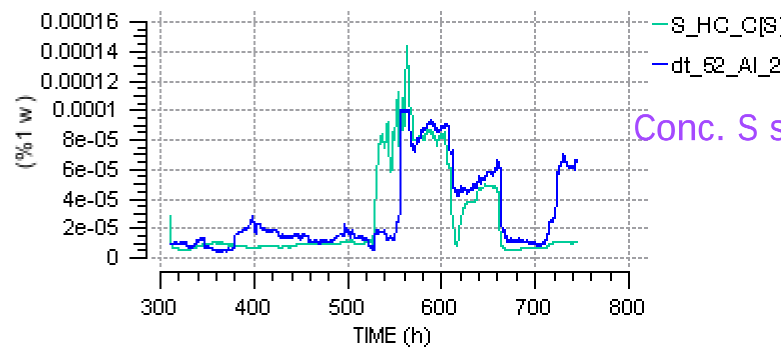
Flujo reciclo



Flujo BP



Aporte CBP



Conc. S salida

18 days

Conclusions

- ✓ Data reconciliation is a model based approach to obtain coherent information from the plant.
- ✓ It allows to compute KPI to follow the time evolution of the process operation.
- ✓ Formulated as an optimization problem.
- ✓ Open problems:
  - – Gross error detection
  - – Speed, batch, non-independent variables,…