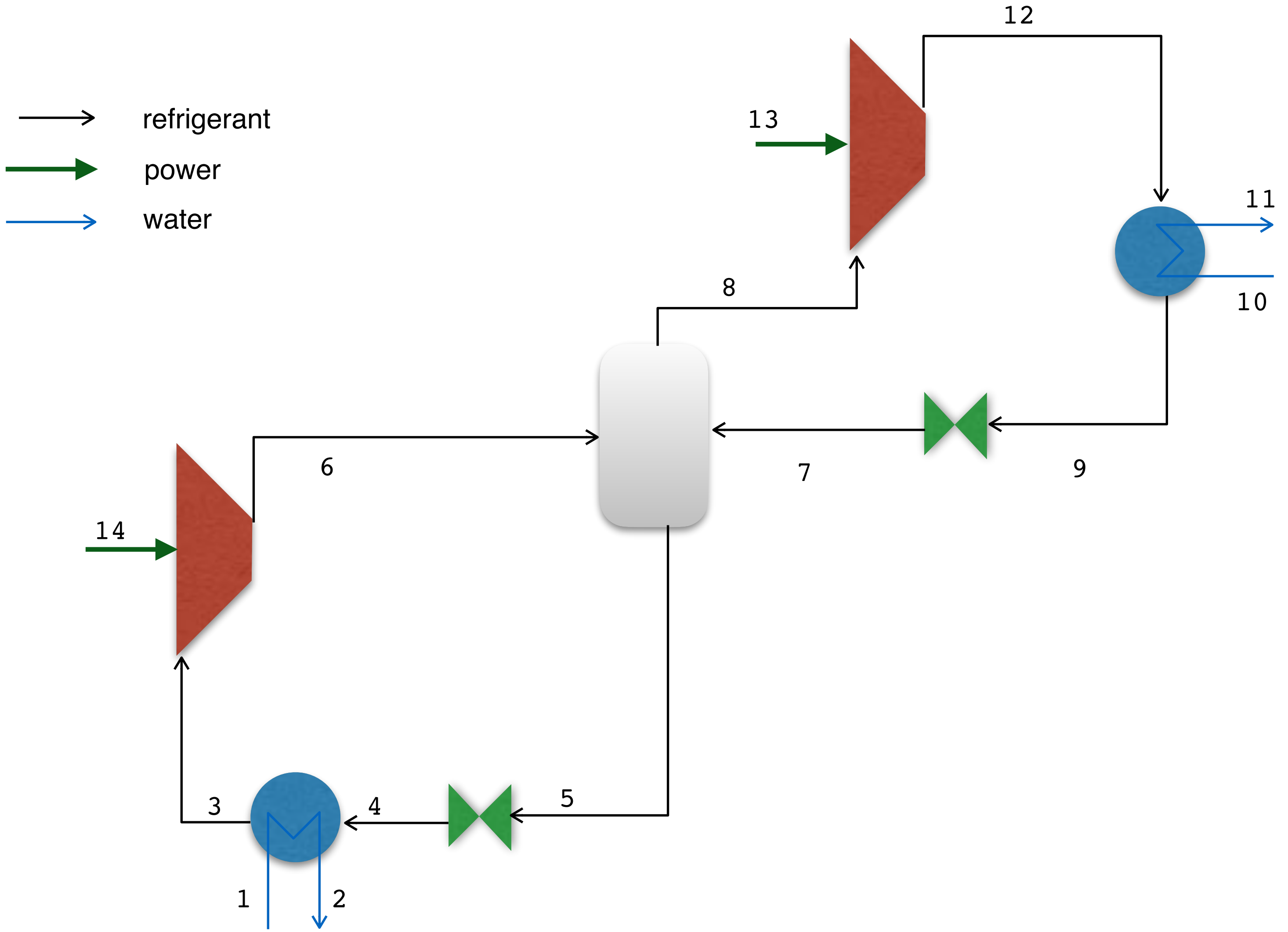
Measurement/specification analysis,

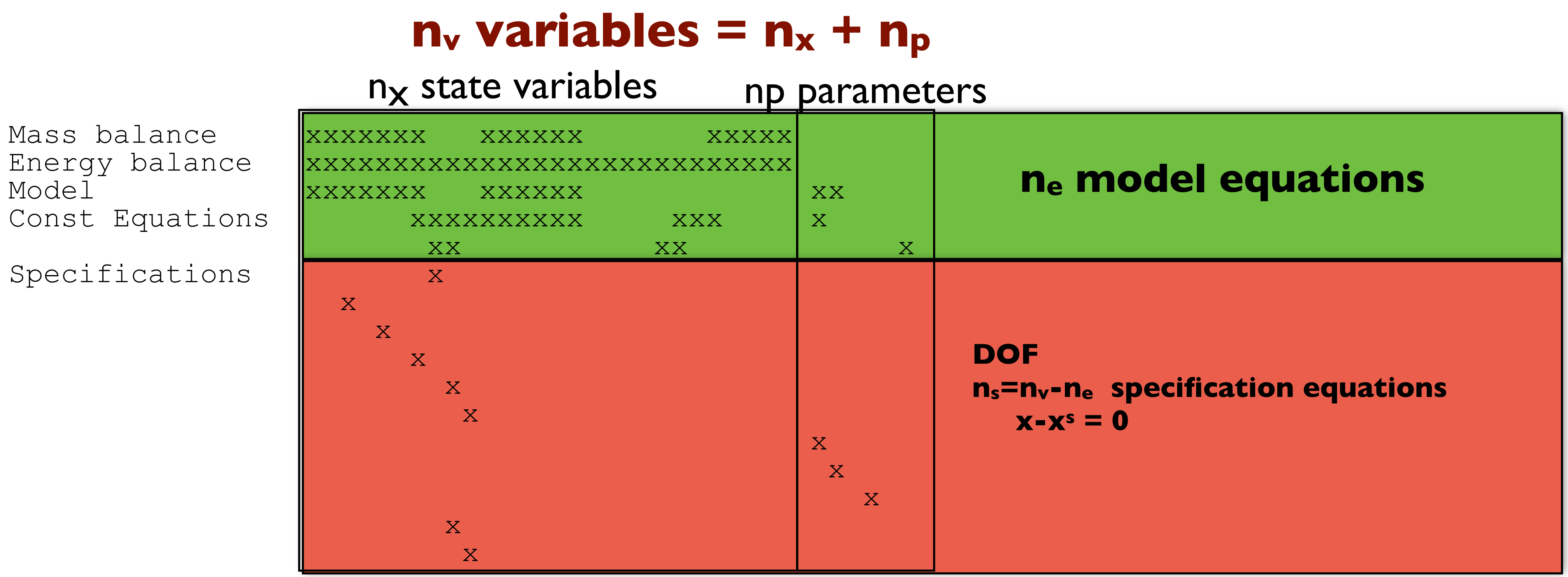
Data reconciliation and Parameter identification

François Marechal
IPESE
EPFL-STI-IGM

Two stages heat pump simulation



Incidence Matrix of a Unit model



To solve the problem :

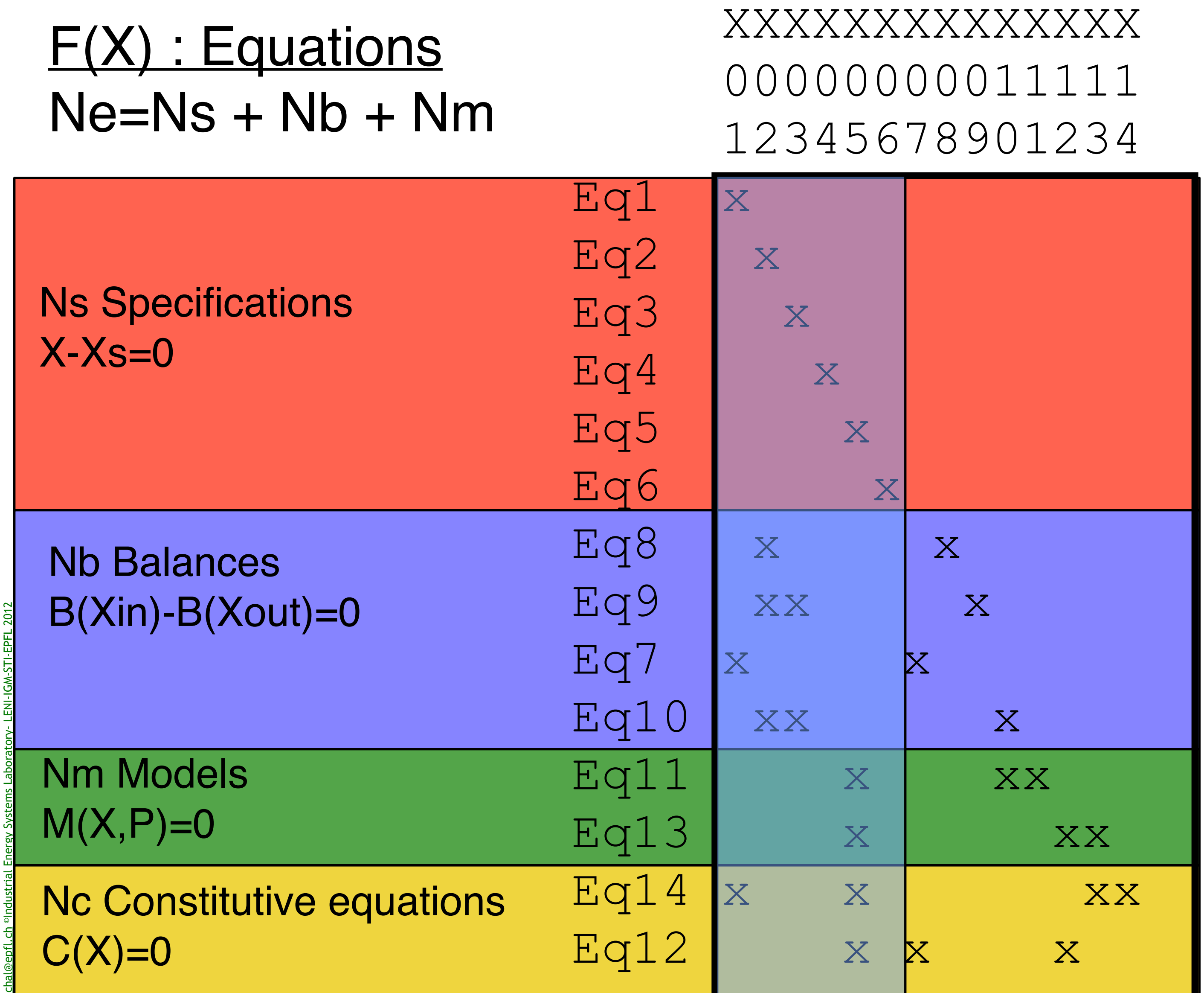
1) square matrix
2) independent equations

In the incidence matrix, the element (i,j) is equal to 1 if variable i is in equation j
It indicates the presence (incidence) of a variable (i) in the equation (j)

Unit model : Incidence matrix rearranged

F(X) : Equations

Ne=Ns + Nb + Nm



X : Variables

Nv state

Ni intermediate

Np parameters

Nx=Nv+Ni+Np

DOF analysis

Ne=Nx

Specified variables : e.g. parameters + context+decisions

Process models & decision support

π^s_{unit} →

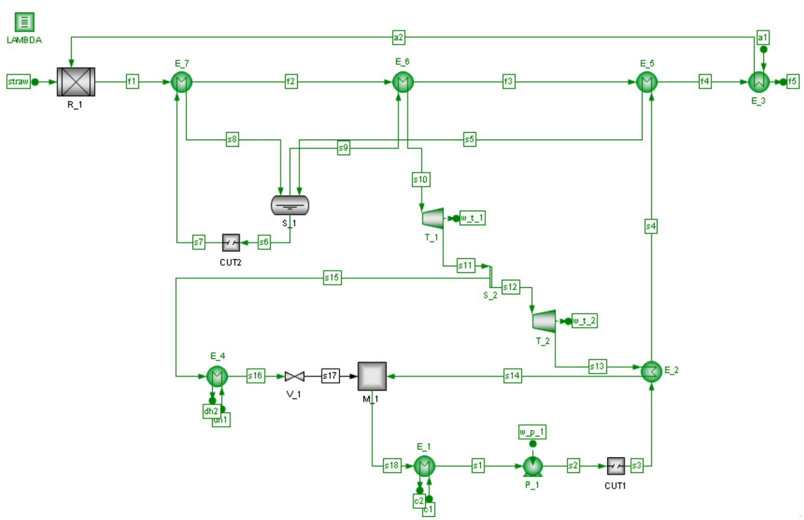
X^s →

Simulation

Model	$F(X, \pi_{unit}) = 0$
Set points	$X - X^s = 0$
Specifications	$\pi_{unit} - \pi^s_{unit} = 0$

→ **Performances**

Optimization values of decisions



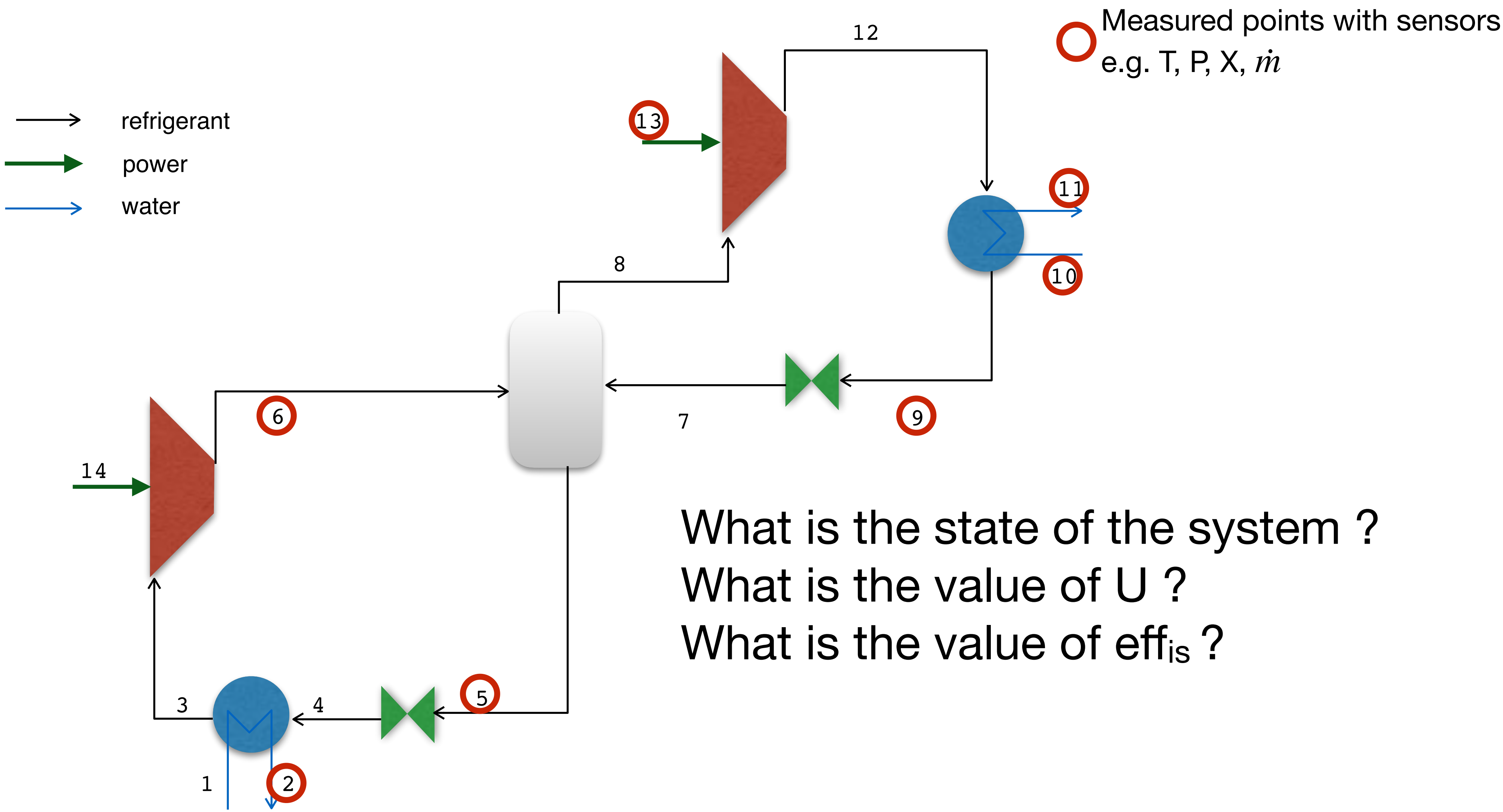
Model is defining the level of detail

What are the X we want to know ?

- Streams ?
- Unit parameters ? π_{unit}

- The process model and the unit models define the expected level of detail
  - i.e. the data we want to generate with the model
- Unit models require Parameters with fixed values
  - What are the values of the parameters ?
    - Literature => correlations, experience
    - From experiments/observation
      - sensors => measured values
        => Observed states
        => Calculated parameters
  - Calibration on existing equipment
    - Parameter fitting

Two stage heat pump : measures and system state



What is the state of the system ?

What is the value of U ?

What is the value of eff_{is} ?

Goals of the lecture

- How to calibrate models using measurements
  - Where to place measurements
  - Virtual sensors by process models
  - Data reconciliation
    - correct the values of the measurement
  - Parameter identification

Measurement and parameter identification

1. Measured values

$$X^m$$

2. Identification

Model	$F(X, \pi_{unit}) = 0$
Measurements	$X - X^m = 0$

3. Identified parameters

$$\pi_{unit}$$

$$\pi^s_{unit} = \pi_{unit}$$

4. Specified parameters

$$\pi^s_{unit}$$

$$X^s$$

5. Simulation

Model	$F(X, \pi_{unit}) = 0$
Set points	$X - X^s = 0$
Specifications	$\pi_{unit} - \pi^s_{unit} = 0$

6. Performances

7. Optimization

EPFL

Unit model : Incidence matrix rearranged

F(X) : Equations

$Ne = N^{obs} + Nb + Nm$

X : Variables

- Nv state
- Ni intermediate
- Np parameters

$Nx = Nv + Ni + Np$

		X01	X02	X03	X04	X05	X06	X07	X08	X09	X10	X11	X12	X13	X14
N^{obs} Measures $X - X^{obs} = 0$	Eq1	x													
	Eq2		x												
	Eq3			x											
	Eq4				x										
	Eq5					x									
	Eq6						x								
Nb Balances $B(Xin) - B(Xout) = 0$	Eq8		x						x						
	Eq9		x	x						x					
	Eq7	x						x							
	Eq10		x	x							x				
Nm Models $M(X,P) = 0$	Eq11					x					x	x			
	Eq13					x							x	x	
Nc Constitutive equations $C(X) = 0$	Eq14	x				x								x	x
	Eq12					x		x					x		

Measured variables: X01–X06. Non Measured variables: X07–X14.

DOF analysis

$$Ne = Nx$$

Analysing the specification or measurements sets

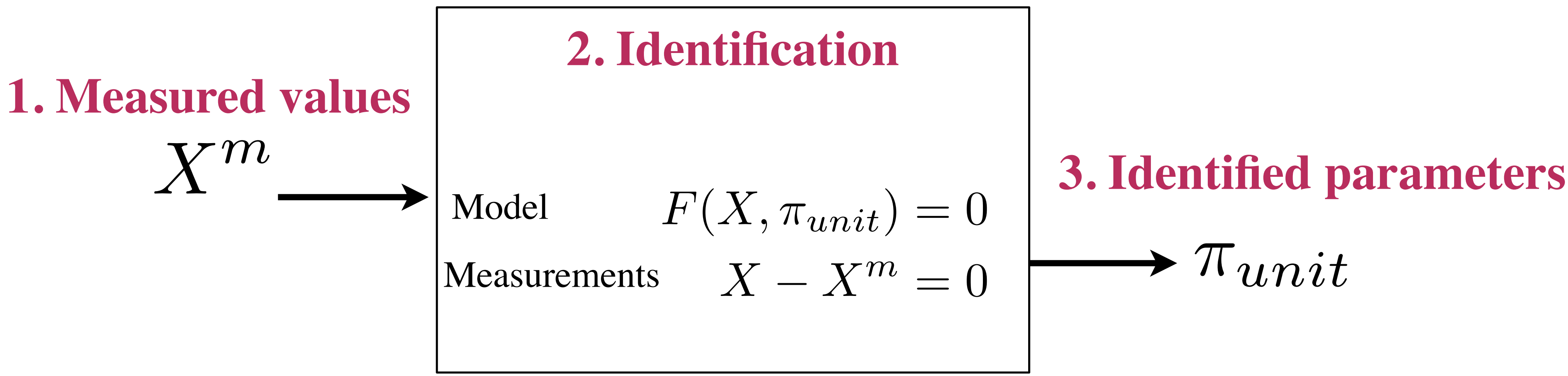
- Goals
  - From a flowsheet flowsheet with a pre-specified set of specification
    - what are the DOF, are there enough specs?
    - if no where to place the missing specifications?
    - If yes what are the extra specifications ?
  - Model is defined by :

$$F(X_{state}) = 0 \Rightarrow \text{equipment model}$$
$$L(X_{state}) = 0 \Rightarrow \text{linking equations}$$
$$T(X_{state}) = 0 \Rightarrow \text{constitutive equations}$$
$$S(X_{state}) = 0 : X_{state} - X_{state}^{specified} = 0 \Rightarrow \text{Specification of the value of state variables}$$

$where$

$$X_{state} = \{x_{StateVariables}, x_{UnitParameters}, y_{decision} \epsilon \{0,1\}\}$$

  - $S(X_{state})$ is the set of specification
    - context
    - operating set points
    - Market specifications
    - Model parameters
  - $S(X_{state})$ needs to be consistent with the model

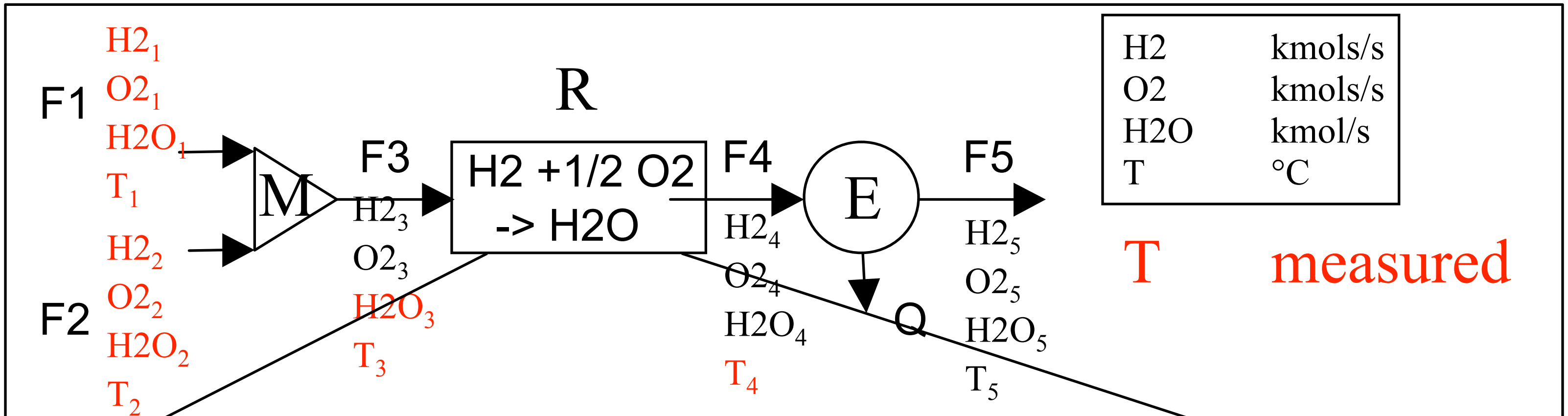


- Do we have enough measurement
  - can the model be solved ?
  - do we need more measurements ?
  - what do we do if we have more measurements ?

Do we have enough measurement/specifications ?

Example of a simplified system

Hydrogen combustion with pure oxygen



Unité R

Mass balance:

$$H2_3 - U - H2_4 = 0$$

$$O2_3 - 1/2\ U - O2_4 = 0$$

$$H2O_3 + U - H2O_4 = 0$$

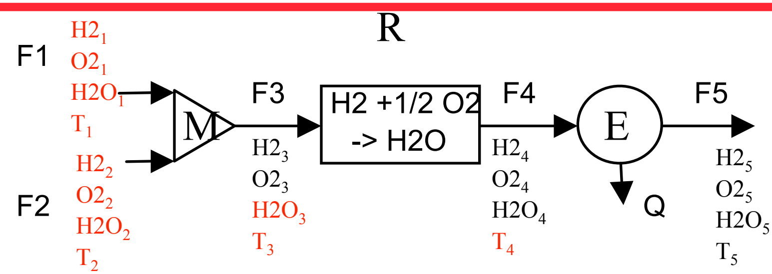
Energy Balance :

$$\Sigma\ x_3 * (h°_{x_3} + h_{x_3}(T_3)) - \Sigma\ x_4 * (h°_{x_4} + h_{x_4}(T_4)) = 0$$

Canonical form : $F(x) = 0 \quad => \quad A\ x = c$

Incidence matrix

Combustion



H2	kmols/s
O2	kmols/s
H2O	kmol/s
T	°C

T mesures

Incidence Matrix : $a_{i,j} = 1$ if variable j occures in equation i

$$A\,x = c$$

Variables : 22 in which 11 measures $\Delta = 11$

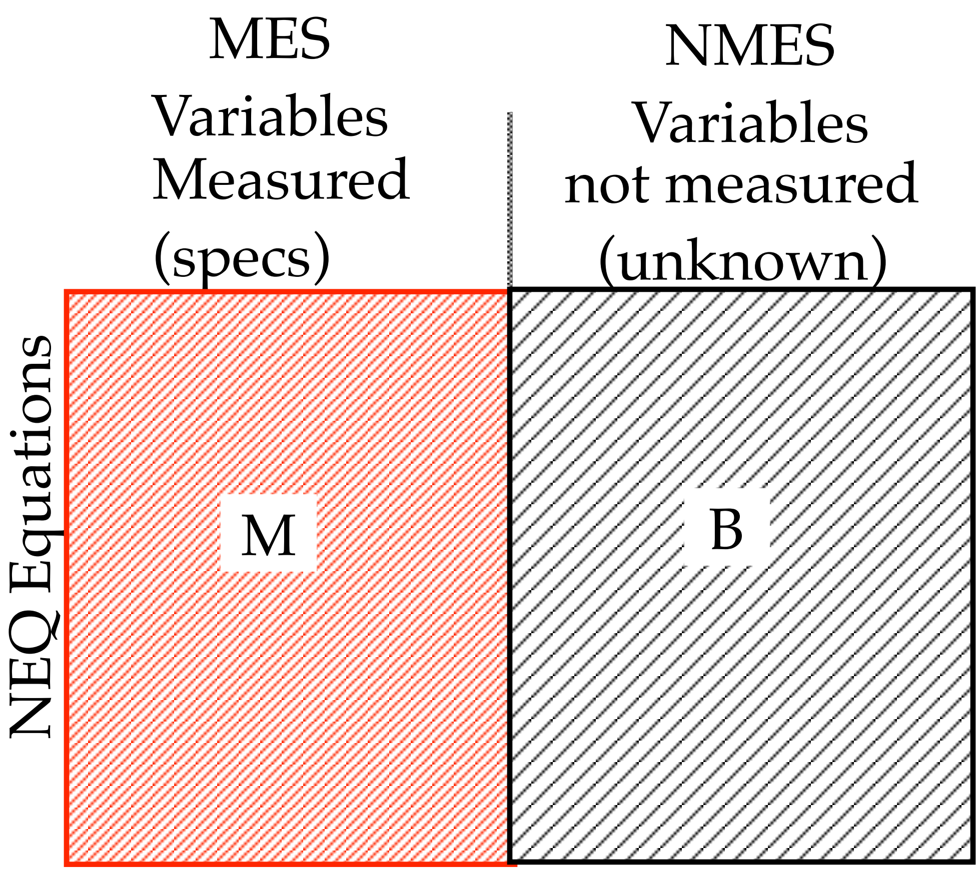
only 10 are needed

Equations 12

			O2 1	H2 1	H2O 1	T 1	O2 2	H2 2	H2O 2	T 2	O2 3	H2 3	H2O 3	T 3	O2 4	H2 4	H2O 4	T 4	[illegible]	O2 5	H2 5	H2O 5	T 5	Q
Bilan Matière	M	O2	X				X				X													
		H2		X				X				X												
		H2O			X				X				X											
Bilan thermique	M		X	X	X	X	X	X	X	X	X	X	X	X										
Bilan Matière	R	O2									X				X				X					
		H2										X				X			X					
		H2O											X				X		X					
Bilan Thermique	R										X	X	X	X	X	X	X	X						
Bilan Matière	E	O2													X					X				
		H2														X					X			
		H2O															X					X		
Bilan thermique	E														X	X	X	X		X	X	X	X	X

Structural analysis : re-arrange the matrix

variables : measured (= specified) or not (to be calculated)



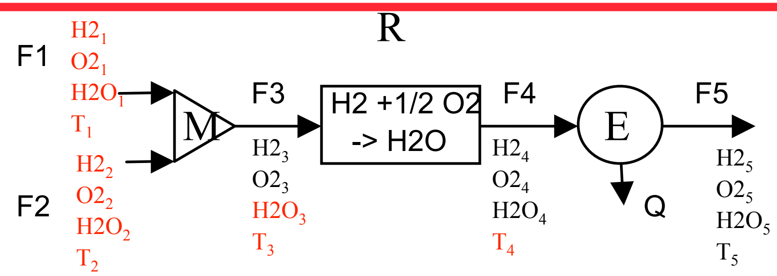
1) **NEQ < NMES** :no solution
(NMES-NEQ) Equations are missing
to calculate unknown variables

2) **NEQ = NMES** : all the unknows can be calulated (just calculable system)

3) **NEQ > NMES :** too many equations **(redundant system)**
in this case some measured values can be recalculated using the value of the other

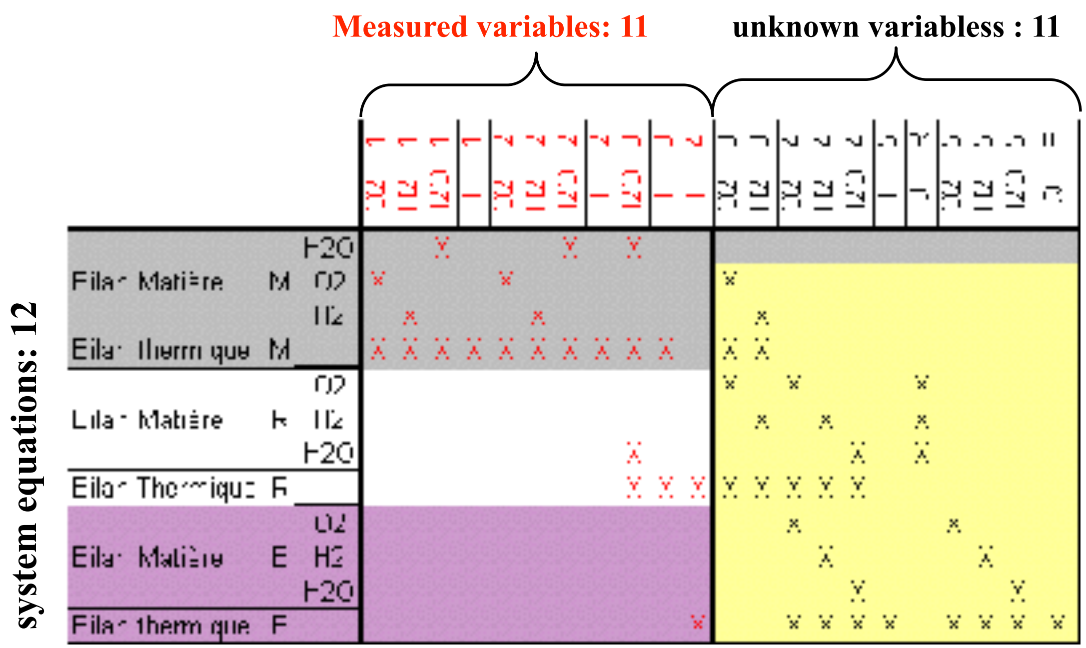
Incidence Matrix

Example : combustion



H2	kmols/s
O2	kmols/s
H2O	kmol/s
T	°C

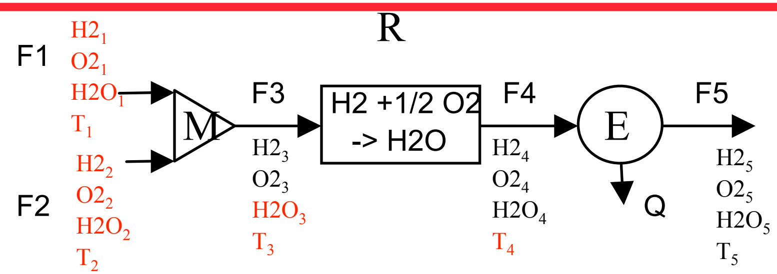
T measured



Square system ?

Rearrange the matrix

Exemple : combustion



H2	kmols/s
O2	kmols/s
H2O	kmol/s
T	°C

T mesures

1) regroup measured and unknowns (**M**+**B**)

2) Reorganise the B matrix (unknowns) by line and column permutations in order to have :

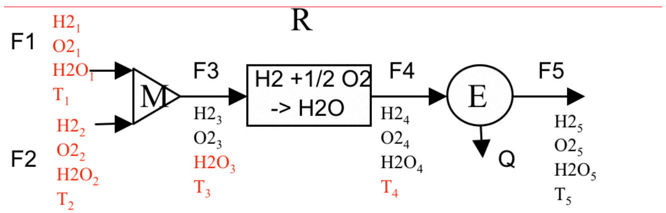
- 1 element on each diagonal position
- regroup in sub-systems (square or rectangles)

measured/specified variables: 11 **Non measured : 11**

system equations: 12

			H2O 3	H2O 1	H2O 2	O2 1	H2 1	O2 2	H2 2	T 1	T 2	T 3	T 4	O2 3	H2 3	U R	O2 4	H2 4	H2O 4	O2 5	H2 5	H2O 5	T 5	Q E
Bilan Matière	M	H2O	X	X	X																			
Bilan Matière	M	O2				X		X						X										
Bilan Matière	M	H2					X		X						X									
Bilan thermique	M		X	X	X	X	X	X	X	X	X	X		X	X									
Bilan Matière	R	O2												X		X	X							
Bilan Thermique	R		X									X	X	X	X		X	X	X					
Bilan Matière	R	H2													X	X		X						
Bilan Matière	R	H2O	X													X			X					
Bilan Matière	E	O2															X			X				
Bilan Matière	E	H2																X			X			
Bilan Matière	E	H2O																	X			X		
Bilan thermique	E												X				X	X	X	X	X	X	X	X

Incidence matrix analysis



			H2O 3	H2O 1	H2O 2	O2 1	H2 1	O2 2	H2 2	T 1	T 2	T 3	T 4	O2 3	H2 3	U R	O2 4	H2 4	H2O 4	O2 5	H2 5	H2O 5	T 5	Q E
Bilan Matière	M	H2O	X	X	X																			
Bilan Matière	M	O2				X		X						X										
Bilan Matière	M	H2					X		X						X									
Bilan thermique	M		X	X	X	X	X	X	X	X	X	X		X	X									
Bilan Matière	R	O2												X		X	X							
Bilan Thermique	R		X									X	X	X	X		X	X	X					
Bilan Matière	R	H2													X	X		X						
Bilan Matière	R	H2O	X													X			X					
Bilan Matière	E	O2															X			X				
Bilan Matière	E	H2																X			X			
Bilan Matière	E	H2O																	X			X		
Bilan thermique	E												X				X	X	X	X	X	X	X	X

Redundant = 1 (nb equations - nb unmeasured variables)

Redundant
nbeq > nb var => possibility to correct measures/ eliminate specification

just calculable
NEQ (7) =NMES(7)

T4 can not be corrected/eliminated

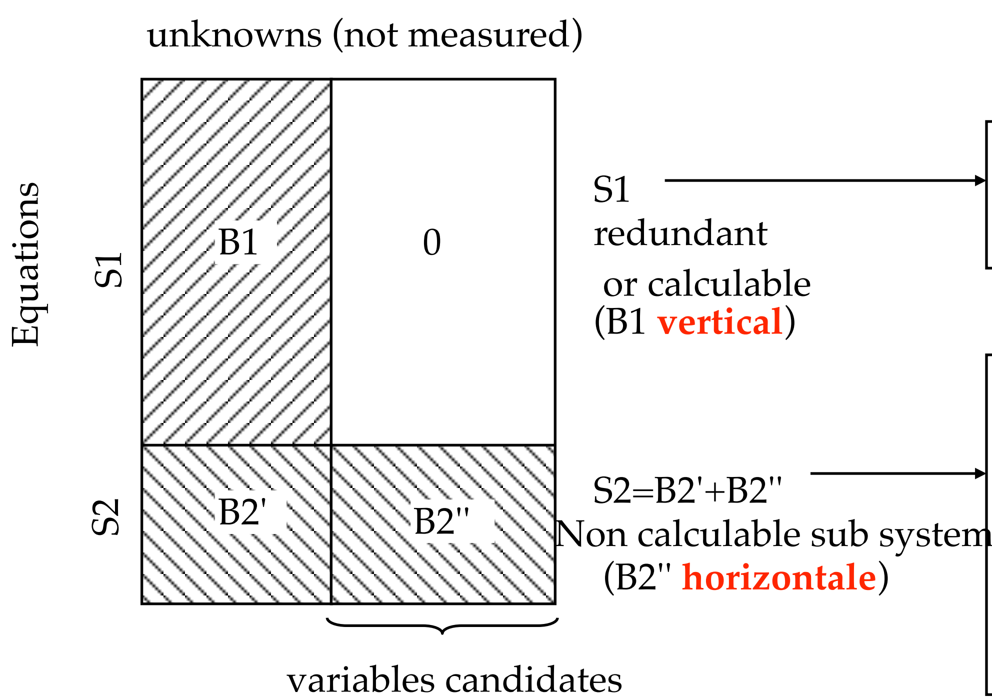
not calculable
NEQ (1) < NMES(2)
Add at least 1 measure (2-1)
T_5 or Q

EPFL

Generalisation : In case of complex systems

1) Reorganise the B matrix (unknowns - equations)
Reorganise the B matrix (unknowns) by line and column permutations in order to have:

- 1 element on each diagonal position
- regroup in sub-systems (square or rectangles)



validable sub-system (rectangle is vertical)
more equations than unknowns

Missing measurement (rectangle is horizontal)
•add measures to make B2'' square or vertical rectangle ($ncolonnes_{B2''}$ - $nlignes_{B2''}$)
•this influences the other permutations => iterations

Analogy measurements and DOF analysis

DoF analysis

- Specifications
- Over-specified
  - Specs to be suppressed
- under specified

  Add specs

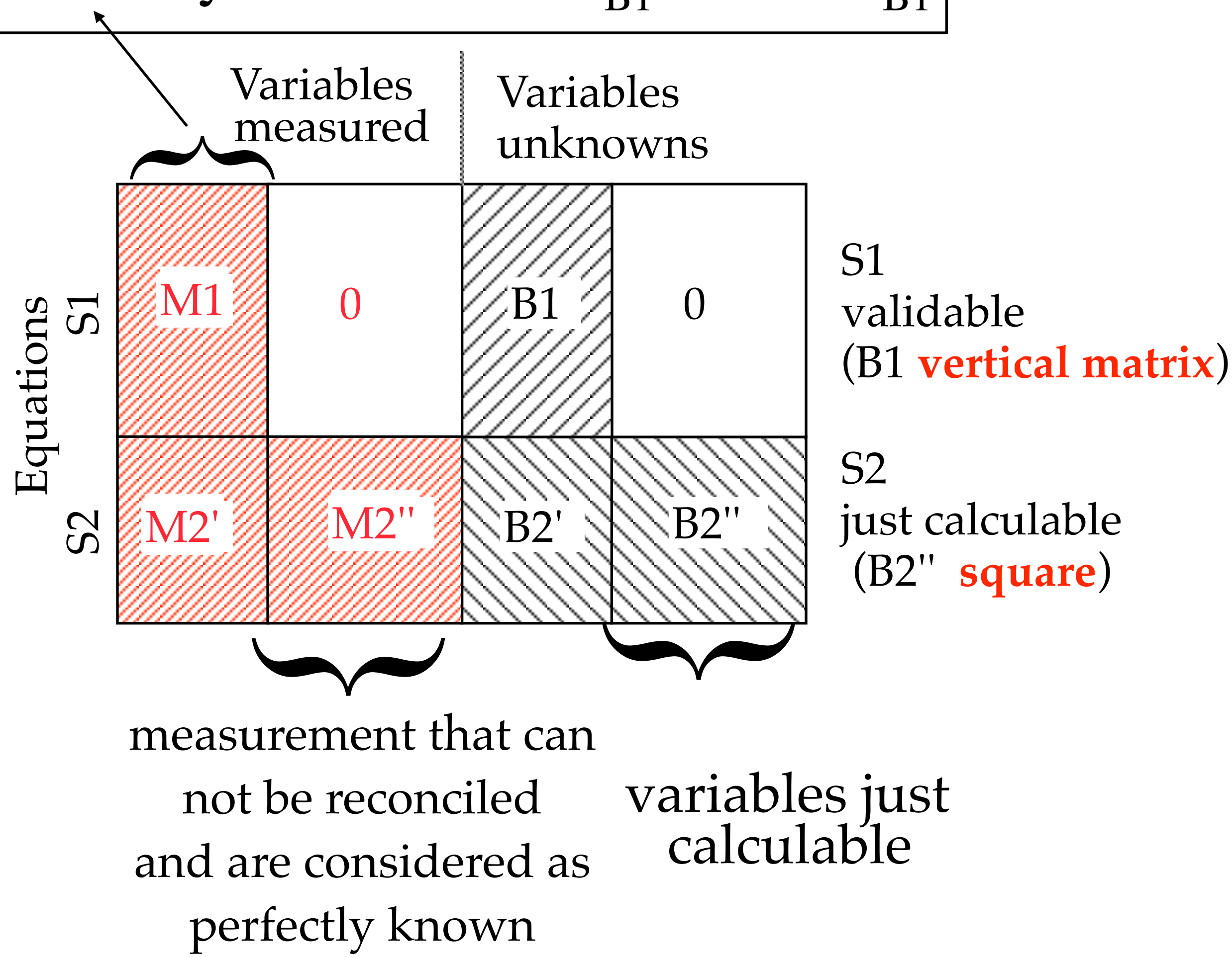
Measurements systems analysis

- Measures
- Redundancy
  - more information available
- Missing measurements
  - add measures

Redundant measurements

Redundant measurement may be reconciled

Redundancy number= $n^{lines}_{B1} - n^{columns}_{B1}$



A redundant measurement can be corrected using the values of the other measurements and the model equations

EPFL

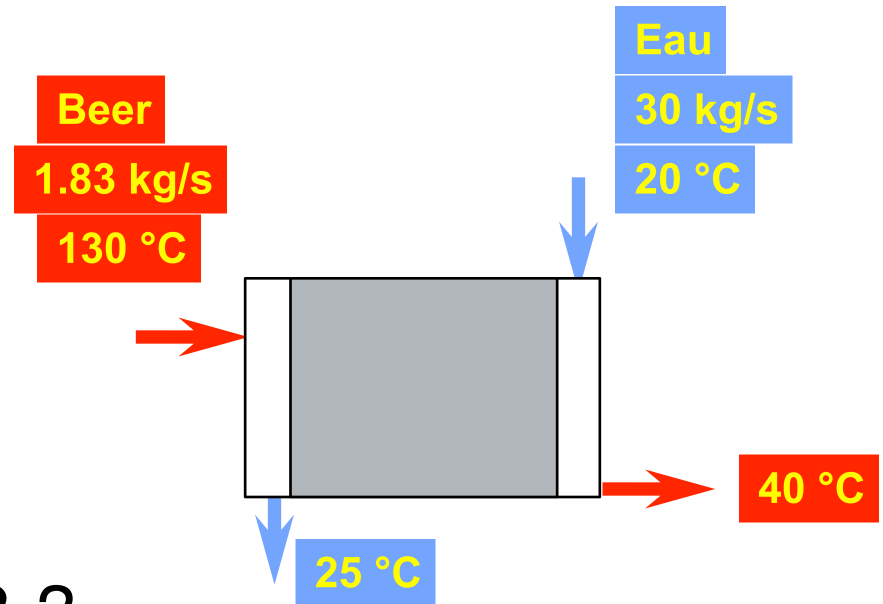
Data reconciliation

What is happening when I have more measures than the minimum number needed ?

What is the heat transfer coefficient of the heat exchanger?

Are the measurements consistent ?

- Equations: 3
  - 2 energy balances
  - $Q=UA\,\Delta T_{lm}$
- State variables: 8
  - 4 temperatures
  - 2 flows
  - 2 parameters Q, U
- Degrees of Freedom : 5 = 8-3
- Measures : 6
  - do not add losses in as a DOF !



Cp = water

Choosing the good measure

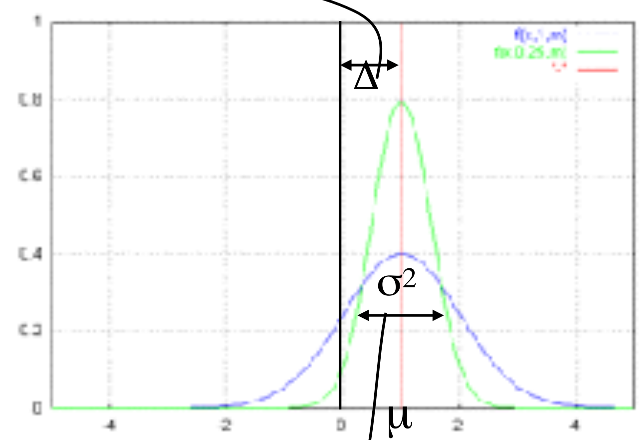
8 variables - 3 equations => 5 measures over 6 have to be fixed

		Measure	1	2	3	4	5	6
Flow 1	kg/s	30.00	32.95	30.00	30.00	30.00	30.00	30.00
T in	°C	20.00	20.00	19.51	20.00	20.00	20.00	20.00
T out	°C	25.00	25.00	25.00	25.49	25.00	25.00	25.00
Q 1	kW	627.	689.	689.	689.	627.	627.	627.
Flow 2	kg/s	1.83	1.83	1.83	1.83	1.67	1.83	1.83
T in	°C	130.	130.	130.	130.	130.	121.9	130.
T out	°C	40.00	40.00	40.00	40.00	40.00	40.00	48.07
Q 2	kW	689.2	689.2	689.2	689.2	627.4	627.4	627.4
ΔT ML	°C	51.3	51.3	51.7	51.1	51.3	48.7	58.3
U	W/m2/K		134	133	135	122	129	108

Measure corrected Specified Calculated

Measurement system

- Classify variables
  - Measured - non measured
  - Redundant - non redundant
  - Calculable - non calculable
  - Specified
- Measures => sensors
  - Exact (mean value)
  - Precision-Accuracy (standard deviation)
- Redundancy
  - Multiple sensors
  - Mass and energy balances



Data reconciliation problem

State variable value — measured value — standard deviation

$$min_{X,Y} \sum_{i=1}^{n_{mes}} (\frac{y_i - y_i^*}{\sigma_i})^2$$

$$
\begin{aligned}
s.t. \quad MassBalance(X,Y) &= 0 \\
EnergyBalance(X,Y) &= 0 \\
Thermodynamic(X,Y) &= 0 \\
ConstituveEquations(X,Y) &= 0 \\
Performance(X,Y,\pi) &= 0 \\
Inequalities(X,Y) &\geq 0
\end{aligned}
$$

$F(Y,X) = 0$

Knowledge about the process

Virtual sensors

Problem resolution : constrained NLP Optimisation

Lagrange multiplier

$$\underset{x_i, y_i, \lambda_i}{Min}\ L = \sum_i \left(\frac{y_i - y_i^*}{\sigma_i}\right)^2 + 2 * \sum_j \lambda_j * f_j(y_i, x_i)$$

Lagrange Formulation

virtual sensor

$$\underset{X, Y, \Lambda}{Min}\ L = (Y - Y^*)^t \mathrm{P}(Y - Y^*) + 2 * \Lambda * F(X, Y)$$

Matrix representation

$$\Rightarrow \nabla L = 0$$

Gradient set to zero

$$soit \quad \frac{\delta L}{\delta \Lambda} = F(Y, X) = 0$$

$$\frac{\delta L}{\delta X} = 2 * \Lambda * B = 0 \quad avec \quad b_{i,j} = \frac{\delta f_i(Y, X)}{\delta x_j}$$

$$\frac{\delta L}{\delta Y} = (Y - Y^*) * P + \Lambda * A = 0 \ avec \quad a_{i,j} = \frac{\delta f_i(Y, X)}{\delta y_j}$$

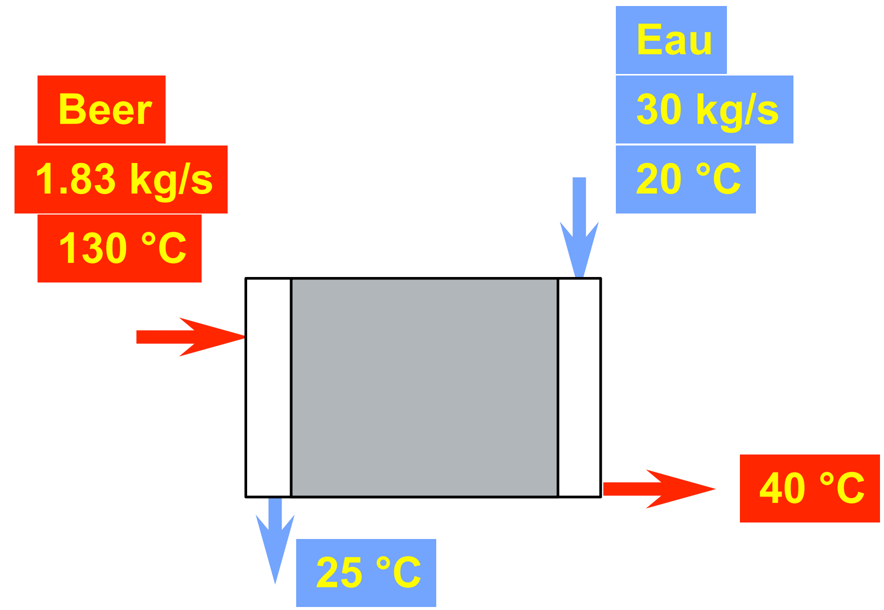
X = non measured, Y = measured

$F(Y, X) = 0$: Set of modeling+ specification equations

EPFL

What is the heat transfer coefficient of the heat exchanger?

- Equations: 3
  - 2 energy balances
  - Q=UA ΔTlm
- State variables: 8
  - 4 temperatures
  - 2 flows
  - 2 parameters Q, U
- Degrees of Freedom : 5 = 8-3
- Measures : 6



data reconciliation results

			Mes.	σ	Vali.	(M-V)/ σ
Flow 1	kg/s	M1	30.00	1.50	30.30	-0.197
T in	°C	T1	20.00	0.50	19.81	0.371
T out	°C	T2	25.00	0.50	25.19	-0.371
Q 1	kW		627.4		680.6	
Flow 2	kg/s	M2	1.83	0.10	1.81	0.215
T in	°C	T3	130.00	1.00	129.96	0.044
T out	°C	T4	40.00	1.00	40.04	-0.044
Q 2	kW		689.2		680.6	
			A	m2	100	
			ΔT LM	°C	51.40	
			U	W/m2/K	132	
					SSQ=	0.3643

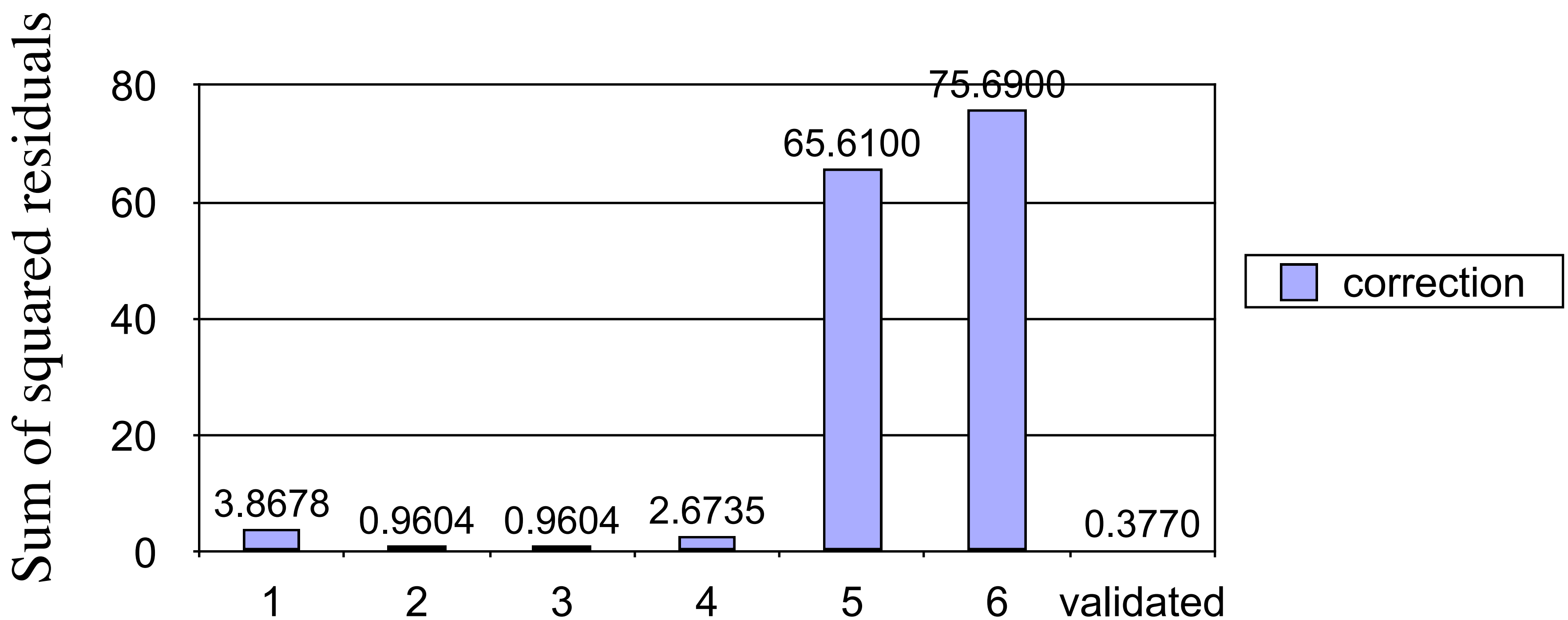
What are the most probable values of the measured values ?

All measures are considered

		Mesures	1	2	3	4	5	6	
Flow 1	kg/s	30.00	32.95	30.00	30.00	30.00	30.00	30.00	30.30
T in	°C	20.00	20.00	19.51	20.00	20.00	20.00	20.00	19.81
T out	°C	25.00	25.00	25.00	25.49	25.00	25.00	25.00	25.19
Q 1	kW	627.	689.	689.	689.	627.	627.	627.	680.6
Flow 2	kg/s	1.83	1.83	1.83	1.83	1.67	1.83	1.83	1.81
T in	°C	130.	130.	130.	130.	130.	121.9	130.	129.96
T out	°C	40.00	40.00	40.00	40.00	40.00	40.00	48.07	40.04
Q 2	kW	689.2	689.2	689.2	689.2	627.4	627.4	627.4	680.6
ΔT ML	°C	51.3	51.3	51.7	51.1	51.3	48.7	58.3	51.40
U	W/m2/K		134	133	135	122	129	108	132

Measure Corrected Specification Calculated Validated

Results validity



Results analysis

- How to use the results
  - Sum of square residuals
    - is there a lot of corrections ?
    - Is the model (what we know) valid ?
      - e.g. a leakage is apriori not modeled
  - Are the bounds activated
    - Is the model valid
  - Sensitivity analysis
    - One can calculate the precision of the value of measured and unmeasured values
  - Corrections analysis
    - Failling sensors => Gross errors (if big corrections => remove the sensor)
    - Sensor calibration
  - Importance of the sensors on the results

Sensitivity

When the solution is obtained, we have

weight x measure

$$\nabla L = 0 \qquad \equiv \begin{bmatrix} P & 0 & A^T \\ 0 & 0 & B^T \\ A & B & 0 \end{bmatrix} * \begin{bmatrix} Y \\ X \\ \Lambda \end{bmatrix} = \begin{bmatrix} P\,Y^* \\ 0 \\ -C \end{bmatrix}$$

Or $\quad MV = D \quad$ D is the set of measured values

And $\quad V = M^{-1}D \quad$ Sensitivity of the calculated variable w.r.t to D

P is the weight of the measures $(\frac{1}{\sigma^2})$

$A = \frac{\delta F(X,Y)}{\delta Y} \qquad B = \frac{\delta F(X,Y)}{\delta X} \qquad F(X,Y)$ process model

Sensitivity analysis : Variance of the results

In detail

The variance is calculated as a sensitivity to the variance of the measurement

Measurement $$Y_i = \sum_{j=1}^{m+n+p} \left(M^{-1}\right)_{ij} D_j$$

$$= \sum_{j=1}^{m} \left(M^{-1}\right)_{ij} P_{jj}\, y_j^* - \sum_{k=1}^{p} \left(M^{-1}\right)_{i\ \ n+m+k} C_k$$

Sensitivity of the measured value
Sensitivity of the precision

Calculated $$X_i = \sum_{j=1}^{m+n+p} \left(M^{-1}\right)_{n+i\ \ j} D_j$$

$$= \sum_{j=1}^{m} \left(M^{-1}\right)_{n+i\ \ j} P_{jj}\, y_j^* - \sum_{k=1}^{p} \left(M^{-1}\right)_{n+i\ \ n+m+k} C_k$$

Sensitivity of the measured value
Sensitivity of the precision

Variance calculation if $Z = \sum_{j=1}^{m} a_j X_j$ then $\operatorname{var}(Z) = \sum_{j=1}^{m} a_j^2 \operatorname{var}(X_j)$

Sensitivity of solutions

to measurement

How much a calculated value is influenced by the value of a measurement

$$M\frac{\delta V}{\delta Y^*}+\frac{\delta M}{\delta Y^*}V-\frac{\delta D}{\delta Y^*}=0\Rightarrow\frac{\delta V}{\delta Y^*}=M^{-1}\begin{bmatrix}P\\0\\0\end{bmatrix}$$

Sensitivity of solutions

How much a calculated value is influenced by the accuracy of a measure

to measurement accuracy

$$M\frac{\delta V}{\delta P}+\frac{\delta M}{\delta P}V-\frac{\delta D}{\delta P}=0$$

$$\Rightarrow \begin{bmatrix} \dfrac{\delta Y}{\delta P} \\ \dfrac{\delta X}{\delta P} \\ \dfrac{\delta \Lambda}{\delta P} \end{bmatrix} = M^{-1}\left[\begin{bmatrix} Y^* \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} Y \\ X \\ \Lambda \end{bmatrix}^*\right]$$

A posteriori variance

Standard deviation of the calculated variables =f(P,Y*)

$$\operatorname{var}(Y_i) = \sum_{j=1}^{m} \left\{ \left(M^{-1}\right)_{ij} P_{jj} \right\}^2 \operatorname{var}\left(y^*_j\right)$$

$$\operatorname{var}(X_i) = \sum_{j=1}^{m} \left\{ \left(M^{-1}\right)_{n_{mes}+i\,j} P_{jj} \right\}^2 \operatorname{var}\left(y^*_j\right)$$

With $\operatorname{var}\left(y^*_j\right) = 1/P_{jj}$

$$\operatorname{var}(Y_i) = \sum_{j=1}^{m} \frac{\left(M^{-1}\right)^2_{ij}}{\operatorname{var}\left(y^*_j\right)}$$

$$\operatorname{var}(X_i) = \sum_{j=1}^{m} \frac{\left(M^{-1}\right)^2_{n_{mes}+i\,j}}{\operatorname{var}\left(y^*_j\right)}$$

Data reconciliation : conclusion

- Corrects the measurement values (most probable consistent) value
- Consistent with heat and mass balances & thermodynamics
- Considers balances as additional measures (virtual sensors)
- A posteriori precision of each value (measured and non measured)
- Precision of performance indicators
- Sensitivity of measurements on performance indicators
- Quality of sensors

Parameter identification

Measurement and parameter identification

1. Measured values

X^m

2. Identification

Model $F(X, \pi_{unit}) = 0$

Measurements $X - X^m = 0$

3. Identified parameters

π_{unit}

4. Specified parameters

$\pi^s_{unit} = \pi_{unit}$

π^s_{unit}

X^s

5. Simulation

Model $F(X, \pi_{unit}) = 0$

Set points $X - X^s = 0$

Specifications $\pi_{unit} - \pi^s_{unit} = 0$

6. Performances

7. Optimization

In a perfect world

measured value of X_i in experiment e

1. Measured values

$X_{i,e}^m$

2. Identification

$$\text{Model } F_u(X_{i,e}, \pi_{p,u,e}) = 0 \quad \forall e \in \{n_e\} \quad \forall u \in \{n_u\}$$

$$\text{Measurements } X_{i,e} - X_{i,e}^m = 0$$

$\pi_{p,u,e}$

value of parameter π_u in experiment e

3. Identified parameters

Parameter identification from a set of experiments

1. Measured values $X_{i,e}^{m}$

2. Identification

$$\min_{X_{i,e},\pi_{p,u}} \sum_{e=1}^{n_e} \sum_{i=1}^{n_m} \frac{(X_{i,e} - X_{i,e}^{m})^2}{\sigma_i^2}$$

$$\text{s.t. } F_u(X_{i,e}, \pi_{p,u}) = 0 \quad \forall e \in \{n_e\} \quad \forall i \in \{n_s\} \quad \forall u \in \{n_u\}$$

n_m : number of measured values

n_e : number of set of experiments

n_u : number of units

n_s : number of state variables in the model

3. Identified parameters $\pi_{p,u}\ \forall u \in \{n_u\}$ and $X_{i,e},\ \forall i \in \{n_s\}\ \forall e \in \{n_e\}$

Validity of the parameter identification

- Number of parameters (p)
- Number of measurement set (n)
- Regression coefficient

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- Regression validity : Fischer test

Fisher value from a table

$$F = \frac{(n-p)R^2}{(p-1)(1-R^2)} > F(p-1, n-p, 1-\alpha)$$

α : significativity level