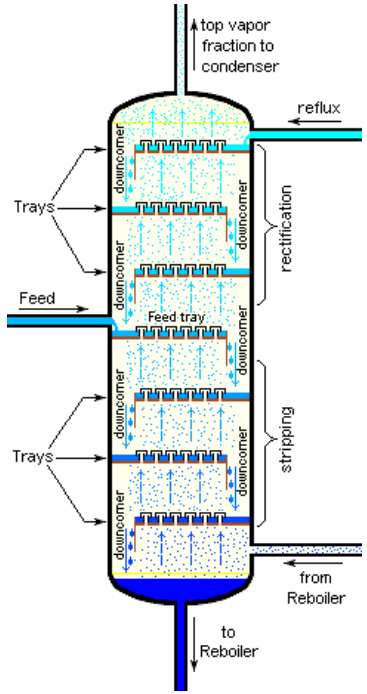
TOWARDS A CLEVER OPTIMIZATION OF MODELS

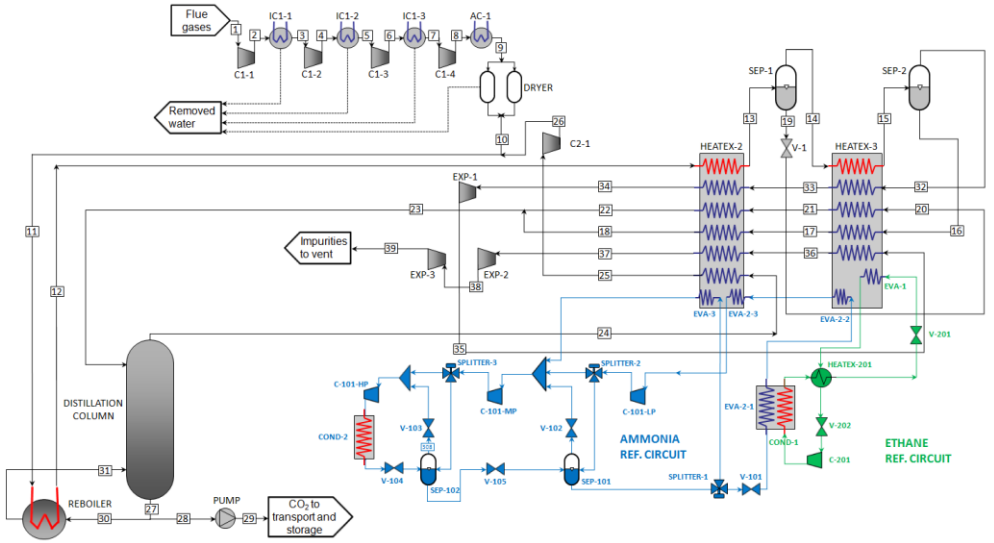
Silvia Lasala
1st April 2016

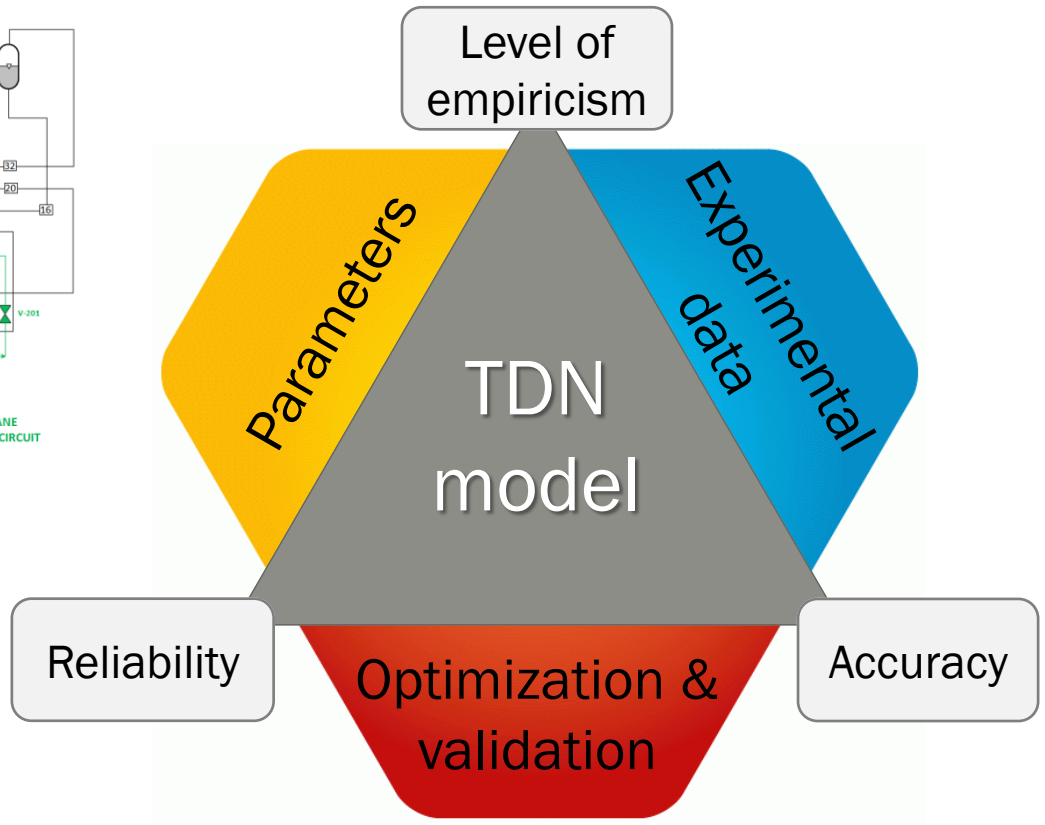
silvia.lasala@univ-lorraine.fr
ENSIC - Nancy (FR)

IFP Energies nouvelles
Rueil-Malmaison (FR)

Introduction







Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Introduction

Cascade of pivotal questions …

? What types of data should be used to perform the model optimization?

? How do characteristics of data influence the optimization problem?

? How to set up a clever optimization, considering the characteristics of available data?

? What is the role played by experimental uncertainties? How important they are in the optimization process?

? How to account for the uncertainty of models?

? …

This presentation is an attempt to answer these questions

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

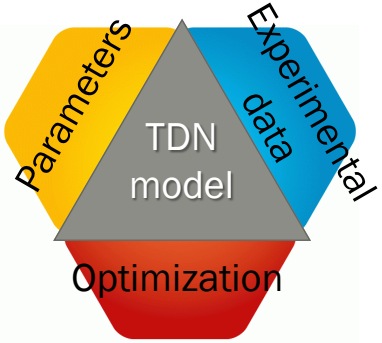
Outline

- **Key characteristics of experimental data**
- **The optimization process. Main considerations**
  - *Types of optimization processes*
  - *The role played by uncertainties of molar fractions*
  - *The uncertainty of model's parameters*
  - *The effect of the numerosity and dispersion of experimental data on optimization results of different types of systems*
- **In conclusion … how to set up a clever optimization of models?**

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of model parameters

Experimental data for parameters optimization



What *kind of parameters* to be optimized?

⇩

What *type of data* for model's optimization?

- Empirical parameters

  Optimization over different types of experimental data (multi-property data: phase-equilibrium, p-ρ-T, h^M, $c_p{}^M$, …, of pure, binary and multi-component systems)

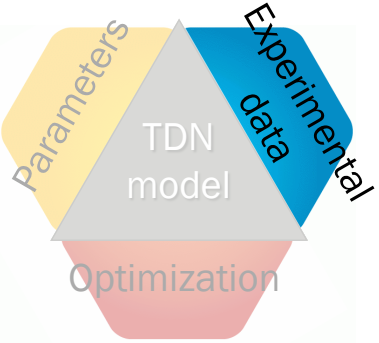
- Theoretically-based parameters

  Optimization over experimental data that macroscopically reflect the theoretical significance of parameters (typical example: saturation VLE data for regressing binary interaction parameters)

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of model parameters

General characteristics of experimental data

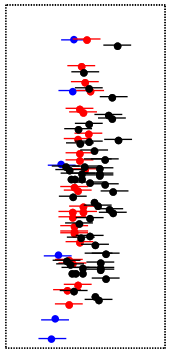


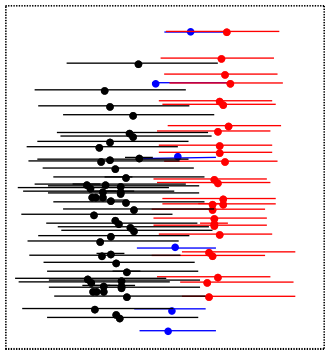
Analysis of general characteristics of experimental data

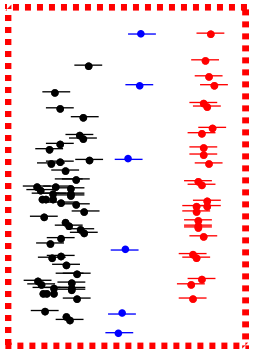
- The numerosity
- The dispersion
- The (availability of) experimental uncertainties

→ Enable the analysis of the **compatibility** between measurements and determination of the **uncertainty of the whole sample** of data.

Different types of samples of experimental data







THE KNOWLEDGE OF EXPERIMENTAL UNCERTAINTIES IS FUNDAMENTAL TO COMPREHEND AND TO QUANTIFY THE CAUSE OF THE DISPERSION OF EXPERIMENTAL DATA

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of model parameters

The role of experimental uncertainties

Do available *data* sets incorporate *uncertainties*?

Yes

⇩

It is possible to:

- compare experimental data on the basis of their compatibility
- optimize model parameters by weighting data over their uncertainties
- define a reasonable lower bound of the residuals between model's calculations and experimental data

No

⇩

It would be preferable to select only data that:

- belong to numerous populations
- are measured by different authors and with different experimental equipment

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

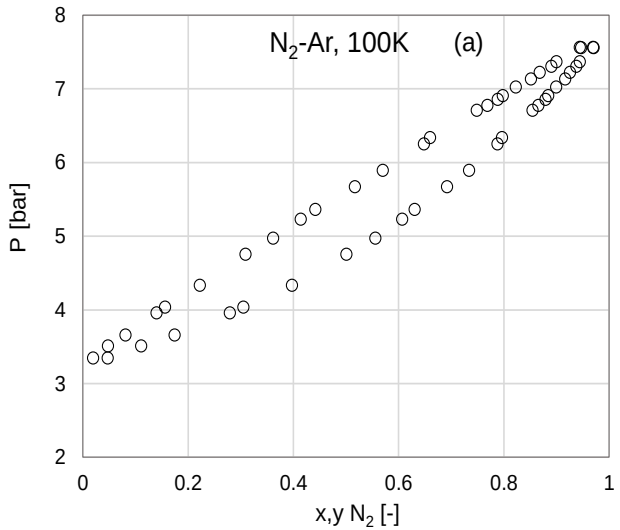
Optimization of model parameters

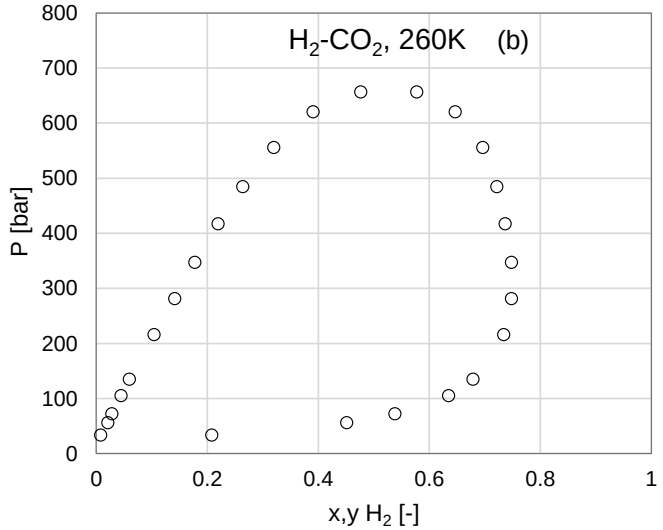
Preliminary analysis. Treating VLE data

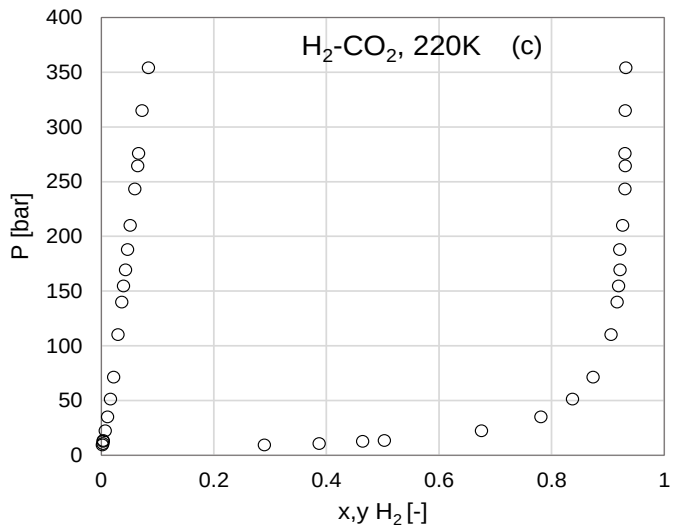
- The numerosity
- The dispersion
- The (availability of) experimental uncertainty
- The type of data (and, thus, of system)

More in particular, characteristics of VLE data

- The *"shape" of the phase diagram*
- The *homogeneity of the distribution of experimental data* along the whole phase diagram
- The availability of both *dew and bubble points* or just one of the two



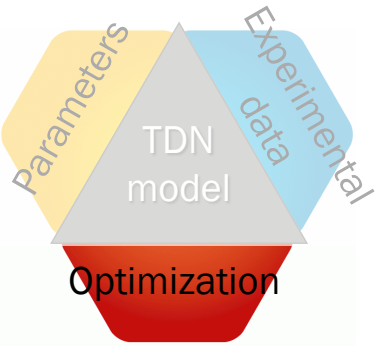




Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of model parameters

The selection of the optimization process



The results of **any optimization process** depend on:

- characteristics of experimental data
- the complexity of the analysed thermodynamic system
- the thermodynamic model under optimization
- the selected optimization process

⇨ ***Is there a specific optimization process that allows the optimal treatment of a group of data characterized by specific features ?***

The best is to apply methods (objective functions to be optimized) that account for the uncertainty of experimental data and to determine the uncertainty of optimal model's parameters.

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The selection of the objective function.

- Simple or weighted least squared methods

Optimization method	Example of LS objective functions $S(\bar{\beta})$
Simple least squared method	$\frac{1}{n_e}\sum_{i=1}^{n_e}(\hat{z}_{1i}-\tilde{z}_{1i})^2$ *with z being either x or y*
Simple least squared method	$\frac{1}{n_e}\sum_{i=1}^{n_e}(\hat{x}_{1i}-\tilde{x}_{1i})^2+\frac{1}{n_e}\sum_{i=1}^{n_e}(\hat{y}_{1i}-\tilde{y}_{1i})^2$
Weighted least squared method	$\frac{1}{n_e}\sum_{i=1}^{n_e}\left(\frac{\hat{z}_{1i}-\tilde{z}_{1i}}{u^e_{\hat{z}_{1i}}}\right)^2$ *with z being either x or y*
Weighted least squared method	$\frac{1}{n_e}\sum_{i=1}^{n_e}\left(\frac{\hat{x}_{1i}-\tilde{x}_{1i}}{u^e_{\hat{x}_i}}\right)^2+\frac{1}{n_e}\sum_{i=1}^{n_e}\left(\frac{\hat{y}_{1i}-\tilde{y}_{1i}}{u^e_{\hat{y}_{1i}}}\right)^2$

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The selection of the objective function

$$S(\bar{\beta}) = \frac{1}{n_e}\sum_{i=1}^{n_e}\left[\left(\frac{\widehat{P}_i - \widetilde{P}_i}{u^e_{\widehat{P}_i}}\right)^2 + \left(\frac{\widehat{T}_i - \widetilde{T}_i}{u^e_{\widehat{T}_i}}\right)^2 + \frac{1}{NC}\sum_{j=1}^{NC}\left(\left(\frac{\hat{x}_{ji} - \tilde{x}_{ji}}{u^e_{\hat{x}_{j,i}}}\right)^2 + \left(\frac{\hat{y}_{ji} - \tilde{y}_{ji}}{u^e_{\hat{y}_{j,i}}}\right)^2\right)\right]$$

Complete weighted least squared method

≡

Maximum likelihood method

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The selection of the objective function

- The maximum likelihood method for a binary mixture:

$$S(\bar{\beta}) = \frac{1}{n_e}\left[\sum_{i=1}^{n_e}\left(\frac{\hat{x}_{1i}-\tilde{x}_{1i}}{\sigma_{x_{1,i}}}\right)^2 + \sum_{i=1}^{n_e}\left(\frac{(\hat{y}_{1i}-\tilde{y}_{1i})\sigma_{x_{1,i}}^2}{\sigma_{x_{1,i}}\sqrt{\sigma_{x_{1,i}}^2\sigma_{y_{1,i}}^2-cov(x_{1,i},y_{1,i})^2}} - \frac{(\hat{x}_{1i}-\tilde{x}_{1i})cov(x_{1,i},y_{1,i})}{\sigma_{x_{1,i}}\sqrt{\sigma_{x_{1,i}}^2\sigma_{y_{1,i}}^2-cov(x_{1,i},y_{1,i})^2}}\right)^2\right]$$

$$\sigma_{y_{1,i}}^2 = \left(u_{y_{1,i}}^e\right)^2 + \left(\frac{\partial y_1}{\partial P}\right)_{T,\beta}^2 \left(u_{P_i}^e\right)^2 + \left(\frac{\partial y_1}{\partial T}\right)_{P,\beta}^2 \left(u_{T_i}^e\right)^2$$

$$\sigma_{x_{1,i}}^2 = \left(u_{x_{1,i}}^e\right)^2 + \left(\frac{\partial x_1}{\partial P}\right)_{T,\bar{\beta}}^2 \left(u_{P_i}^e\right)^2 + \left(\frac{\partial x_1}{\partial T}\right)_{P,\bar{\beta}}^2 \left(u_{T_i}^e\right)^2$$

$$cov(x_{1,i},y_{1,i}) = \left(\frac{\partial x_1}{\partial T}\right)_{P,\bar{\beta}}\left(\frac{\partial y_1}{\partial T}\right)_{P,\bar{\beta}} \left(u_{T_i}^e\right)^2 + \left(\frac{\partial x_1}{\partial P}\right)_{T,\bar{\beta}}\left(\frac{\partial y_1}{\partial P}\right)_{T,\bar{\beta}} \left(u_{P_i}^e\right)^2$$

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

Experimental uncertainties of molar fractions

Considering a stream of NC components, **chromatography** allows the measurements of:

- *The number of moles of each specie, n_i*
- *The uncertainty, $u(n_i)$, of each measured n_i: $u(n_i) = \overline{b_i} \cdot n_i + \overline{c_i}$*

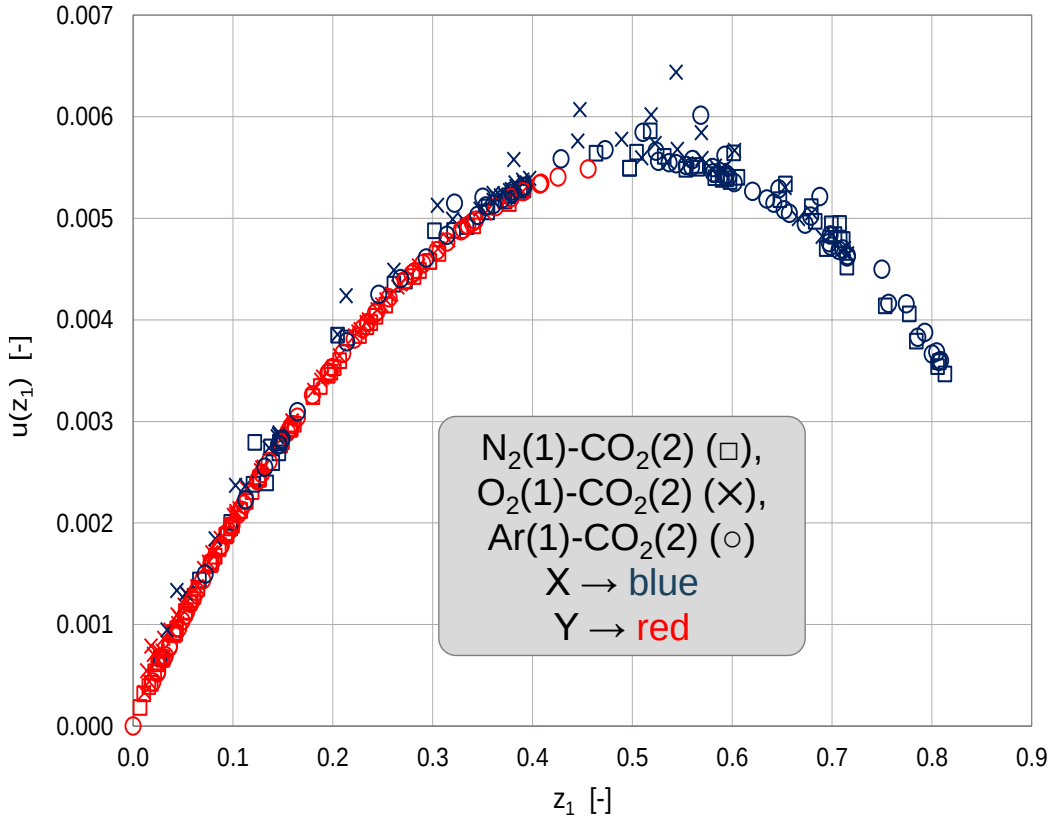
The uncertainty of molar fraction z_i is determined applying the Error Propagation formula:

$$u^2(z_i) = \sum_{j=1}^{NC} \left[\frac{\partial f(n_1, \ldots, n_{NC})}{\partial n_j} \right]^2_{n_{i=j}} \cdot u^2(n_j) \quad ; \quad f: z_i = \frac{n_i}{\sum_j^{NC} n_j}$$

For a binary mixture

$$u_{z_1}^2 = \frac{n_2^2}{(n_1+n_2)^4} u_{n_1}^2 + \frac{n_1^2}{(n_1+n_2)^4} u_{n_2}^2$$

$$u_{z_2}^2 = \frac{n_2^2}{(n_1+n_2)^4} u_{n_1}^2 + \frac{n_1^2}{(n_1+n_2)^4} u_{n_2}^2$$



Similar considerations are derived for **gravimetric compositional measurements**

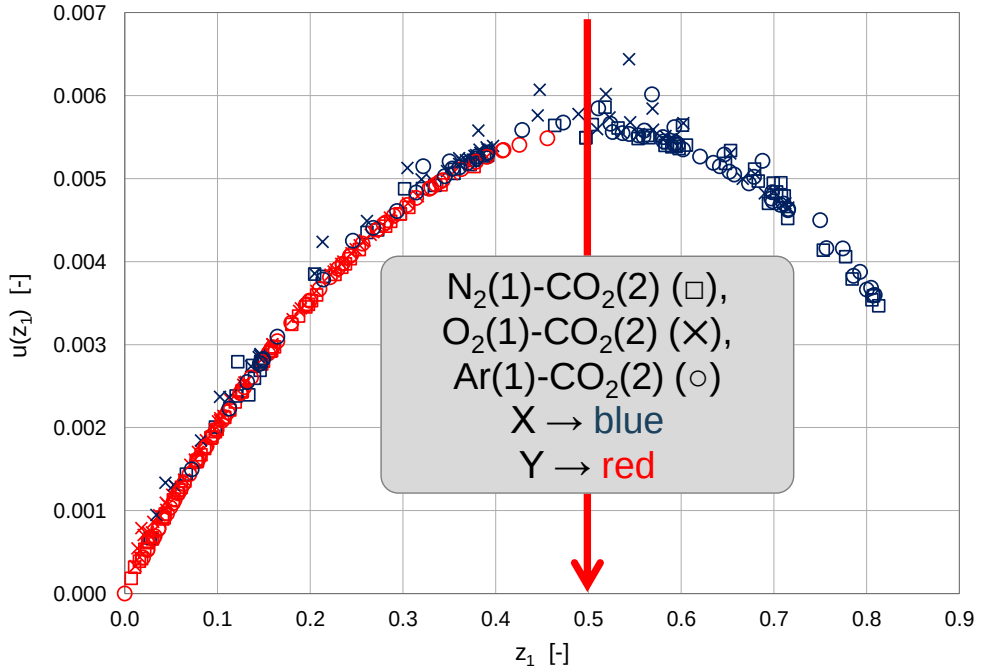
Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

Experimental uncertainties of molar fractions

The analysis of our results has shown that:

1. In general, $\boldsymbol{u(z_i) = a \cdot z_i^2 + b \cdot z_i}$
2. Molar fractions of 0.5 have a maximum uncertainty: $\boldsymbol{(du(z_i)/dz_i)_{z_i=0.5} = 0}$
3. At the critical point ($x_i^{cr} = y_i^{cr}$):
$$\boldsymbol{u(x_i^{cr}) = u(y_i^{cr})}$$
4. In general, at non-critical conditions:
$$\boldsymbol{u(x_i) \neq u(y_i)}$$



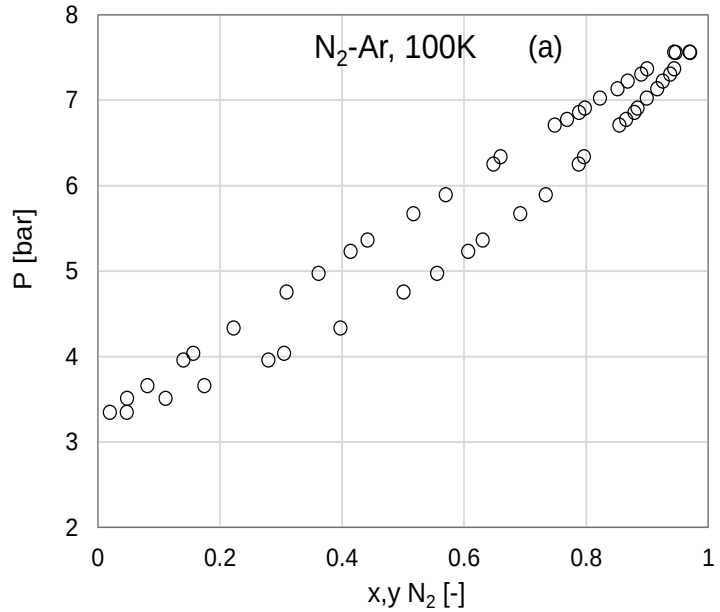
The application of these outcomes enables the estimation of molar fractions uncertainty ($u(z_i)$) of literature data for which it is only specified a **maximum molar fraction uncertainty, associated to the specific mole fraction z***, $u_{max}(z^*)$:

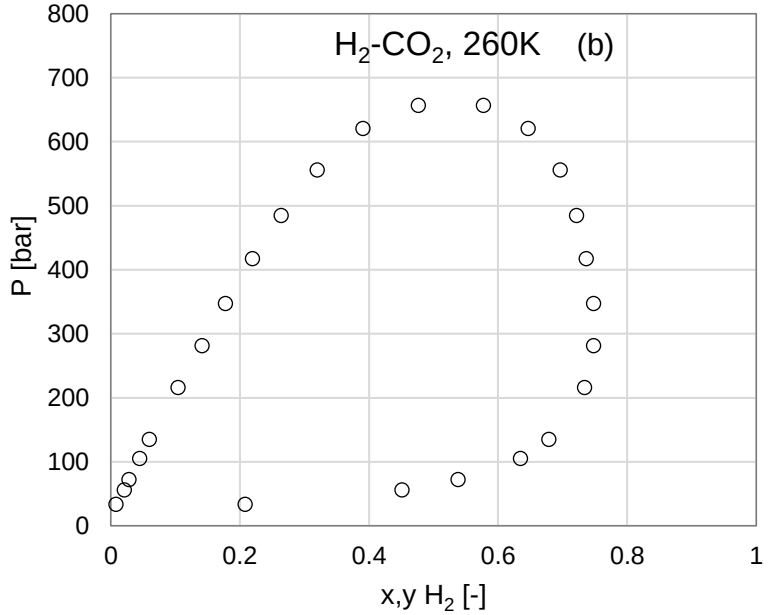
$$u(z_i) = u_{max}(z_i^*)\frac{z_i^2 - z_i}{z_i^{*2} - z_i^*}$$

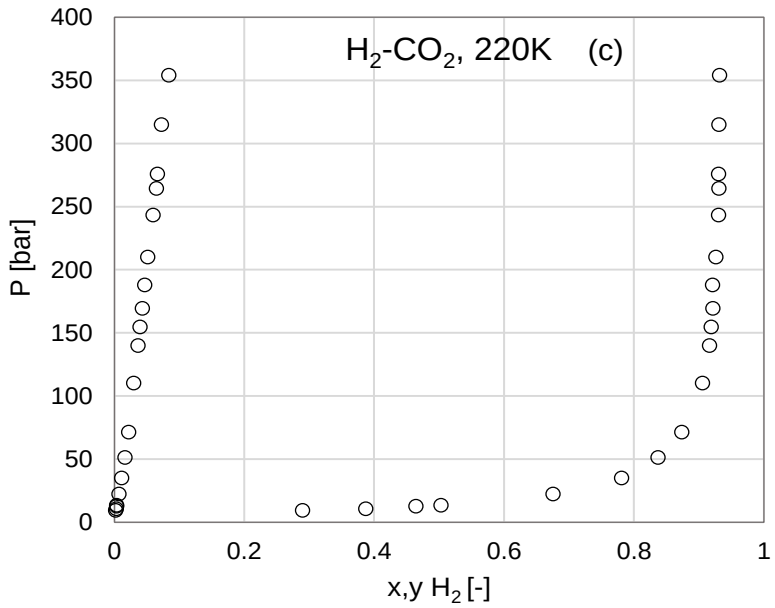
Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

Maximum likelihood method and experimental uncertainties







$$\sigma^2_{y_{1,i}} = \left(u^e_{y_{1,i}}\right)^2 + \left(\frac{\partial y_1}{\partial P}\right)^2_{T_i,\beta} \left(u^e_{P_i}\right)^2 + \left(\frac{\partial y_1}{\partial T}\right)^2_{P_i,\beta} \left(u^e_{T_i}\right)^2$$

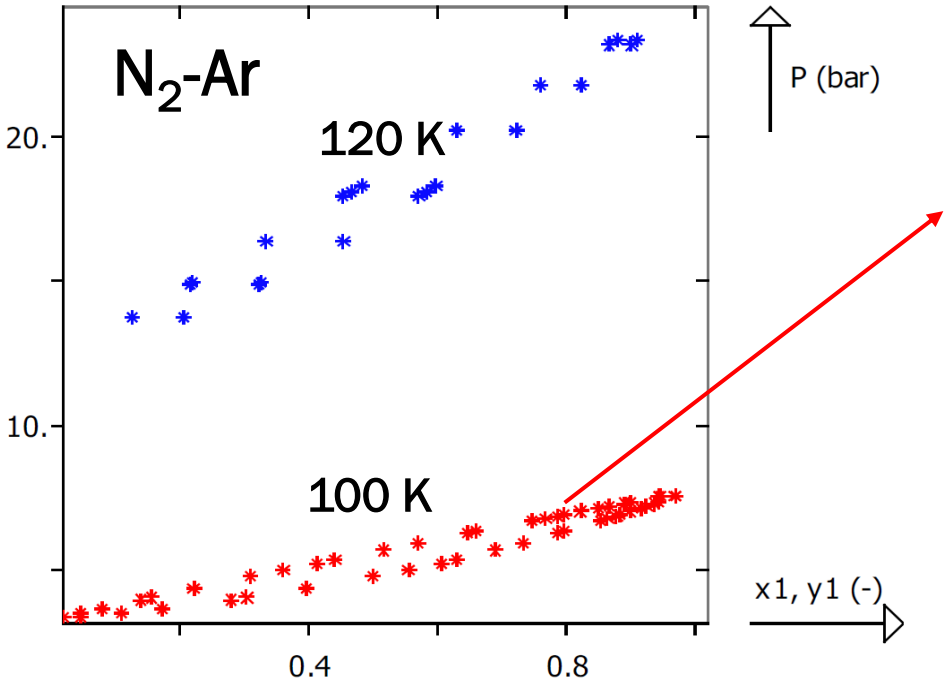
$$\sigma^2_{x_{1,i}} = \left(u^e_{x_{1,i}}\right)^2 + \left(\frac{\partial x_1}{\partial P}\right)^2_{T_i,\overline{\beta}} \left(u^e_{P_i}\right)^2 + \left(\frac{\partial x_1}{\partial T}\right)^2_{P_i,\overline{\beta}} \left(u^e_{T_i}\right)^2$$

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

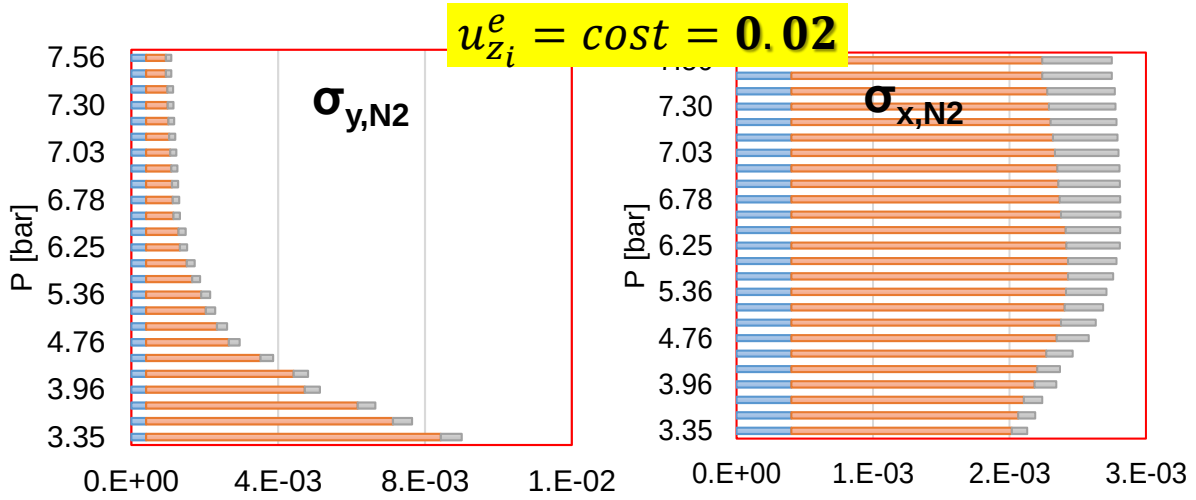
Optimization of models over VLE data

Maximum likelihood method and experimental uncertainties

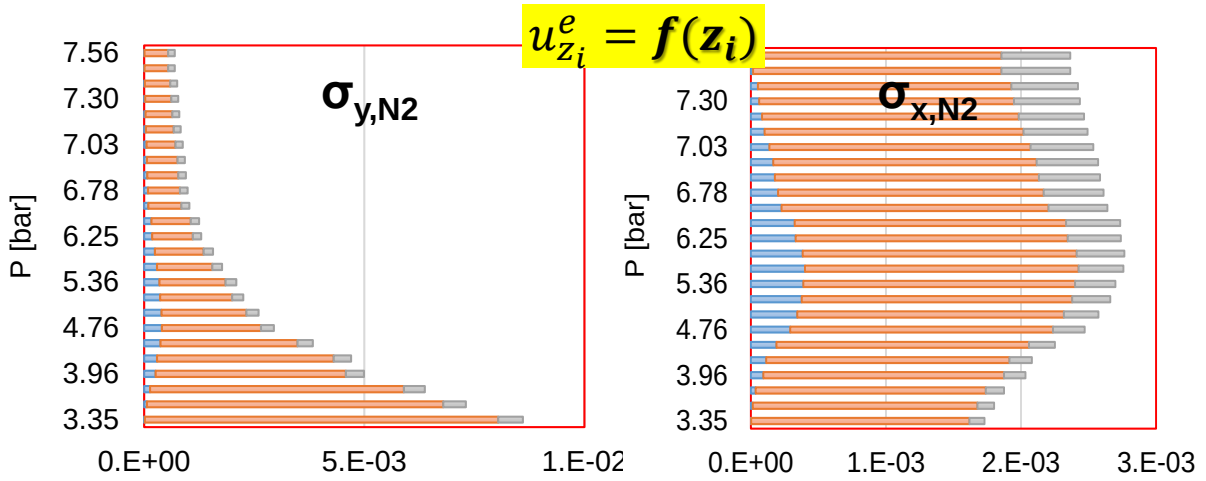
$$\sigma^2_{z_{1,i}} = \left(u^e_{z_{1,i}}\right)^2 + \left(\frac{\partial z_1}{\partial P}\right)^2_{T_i,\bar{\beta}} \left(\underbrace{u^e_{P_i}}_{\mathbf{0.2bar}}\right)^2 + \left(\frac{\partial z_1}{\partial T}\right)^2_{P_i,\bar{\beta}} \left(\underbrace{u^e_{T_i}}_{\mathbf{0.2K}}\right)^2$$



$u^e_{z_i} = cost = \mathbf{0.02}$



$u^e_{z_i} = \boldsymbol{f(z_i)}$

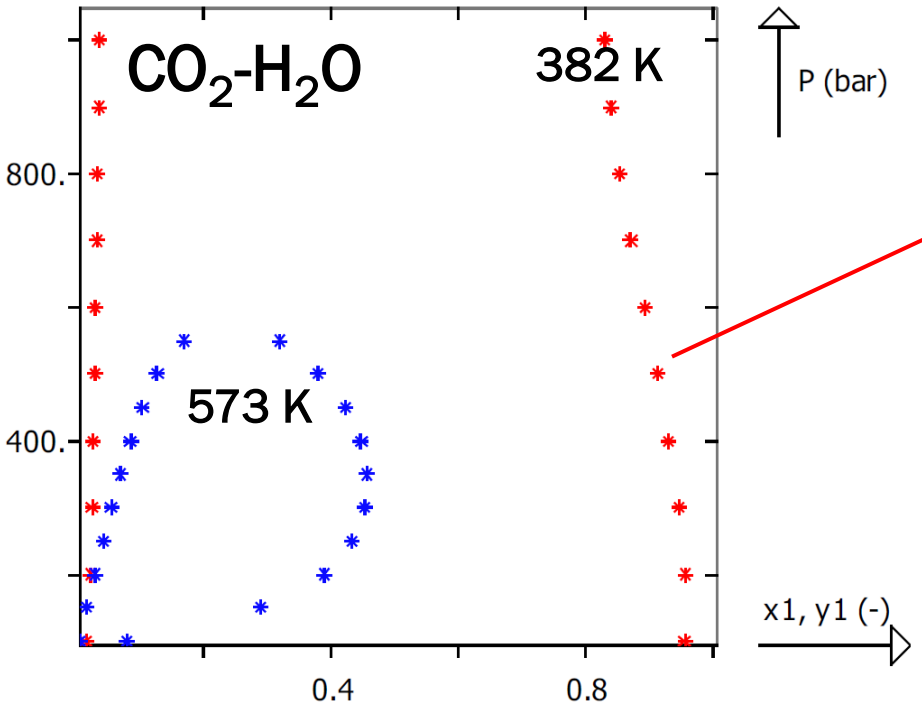


Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

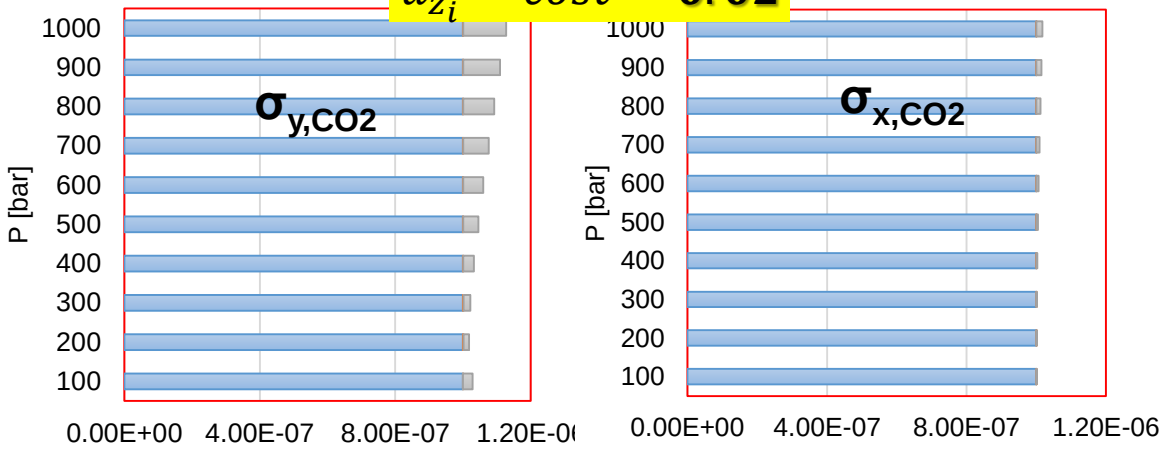
Optimization of models over VLE data

Maximum likelihood method and experimental uncertainties

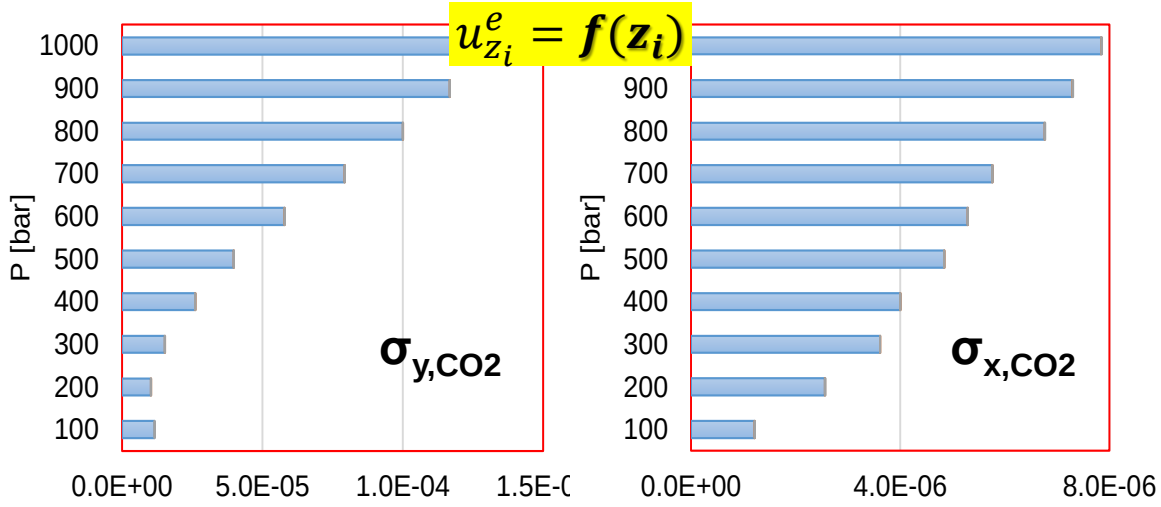
$$\sigma^2_{z_{1,i}} = \left(u^e_{z_{1,i}}\right)^2 + \left(\frac{\partial z_1}{\partial P}\right)^2_{T_i,\bar{\beta}} \left(\underbrace{u^e_{P_i}}_{0.2bar}\right)^2 + \left(\frac{\partial z_1}{\partial T}\right)^2_{P_i,\bar{\beta}} \left(\underbrace{u^e_{T_i}}_{0.2K}\right)^2$$



$u^e_{z_i} = cost = 0.02$



$u^e_{z_i} = f(z_i)$



Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

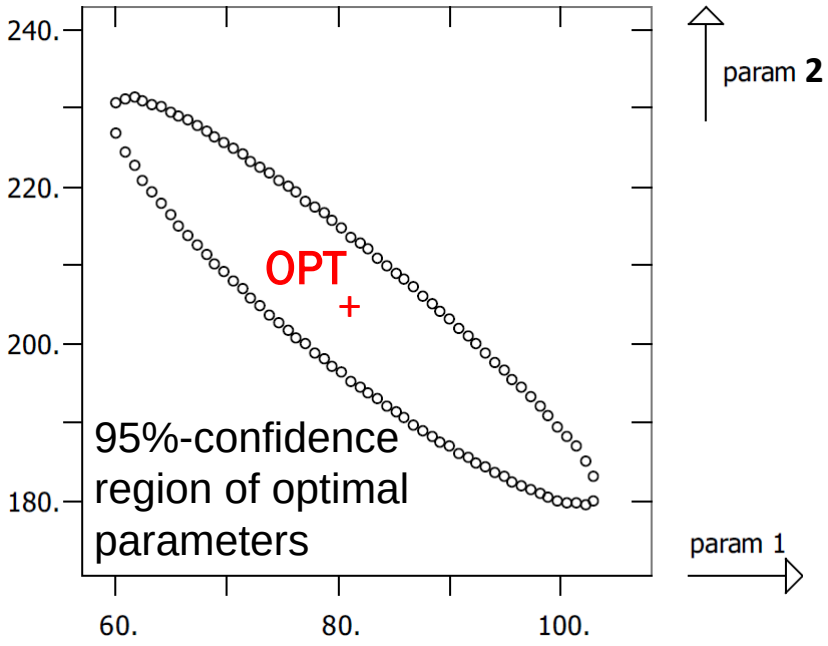
Optimization of models over VLE data
Maximum likelihood method and experimental uncertainties

- It can be stated that the application of MLM positively damps the highest non-linearities of the optimization problem, under-weighting in a more relevant way the deviations at the critical region of the phase diagram.
- Thus, we have seen that the contribution of molar fraction uncertainties:
  - *is comparable to that of pressure and temperature* ***in binary VLE systems containing two subcritical components***
  - *is much greater than that of pressure and temperature uncertainties, in the definition of molar fraction variances,* ***in binary VLE systems containing supercritical components***
- The use of uncertainties specific to each molar fraction, rather than overall constant uncertainties defined for a global set of data, leads to different results of the optimization (different optimal parameters).

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The uncertainty of optimal parameters β_{opt}



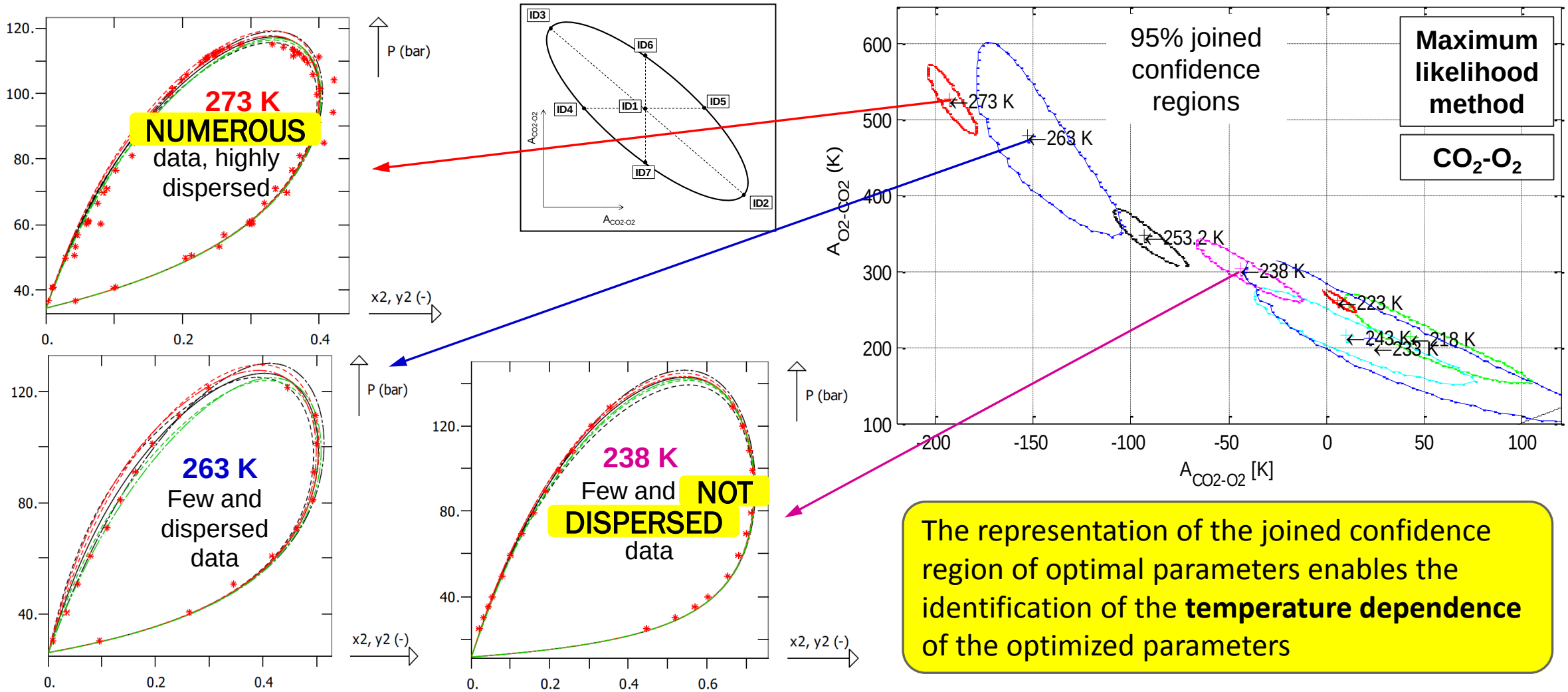
$$S(\boldsymbol{\beta})_\alpha = S(\boldsymbol{\beta}_{opt})\left[1 + \frac{NP}{n_e - NP} F_\alpha(NP, n_e - NP)\right]$$

- ⇧ **numerosity (NP)** $\Longrightarrow$ ⇩ $S(\beta)_\alpha$
- ⇩ $\boldsymbol{S(\beta_{opt})}$ $\Longrightarrow$ ⇩ $S(\beta)_\alpha$
- ⇧ **data uncertainty** $\Longrightarrow$ ⇧ difference between the curvature of the MLM and LSM objective functions around the minimum:
  - when the EoS is able to represent experimental data, MLM and LSM optimal parameters are similar but **MLM** leads to **more extended regions of uncertainties**
  - when the model is not able to represent experimental data, the conclusion is not straightforward. However, **MLM** leads **more regular uncertainty regions**.

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The effect of the numerosity and dispersion of data

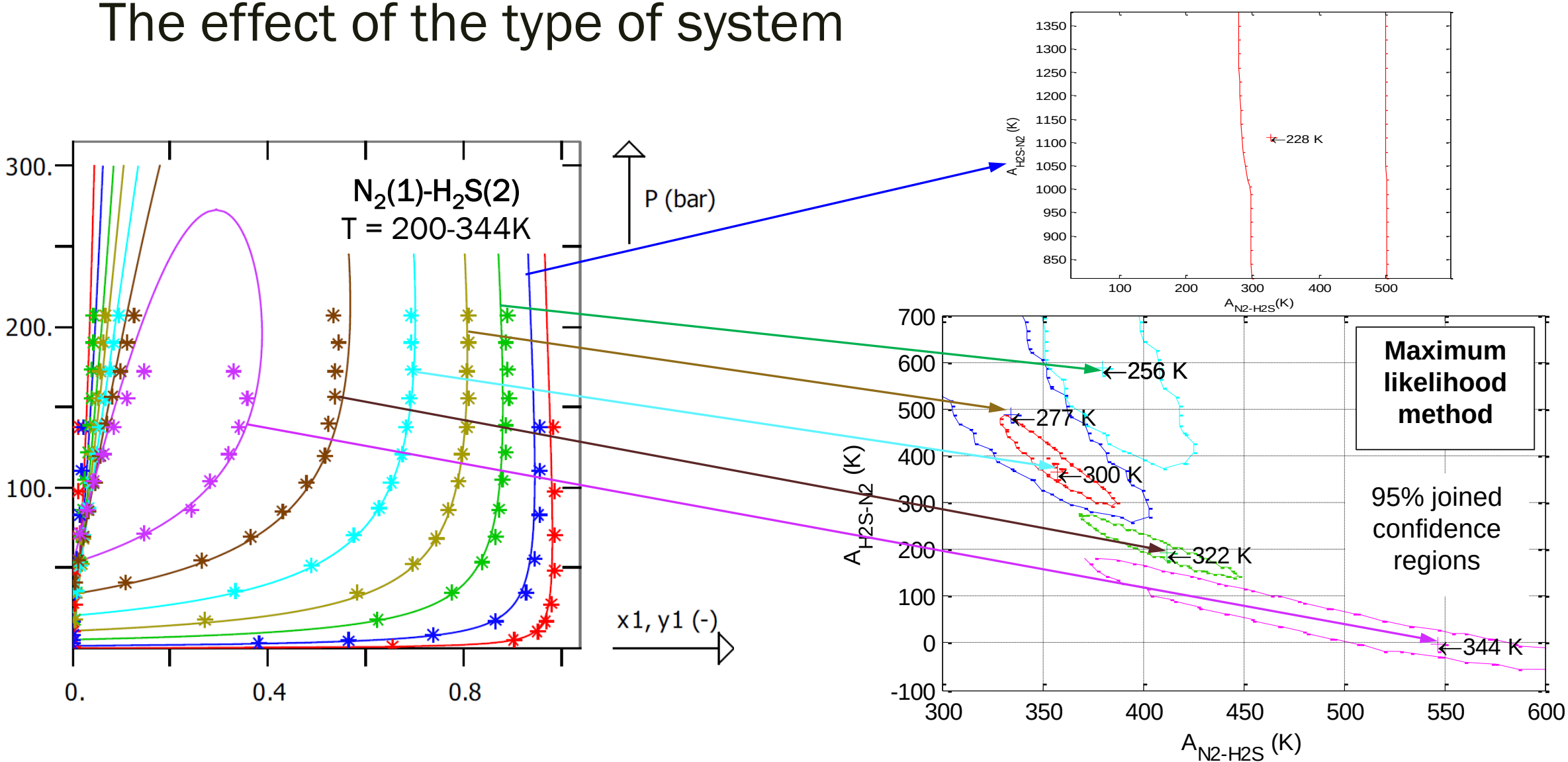


The representation of the joined confidence region of optimal parameters enables the identification of the **temperature dependence** of the optimized parameters

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

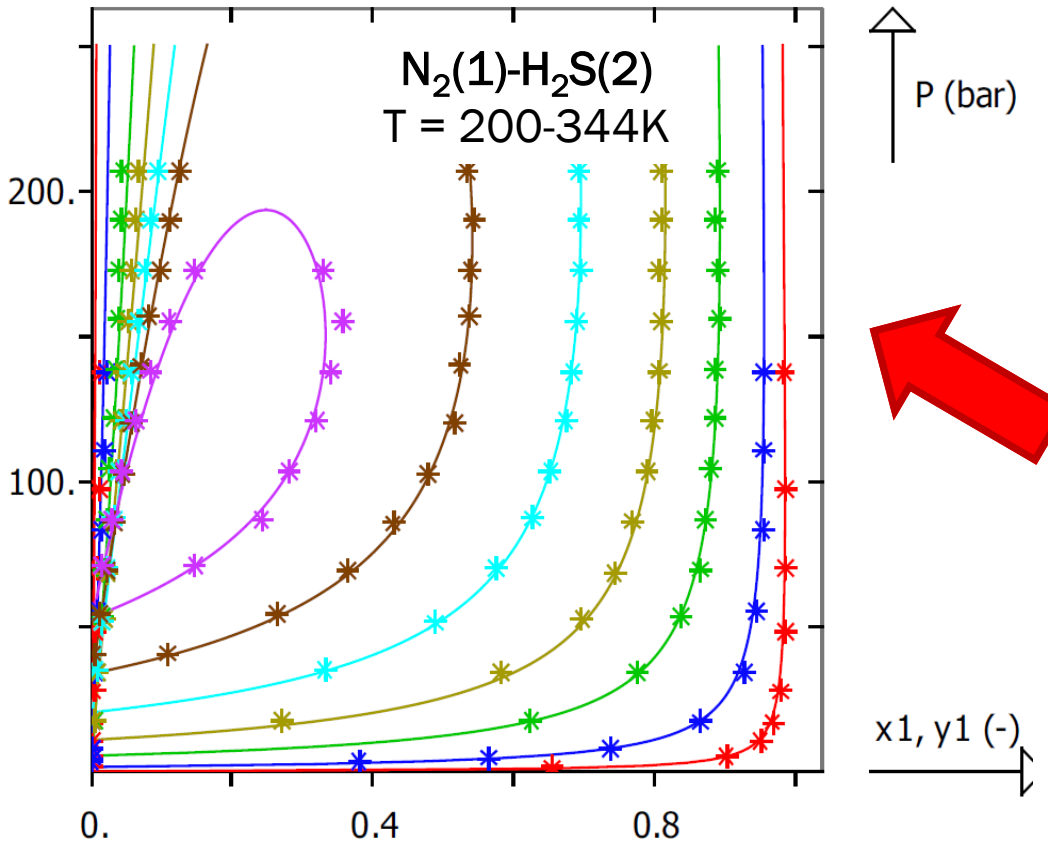
The effect of the type of system

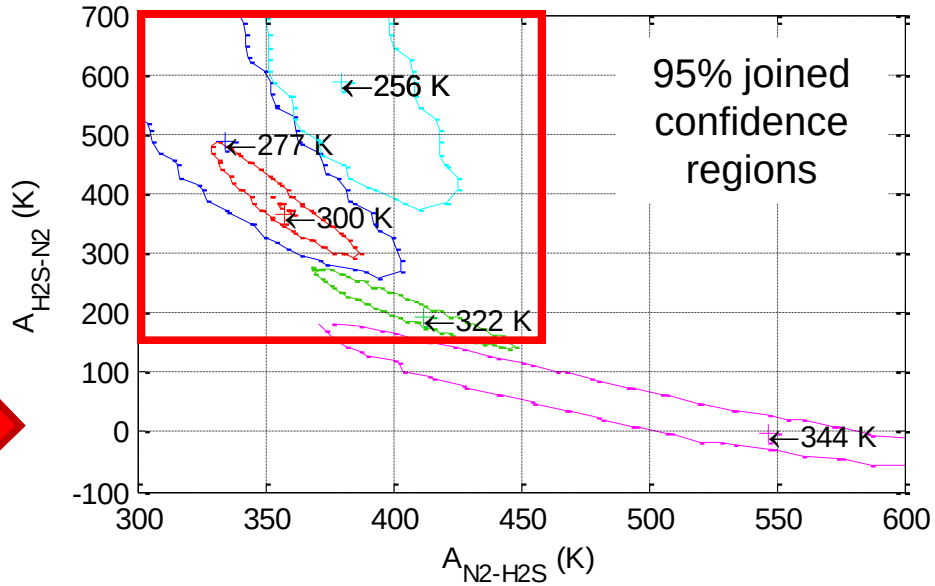


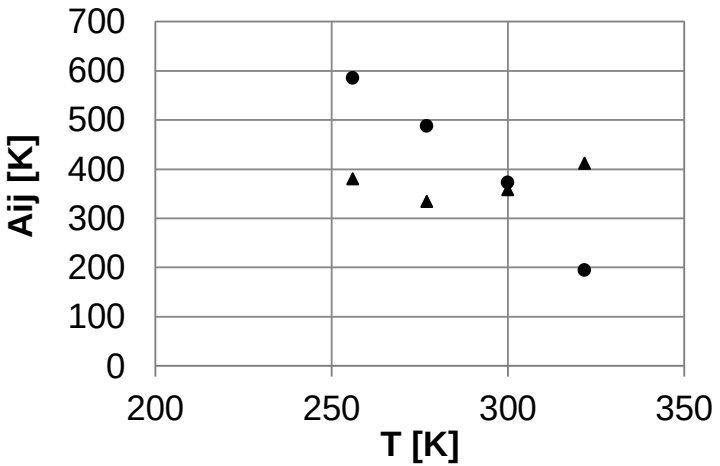
Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of models over VLE data

The effect of the type of system

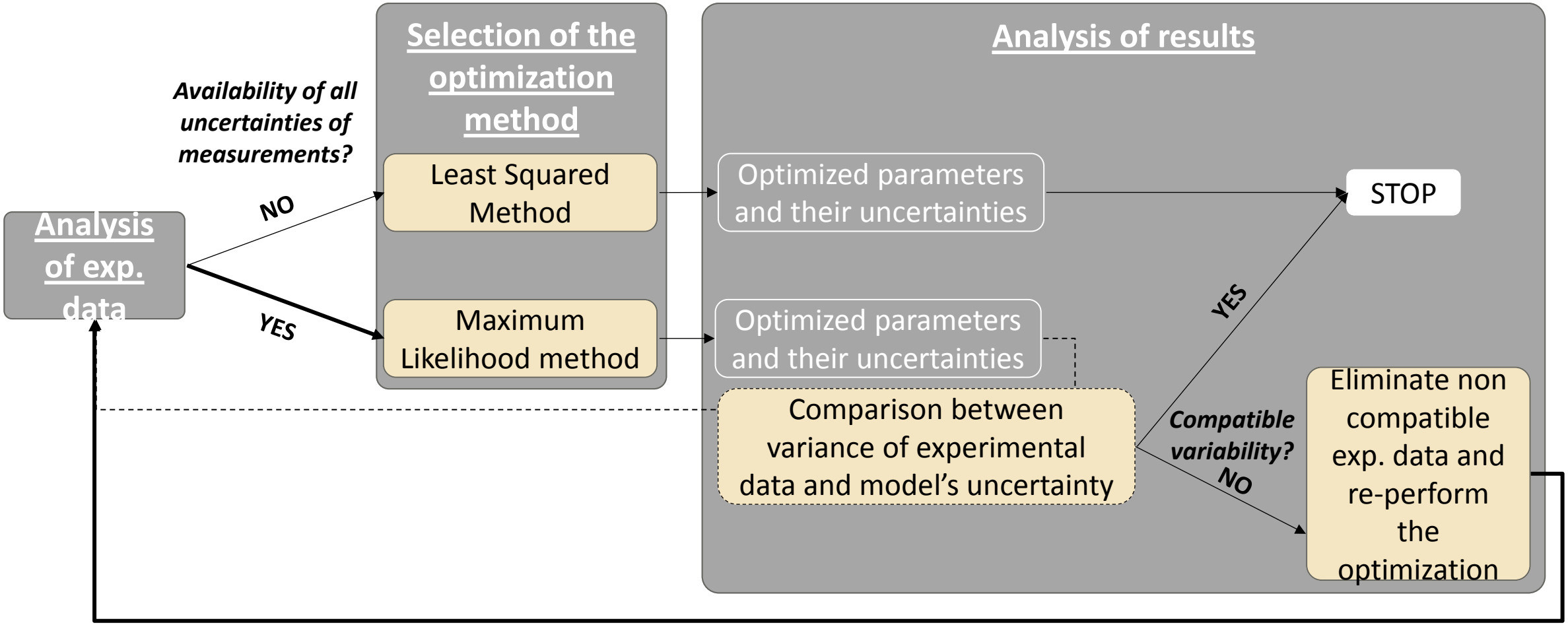






Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

Optimization of model parameters
Conclusions



Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC

THANKS

TO ALL FOR YOUR KIND ATTENTION

and

to Jean-Noël Jaubert & Romain Privat
for helpful discussions

Silvia Lasala - silvia.lasala@univ-lorraine.fr - ENSIC